

UNIVERSIDADE FEDERAL DE ALAGOAS  
INSTITUTO DE COMPUTAÇÃO  
MESTRADO EM MODELAGEM COMPUTACIONAL  
DO CONHECIMENTO

BRUNO FERNANDO COSTA DO NASCIMENTO  
bruno.f.c.nascimento@gmail.com

**Um Sistema de Informação para o Gerenciamento de Banco de Dados do  
Germoplasma da Cana-de-Açúcar da UFAL/RIDESA**

Maceió  
2012

BRUNO FERNANDO COSTA DO NASCIMENTO

**Um Sistema de Informação para o Gerenciamento de Banco de Dados do  
Germoplasma da Cana-de-Açúcar da UFAL/RIDESA**

Dissertação de Mestrado apresentada ao mestrado em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas, como requisito parcial para obtenção do grau de Mestre em Modelagem computacional de Conhecimento.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Roberta Vilhena Vieira Lopes

Coorientador: Manoel Agamemnon Lopes

Maceió  
2012

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**  
**Bibliotecário Responsável: Valter dos Santos Andrade**

N244s Nascimento, Bruno Fernando Costa do.  
Um sistema de informação para o gerenciamento de banco de dados do Germoplasma da cana-de-açúcar da UFAL/RIDESA / Bruno Fernando Costa do Nascimento. – 2012.  
102 f. : il.

Orientadora: Roberta Vilhena Vieira Lopes.

Coorientadora: Manoel Agamemnon Lopes.

Dissertação (Mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Programa de Pós-Graduação em Modelagem Computacional de Conhecimento. Maceió, 2012.

Bibliografia: f. 98-101.

Apêndices: f. 102.

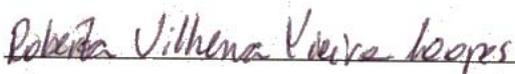
1. Banco de germoplasma. 2. Genótipo *Saccharum* SSP. 3. Melhoramento vegetal - Cana-de-açúcar. 4. Modelagem. 5. Banco de dados – Gerenciamento. 6. Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA).  
I. Título.

CDU: 004.78:633.61



Membros da Comissão Julgadora da Dissertação de Mestrado de Bruno Fernando Costa do Nascimento, intitulada: “Um Sistema de Gerenciamento de Banco de Dados para o Germoplasma da Cana-de-Açúcar”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas em 22 de junho de 2012, às 14h00min, na sala de aula do Mestrado em Modelagem Computacional de Conhecimento.

**COMISSÃO JULGADORA**



**Profa. Dra. Roberta Vilhena Vieira Lopes**

UFAL – Instituto de Computação

Orientadora



**Prof. Dr. Patrick Henrique da Silva Brito**

UFAL – Instituto de Computação

Examinador



**Prof. Dr. Manoel Agamemnon Lopes**

UFAL – Centro de Ciências Agrárias

Examinador



**Prof. Dr. Paulo Vanderlei Ferreira**

UFAL – Centro de Ciências Agrárias

Examinador



**Prof. Dr. Fernando da Fonseca de Souza**

UFPE – Centro de Informática

Examinador

Maceió, junho de 2012.

A Deus que em sua plenitude me concedeu forças para essa caminhada, aos meus pais, irmãos e a minha noiva, por toda força que me deram nos momentos difíceis.

## AGRADECIMENTOS

Nesta página muito especial deste trabalho, gostaria de agradecer a algumas pessoas, dentre as muitas que me ajudaram a realizá-lo.

Agradeço a Deus pela graça concedida de viver em saúde e paz, pela família que me concedeu e pelos amigos que colocou no meu caminho.

Aos meus pais, irmãos e noiva, por acompanharem e torcerem por esse importante passo da minha vida profissional. Sabendo que são os maiores motivadores deste meu desafio, pela capacidade de me ajudar a construir meu desenvolvimento pessoal e profissional.

A minha caríssima **Orientadora Prof. Dr<sup>a</sup>. Roberta Vilhena Vieira Lopes**, pela dedicação, paciência, sempre disposta a ouvir e ajudar. Por assumir a responsabilidade de orientar um aluno que não pertencia à área de computação. Por ajudar a transformar um simples desejo em realidade. Por todas as vezes que escutei “não tem problema ou tudo bem” quando não entregava os trabalhos solicitados em prazo. Por ser o espelho modelo de professor que desejo seguir, pois espero ser compreensivo, atencioso e principalmente amigo dos meus alunos.

Ao meu caríssimo **Coorientador Prof.Dr. Manoel Agamemnon Lopes**, por sua paciência, sempre aconselhando. Por ter despertado o desejo de um mestrado em computação. Por apoiar a investida em uma área diferente da minha nativa ciência agrônoma.

Aos amigos que conquistei no mestrado, Valter, Bruno Raphael, Hélio, Leonardo, Henrique e Douglas que sempre estiveram por perto me ajudando e me dando apoio.

Ao Mestrado de Modelagem Computacional do Conhecimento e ao Instituto de Computação, por me acolher sem distinção de área de estudo.

Aos professores, tanto do mestrado quanto da graduação do IC, por contribuírem com minha formação acadêmica.

À FAPEAL pelo apoio financeiro.

Enfim, agradeço a todos aqueles que contribuíram para a execução deste trabalho, reconheço calorosamente o apoio.

“Seu tempo é limitado, então não percam tempo vivendo a vida de outro. Não sejam aprisionados pelo dogma – que é viver com os resultados do pensamento de outras pessoas. Não deixe o barulho da opinião dos outros abafar sua voz interior. E mais importante, tenha a coragem de seguir seu coração e sua intuição. Eles de alguma forma já sabem o que você realmente quer se tornar. Tudo o mais é secundário.”

Steven Jobs, (empresário e inventor americano, fundador e ex-presidente da empresa Apple, de eletrônicos e do estúdio de animação).

## RESUMO

Com a incessante busca por novas fontes de energia renováveis, o Brasil vem se destacando devido as suas várias fontes alternativas para produção de energia, na qual a cana-de-açúcar destaca-se por sua alta produção de energia tanto com o álcool combustível, como na reutilização de seus subprodutos. Os países investem em técnicas e tecnologias que visam produzir energia com uma menor taxa de poluição e reaproveitamento de materiais que poderiam servir de fonte de poluição para o meio ambiente. Contudo, essas fontes utilizadas, por outros países, não conseguem sobrepor a produção e a viabilidade econômica para produção energética que a cana-de-açúcar possui. Desta forma, o Brasil se destaca no âmbito internacional de produção de energia renovável, pois é o maior produtor mundial de cana-de-açúcar. O sistema proposto nesta dissertação é um sistema de informação para o gerenciamento de banco de dados, que armazenará os dados do germoplasma da cana-de-açúcar, visando auxiliar no desenvolvimento de novas variedades de cana-de-açúcar, armazenando os dados obtidos nas diversas fases de produção de uma variedade, bem como, armazenar os dados tecnológicos destas variedades que são produzidos nas usinas e destilarias, de cana-de-açúcar, ligados a Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA). Desta forma, o processamento destes dados possibilitará uma redução no tempo de produção de uma nova variedade, desta forma, diminuindo o custo para obtenção de uma variedade, bem como, aumentando a velocidade de resposta e dando confiabilidade nas informações processadas dos dados.

**Palavras chaves:** Banco de germoplasma. Genótipo de *Saccharum ssp.* Melhoramento vegetal. Modelagem de dados. RIDESA.

## ***ABSTRACT***

With the incessant search for new sources of renewable energy, Brazil has been highlighted due to its various alternative sources for energy production, in which the sugar cane stands out for its high energy production with both alcohol fuel as reuse of byproducts. Countries invest in techniques and technologies to produce energy with a lower rate of pollution and reuse of materials that could serve as a source of pollution to the environment. However, these sources used by other countries, can not override the production and economic viability for energy production that cane sugar has. Thus, Brazil stands out internationally for producing renewable energy, it is the world's largest producer of cane sugar. The proposed system in this paper is an information system for managing the database, which will store the data of germplasm of cane sugar, aiming to assist in the development of new varieties of cane sugar, storing the data obtained in various stages of production of a variety, as well as store data technology that these varieties are produced in mills and distilleries, cane-sugar, bound Interuniversity Network for the Development of sugarcane industry (RIDESA). Thus, the processing of these data will allow a reduction in production time of a new variety thus reducing the cost for obtaining a variety as well as increasing the response speed and reliability of the information providing processed data.

**Keywords:** Genebank. Genotype *Saccharum ssp.* Plant breeding. Data modeling. RIDESA.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> – Foto área da EFCSO .....	20
<b>Figura 2</b> – Arquitetura básica de um SGBD.....	28
<b>Figura 3</b> – Fases de um Projeto de Banco de Dados. ....	29
<b>Figura 4</b> – Esquema conceitual do SG-Can.....	31
<b>Figura 5</b> – Modelo Hierárquico .....	32
<b>Figura 6</b> – Modelo de rede .....	33
<b>Figura 7</b> – Modelo Relacional .....	34
<b>Figura 8</b> – Exemplo de entidade e atributos do SG-Can.....	35
<b>Figura 9</b> – Exemplo de relacionamento 1:1 do SG-Can.....	37
<b>Figura 10</b> – Exemplo de relacionamento 1:N do SG-Can.....	38
<b>Figura 11</b> – Exemplo de Tabela do SG-Can.....	40
<b>Figura 12</b> – Exemplo de PK do SG-Can. ....	41
<b>Figura 14</b> – Representação da 1FN no SG-Can.....	49
<b>Figura 15</b> – Exemplo do Fator integridade.....	52
<b>Figura 16</b> – Esquema de armazenamento dos dados dos genótipos.....	64
<b>Figura 17</b> – Exemplo de Cruzamento BP.....	66
<b>Figura 18</b> – Exemplo de Cruzamento MP .....	67
<b>Figura 19</b> – Exemplo de Cruzamento MPE.....	68
<b>Figura 20</b> – Esquema das tabelas de cruzamento .....	69
<b>Figura 21</b> – Exemplo de Semeio .....	70
<b>Figura 22</b> – Esquema interno da tabela pmgca_semeio .....	71
<b>Figura 23</b> – Exemplo de repicagem e aclimação das plântulas. ....	72
<b>Figura 24</b> – Esquema interno da tabela de pmgca_repicagem. ....	72
<b>Figura 25</b> – Exemplo de plantio de T1. ....	73
<b>Figura 26</b> – Esquema interno da tabela pmgca_plantiot1. ....	74
<b>Figura 27</b> – Esquema interno da tabela de pmgca_selecaot1.....	75
<b>Figura 28</b> – Exemplo de plantio de T2 .....	75
<b>Figura 29</b> – Esquema interno da tabela de pmgca_plantiot2.....	76
<b>Figura 30</b> – Esquema interno da tabela pmgca_selecaot2.....	77
<b>Figura 31</b> – Exemplo de plantio de T3. ....	78
<b>Figura 32</b> – Esquema interno da tabela pmgca_plantiot3. ....	79
<b>Figura 33</b> – Esquema interno da tabela pmgca_selecaot3.....	80
<b>Figura 34</b> – Estrutura interna da tabela pmgca_fexperimental.....	81
<b>Figura 35</b> – Estrutura da tabela pmgca_c maturacao.....	83

<b>Figura 36</b> – Diagrama ER dos setores de controle de usuário e grupos.....	86
<b>Figura 37</b> – Interface para escolha do relatório de seleção de T1.....	87
<b>Figura 38</b> – Relatório de T1 .....	88
<b>Figura 39</b> – Classificação da visualização dos dados.....	88
<b>Figura 40</b> – Demonstração dos dados filtrados.....	89
<b>Figura 41</b> – Modelo de Relatório de seleção de T1.....	89
<b>Figura 42</b> – Fragmento do código do arquivo httpd.conf.....	92
<b>Figura 43</b> – Modelo HttpRequest de um SGBD.....	94

## LISTA DE ABREVIATURAS

ANSI	American National Standards Institute
APIS	Application Programming Interface
AVA	Argissolo Vermelho Amarelo
BDR	Banco de Dados Relacional
CM	Curva de Maturação
CWI	Instituto de Pesquisa Nacional para Matemática e Ciência da Computação
DBMS	Data Base Management System
DDL	Linguagem de Definição de Dados
DRY	Don't Repeat Yourself
EFCSO	Estação de Floração e Cruzamento Serra do Ouro
E-R	Modelo Entidade-Relacionamento
ESC	Escaldadura.
FAL	Ferrugem Alaranjada;
FAM	Ferrugem Amarelada;
FE	Fase de experimentação
FK	Foreign Key
FM	Fase de Multiplicação
GUIs	Interfaces gráficas para os usuários
IAA	Instituto do Açúcar e do Alcool
IAC	Instituto Agrônomo de Campinas
ISO	International Standards Organizations
MAM	Mancha Amarelada
MAO	Mosaico
MAP	Mancha Parda
MVC	Model View Controller
NCSA	National Center for Supercomputing Applications
ORM	Mapeador objeto-relacional
PCs	Computadores Pessoais
PDV	Podridão Vermelha;
PHP	Hypertext Preprocessor
PK	Chaves Primárias
PLANALSUCAR	Programa Nacional de Melhoramento da Cana-de-açúcar
PMGCA	Programa de Melhoramento Genético da Cana-de-Açúcar
PROALCOOL	Programa Nacional do Alcool
RAQ	Raquitismo da Soqueira;
RB	República do Brasil
RIDESA	Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro;
RSS	Really Simple Syndication
SG-Can	Sistema de Informação para o Gerenciamento do Banco de Dados do Germoplasma da Cana-de-açúcar;
SQL	Structured Query Language
T1	Fase de Teste 1
T2	Fase de Teste 2
T3	Fase de Teste 3

TCH	Tonelada de Cana por Hectare
TPH	Tonelada de POL por Hectare
UFAL	Universidade Federal de Alagoas
UFPR	Universidade Federal do Paraná
UFRPE	Universidade Federal Rural de Pernambuco
UFRRJ	Universidade Federal Rural do Rio de Janeiro
UFS	Universidade Federal de Sergipe
UFSCar	Universidade Federal de São Carlos
UFV	Universidade Federal de Viçosa
VSAM	Virtual Storage Access Method

## SUMÁRIO

1.0	INTRODUÇÃO.....	15
1.1	Processo Histórico no Brasil .....	15
1.2	Evoluções da pesquisa da Cana-de-Açúcar .....	16
1.2.1	A Pesquisa no Brasil.....	17
1.2.2	O Início do Melhoramento Genético de Cana-De-Açúcar No Brasil.....	18
1.2.3	Programa de Melhoramento da COPERSUCAR .....	18
1.2.4	Programa Nacional de Melhoramento da Cana-de-açúcar - PLANALSUCAR.....	19
1.3	Estação de Floração e Cruzamento Serra do Ouro (EFCSO).....	20
1.4	Características botânicas da cana-de-açúcar.....	21
1.5	Motivação .....	21
1.6	Objetivos.....	23
1.6.1	Geral .....	23
1.6.2	Específico .....	23
1.7	Organização .....	23
2.0	SISTEMA DE BANCO DE DADOS.....	25
2.1	Introdução.....	25
2.2	Modelos de Sistemas de Banco de Dados .....	25
2.2.1	Projeto de Banco de Dados.....	28
2.2.2	Modelagem de (banco de) dados .....	30
2.2.3	Modelos de Banco de Dados .....	32
2.3	Banco de Dado Relacional (BDR) .....	34
2.3.1	Modelagem de Dados Utilizando o Modelo Entidade-Relacionamento .....	35
2.3.2	Por que utilizar o Modelo Relacional? .....	39
2.3.3	Conceito do Modelo Relacional .....	39
2.3.4	Estrutura de um Banco de Dados Relacional .....	40
2.3.5	Álgebra Relacional .....	42
2.3.6	A Linguagem SQL .....	45
2.3.7	Restrições de integridade do modelo relacional.....	46
2.3.8	Restrições de Domínio .....	47
2.3.9	Restrições de Chave e em Valores Null .....	47
2.3.10	Restrições de integridade de entidade .....	48
2.3.11	Restrições de integridade referencial.....	48
2.3.12	Normalização de BDR utilizada para o desenvolvimento do SG-Can.....	48
2.4	Escolha do Ambiente Computacional .....	50

2.4.1	Parte conceitual .....	50
2.4.2	Fator integridade.....	51
2.4.3	Fator desempenho.....	52
2.4.4	Fator disponibilidade.....	53
2.4.5	Fator invulnerabilidade.....	53
2.4.6	Coexistência dos quatro fatores fundamentais .....	54
2.4.7	Programas Utilizados.....	54
3.0	BANCO DE GERMOPLASMA DA CANA-DE-AÇÚCAR .....	61
3.1	Introdução.....	61
3.2	O Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA).....	61
3.3	RIDESA.....	62
3.4	O banco de Germoplasma da RIDESA .....	63
3.5	Processos para obtenção de novas Variedades de Cana-de-Açúcar.....	65
3.5.1	Quarentena.....	65
3.5.2	Cruzamentos .....	65
3.5.3	Secagem e Deslintamento.....	69
3.5.4	Semeio .....	70
3.5.5	Repicagem .....	71
3.5.6	Plantio da fase T1 .....	73
3.5.7	Plantio da fase T2 .....	75
3.5.8	Seleção de T2 .....	76
3.5.9	Plantio da fase T3 .....	77
3.5.10	Avaliação de T3.....	79
3.5.11	Avaliação da Fase Experimental – FE.....	80
3.5.12	Avaliação da Curva de Maturação – CM .....	82
3.6	Desenvolvimento do SG-Can .....	83
3.6.1	O SGBD SG-Can.....	85
3.6.2	Exemplo de utilização do SG-Can.....	87
4.0	O MODELO DO BANCO .....	90
4.1	Instalações da Infraestrutura.....	90
4.2	Implementação do modelo de dados .....	90
4.3	Conectividades ao banco de dados .....	91
4.4	Por que utilizar Python e Django?.....	92
4.5	A interação do Django e Python com um SGBD? .....	92
4.6	Segurança do Banco de Dados utilizando Django .....	93
5.0	CONCLUSÃO.....	96
5.1	Gerais.....	96

5.2	Específicos.....	96
5.3	Trabalhos Futuros .....	97
6.0	REFERÊNCIAS .....	98

## 1.0 INTRODUÇÃO

### 1.1 Processo Histórico no Brasil

A cana-de-açúcar chegou ao Brasil por volta de 1500, junto com os portugueses. As primeiras mudas vieram na expedição marítima de Martim Afonso de Souza. Aqui a planta espalhou-se em solo fértil, com a ajuda do clima tropical quente e úmido e da mão-de-obra escrava vinda da África. O registro dessa nova colônia enriqueceu Portugal e espalhou o açúcar brasileiro – assim como aquele produzido na América Central, na colonização de franceses, espanhóis e ingleses – por toda a Europa (COMCIENCIA, 2011) (SUAPESQUISA, 2011).

A crise de 1929, com a queda dos preços internacionais prejudicando o desempenho das exportações do açúcar, abriu espaço para a intervenção do Estado na economia açucareira. O governo brasileiro incentivou o consumo de álcool combustível e tornou obrigatória a mistura de 5% de etanol na gasolina utilizada no país, em 1931 (SUAPESQUISA, 2011).

O crescimento da produção de etanol expandiu a cultura de cana-de-açúcar no Sudeste, especialmente em São Paulo, com o produto ganhando mais espaço na mistura carburante diante das dificuldades de importação de petróleo, na Segunda Guerra Mundial (SUAPESQUISA, 2011).

Em 1933, é criado o Instituto do Açúcar e do Alcool (IAA) pelo então presidente do Brasil Getúlio Vargas. O IAA visava regular a produção do setor canavieiro. Devido à expansão da produção de cana-de-açúcar ocorrida no Nordeste e no Sudeste. As usinas nordestinas (mais concentradas nos estados de Pernambuco e Alagoas) eram responsáveis por toda a exportação brasileira e ainda complementavam a demanda dos estados do sul. A produção do Nordeste somada à de Campos, no norte fluminense, e a rápida expansão das usinas paulistas acenavam para um risco eminente de superprodução. Desta forma, o IAA adotou o regime de cotas, que atribuía a cada usina uma quantidade de cana a ser moída, a produção de açúcar e também a de álcool. A aquisição de novos equipamentos ou a modificação dos existentes também precisavam de autorização do IAA (EMBRAPA, 2011).

Em 1975, visando autonomia energética, o Brasil desenvolveu o Programa Nacional do Alcool (PROÁLCOOL). A solução operacional foi criar procedimentos, incentivos e facilidades que permitisse, num primeiro momento, misturar etanol à gasolina consumida no

país e posteriormente contar quase exclusivamente com esse combustível para abastecer a frota de veículos leves nacionais (COMCIENCIA, 2011).

Em 1979, surgiu o carro a álcool brasileiro e com ele as raízes de um parque produtor com capacidade anual instalada de 18 bilhões de litros de etanol combustível, equivalente a 100 milhões de barris de gasolina por ano. O desenvolvimento de tecnologia específica conquistou os brasileiros e a frota nacional chegou a ser formada por 85% de veículos leves movidos a etanol, no final dos anos 1980 (COMCIENCIA, 2011) (SUAPESQUISA, 2011).

O apoio do governo à produção de álcool se intensificou com as duas crises internacionais do petróleo, em 1973 e 1979. Mas a interferência estatal estava com os dias contados. Ao final dos anos 1990, o mercado estava livre e, desde então, desenvolve sua auto regulação (COMCIENCIA, 2011).

Na década de 80 ocorreu o desabastecimento de álcool combustível no Brasil, ocasionado, em grande parte, pelo aumento do consumo e pela força política dos usineiros, o que mostrou a fragilidade do programa (EMBRAPA, 2011).

No início da década de 90 foi extinto o IAA e, com isso, inicia-se o período de desregulamentação do setor, com a liberação das exportações e dos preços do açúcar e do álcool.

Em 1992, extinção do PLANALSUCAR<sup>1</sup>. O então presidente do Brasil Fernando Collor extingue o PLANALSUCAR e deixa todo o corpo técnico parado.

Em 1991 é criada a Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA). O corpo técnico do extinto PLANALSUCAR é absorvido pelas universidades.

De 1994 a 1999 os preços do açúcar e do álcool passam a ser regidos exclusivamente pelo mercado, sem qualquer intervenção do estado, promovendo uma expansão no setor canavieiro.

## ***1.2 Evoluções da pesquisa da Cana-de-Açúcar***

Classicamente, o registro definitivo de que plantas no campo cresciam a partir de sementes de cana-de-açúcar foi feito por um administrador de fazenda em Barbados, em maio de 1858. Posteriormente, em 1869, ainda em Barbados, relatou-se a realização de polinização artificial. Antes disso, em 1862, em Java, foi registrada a germinação espontânea de sementes de cana-de-açúcar e também em 1871, em Reunião. Dentre alguns trabalhos que

---

<sup>1</sup> PLANALSUCAR - Programa Nacional de Melhoramento da Cana-de-açúcar

definitivamente marcaram o início do melhoramento genético da cana-de-açúcar citam-se que os desenvolvidos por *Soltwedel*, em Java, a partir de 1885, quando conseguiu fazer germinar sementes de *S. spontaneum*, primeiramente, e depois, em 1887, realizou cruzamentos entre aquela espécie e *S. officinarum*. Simultânea e independentemente, em Barbados, *Harrison* e *Bovell*, àquela mesma época, também tinham conseguido obter plântulas a partir de sementes de cana-de-açúcar coletadas no campo (MARTINS, 2005) (MOZAMBANI, 2006).

### ***1.2.1 A Pesquisa no Brasil***

Em 1948 o médico Gervásio Caetano Peixoto Lima afirmou, quando defendeu tese na Faculdade de Medicina do Rio de Janeiro, o que parecia ser um conhecimento bastante estabelecido: que a cana-de-açúcar se reproduzia a partir de sementes sexuais nos locais de origem. Sornay, em 1920, citado por MIOCQUE (1993) menciona que o cientista capanemense comunicou, em dezembro de 1869, na Câmara de Agricultura de Maurício, que tinha observado plântulas de cana crescendo espontaneamente em canaviais de Caiana, tendo enviado para aquela ilha uma coleção de variedades de canas do Brasil, entre as quais constava a São Julião. Esta teria resultado de cruzamento (que na época chamavam de “enxerto”) da Caiana com a Cana-Mole, realizado pelo Comendador Julião Ribeiro de Castro, em Campos, RJ, publicado no jornal O Globo, em 21 de outubro de 1874 (MIOCQUE, 1993). Naquela mesma época, o Barão de Villa Franca escreveu que a fertilidade de sementes de cana era um “fato de conhecimento comum no Brasil”.

Esse mesmo Barão teria sido quem obteve a cana enxerto a partir de cruzamento da cana Ubá com a Roxa. Há também o registro de que em 27 de setembro de 1897 o Imperial Instituto Fluminense de Agricultura enviou para Pernambuco uma grande remessa de mudas de cana, entre as quais se incluíam a Enxerto, a *Quissaman* e a *Caetano*. A *Quissaman* pode ter sido uma variedade obtida no Engenho Central de *Quissaman*, em Campos, mas não se encontrou nenhuma menção sobre isso; a variedade *Caetano* teria sido uma variedade obtida pelo médico Gervásio Caetano Peixoto Lima (MIOCQUE, 1993) (MARTINS, 2005) (MOZAMBANI, 2006).

O final do século XIX e o início do XX caracterizaram-se pelo aumento crescente das preocupações no setor açucareiro, no exterior e no Brasil, tanto com a baixa produtividade dos canaviais e epidemias de doenças quanto com a melhoria da qualidade das canas. Passou a ocorrer, então, intenso intercâmbio de tipos de cana entre os países, especialmente por iniciativa

de proprietários de engenhos. No Brasil, estes foram também pioneiros na criação de novas variedades, mas sempre eram iniciativas dispersas e eventuais, entregues à curiosidade de alguns empresários (AMARAL, 2003).

### ***1.2.2 O Início do Melhoramento Genético de Cana-De-Açúcar No Brasil***

A consciência da necessidade de pesquisa sistematizada acabou por levar a reivindicações pela criação de estações experimentais, que tiveram como principais atividades a experimentação com variedades introduzidas e a criação de outras variedades locais. No Brasil, isso ocorreu em ciclos, sempre em momentos em que os produtores se sentiam pressionados por crise no setor. Esses ciclos grosseiramente podem ser divididos em três: o primeiro, nas duas primeiras décadas do século XX; o segundo, na década de 30; e o terceiro, no final da década de 60 e início da de 70 (MARTINS, 2005) (MOZAMBANI, 2006).

### ***1.2.3 Programa de Melhoramento da COPERSUCAR***

O Programa de Melhoramento da COPERSUCAR foi iniciado em 1968 pela Cooperativa Central dos Produtores de Açúcar e Álcool do Estado de São Paulo, também sob a influência do Dr. J. Mangelsdorf, contratado como assessor do programa. Logo em seguida, a COPERSUCAR incorporou a Estação Experimental de Sertãozinho, da Cooperativa dos Usineiros do Oeste do Estado de São Paulo – COPERESTE, que já vinha conduzindo um modesto programa de melhoramento. Criou uma boa estrutura de pessoal e de estações experimentais regionais (quatro em São Paulo e uma no Paraná), além da sede no Centro de Tecnologia em Piracicaba. Das variedades liberadas, as mais bem sucedidas foram as da primeira liberação, destacando-se SP70-1143, SP71-1406 e SP71-6163, que, juntas, tiveram domínio absoluto em toda a década de 80 e parte da década de 90 do século passado, em toda a região centro-sul do País, chegando também a ter significativa participação nos demais estados produtores, principalmente as duas primeiras (COPERSUCAR, 1995).

Este programa tem também dado ênfase ao desenvolvimento de técnicas biotecnológicas, inclusive com financiamento de grandes projetos internacionais e também tem firmado convênios com empresas ou associações não filiadas para o desenvolvimento de novas variedades (SILVA, 2010).

### ***1.2.4 Programa Nacional de Melhoramento da Cana-de-açúcar - PLANALSUCAR***

A partir da segunda metade da década de 60, surgiram preocupações com a base tecnológica da agroindústria canavieira no Brasil. O fator que motivou o início desta preocupação foi devido a franca expansão do setor canavieiro brasileiro, especialmente em São Paulo. A preocupação com a agricultura canavieira foi motivada por uma série de conjunturas externas e internas, pela baixa produtividade brasileira e pela vulnerabilidade a que essa atividade estava sujeita, por falta de uma sólida base de pesquisa e de desenvolvimento tecnológico. Ainda que nunca citado, talvez porque também nunca percebido, mesmo pelos próprios participantes, provavelmente a edição do livro “Cultura e Adubação da Cana-de-açúcar” pelo então Instituto Brasileiro de Potassa (MALAVOLTA, 1964), reunindo uma plêiade de especialistas da época para uma análise global da base tecnológica que sustentava a atividade sucroalcooleira no Brasil, contribuiu decisivamente para o reconhecimento sobre a urgente necessidade de investimentos em pesquisa e tecnologia canavieira. Provavelmente, também devido à tomada de tal consciência é que resultou o convite, por parte do IAA, ao melhorista de cana-de-açúcar do Havaí, A.J. Mangelsdorf, para visitar o Brasil e realizar um diagnóstico das necessidades técnicas do setor canavieiro (MANGELSDORF, 1967).

Desta análise resultou um relatório que acabou sendo decisivo para a criação do PLANALSUCAR.

A atuação do PLANALSUCAR estava ligada a todas as áreas de produção do setor canavieiro. Além de financiar a modernização do parque de usinas e destilarias, em 1971, a poupança criada pelo fundo de exportação após a disparada nos preços do açúcar também financiou o PLANALSUCAR. As ações do órgão iam desde o desenvolvimento de novas variedades de cana até a projeção de safras. Segundo Luiz Carlos Corrêa Carvalho, ex-superintendente do PLANALSUCAR, os estudos do avanço da cana estavam prontos antes mesmo de surgir o programa que alavancou definitivamente a produção do etanol no País (ESTADÃO, 2010).

Em 1990, com a extinção do PLANALSUCAR, houve momentos de grande incerteza sobre os destinos do seu patrimônio físico, pessoal e genético. Porém, um ano depois, grande parte de todo esse patrimônio foi absorvida por sete universidades federais (UFAL – Universidade Federal de Alagoas, UFS – Universidade Federal de Sergipe, UFRPE – Universidade Federal Rural de Pernambuco, UFRRJ – Universidade Federal Rural do Rio de Janeiro, UFV – Universidade Federal de Viçosa, UFSCar – Universidade Federal de São Carlos

e UFPR – Universidade Federal do Paraná) que compõem a Rede Interinstitucional de Desenvolvimento do Setor Sucroalcooleiro (RIDESA) e, assim, a maioria dos melhoristas foi para as novas instituições, o que possibilitou resgatar o material genético e dar continuidade aos trabalhos de cruzamentos, seleções e avaliações (BARBOSA, 2002).

### **1.3 Estação de Floração e Cruzamento Serra do Ouro (EFCSO)**

A EFCSO, comumente chamada apenas de Serra do Ouro, é o local onde está armazenada uma grande variedade de cana-de-açúcar de toda parte do mundo. Este acervo é conhecido como Banco de Germoplasma da cana-de-açúcar e possui atualmente, ativos para cruzamento, mais de 2600 acessos de germoplasma de cana, além de, ter um museu de variedades de cana de diversas partes do mundo (Figura 1).

**Figura 1** – Foto área da EFCSO



**Fonte:** PMGCA

A Serra do Ouro é uma estação de cruzamento da UFAL/CECA/PMGCA/RIDESA, e está localizada em Murici-AL, a uma latitude de 09°13'S, longitude de 35°50'W, altitude de 450 m, a 35 km de distância geográfica do litoral (BARBOSA, 2002). Apresenta área de 80 ha, o solo é classificado como argissolo vermelho amarelo<sup>2</sup> (AVA), de relevo ondulado (10 a 12%

---

<sup>2</sup> Argissolo - São solos medianamente profundos a profundos, moderadamente drenados, com horizonte B textural (horizonte diagnóstico que caracteriza a classe de solo), de cores vermelhas a amarelas e textura argilosa, abaixo de um horizonte A ou E de cores mais claras e textura arenosa ou média, com baixos teores de matéria orgânica.

de declividade). O clima é As' na classificação de Koeppen<sup>3</sup> com chuvas de outono a inverno (março a agosto), apresentando estação seca no verão (setembro a fevereiro), com pluviosidade anual média de 2.363mm, e temperatura média mínima de 18,9° e média máxima de 27,1°C. Essas condições tornam esta localidade muito propícia para o florescimento da maioria dos genótipos de cana-de-açúcar, como também para boa fertilidade de pólen (BARRETO, 2002) (ROCHA, 2008).

Atualmente, a Serra do Ouro está produzindo mais de 300.000 plântulas de cana-de-açúcar por ano (BARBOSA, 2008).

#### **1.4 Características botânicas da cana-de-açúcar**

A cana-de-açúcar é uma planta monocotiledônea, *alógama* e semi-perene, provavelmente originária das regiões da Indonésia e Nova Guiné. Pertencem à família *Poaceae*, tribo *Andropogoneae* e gênero *Saccharum*. Seus atuais cultivares são híbridos interespecíficos, sendo que nas constituições genéticas participam principalmente as espécies *S. officinarum* e *S. spontaneum*. Trata-se de uma planta de reprodução sexuada, porém, quando cultivada comercialmente, é multiplicada assexuadamente, por propagação vegetativa.

É caracterizada pela inflorescência do tipo panícula, flor hermafrodita, caule em crescimento cilíndrico composto de nós e entrenós, e folhas alternas, opostas, presas aos nós dos colmos, com lâminas de sílica em suas bordas, e bainha aberta (SILVA, 2010).

Ao longo dos anos os programas de melhoramento genéticos da cana-de-açúcar, ligados a RIDESA, vêm gerando um grande número de dados provenientes de estudos científicos e análises industrial.

#### **1.5 Motivação**

Esta dissertação pela sua natureza multidisciplinar apresenta uma motivação computacional, que é a necessidade de desenvolver um sistema de informação que seja capaz de armazenar os dados referentes ao banco de germoplasma de cana-de-açúcar do Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA).

---

<sup>3</sup> Classificação de Koeppen - Classificação climática de Köppen-Geiger, mais conhecida por classificação climática de Köppen, é o sistema de classificação global dos tipos climáticos mais utilizada em geografia, climatologia e ecologia.

Este sistema de informação deve ser capaz de armazenar todos os dados, gerados na cadeia de obtenção de novas variedades da cana. Esta nova ferramenta deve auxiliar na escolha dos genótipos a serem cruzados, levando em conta as características agroindústrias de cada genótipo. Também servirá de base de pesquisa para todos que precisarem de informações acerca do germoplasma da cana-de-açúcar. Sendo assim, este software será responsável por gerenciar todos os dados de cana-de-açúcar que são conduzidos em âmbito nacional pela Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro (RIDESA), que é a responsável pela obtenção das variedades da sigla RB. Estas variedades obtidas pela RIDESA têm prestado relevantes serviços ao país, com a liberação de mais de trinta novas variedades RB em apenas dezessete anos, que atualmente ocupam mais de 50% da área canavieira do país, contribuindo com significativos aumentos da produtividade e qualidade dessa agroindústria (BARBOSA, 2008).

Além da motivação computacional é possível destacar também uma motivação econômica, pois a cana-de-açúcar é uma cultura de grande expressão no Brasil, ocupando segundo a CONAB (2010), uma área estimada na safra 2010/2011 de 8,09 milhões de hectares. Essa matéria-prima é utilizada principalmente para a fabricação de açúcar, etanol e bioeletricidade, e é a cadeia produtiva de maior importância socioeconômica e estratégica em termos energéticos (BARBOSA, 2008).

Nos últimos anos, devido à crise energética e o aquecimento global, o etanol e o bagaço vêm ganhando destaque mundial. O Brasil destaca-se como o maior produtor mundial dessa matéria-prima, seguido pela Índia, China, Paquistão, México, Tailândia, Colômbia, Austrália, Indonésia, Estados Unidos, África do Sul, Filipinas, dentre outros (EMBRAPA, 2010) (UNICA, 2010).

Por fim é preciso destacar a motivação ecológica, pois diante da pressão ambientalista em todo o mundo para o uso de combustíveis de fontes renováveis e não poluentes, o Brasil está conseguindo uma posição de destaque mundial, pois é o maior produtor de cana-de-açúcar do mundo. A cana de açúcar destaca-se como fonte de energia renovável e de combustível menos poluentes. Este fato se dá porque o etanol, que é um combustível carburante de fonte renovável, e o bagaço, pela geração de bioeletricidade, estão contribuindo para o equilíbrio ambiental, pois a cultura possui alta taxa de sequestro de carbono, reduzindo os efeitos da emissão de poluentes (UNICA, 2010).

De acordo com a Agência Internacional de Energia, a produção e o uso do etanol no Brasil permitem reduzir em mais de 80% a emissão de gases do efeito estufa em relação à gasolina (JUNIOR, 2009).

## **1.6 Objetivos**

### **1.6.1 Geral**

Desenvolver o SG-Can para o PMGCA/UFAL, tornando-o, uma ferramenta confiável aos melhoristas, no processamento das informações dos genótipos da cana-de-açúcar.

Deixar o SG-Can apto para receber novos módulos de processamento de dados Genéticos tornando-o capaz de fazer interações entre a morfologia e a genética da cana-de-açúcar.

Tornar o SG-Can base para o desenvolvimento de novos sistemas de processamento de dados agronômicos.

### **1.6.2 Específico**

Construir um sistema de informação para o gerenciamento de gerenciamento de banco de dados modernos e tecnologicamente avançado para o banco de germoplasma de cana-de-açúcar do PMGCA / CECA / UFAL / RIDESA, com as seguintes atribuições:

1. Permitir manusear, com eficiência, o acervo do germoplasma da cana-de-açúcar da Serra do Ouro.
2. Permitir a inserção de adição de novos módulos computacionais que sejam capazes de simular diversas etapas das atividades do acervo da Serra do Ouro.
3. Aperfeiçoar o processo de obtenção de novas variedades.
4. Permitir o tratamento, processamento e mineração de dados morfológicos e industriais da cana-de-açúcar.

## **1.7 Organização**

Para cumprir os objetivos definidos na Seção 1.6, organizou-se este trabalho em seis capítulos, incluindo esta introdução.

No Capítulo 2, será feita uma abordagem sobre os sistemas de banco de dados, sua importância contextual em uma instituição, bem como uma abordagem sobre os modelos e padrões para desenvolvimento de bancos de dados. Será realizada uma descrição sobre os modelos de banco de dados. Também terá uma melhor explanação sobre o modelo de Banco de

Dados Relacional (BDR), o qual foi escolhido para o desenvolvimento do SG-Can, a estrutura de um BDR, as operações e as entidades do SG-Can. Por fim, será demonstrado o ambiente computacional escolhido, com a descrição dos módulos do sistema, bem como as ferramentas utilizadas para criação deste sistema de informação.

No Capítulo 3 será demonstrado um breve histórico sobre o Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA) e sobre a Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro (RIDESA). Será realizada uma descrição sobre as fases de criação de uma variedade, bem como dos dados do PMGCA/CECA/UFAL/RIDESA, para que seja possível demonstrar a importância do desenvolvimento do SG-Can, sendo demonstrada a importância da criação de um sistema que integre as bases da RIDESA.

No Capítulo 4, será apresentado o modelo de dados utilizado, bem como as ferramentas utilizadas para o desenvolvimento. Também será demonstrado o porquê da escolha da linguagem de programação, do banco de dados escolhidos e de outras ferramentas utilizadas, além da instalação destes softwares.

No Capítulo 5, será mostrada a conclusão dessa dissertação e as perspectivas de trabalhos futuros a serem desenvolvidos.

No Capítulo 6, será apresentada a bibliografia utilizada para a construção desta dissertação.

Assim, os capítulos descreveram toda abordagem literária e computacional para o desenvolvimento do SG-Can. Iniciando por uma abordagem geral para a criação do mesmo.

## 2.0 SISTEMA DE BANCO DE DADOS

### 2.1 Introdução

A tecnologia de banco de dados já foi descrita como sendo “uma das áreas de mais rápido crescimento na área da computação e da informação”. Os estudos em banco de dados ainda são comparativamente jovens, levando-se em consideração outras áreas da ciência da computação. Entretanto, apesar de ser recente, esta área tornou-se, rapidamente, de considerável importância, tanto no plano prático como no teórico (DATE, 2003).

Quando se fala de banco de dados deve-se ter em mente os seguintes termos: dados, modelagem dos dados, tipos de banco de dados, relacionamentos dos dados, recuperação de dados, segurança e etc. Uma melhor especificação sobre estes termos, entre outros, será realizada abaixo nos tópicos deste capítulo.

### 2.2 Modelos de Sistemas de Banco de Dados

Na década de sessenta, os sistemas computacionais, embora ainda embrionários sob o ponto de vista da difusão, já davam mostras de que grandes volumes de dados precisavam ser tratados de forma mais criteriosa, para que sua integridade fosse preservada e seu acesso garantido, lhes conferindo assim credibilidade. Sendo assim e contando com a fundamental evolução do hardware, cada vez mais a tendência de armazenar grandes quantidades de dados foi se consolidando, até que no final dos anos sessenta, grandes fabricantes de máquinas como *IBM, Borroughs, Unisys, etc.*, aos poucos disponibilizavam ao cenário internacional, já então caracterizado pelo uso de computadores de grande capacidade de processamento, produtos mais especificamente voltados ao armazenamento e disponibilização dos dados, visto que em áreas estratégicas distintas, como indústria automobilística, siderúrgicas e instituições financeiras, a necessidade era iminente (DATE, 2003).

Desta forma, como contemporâneo do arquivo indexado *Virtual Storage Access Method* (VSAM), tecnologia mais utilizada até então para armazenamento e disponibilização de grandes volumes, surgia o software banco de dados, produto que por intermédio de instruções peculiares normalizadas, permitia que basicamente quatro eventos ocorressem, não simultaneamente: gravação, leitura, atualização e destruição do dado, de forma global ou pontual (CÂNDIDO, 2007). Permitia também, caso houvesse alguma falha de caráter físico

com a máquina, ou lógico, durante a manipulação dos dados, que os mesmos fossem posteriormente recuperados por intermédio de mecanismos de cópias prévias ou ainda de expurgo controlado de dados. Consolidava-se, assim, o banco de dados hierárquico, cujo engajamento no cenário computacional mundial foi único, mesmo porque nada similar existia disponível neste sentido (IBM, 1998).

No início dos anos setenta surge o conceito de banco de dados relacional, um modelo de banco de dados que permite manusear facilmente os dados de forma eficiente e sensivelmente mais acessível, financeiramente, quando comparada aos bancos hierárquicos utilizados até então, embora também capaz de armazenar e disponibilizar grandes volumes de dados e também contemplar os mecanismos de recuperação e expurgo (ALVES, 2004).

Os sistemas de banco de dados são projetados para administrar grandes volumes de dados de informações sobre uma determinada aplicação, promovendo um ambiente das mesmas (SILBERSCHATZ, 2006).

De acordo com Date (2003, p.06),

Um sistema de banco de dados é basicamente um sistema computadorizado de manutenção de registros; em outras palavras, é um sistema computadorizado cuja finalidade geral é armazenar dados e permitir que os usuários busquem e atualizem esses dados quando os solicitar.

Segundo Elmasri e Navathe (2005, p. 04),

“Sistema de banco de dados é uma coleção de dados relacionados”.

O maior benefício de um sistema de banco de dados é proporcionar, ao usuário, uma visão abstrata dos dados. Uma vez que a maioria dos usuários de bancos de dados não é especialista em computação, omite-se deles a complexidade da estrutura interna dos bancos de dados, graças a diversos níveis de abstração que simplificam a interação do usuário com o sistema (SILBERSCHATZ, 2006).

De uma maneira geral, pode-se dizer que um sistema de banco de dados é constituído por um conjunto de programas e/ou aplicações; estes, por sua vez, estão associados a um conjunto de dados por intermédio de um sistema gerenciador de banco de dados (SGBDs) (SILBERSCHATZ, 2006).

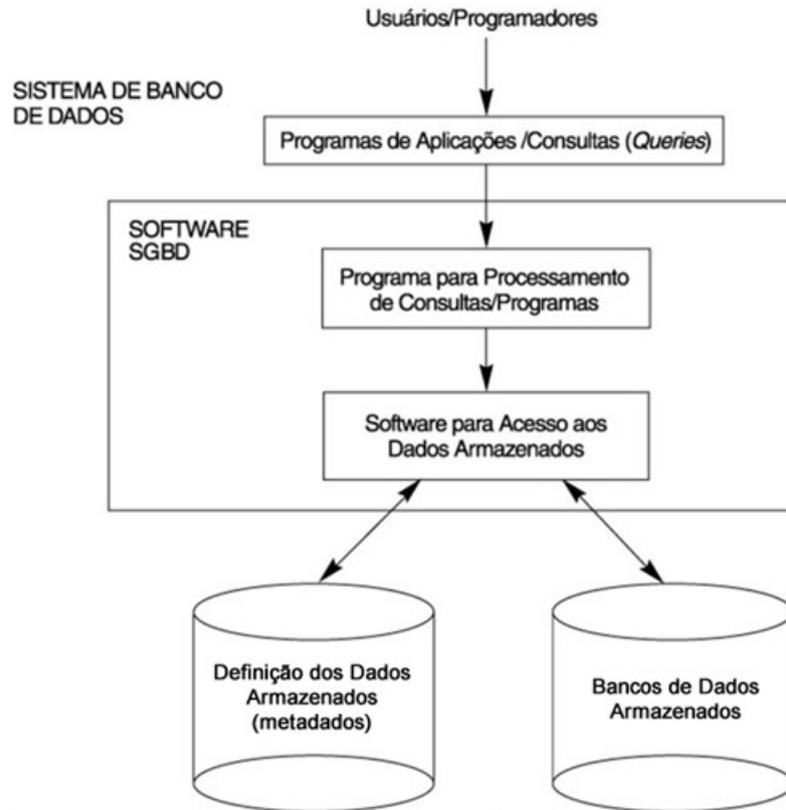
O termo sistema de gerenciamento de banco de dados (SGBD) vem do inglês *Data Base Management System* (DBMS). É um conjunto de programas de computador (*softwares*) responsáveis pelo gerenciamento de uma base de dados. Seu principal objetivo é retirar da aplicação cliente a responsabilidade de gerenciar o acesso, a manipulação e a organização dos dados. O SGBD disponibiliza uma interface para que seus clientes possam incluir, alterar ou consultar dados previamente armazenados. Em bancos de dados relacionais a interface é

constituída pelas *Application Programming Interface* (APIs) ou drivers do SGBD, que executam comandos na linguagem *Structured Query Language* (SQL) (OLIVEIRA, 2002).

A arquitetura de pacotes de SGBDs evoluiu dos sistemas pioneiros e monolíticos, em que o pacote de softwares SGBD era um bloco único formando um sistema fortemente integrado, para os modernos pacotes de SGBDs modulares por projeto, com uma arquitetura cliente/servidor. Essa evolução reflete as tendências na computação, na qual os computadores grandes e centralizados (*mainframes*) estão sendo substituídos por vários computadores pessoais (PCs) e estações de trabalho (*workstations*) conectados via redes de comunicações a vários tipos de servidores – servidor Web, servidor de banco de dados, servidores de arquivos servidores de aplicações e assim por diante (ELMASRI, 2005).

Em uma estrutura básica de SGBD (figura 02) cliente/servidor, as funcionalidades do sistema são distribuídas, normalmente, em dois módulos. O módulo cliente é projetado para ser executado em uma estação de trabalho ou em um computador pessoal. Em geral, os programas de aplicações e as interfaces de usuário, que acessam o banco de dados, são processados no módulo cliente. Desta forma, o módulo cliente trata a interação com os usuários e oferece uma interface amigável, como formulários Interfaces gráficas para os usuários (GUIs) baseados em menus. Já o módulo servidor, trata de armazenamento de dados, acessos, pesquisas e outras funções (ELMASRI, 2005).

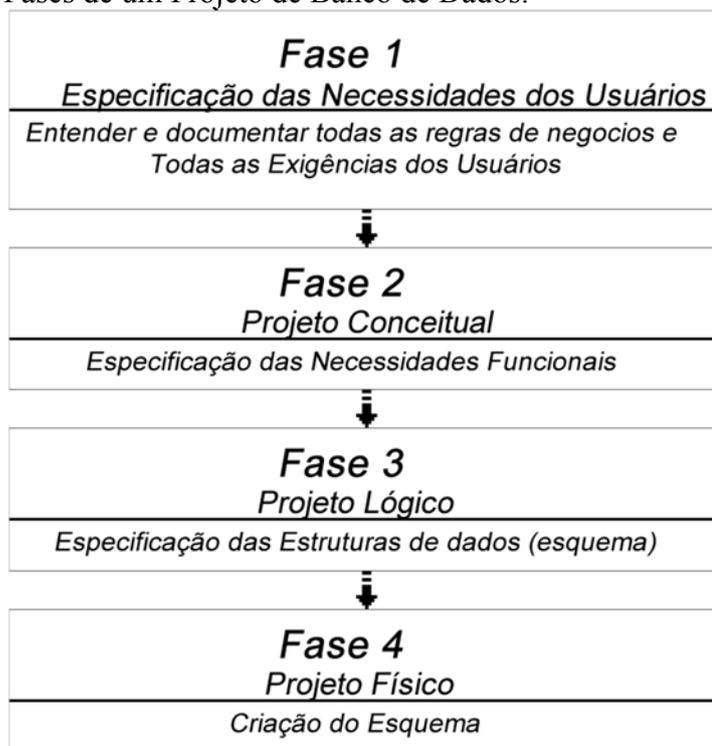
**Figura 2** – Arquitetura básica de um SGBD.



**Fonte:** Elmasri e Navathe (2005).

### ***2.2.1 Projeto de Banco de Dados***

A principal função do projeto de banco de dados é armazenar dados através de um conjunto de relações que permitam armazenar os dados sem redundância desnecessária, e também recuperar facilmente estes dados (SILBERSCHATZ, 2006). Pode-se dizer que as fases necessárias para a construção de um Banco de Dados são quatro, conforme ilustração da Figura 3.

**Figura 3** – Fases de um Projeto de Banco de Dados.

**Fonte:** Autor adaptado de SILBERSCHATZ, (2006).

A primeira fase do projeto, conhecida por “especificação das necessidades dos usuários”, ou “levantamento dos requisitos”, tem por finalidade caracterizar todos os dados necessários na perspectiva dos usuários (SILBERSCHATZ, 2006).

Para isso, coloca-se como primordial: levantar as necessidades dos usuários; possuir um bom entendimento da organização; conhecer a área de negócio na qual a organização está inserida; ficar atento às mudanças organizacionais, com base nas quais o banco de dados será modelado, proporcionando assim, uma maior certeza na tomada de determinadas decisões. É absolutamente essencial entender os requisitos dos usuários e adequar-se às suas necessidades a fim de obter-se um Banco de Dados bem sucedido.

Na segunda fase, o projetista seleciona o tipo de modelo de dados e, por meio da aplicação de seus conceitos, transcreve as necessidades especificadas em um esquema conceitual de banco de dados, chamado “projeto conceitual”, proporcionando uma visão detalhada da organização (SILBERSCHATZ, 2006). Nessa fase, deve-se focar a descrição dos dados e de seus relacionamentos, independentemente das restrições de implementação. Os detalhes físicos de armazenamento não devem ser levados em consideração nesse momento.

A terceira fase, conhecida por “projeto lógico”, tem seu início a partir do modelo conceitual, levando em consideração uma das abordagens de banco de dados possíveis, tais como a abordagem relacional e a abordagem orientada a objetos (SILBERSCHATZ, 2006).

Nesta fase são descritas as estruturas que estarão contidas no banco de dados, de acordo com as possibilidades da abordagem. Contudo, não se considera, ainda, nenhuma característica específica de um sistema gerenciador de banco de dados. A fase de projeto lógico contempla, também, as “especificações das necessidades funcionais”, nas quais os usuários descrevem os tipos de operações a serem realizadas sobre os dados.

Na quarta fase, a do “projeto físico”, o projeto lógico é mapeado para o modelo de implementação de dados; modelo este, especificamente dependente dos recursos do sistema gerenciador de banco de dados que será usado (SILBERSCHATZ, 2006). Tais recursos incluem as formas de organização de arquivos e as estruturas internas de armazenamento, tais como: tamanho de atributos, índices, tipos de atributos, nomenclaturas, dentre outros.

### ***2.2.2 Modelagem de (banco de) dados***

A modelagem de dados responde a um conjunto de questões específicas relevantes para qualquer aplicação de processamento de dados. São vários os fatores que tornam a modelagem de dados fundamental para a construção de Sistemas de Bancos de Dados. Um desses fatores é a necessidade de validar requisitos, e é por intermédio da modelagem que se aumenta a probabilidade de encontrar erros, de se identificar inconsistências e de se descobrir omissões, já nas fases iniciais de projeto (PRESSMAN, 2002).

Por meio da modelagem de dados, consegue-se identificar quais são os principais objetos de dados a serem processados pelo sistema; qual a composição de cada objeto de dados; que propriedades os descrevem; onde os objetos costumam ficar; quais as relações entre cada objeto e os demais; quais as relações entre os objetos e os processos que os transformam (PRESSMAN, 2002).

A modelagem possibilita o alcance de quatro objetivos: ajuda a visualizar o sistema como ele é ou como deve ser; permite especificar a estrutura ou o comportamento de um sistema; proporciona um guia para a construção do sistema e documenta as decisões tomadas (BOOCH, 2000).

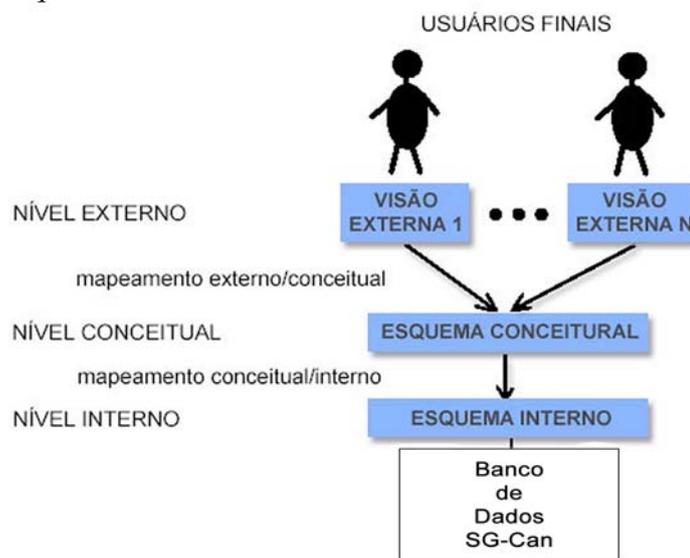
Para iniciar um processo de modelagem, deve-se definir qual a principal fonte de dados. Duas fontes podem ser consideradas: “o conhecimento de pessoas”, em que novos sistemas devem ser concebidos, não havendo descrição inicial dos dados; e “descrições de dados existentes”, em que, para a manutenção de um sistema, utilizasse como descrição inicial dos dados as descrições dos arquivos do sistema já em uso (engenharia reversa).

Um modelo de dados está sob a estrutura do banco de dados. Trata-se de um conjunto de ferramentas conceituais usadas para a descrição de dados e seus relacionamentos, da semântica de dados e das regras de consistência (SILBERSCHATZ, 2006).

Segundo Silberschatz (2006), a arquitetura para Banco de Dados (figura 4) possui três diferentes níveis:

1. **Esquema Interno:** O esquema interno, que é considerado o nível de abstração mais baixo e que descreve como os dados são realmente armazenados, é representado pelos modelos físicos de dados. Por meio desses modelos, é descrita a estrutura física de armazenamento do banco de dados, a sua organização de arquivos, os seus métodos de acesso, os meta-dados, os arquivos de índices e outros dados do sistema.
2. **Esquema Externo:** O esquema externo, também conhecido como visões de usuário, é considerado o nível de abstração mais alto. Nele, são descritas as partes do banco de dados que são do interesse de cada grupo específico de usuários, já que em sua maioria os usuários não precisam acessar todos os dados contidos em um banco de dados. Esse nível de visão existe para simplificar a interação do usuário com os Sistemas de Bancos de Dados.
3. **Esquema Conceitual:** O esquema conceitual é representado pelos modelos conceituais e lógicos. Por meio desses modelos, é descrita a estrutura do banco de dados completo sob o ponto de vista do usuário, omitindo-se detalhes de armazenamento de dados e concentrando-se na descrição de entidades, atributos, relacionamentos, operações de usuário e restrições sobre os dados.

**Figura 4** – Esquema conceitual do SG-Can.



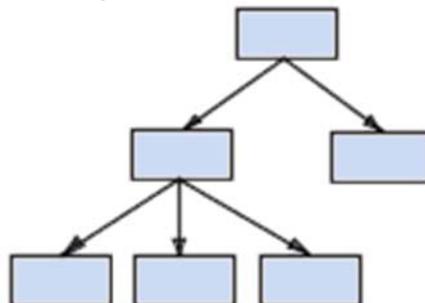
Fonte: Autor adaptando Silberschatz (2006)

### 2.2.3 Modelos de Banco de Dados

Segundo Elmasri (2005), os modelos de banco de dados podem ser divididos em três grandes grupos:

1. Modelo Hierárquico: é um tipo de sistema de gerenciamento de banco de dados que conecta registros numa estrutura de dados em árvore através de ligações de tal modo que cada tipo de registro tenha apenas um possuidor como demonstra a figura 5. Os dados são classificados hierarquicamente, de acordo com uma arborescência descendente. Este modelo utiliza apontadores entre os diferentes registros. Trata-se do primeiro modelo de SGBD.
  - Características:
    - Os dados são organizados em vários tipos de registros;
    - Relacionamentos explícitos entre registros: pai, filhos, irmãos;
    - Registros devem ser vistos dentro de um contexto.
  - Restrições de integridade inerentes ao modelo:
    - Ocorrências registros filhos só podem existir ligadas a algum pai;
    - Se um registro filho tem dois ou mais pais do mesmo tipo de registro, o registro filho deve ser duplicado para cada pai;
    - Para um registro ter mais de um pai de tipos de registros diferentes, um deve ser o pai real e os outros devem ser pais virtuais.

**Figura 5** – Modelo Hierárquico



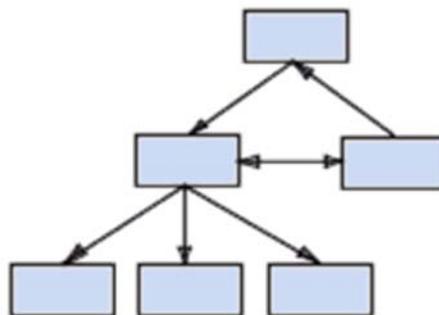
**Fonte:** Autor

2. Modelo de Rede: O modelo em redes surgiu como uma extensão ao modelo hierárquico, eliminando o conceito de hierarquia e permitindo que um mesmo registro estivesse envolvido em várias associações (figura 6). No modelo em rede, os registros são organizados em grafos onde aparece um único tipo de associação que define uma relação 1:N entre 2 tipos de registros: proprietário e

membro. Como o modelo hierárquico, este modelo utiliza apontadores para os registros. Contudo, a estrutura já não é necessariamente arborescente no sentido descendente.

- Características:
  - Apresentação de dados estruturados como grafo;
  - Modela comportamento de relacionamentos, especificando restrições de inserção/remoção/modificação;
  - Otimização através de “navegação inteligente”.
- Restrições de integridade inerentes ao modelo:
  - Permite especificar restrições de integridade relativas a inserção de registros em relacionamentos;
  - Permanência de registros no banco de dados, dependendo de seus relacionamentos.

**Figura 6** – Modelo de rede

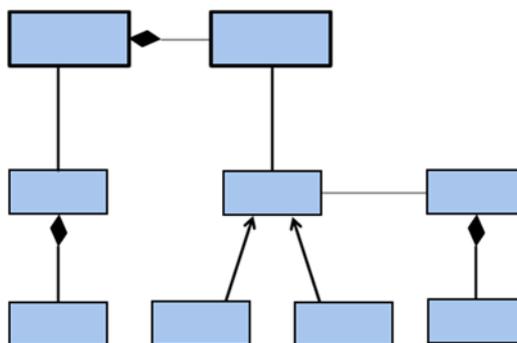


**Fonte:** Autor

3. Modelo Relacional: O modelo relacional (figura 7) é um modelo de dados, adequado a ser o modelo subjacente de um SGBD. Baseia-se no princípio em que todos os dados estão guardados em tabelas ou, matematicamente falando, relações. Toda sua definição é teórica e baseada na lógica de predicados e na teoria dos conjuntos.
- Características:
    - Base de dados é visualizada como um conjunto de tabelas, cada uma representando uma relação;
      - ❖ Relacionamentos representados por valores de dados;
      - ❖ Tabelas consistem de:
        - Linhas → *tuplas*;
        - Colunas → Atributos.
      - ❖ Simetria nas consultas

- ❖ Linguagens definidas:
  - Álgebra e cálculo relacionais, “transformados” em SQL com ampla aceitação;
  - SQL - linguagem padrão para bancos de dados relacionais.

**Figura 7** – Modelo Relacional



**Fonte:** Autor

De acordo com a pesquisa realizada dos modelos de banco de dados, foi definido que o melhor modelo para o desenvolvimento do SGBD do germoplasma da cana-de-açúcar seria o modelo relacional. Desta forma, este modelo será melhor descrito na seção abaixo. Onde será feita uma abordagem mais bem detalhada dos bancos de dados relacionais.

### 2.3 Banco de Dado Relacional (BDR)

A estruturação de banco de dados de forma relacional foi desenvolvida por Ted Codd, no laboratório da IBM *Reseracch*, em 1970, em um artigo clássico (Codd, 1970) que imediatamente atraiu a atenção em virtude de sua simplicidade e base matemática. O modelo usa o conceito de uma relação matemática – algo como uma tabela de valores – como seu bloco de construção básica e tem sua base teórica na teoria dos conjuntos e na lógica de predicados de primeira ordem (ELMASRI, 2005).

Este modelo, por suas características e por sua completude, mostrou ser uma excelente opção, superando os modelos mais usados àquela época: o de redes e o hierárquico. A maior vantagem do modelo relacional sobre seus antecessores é a representação simples dos dados e a facilidade com que consultas complexas podem ser expressas.

Este modelo também é conhecido como modelo de dados entidade-relacionamento, sendo baseado na percepção do mundo real que consiste em um conjunto de objetos, denominadas entidades, e nos relacionamentos entre esses objetos (SILBERSCHATZ, 2006).

### 2.3.1 Modelagem de Dados Utilizando o Modelo Entidade-Relacionamento

O modelo Entidade-Relacionamento (E-R) é modelo de dados conceitual de alto nível, cujos conceitos foram projetados para estar o mais próximo possível da visão que o usuário tem dos dados, não se preocupando em representar como estes dados estarão realmente armazenados. O modelo E-R é utilizado, principalmente, durante o processo de projeto de banco de dados.

#### 2.3.1.1 Entidades e Atributos

O objeto básico tratado pelo modelo E-R é a “entidade”, que pode ser definida como uma representação de um objeto existe no mundo real, concreto (automóvel, casa, professor) ou abstrato (duplicata, venda, entre outros) que apresente de forma distinguível de outros objetos. Cada entidade possui um conjunto particular de propriedades (características) que a descreve, chamado de atributo (SILBERSCHATZ, 2006). O conjunto de todas as variedades de cana-de-açúcar, por exemplo, pode ser definido como o conjunto de entidade para os biólogos, melhoristas e etc., e as respectivas informações sobre BRIX, TCH, TPH e etc., podem ser considerados como atributos, como demonstra a figura 8.

**Figura 8** – Exemplo de entidade e atributos do SG-Can.

genotipo	prog_f	prog_m	procedencia
RB92579	RB72454	RB867515	UCO
SP79-1011	SP11-8675	SP42-5086	SP

Onde:

E - Entidade;  
A - Atributo

**Fonte:** Autor

Um atributo pode ser dividido em diversas subpartes com significado independente entre si, recebendo o nome de atributo composto, como é o caso do atributo endereço, composto de rua, bairro, cidade, estado e CEP. Um atributo que não pode ser subdividido é chamado de atributo simples ou atômico, por exemplo, o atributo CPF.

É preciso fazer uma ressalva sobre atributos atômicos, por que para os brasileiros o atributo nome é atômico, porém nos Estados Unidos, por exemplo, o atributo nome é composto,

pois se divide em *first name*, *midle name* e *last name*, desta forma, é importante salientar que certos atributos são atômicos ou não, dependendo da interpretação de cada analista.

Os atributos, que podem assumir apenas determinado valor em uma determinada instância, são denominados atributos simplesmente valorados, como é o caso do atributo salário de um empregado. Um atributo que pode assumir diversos valores em uma mesma instancia é denominado multivalorado, como por exemplo, telefone de uma pessoa – supondo que ela possua mais telefone. Um atributo que é gerado a partir de outro é chamado de atributo derivado, são os atributos cujos valores não são armazenados no BD, por não haver necessidade de fazê-lo, como é o caso salário médio dos empregados, que pode ser derivado do atributo salário.

### **2.3.1.2 Tipos de Entidades**

Um banco de dados costuma conter grupos de objetos que são similares possuindo, os mesmos atributos, porém, cada entidade com seus próprios valores para cada atributo. Estes conjuntos de objetos similares definem um “tipo entidade”. Cada tipo entidade é identificado por seu nome e pelo conjunto de atributos, que definem suas propriedades. A descrição do tipo entidade é chamada de “esquema do tipo entidade”, especificando o nome do tipo entidade, o nome de cada um de seus atributos e qualquer restrição que incida sobre as entidades.

### **2.3.1.3 Relacionamento**

O relacionamento é a associação entre diversas entidades, onde na maior parte dos bancos de dados relacionam apenas duas entidades.

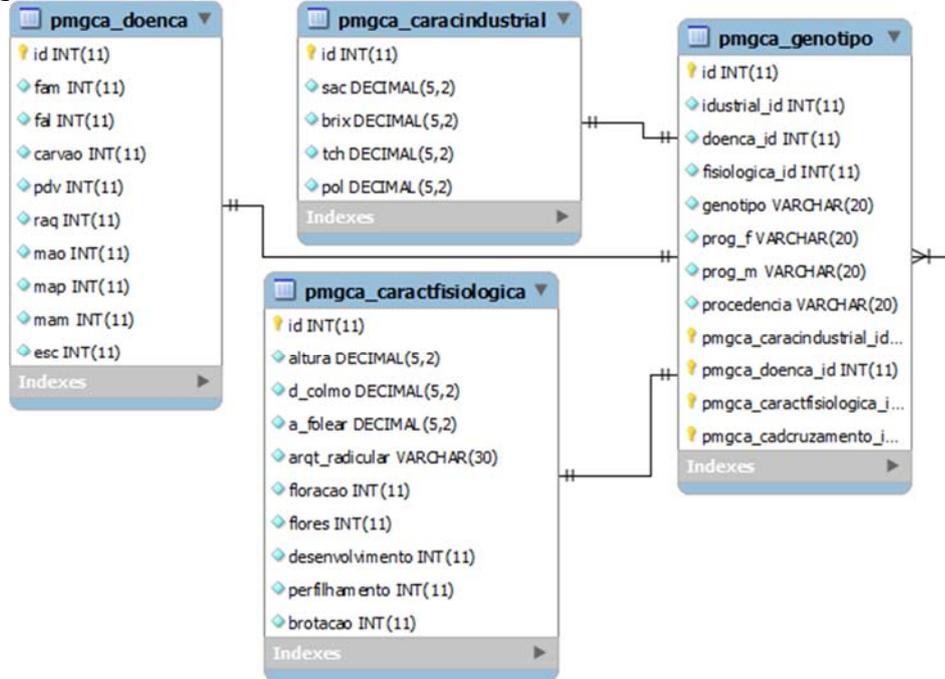
Segundo Setzer (1986), num conjunto de relacionamentos o número de entidades às quais outras entidades podem estar associadas, é definido de cardinalidade, onde, para um conjunto de relacionamentos binários entre conjunto de entidade X e Y, a cardinalidade pode ser uma das seguintes:

1. Um para um (1:1): uma entidade de X está associada à no máximo uma entidade de Y, e uma entidade de Y está associada à no máximo uma entidade de X. Um exemplo deste relacionamento dentro do banco de dados do SG-Can (figura 9).

Esta relação existe quando os campos que se relacionam são ambos do tipo Chave Primária, em suas respectivas tabelas. Cada um dos campos não apresenta valores repetidos.

Na prática existem poucas situações onde pode-se utilizar um relacionamento deste tipo. O campo que relaciona duas tabelas deve fazer parte, ter sido definido, na estrutura das duas tabelas.

**Figura 9** – Exemplo de relacionamento 1:1 do SG-Can.



**Fonte:** Autor

- Um para vários (1:N): uma entidade de X está associada a qualquer número de entidade de Y, e uma entidade de Y está associada a no máximo a uma entidade de X. A figura 10 ilustra esta relacionamento dentro do banco de dados do SG-Can.

Este é o tipo de relacionamento mais comum entre duas tabelas. Uma das tabelas (o lado um do relacionamento) possui um campo que é a Chave Primária e a outra tabela (o lado vários) se relaciona através de um campo cujos valores relacionados podem se repetir várias vezes.

**Figura 10** – Exemplo de relacionamento 1:N do SG-Can.



**Fonte:** Autor

3. Vários para Vários (N:N): uma entidade de X está associada a qualquer número de entidade de Y, e uma entidade de Y pode estar associada a qualquer entidade de X.

Este tipo de relacionamento "aconteceria" em uma situação onde em ambos os lados do relacionamento os valores poderiam se repetir. Na prática não é possível implementar um relacionamento deste tipo, devido a uma série de problemas que seriam introduzidos no modelo do banco de dados. Para evitar este tipo de problema é bastante comum "quebrar" um relacionamento do tipo Vários para Vários em dois relacionamentos do tipo Um para Vários. Isso é feito através da criação de uma nova tabela, a qual fica com o lado Vários dos relacionamentos – esse tipo de relacionamento não foi utilizado no SG-Can.

Vários são os fornecedores atuais de bancos de dados, dentre os quais pode-se citar IBM, *Microsoft*, *Sybase* e *Oracle*, este último uma empresa que começou naquela época e que hoje é provavelmente a líder do mercado de banco de dados relacionais. Existem também sistemas como *FireBird*, *MySQL* e *PostGres*, que com a nova tendência de software livre, vêm sendo utilizados por várias empresas.

Ainda existem sistemas legados mantidos nos modelos antigos, como o IMS DBMS da IBM no modelo hierárquico e o IDS e IDMS no modelo de redes. Há também uma nova abordagem que vem sendo bastante explorada: o modelo orientado a objetos, que tem com exemplos o *Objectstore* e o *Versant*. Apesar da orientação a objetos não ser um conceito novo, apenas recentemente, com a proposição de novas metodologias, como OMT e UML, e de

linguagens de programação como C++ e Java, seu uso vem se intensificando, tornando-se um padrão em várias aplicações de software. Com isso, a existência de banco de dados orientados a objetos está sendo colocada como uma das revoluções na área de banco de dados. Assim, numa tentativa de reunir vantagens de objetos com as inerentes ao modelo relacional, têm-se os bancos de dados objeto-relacional. Mesmo assim, o modelo relacional ainda é o modelo dominante em seu mercado.

### ***2.3.2 Por que utilizar o Modelo Relacional?***

O modelo relacional será utilizado pelas seguintes razões:

- A. A grande utilização do modelo relacional para o desenvolvimento de SGBDs deve-se em grande parte ao fato de que existe uma representação uniforme dos dados, tornando simples a representação dos dados (padronização), ou seja, facilita o entendimento da estrutura de dados tabular. Dando ao sistema maior integridade e consistência aos dados;
- B. A linguagem SQL utilizada é interativa e muito mais próxima da linguagem natural escrita (inglês). Dando uma maior possibilidade de utilização direta pelo usuário final;
- C. Redução no tamanho dos códigos de programa;
- D. Controle de redundância;
- E. Maior segurança;
- F. Maior flexibilidade para acréscimo de novos dados no BD;
- G. Possibilidade de criação de gatilhos (*Triggers*) e procedimentos armazenados (*Stored Procedure*).

### ***2.3.3 Conceito do Modelo Relacional***

O modelo relacional representa o banco de dados como uma coleção de relações. Informalmente, cada relação se parece com uma tabela de valores ou, em alguma extensão, como um arquivo de registros ‘planos’. Entretanto, há importantes diferenças entre relações e arquivos.

Quando uma relação é pensada como uma tabela de valores, cada linha na tabela representa uma coleção de valores de dados relacionais. No modelo relacional, cada linha na

tabela representa um fato correspondente a uma entidade ou relacionamento do mundo real. O nome das tabelas e os nomes das colunas são usados para ajudar na interpretação do significado dos valores em cada linha. Os nomes das colunas especificam como interpretar os valores de dados em cada linha, com base na coluna em que cada valor está. Todos os valores em uma coluna são do mesmo tipo de dado (ALVES, 2004).

Na terminologia do modelo relacional formal, uma linha a chamada *tupla*, um cabeçalho de coluna é conhecido como atributo, e a tabela é chamada de relação. O tipo de dado que descreve os tipos de valores que podem aparecer em cada coluna é representado pelo domínio de valores possíveis (ALVES, 2004).

### 2.3.4 Estrutura de um Banco de Dados Relacional

#### 2.3.4.1 Tabelas (Relações)

Nos modelos de bases de dados relacionais, a tabela é um conjunto de dados dispostos em número finito de colunas e número ilimitado de linhas (ou *tuplas*). As colunas são tipicamente consideradas os campos da tabela, e caracterizam os tipos de dados que deverão constar na tabela (numéricos, alfanuméricos, datas, coordenadas, etc.). O número de linhas pode ser interpretado como o número de combinações de valores dos campos da tabela, e pode conter linhas idênticas, dependendo do objetivo. A figura 11 traz um exemplo de tabela que é utilizada no SG-Can.

**Figura 11** – Exemplo de Tabela do SG-Can.

**Tabela Genótipo no modo estrutural**

Campo	Tipo	Collation	Atributos	Nulo	Padrão	Extra
id	int(11)			Não		auto_increment
industrial_id	int(11)			Não		
doenca_id	int(11)			Não		
fisiologica_id	int(11)			Não		
genotipo	varchar(20)	utf8_general_ci		Não		
prog_f	varchar(20)	utf8_general_ci		Não		
prog_m	varchar(20)	utf8_general_ci		Não		
procedencia	varchar(20)	utf8_general_ci		Não		

**Tabela Genótipo no modo visualização dos dados**

id	industrial_id	doenca_id	fisiologica_id	genotipo	prog_f	prog_m	procedencia
1	1	1	1	RB92579	RB72454	RB867515	UCO
2	1	1	1	SP79-1011	SP11-8675	SP42-5086	SP

← Tupla

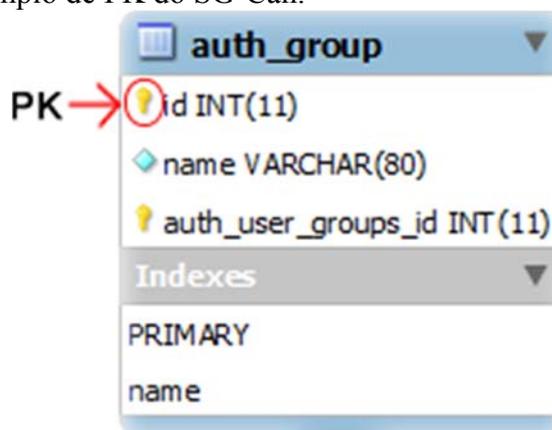
Fonte: Autor

### 2.3.4.2 Atributo Chave

Cada *tupla* de uma tabela tem que estar associada a uma chave única que permita identifica-la. Uma chave pode estar composta por um ou mais atributos. Uma chave tem que ser única dentro de sua tabela e não se pode descartar nenhum atributo da mesma para identificar uma fila. As chaves podem ser divididas, normalmente, em dois tipos:

1. Chave Primária (*Primary Key* – PK) – Atributo ou combinação de atributos que possuem a propriedade de identificar de forma única uma linha da tabela. Corresponde a um atributo determinante. Cada tabela deve incluir um campo ou conjunto de campos que identifique de forma exclusiva, cada registro armazenado na tabela (figura 12). Esses atributos são chamados de chave primária da tabela (ALVES, 2004). Desta forma, com a chave primária cria-se uma identificação única, o que dá total segurança para que aplicações possam acessar alterar e excluir dados sem correr o risco de apagar ou alterar dois campos da tabela ao mesmo tempo. Chave primária é um importante objeto quando se aplica regras de normalização de dados, muitas das formas normais são baseadas nas relações dos demais atributos com a chave primária da tabela.

**Figura 12** – Exemplo de PK do SG-Can.

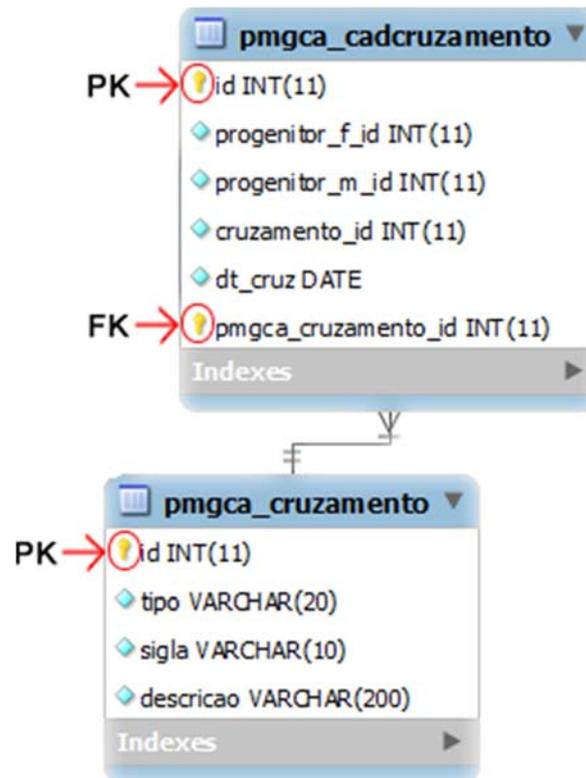


**Fonte:** Autor

2. Chave Estrangeira (*Foreign Key* – FK) – Uma chave externa ou estrangeira é um atributo ou uma combinação de atributos numa relação R2, cujos valores são necessários para equivaler à chave primária de uma relação R1 (ALVES, 2004). Uma chave estrangeira é um campo, que aponta para a chave primária de outra tabela ou da mesma tabela (figura 13). Ou seja, passa a existir uma relação entre *tuplas* de duas tabelas ou de uma única tabela. A finalidade da chave estrangeira

é garantir a integridade dos dados referenciais, pois apenas serão permitidos valores que supostamente vão aparecer na base de dados (ALVES, 2004). Esse tipo de atributo não permite exclusão, modificação e/ou inserção de dados em tabelas que estejam dependentes umas das outras, o que requer modificadores especiais. Isso também exige uma maior atenção do administrador da base de dados, quanto à própria manipulação dos dados.

Figura 13 – Exemplo de FK do SG-Can.



Fonte: Autor

### 2.3.5 Álgebra Relacional

A álgebra relacional é uma estrutura relacional com uma única relação, a relação de igualdade. Ela consiste em um conjunto de operações que tornam uma ou duas relações como entrada e produzem uma nova relação como resultado (“saída”).

As operações fundamentais na álgebra relacional são: selecionar, projetar, renomear, (unárias) – produto cartesiano, união e diferença de conjuntos (binárias).

Além das operações fundamentais, existem outras operações: interseção de conjuntos, ligação natural, dentre outras, que são definidas em termos das operações fundamentais.

### 2.3.5.1 Operação Selecionar

Seleciona *tuplas* que satisfazem um dado predicado (condição), descartando as outras. Foi padronizado para representar a seleção, a letra minúscula grega sigma ( $\sigma$ ). O predicado aparece subscrito em s. A relação argumento aparece entre parênteses seguindo o  $\sigma$ . A forma geral de uma seleção é:

$$\sigma\langle\text{condição}\rangle(\text{Relação})$$

As comparações são permitidas usando =,  $\neq$ , <,  $\leq$ , > e  $\geq$  e os conectivos e (^) e ou (v) e que envolvam apenas os atributos existentes na relação.

### 2.3.5.2 Operação Projetar

A operação projetar é uma operação unitária que retorna sua relação argumento, com certas colunas deixadas de fora. A projeção é representada pela letra grega pi ( $\pi$ ). A forma geral é:

$$\pi\langle\text{atributo da relação}\rangle(\text{Relação})$$

### 2.3.5.3 Operação Produto Cartesiano

Esta operação combina atributos (colunas) a partir de diversas relações. Trata-se de uma operação binária muito importante. Esta operação mostra todos os atributos das relações envolvidas. A forma geral é:

$$\text{Relação1 X Relação2}$$

### 2.3.5.4 Operação Renomear

A operação de renomear uma tabela é usada sempre que uma relação aparece mais de uma vez em uma consulta. É representada pela letra grega rho minúscula ( $\rho$ ). A forma geral é:

$$\rho\langle\text{novo nome}\rangle(\text{Relação})$$

Outra forma de renomear uma relação é atribuí-la a uma outra. Isto é feito com o símbolo  $\leftarrow$ .

$$\text{Relação2} \leftarrow \text{Relação1}$$

### 2.3.5.5 *Operação União (binária)*

A operação binária união é representada, como na teoria dos conjuntos pelo símbolo  $\cup$ . A forma geral é:

$$\textit{Relação1} \cup \textit{Relação2}$$

### 2.3.5.6 *Operação Diferença de Conjuntos*

A operação diferença de conjuntos permite-nos encontrar *tuplas* que estão em uma relação e não em outra. A expressão  $r-s$  resulta em uma relação que contém todas as *tuplas* que estão em  $r$  e não em  $s$ . A forma geral é:

$$\textit{Relação1} - \textit{Relação2}$$

### 2.3.5.7 *Operação Interseção de Conjuntos*

É representado pelo símbolo  $\cap$ . A forma geral é:

$$\textit{Relação1} \cap \textit{Relação2}$$

### 2.3.5.8 *Operação junção Natural*

A ligação natural é uma operação binária que permite combinar certas seleções e um produto cartesiano em uma única operação. É representada pelo símbolo  $\bowtie$ .

A operação ligação natural forma um produto cartesiano de seus dois argumentos, faz uma seleção forçando uma equidade sobre os atributos que aparecem em ambos os esquemas relação. A forma geral é:

$$\textit{Relação1} \bowtie (\textit{atributoA}, \textit{atributoB}) \textit{Relação2}$$

Que equivale a:

$$\sigma_{\textit{atributoA} = \textit{atributo}}(\textit{Relação1} \times \textit{Relação2})$$

### 2.3.5.9 *Operação Divisão*

A operação divisão, representada por  $\div$ , serve para consultas que incluem frases com “para todos”.

### 2.3.5.10 Operação de Inserção

Esta operação é usada para inserir dados em uma relação.

$$\text{Relação1} \leftarrow \text{Relação1} \cup \text{Relação2}$$

### 2.3.5.11 Remoção

A remoção é expressa da seguinte forma:

$$\text{Relação1} \leftarrow \text{Relação1} - \text{Relação2}$$

### 2.3.5.12 Operação Atualização

Em alguns casos, pode-se desejar mudar um valor em uma *tupla* sem mudar todos os valores da *tupla*. Se esta modificação for feita utilizando remoção ou inserção pode-se não ser capaz de reter os valores que não se quer modificar. Neste caso, utiliza-se o operador atualização, representado pela letra grega  $\delta$ .

A atualização é uma operação fundamental da álgebra relacional  $\delta \text{atributo} \leftarrow \text{valor}(\text{Relação})$  onde atributo deve ser um atributo de Relação e valor deve ter um domínio compatível.

## 2.3.6 A Linguagem SQL

A Linguagem SQL foi desenvolvida originalmente pela IBM para consulta ao seu Sistema-R evoluiu, vindo a tornar-se a linguagem mais usada para criação, manipulação e consultas em Sistemas Gerenciadores de Banco de Dados (SGBD) relacionais.

Tal fato levou à definição de uma SQL padrão, permitindo aos seus usuários avaliar a completude da linguagem oferecida pelos sistemas disponíveis e a identificar diferenças entre características específicas de produtos. Além disso, a padronização garante portabilidade, já que produtos que se baseiam em características padrão tendem a ser mais portáveis.

A primeira versão da SQL padrão foi desenvolvida em 1986 pela *American National Standards Institute* (ANSI) e recebeu o nome de SQL-86. Em 1989 foi lançada a SQL-89, contendo uma revisão pouco expressiva da linguagem, com a incorporação de restrições de integridade. Já em 1992, a ANSI, em conjunto com a *International Standards Organizations*

(ISO) lançou uma revisão mais significativa, ainda utilizada por muitos dos sistemas disponíveis. Em 1999 foram adicionadas algumas outras extensões, como os gatilhos (*triggers*), gerando a SQL:1999. Apesar desta ser a linguagem padrão atual, já existe uma nova versão, a SQL:2003 também chamada de SQL:200n (SILBERSCHATZ, 2006).

Um subconjunto desta linguagem forma a Linguagem de Definição de Dados (DDL) que compreende comandos básicos para a criação, a remoção e a modificação de relações.

### ***2.3.7 Restrições de integridade do modelo relacional***

A maioria das aplicações de um BD tem certas restrições de integridade que devem complementar os dados. Os SGBDs devem prover funcionalidades para a definição e a garantia dessas restrições. O tipo mais simples de restrição de integridade envolve a especificação de um tipo de dado para cada item de dado. Os tipos mais complexos de restrições podem ocorrer, com frequência, envolvendo a definição de que o registro em um arquivo deve estar relacionado aos registros de outros arquivos.

Em um BDR, normalmente existirão muitas relações e as *tuplas* dessas relações estão, em geral, relacionadas de várias maneiras. O estado do BD como um todo corresponderá aos estados de todas as relações em um determinado instante. Há, geralmente, muitas limitações ou restrições para os valores reais em um estado do BD. Essas restrições são derivadas de regras do minimundo que o BD representa.

De acordo com Elmasri (2005), as restrições em BDs podem, geralmente, ser divididas em três categorias principais:

1. Restrições inerentes baseadas em modelo – são restrições que são inerentes ao BD.
2. Restrições baseadas em esquema – são restrições que podem ser expressas diretamente no esquema do modelo de dado, normalmente por suas especificações em DLL.
3. Restrições baseadas em aplicações – são restrições que não podem ser expressas diretamente no esquema do modelo de dados e, por isso, devem ser expressas e impostas pelos programas de aplicação.

De acordo com a explanação acima, serão discutidas as principais restrições que podem ser expressas no modelo relacional – as restrições baseadas em esquema. Incluem as restrições

de domínio, restrições de chave, restrições em *null*, restrições de integridade de entidade e restrições de integridade referencial.

### **2.3.8 Restrições de Domínio**

As restrições de domínio especificam que, dentro de cada *tupla*, o valor de cada atributo A deve ser um valor atômico do domínio dom (A). Os tipos de dados associados os domínios incluem os tipos de dados numéricos padrões para inteiros (como inteiro curto, inteiro e inteiro longo) e números reais (ponto flutuante e flutuante de precisão dupla). Os caracteres, booleanos e as cadeias de caracteres de comprimento fixo e de comprimento variável também estão disponíveis, como data, hora, *timestamp* (marca do tempo) e, em alguns casos, os tipos de dado moeda. Outros domínios possíveis podem ser descritos por um subconjunto de valores de um tipo de dado ou por um tipo de dado enumerado, por meio dos quais todos os possíveis valores serão explicitamente relacionados.

### **2.3.9 Restrições de Chave e em Valores Null**

A restrição de chave serve para garantir que as *tuplas* de uma relação sejam únicas. Para isso, identifica um conjunto mínimo de atributos que devem ter valores diferentes em todas as *tuplas* de uma instância da relação [ALVES, 04]. Este conjunto de atributos denomina-se chave candidata da relação e deve satisfazer os seguintes requisitos:

1. Não podem existir 2 *tuplas* diferentes com os mesmos valores para estes atributos, ou seja, a chave identifica unicamente qualquer *tupla* da relação válida;
2. Ao retirar-se qualquer atributo componente da chave, ela deixa de identificar unicamente as *tuplas*;
3. Se o segundo requisito for violado, então a chave candidata é uma super chave.

Pela definição de relação, é sempre garantida a existência de uma chave. Entretanto, cada relação pode conter várias chaves candidatas. Cabe ao DBA escolher dentre elas aquela que será a chave primária, a ser usada pelo banco de dados em operações de otimização (ALVES, 2004). A escolha desta chave é muito importante e deve ser feita visando garantir a qualidade dos dados. A chave primária não pode assumir valor *null*.

### ***2.3.9.1 Restrições de Chave Estrangeira***

No modelo relacional é comum que o dado de uma relação esteja ligado a dados de outras relações. Se uma delas é modificada a outra também deve ser checada e modificada, se for o caso, de maneira a garantir a consistência dos dados. Para que o banco de dados possa fazer esta checagem, é especificada uma restrição envolvendo ambas as relações. Esta restrição denomina-se restrição de chave estrangeira.

A chave estrangeira é um conjunto de atributos de uma relação que é usado para fazer referência a uma *tupla* de outra relação, correspondendo à chave primária da relação referenciada (ALVES, 2004). A chave estrangeira deve conter o mesmo número de atributos da chave primária da outra relação, e seus respectivos domínios, mas não necessariamente os mesmos nomes. Além disso, diferente da chave primária, pode assumir valor *null*.

### ***2.3.10 Restrições de integridade de entidade***

A restrição de integridade de entidade estabelece que nenhum valor de chave primária pode ser *null*. Isso porque o valor da chave primária é usado para identificar as *tuplas* individuais em uma relação. Ter valores *null* para a chave primária implica não poder identificar alguma *tupla*.

### ***2.3.11 Restrições de integridade referencial***

A restrição de integridade referencial é classificada entre duas relações e é usada para manter consistência entre as *tuplas* nas duas relações. Informalmente, a restrição de integridade referencial declara que uma *tupla* em uma relação, que faz referência a outra relação, deve-se referir a uma *tupla* existente nessa relação (ALVES, 2004).

### ***2.3.12 Normalização de BDR utilizada para o desenvolvimento do SG-Can***

A normalização é um processo formal passo a passo que examina os atributos de uma entidade, com o objetivo de evitar anomalias observadas na inclusão, exclusão e alteração de *tuplas* específicas.

Esse processo causa a simplificação do atributo dentro da respectiva *tupla*, eliminando grupos repetitivos, dependências parciais de chaves concatenadas, dependências transitivas, dados redundantes e dependências multivaloradas.

Para se atingir esse estágio, é necessário que as *tuplas* sejam analisadas de forma a verificar se seus atributos apresentam relações não normalizadas, submetendo-os aos conceitos subsequentes de primeira, segunda, terceira, quarta e quinta forma normal.

### 2.3.12.1 Primeira Forma Normal (1FN)

Uma *tupla* está na primeira forma normal se é somente se o relacionamento entre a chave primária e os atributos não chaves forem unívocos, ou seja, uma relação está na 1FN se e somente se todos os domínios básicos contiverem somente valores atômicos. Aplicar a 1FN significa retirar os elementos repetitivos. Esta definição simplesmente afirma que qualquer relação normalizada está na 1FN, o que é naturalmente correto. Uma relação que esteja somente na primeira forma normal tem uma estrutura que é indispensável por uma série de razões (DATE, 2003).

Utilizando a Figura 14, que representa a tabela de genótipos do SG-Can, pode-se observar que campo *id* identifica cada *tupla* da tabela, impedindo que a entidade genótipo seja repetida. Os campos *id\_industrial*, *id\_doenca* e *id\_fisiologica* são FK de outras tabelas.

**Figura 14** – Representação da 1FN no SG-Can.

id	idustrial_id	doenca_id	fisiologica_id	genotipo	prog_f	prog_m	procedencia
1	1	1	1	RB92579	RB72454	RB867515	UCO
2	1	1	1	SP79-1011	SP11-8675	SP42-5086	SP

**Fonte:** Autor

### 2.3.12.2 Segunda Forma Normal (2FN)

Consiste em retirar da estrutura dos dados, que possuem chaves compostas, todos os dados que são funcionalmente dependentes de somente alguma parte dessa chave, ou seja, uma *tupla* está na Segunda Forma Normal 2FN se ela estiver na 1FN e todos os atributos não chave forem totalmente dependentes da chave primária (dependente de toda a chave e não apenas de parte dela). Elimina-se a dependência parcial, ou seja, os atributos não chaves devem depender integralmente a chave primária composta (DATE, 2003).

Ainda utilizando a Figura 14, é possível identificar as FK's *id\_industrial*, *id\_doenca* e *id\_fisiologica* que representam as características industriais, suscetibilidade a doenças e ainda as características fisiológicas de cada genótipo. Onde as FK's estão na 1FN em suas tabelas e na 2FN na Tabela Genótipo.

### ***2.3.12.3 Terceira Forma Normal (3FN)***

A terceira forma normal determina que não devam existir atributos com dependência funcional transitiva em uma tabela, porque podem provocar anomalias de inclusão, manutenção e deleção, ou seja, uma relação **R** está na terceira forma normal se é somente se, por todo o tempo, cada *tupla* de **R** consistir de um valor de chave primaria que identifique alguma entidade, juntamente com um conjunto de valores de atributos mutuamente independentes que descrevam aquela entidade de alguma forma (DATE, 2003). A aplicação da terceira forma Normal consiste em retirar das estruturas, os campos que são funcionalmente dependentes dos outros campos que não são chaves, ou seja, nas tabelas não pode haver campos gerados por equação de outros campos. Na Figura 14, por exemplo, é possível ver que não existe um campo equacional, ou seja, formado por uma equação de dados.

## ***2.4 Escolha do Ambiente Computacional***

### ***2.4.1 Parte conceitual***

Ao longo dos últimos quarenta e cinco anos, a evolução contínua do hardware, alavancada pelo advento dos microprocessadores no início dos anos oitenta, vem provendo proporcional avanço aos sistemas que destes dependem, particularmente sistemas que interagem com dados. Para estes sistemas em particular, sobretudo para os que acessam grandes volumes de dados sigilosos, como é o caso deste trabalho, é nítida a existência de quatro fatores implícitos, principais. Estes fatores independem do produto e/ou fabricante do SGBD, surgiram por demanda, separadamente, em épocas distintas, de forma complementar entre si e hoje, configuram conjunta e irreversivelmente os sistemas gerenciadores de bancos de dados. Na sequência de surgimento, estes fatores são:

- Integridade;
- Desempenho;

- Disponibilidade;
- Invulnerabilidade.

Os fatores integridade e desempenho estão integralmente contidos no modelo de dados. O modelo de dados é o processo através do qual, de forma ainda teórica, mensura-se as entidades do banco de dados e seus respectivos relacionamentos (ELSMASRI, 2005).

Estas entidades e seus relacionamentos traduzem as peculiaridades dos dados que armazenaram e, conseqüentemente, sua confiabilidade. Num segundo momento, mais especificamente na ocasião da implementação física do modelo de dados, as entidades tornar-se-ão tabelas e os relacionamentos, restrições ou simplesmente, *constraints* (ELSMASRI, 2005). Neste momento, tabelas e *constraints* passam a ser tratadas fisicamente, e recebem o status de objeto de banco de dados. Já os fatores disponibilidade e invulnerabilidade independem do modelo lógico de dados.

O fator disponibilidade está ligado à infraestrutura, onde cópias do banco de dados coexistem em máquinas diferentes para suprir eventual falha, física ou lógica, de alguma delas.

O fator invulnerabilidade, também ligado à infraestrutura, está mais intimamente relacionado a algoritmos complexos de transmissão de dados, que visam de forma resumida, identificar e impedir acessos indevidos aos dados.

A seguir, a abrangência destes fatores.

#### **2.4.2 Fator integridade**

Este fator é representado pelos relacionamentos e *constraints* existentes no modelo lógico de dados. É no modelo de dados, processo através do qual, projetamos as entidades, que encontramos a integridade do banco de dados como um todo, ainda que teoricamente (ELSMASRI, 2005). A implementação do modelo de dados resulta invariavelmente no banco de dados físico, propriamente dito.

O banco de dados relacional, como o utilizado neste trabalho, é composto fisicamente de tabelas que, ao se relacionarem por intermédio de regras e *constraints*, contemplam a integridade dos dados, sua unicidade. O processo através do qual a unicidade é garantida denomina-se normalização que, por sua vez, é composta de sub etapas denominadas formas normais e pertence, em um âmbito superior, ao processo denominado modelagem de dados.

Assim, aplicando as formas normais, dizemos que o modelo lógico de dados está normalizado, garantindo com isso a unicidade de seus dados e por consequência, sua integridade.

Um banco de dados possui três formas de integridade (figura 15):

1. Integridade de Domínio – Neste tipo de integridade, as validações ocorrem em cada campo da tabela (ou entidade), como se um campo deve aceitar valores *Null* ou apenas uma faixa de valores;
2. Integridade de Entidade – Já na integridade de entidade têm-se validações um nível acima, onde é preciso definir quais campos da tabela são chaves primárias (*PK*), ou seja, um campo que nunca pode se repetir.
3. Integridade Referencial – A integridade referencial é onde se cria referências a campos do tipo PK de outras tabelas, as chaves estrangeiras. Este tipo de integridade é fundamental para verificar se um dado será inserido de forma correta e, como é uma chave estrangeira, se este dado realmente existe.

**Figura 15** – Exemplo do Fator integridade.

Campo	Tipo	Collation	Nulo	Padrão	Extra
<u>id</u>	int(11)		Não		auto_increment
industrial_id	int(11)		Não		
doenca_id	int(11)		Não		
fisiologica_id	int(11)		Não		
genotipo	varchar(20)	utf8_general_ci	Não		
prog_f	varchar(20)	utf8_general_ci	Não		
prog_m	varchar(20)	utf8_general_ci	Não		
procedencia	varchar(20)	utf8_general_ci	Não		

**Fonte:** Autor

### 2.4.3 Fator desempenho

Com o modelo de dados normalizado e o banco de dados fisicamente implementado, a próxima etapa é providenciar para que o seu desempenho seja satisfatório, sem que para isso haja qualquer tipo de comprometimento à integridade dos dados, outrora garantida pela normalização. O fator que mede o desempenho de um banco de dados, relacional ou não, é denominado tempo de resposta. Tempo de resposta é o tempo medido entre o início da manipulação do dado e a respectiva efetivação desta (ELSMASRI, 2005). Este fator é relativo,

variando assim de sistema para sistema. Então, o tempo de resposta satisfatório para um sistema pode ser insatisfatório para outro. Aperfeiçoar o desempenho requer cuidados múltiplos, ora por parte da readequação do modelo de dados, desnormalizando-o se necessário, ora por parte da aplicação que acessa os dados, através da otimização das consultas, ora baseando-as em indicadores fornecidos pelo próprio SGBD tais como estatísticas de operação, adequação de índices e contenção de *Input/Output*, I/O, ora por armazenamento de consultas recentes (ELSMASRI, 2005).

Se necessário for, a princípio, este fator pode ser aperfeiçoado através da desnormalização. Outra forma nem sempre eficiente, mais dispendiosa, porém de mais simples implementação é a melhoria dos recursos de hardware através da otimização de processadores, aumento da memória física dedicada ao banco de dados, redes mais eficientes no que tange a transferência dos dados e discos magnéticos mais rápidos.

#### ***2.4.4 Fator disponibilidade***

Independente da modelagem dos dados, o fator disponibilidade resume-se em fazer com que o mesmo dado esteja disponível em mais de um local fisicamente, mas que seja tratado logicamente como único (ELSMASRI, 2005). Com isso, não se interfere na integridade e desempenho, e ao mesmo tempo se garante a disponibilidade do dado, no caso de eventual falha de um dos locais onde fisicamente ele está. Hoje em dia, esta disponibilidade é provida por sistemas gerenciadores independentes do banco de dados, responsáveis por equalizar as bases de dados em tempo real e mantê-las em funcionamento simultâneo e sincronizado.

Esta prática é conhecida como contingência e/ou redundância, caracteriza-se por seu alto custo operacional e diferencia-se de processos nativos de bancos de dados pelo discernimento entre duplicação do dado e replicação deste.

#### ***2.4.5 Fator invulnerabilidade***

Também independente da modelagem dos dados, este importante fator prima para que o dado saia de sua origem e chegue a seu destino, independentemente do sentido, sem que no trajeto haja qualquer tipo de interferência indesejada no sentido de manipular estes dados.

Para isso, algoritmos de criptografia aliados a equipamentos de segurança, como *Firewalls*, *Proxies*, redes protegidas, etc. atuam em conjunto. Alguns equipamentos providos

de algoritmos mais sofisticados, além de impedirem a interferência ao dado, mapeiam os potenciais invasores e por intermédio de técnicas heurísticas, os identificam (ELSMASRI, 2005).

#### ***2.4.6 Coexistência dos quatro fatores fundamentais***

Para que se possa efetivamente entender a importância destes fatores e, sobretudo, a necessidade da sua utilização, faz-se necessário uma breve retrospectiva utilizando-se um exemplo real: No início dos anos oitenta, grandes instituições financeiras, mais especificamente bancos privados, iniciavam campanhas de popularização de seus serviços. Surge então o conceito do Banco 24 horas, onde era possível efetuar fora do expediente bancário convencional, algumas tarefas que até então só eram possíveis serem efetuadas dentro do estabelecimento bancário, no horário convencional de atendimento. Desta forma, a possibilidade de um saldo desatualizado, por conta do uso desta facilidade, era na ocasião tão comum quanto tolerável. No final dos anos oitenta, já com a implementação de processos que refletiam a atualização de dados em tempo real, rotulados on-line e real-time, um saldo desatualizado era inadmissível, assim como também não mais se tolerava períodos muito dilatados para a conclusão de operações desta natureza (ELSMASRI, 2005).

Percebe-se claramente, através deste exemplo, a importância da coexistência destes fatores, em algumas ocasiões antagônicos entre si, mas jamais mutuamente exclusivos no contexto geral. Logo, o fato deste trabalho se engajar nos preceitos de sistemas transacionais, cuja principal característica é a manipulação de grandes volumes de dados sigilosos por um grande número de usuários, ao mesmo tempo, os fatores expostos anteriormente foram considerados, estudados e implementados, ajustando-se assim ao trabalho outrora proposto.

#### ***2.4.7 Programas Utilizados***

Os programas utilizados na elaboração deste trabalho são: o sistema operacional *Windows server 2003 Service Pack 2 (SP2)*, o pacote de aplicativo para desenvolvedores *WEB AppServ Network Version 2.2.10* que instala a linguagem de programação *PHP Script Language Version 5.2.6*, o servidor *WEB Apache Web Server Version 2.2.8* e o utilitário de *SGBD MySQL Database Version 5.0.51b*. Também foi utilizado o *Framework Django Version*

1.2, o ambiente de desenvolvimento *Eclipse SDK Version 3.6.1*, as linguagens de programação *Python 3.0* e *JAVA*.

Com o intuito de melhor explicar a metodologia aplicada, são descritas a seguir algumas características relevantes de cada um dos programas citados, as peculiaridades que justificam suas escolhas e detalhes de suas respectivas implementações.

#### **2.4.7.1 Windows server 2003 Service Pack 2 (SP2)**

O Windows Server 2003 é a mais produtiva plataforma de infraestrutura para potencializar aplicações, redes e serviços da Web conectados a partir de um grupo de trabalho até uma central de dados. Fácil de implantar, gerenciar e usar, o Windows Server 2003 ajuda você a construir uma infraestrutura segura de TI que fornece uma plataforma de aplicação poderosa para construir rapidamente soluções conectadas e uma infraestrutura de informações para comunicação aprimorada e colaboração a qualquer dia e hora (MICROSOFT, 2011).

Sua organização pode tirar proveito dos benefícios de produtividade e economia de custos, além da mais alta qualidade do Windows Server. E uma rede universal de parceiros pode ajudá-lo a definir as soluções necessárias (MICROSOFT, 2011).

O sistema operacional Windows server 2003 Service Pack 2 já estava instalado no servidor, restando apenas utilizar o sistema.

#### **2.4.7.2 AppServ Network Version 2.2.10**

É um pacote de aplicativos para desenvolvedores web, que foi inspirado no *Phanupong Panyadee (AppServ Foundation)*. É um programa gratuito, que tem como principal atrativo facilitar a instalação dos componentes: **Apache 2.2.8**, **PHP 5.2.6** e **MySQL 5.0.51b** (APPSERV, 2011).

Para melhor entendimento será feita uma breve descrição sobre cada componente utilizado.

##### **✓ Apache 2.2.8**

O servidor Apache (ou Servidor HTTP Apache, em inglês: *Apache HTTP Server*, ou simplesmente: Apache) é o mais bem sucedido servidor web livre. Foi criado em 1995 por Rob McCool, então funcionário do *National Center for Supercomputing Applications (NCSA)* da

Universidade de *Illions*. É a principal tecnologia da Apache Software Foundation, responsável por mais de uma dezena de projetos envolvendo tecnologias de transmissão via web, processamento de dados e execução de aplicativos distribuídos (APACHE, 2011).

O servidor é compatível com o protocolo HTTP. Suas funcionalidades são mantidas através de uma estrutura de módulos, permitindo inclusive que o usuário escreva seus próprios módulos - utilizando a API do software (APACHE, 2011).

O projeto Apache é um esforço coletivo de vários colaboradores, para o desenvolvimento de um software robusto, gratuito, e com qualidade, para a implementação de um servidor HTTP.

É disponibilizada em versões para os sistemas Windows, Novell Netware, OS/2 e diversos outros do padrão POSIX (*Unix, Linux, FreeBSD, etc.*) [APACHE, 11].

#### ✓ **PHP Script Language Version 5.2.6**

O engenheiro de software Rasmus Lerdorf, membro da equipe apache é o criador e a força motriz original por trás do PHP. Criado na década de 90, o *Hypertext Preprocessor* (PHP) é uma das mais poderosas linguagens para o desenvolvimento de aplicações na Web. Conquistou espaço devido à facilidade de utilização e a grande versatilidade de recursos que possui. O produto foi originalmente chamado de “*Personal Home Page Tools*”; mas como se expandiu em escopo, um novo nome e mais apropriado foi escolhido por votação da comunidade. O objetivo principal é permitir aos desenvolvedores escrever páginas que serão geradas de maneira dinâmica e rápida (MINORELLO, 2007).

Com uma linguagem inteiramente voltada para a Internet, o PHP possibilita o desenvolvimento de scripts dinâmicos para sites. Os sites criados em HTML ganham interatividade com as técnicas de programação que essa linguagem oferece. Trata-se de uma linguagem extremamente modular, o que a torna ideal para instalação e uso em servidores Web (MINORELLO, 2007).

Atualmente o PHP está na versão 5, chamado de PHP5 ou simplesmente PHP. O PHP tem pouca relação com layout, eventos ou qualquer coisa relacionada a aparência de uma página Web, ou seja, a maior parte do que o PHP realiza é invisível para o usuário final. O resultado final do PHP é HTML. Visualizando uma página de PHP não será capaz de dizer que não foi escrita em HTML.

### ✓ MySQL 5.0.51b

O MySQL é um dos sistemas de gerenciamento de banco de dados mais populares que existe e, por ser otimizado para aplicações Web, é amplamente utilizado na internet. É muito comum encontrar serviços de hospedagem de sites que oferecem o MySQL e a linguagem PHP, justamente porque ambos trabalham muito bem em conjunto (MINORELLO, 2007).

Outro fator que ajuda na popularidade do MySQL é sua disponibilidade para praticamente qualquer sistema operacional, como *Linux*, *FreeBSD* (e outros sistemas baseados em *Unix*), *Windows* e *Mac OS X*. Além disso, o MySQL é um software livre (sob licença GPL), o que significa que qualquer um pode estudá-lo ou alterá-lo conforme a necessidade (DEV, 2011) (MINORELLO, 2007).

O MySQL surgiu na Suécia através de um desenvolvimento em conjunto de três colegas: Allan Larsson, David Axmark e Michael Monty Widenius. Trabalhando com base de dados, eles sentiram a necessidade de fazer determinadas conexões entre tabelas e usaram o mSQL para isso. Porém, não demorou para perceberem que essa ferramenta não lhes atendia conforme o necessário e passaram a trabalhar em uma solução própria. Surgia então o MySQL, cuja primeira versão foi lançada no ano de 1996 (DEV, 2011) (MINORELLO, 2007).

Entre as características técnicas do SGBD MySQL, estão:

1. Alta compatibilidade com linguagens como PHP, Java, Python, C#, Ruby e C/C++;
2. Baixa exigência de processamento (em comparação como outros SGBD);
3. Vários sistemas de armazenamento de dados (*database engine*), como *MyISAM*, *MySQL Cluster*, *CSV*, *Merge*, *InnoDB*, entre outros;
4. Recursos como *transactions* (transações), conectividade segura, indexação de campos de texto, replicação, etc.;
5. Instruções em SQL, como indica o nome.

Por que Utilizar MySQL em Relação a outros SGBDs?

O MySQL é focado na agilidade. Assim, se a aplicação necessita de retornos rápidos e não envolve operações complexas, o MySQL é a opção mais adequada, pois é otimizado para proporcionar processamento rápido dos dados e tempo curto de resposta sem exigir muito do hardware.

### 2.4.7.3 Framework Django Version 1.2

Django é um *framework web* de alto nível escrito em *Python* que estimula o desenvolvimento rápido e limpo. Desenvolvido e utilizado por mais de dois anos em uma operação ágil de notícias online, Django foi projetado para lidar com dois desafios: os prazos apertados de uma redação e os requisitos rígidos dos experientes desenvolvedores web que o criaram. Ele permite a construção rápida de aplicações web de alto desempenho e elegância. Django se concentra no máximo de automatização possível e adere ao princípio *Don't Repeat Yourself* (DRY) (DJANGO, 2011).

Características do framework Django

1. Mapeador objeto-relacional (ORM);
2. Interface de administração automática;
3. Formulários;
4. *URL's* elegantes;
5. Sistema de *templates*;
6. Sistema de cache;
7. Internacionalização.

### 2.4.7.4 Eclipse SDK Version 3.6.1

Eclipse é uma IDE desenvolvida em Java, com código aberto para a construção de programas de computador. O projeto Eclipse foi iniciado na IBM que desenvolveu a primeira versão do produto e doou-o como software livre para a comunidade. Hoje, o Eclipse é a IDE Java mais utilizada no mundo. Possui como características marcantes o uso da SWT e não do Swing como biblioteca gráfica, a forte orientação ao desenvolvimento baseado em plug-ins e o amplo suporte ao desenvolvedor com centenas de plug-ins que procuram atender as diferentes necessidades de diferentes programadores (ECLIPSE, 2011).

Com o uso de *plugins*, pode ser usado não só para desenvolver em Java, mas também em C/C++, *PHP*, *ColdFusion* e até mesmo *Python* (ECLIPSE, 2011).

### 2.4.7.5 Python 3.0

Python é uma linguagem de programação de alto nível, interpretada, imperativa, orientada a objetos, de tipagem dinâmica e forte. Foi lançada por Guido Van Rossum em 1991. Atualmente possui um modelo de desenvolvimento comunitário, aberto e gerenciado pela organização sem fins lucrativos *Python Software Foundation*. Apesar de várias partes da linguagem possuírem padrões e especificações formais, a linguagem como um todo não é formalmente especificada. O padrão de facto é a implementação *CPython* (PYTHON, 2011).

A linguagem foi projetada com a filosofia de enfatizar a importância do esforço do programador sobre o esforço computacional. Prioriza a legibilidade do código sobre a velocidade ou expressividade. Combina uma sintaxe concisa e clara com os recursos poderosos de sua biblioteca padrão e por módulos e frameworks desenvolvidos por terceiros (PYTHON, 2011).

Python foi concebido no final de 1989 por Guido Van Rossum no Instituto de Pesquisa Nacional para Matemática e Ciência da Computação (CWI), nos Países Baixos, como um sucessor da ABC capaz de tratar exceções e prover interface com o sistema operacional AMOEBA através de scripts. Também da CWI, a linguagem ABC era mais produtiva que C, ainda que com o custo do desempenho em tempo de execução. Mas ela não possuía funcionalidades importantes para a interação com o sistema operacional, uma necessidade do grupo. Um dos focos primordiais de Python era aumentar a produtividade do programador (PYTHON, 2011).

Em 1991, Guido publicou o código (nomeada versão 0.9.0) no grupo de discussão *alt.sources*. Nessa versão já estavam presentes classes com herança, tratamento de exceções, funções e os tipos de dado nativos *list*, *dict*, *str*, e assim por diante. Também estava presente nessa versão um sistema de módulos emprestado do Modula-3. O modelo de exceções também lembrava muito o do Modula-3, com a adição da opção *else clause*. Em 1994 foi formado o principal fórum de discussão do Python, *comp.lang.python*, um marco para o crescimento da base de usuários da linguagem (PYTHON, 2011).

A versão 1.0 foi lançada em janeiro de 1994. Novas funcionalidades incluíam ferramentas para programação funcional como *lambda*, *map*, *filter* e *reduce*. A última versão enquanto Guido estava na CWI foi o Python 1.2. Em 1995, ele continuou o trabalho no CNRI em *Reston*, Estados Unidos, de onde lançou diversas versões. Na versão 1.4 a linguagem ganhou parâmetros nomeados (a capacidade de passar parâmetro pelo nome e não pela posição na lista

de parâmetros) e suporte nativo a números complexos, assim como uma forma de encapsulamento (PYTHON, 2011).

Uma grande inovação da versão 2.2 foi à unificação dos tipos Python (escritos em C) e classes (escritas em Python) em somente uma hierarquia. Isto tornou o modelo de objetos do Python consistentemente orientado a objeto. Também foi adicionado *generator*, inspirado em *Icon* (PYTHON, 2011).

A terceira versão da linguagem foi lançada em dezembro de 2008, chamada Python 3.0 ou Python 3000. Com noticiado desde antes de seu lançamento, houve quebra de compatibilidade com a família 2.x para corrigir falhas que foram descobertas neste padrão, e para limpar os excessos das versões anteriores. A primeira versão alfa foi lançada em 31 de agosto de 2007, a segunda em 7 de dezembro do mesmo ano (PYTHON, 2011).

#### **2.4.7.6 JAVA (*Javascript*)**

JAVA é uma linguagem de programação compilável, orientada a objeto, bastante versátil e, por suas características, altamente recomendada para aplicativos que acessam bases relacionais transacionais de grande volume de dados. Para este trabalho, seu principal atrativo é a facilidade de reprogramação do código e, sobretudo a conveniência no que se refere ao desempenho, visto que por ser uma linguagem intrinsecamente não performática, além de não depender da plataforma onde está sendo executado, o desempenho da aplicação depende diretamente do desempenho do banco de dados (POLETTI, 2008).

### **3.0 BANCO DE GERMOPLASMA DA CANA-DE-AÇÚCAR**

#### **3.1 *Introdução***

O setor sucroalcooleiro cada vez mais se moderniza em busca de maiores produtividades, porém vários são os fatores que interferem na produção, tais como as condições ambientais, as variedades cultivadas e seus manejos, pragas, doenças, nutrição, florescimento, entre outros.

Variedades melhoradas apresentam acentuada contribuição para a viabilização econômica da cultura. Elas são obtidas pelos programas de melhoramento genético, através de diversas etapas de pesquisa, desde a realização de cruzamentos, produção de plântulas, seleção, experimentação, entre outras, até ser liberada para o plantio comercial (MARTINS, 2005).

Um bom banco de germoplasma é fator fundamental para o desenvolvimento de novas variedades. A partir deste banco torna-se possível cruzar variedades, mais adaptadas às adversidades edafoclimáticas, e desta forma, aumenta-se a chance de obtenção de uma nova variedade mais adaptada.

Devidos a estes fatores, além da importância econômica da cana-de-açúcar para o Brasil, se faz necessário um sistema de banco de dados que seja capaz de armazenar, processar e responder as questões pertinentes ao germoplasma cana. Este banco de dados deve ser uma ferramenta que auxilie nas rotinas do processo de produção, desenvolvimento, seleção e liberação de uma nova variedade de cana.

#### **3.2 *O Programa de Melhoramento Genético da Cana-de-Açúcar (PMGCA)***

No Brasil o melhoramento genético da cana-de-açúcar surgiu no século XX (início da década de trinta), através do Instituto Agrônomo de Campinas (IAC) e da Estação Experimental de Campos – RJ, que contribuíram com novas variedades (CB e IAC) de forma expressiva para o desempenho da cultura no período, com destaque para as variedades CB41-76 e CB45-3. Ao final da década de sessenta tiveram o surgimento de novos programas de melhoramento genético, o do PLANALSUCAR, responsável pela obtenção de genótipos de sigla RB (República do Brasil) e o da COPERSUCAR, de genótipos de sigla SP. Com a extinção do PLANALSUCAR em 1990, as Universidades Federais de Alagoas (UFAL), Rural

de Pernambuco (UFRPE), Viçosa – MG (UFV), Rural do Rio de Janeiro (UFRRJ), São Carlos – SP (UFSCAR), Goiás (UFGO) e Paraná (UFPR) deram continuidade ao programa para a obtenção de variedades RB, sendo conduzida em rede nacional pela RIDESA – Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro (RAMOS, 2008).

Para que um programa de melhoramento seja eficiente, um dos principais requisitos é possuir um banco de germoplasma localizado em ambiente favorável ao florescimento profuso. Relata-se que as condições de florescimento natural da cana-de-açúcar para a produção de sementes, estão em áreas próximas do litoral, entre 8° e 10° de latitude, com elevada umidade relativa, temperaturas mínimas não inferiores a 18° C e máximas não superiores a 32° C (SILVA, 2010).

São poucos os lugares privilegiados para que ocorra o florescimento natural da cana-de-açúcar e entre esses lugares, destaca-se a Serra do Ouro, uma base de pesquisa do programa de melhoramento genético da cana-de-açúcar (PMGCA), gerenciada pelo Centro de Ciências Agrárias (CECA) da Universidade Federal de Alagoas (UFAL). É a base responsável por manter mais de 2600 genótipos de cana-de-açúcar que são utilizados em cruzamentos para obtenção de novas variedades de cana-de-açúcar da sigla RB, atendendo a RIDESA.

### **3.3 RIDESA**

Formada por Universidades Federais - Ministério da Educação e Desporto foi criada, com a finalidade de incorporar as atividades do extinto PLANALSUCAR, e dar continuidade ao desenvolvimento de pesquisas visando à melhoria da produtividade do setor sucroalcooleiro (RIDESA, 2011).

A RIDESA foi inicialmente instituída por meio de convênio firmado entre sete Universidades Federais (UFPR, UFSCar, UFV, UFRRJ, UFS, UFAL e UFRPE) que estavam localizadas nas áreas de atuação das Coordenadorias do ex-PLANALSUCAR, do qual foi absorvido o corpo técnico e a infraestrutura das sedes das coordenadorias e estações experimentais. Com o apoio de parte significativa do Setor Sucroalcooleiro, por meio de convênio, a rede começou a desempenhar suas funções em 1991, aproveitando a capacitação dos pesquisadores e as bases regionais do ex-PLANALSUCAR, aos quais se juntaram professores das universidades. Tem como base para o desenvolvimento da pesquisa 31 estações experimentais estrategicamente localizadas nos Estados onde a cultura da cana-de-açúcar apresenta maior expressão. Além dessas estações experimentais, a rede também desenvolve

pesquisa nos campus das sete universidades federais, envolvendo, principalmente, pesquisas conduzidas nos diferentes cursos de pós-graduação, em nível de mestrado e doutorado (RIDESA, 2011).

Ao longo dos anos de atuação, as universidades federais deram maior ênfase à manutenção e continuidade da pesquisa relacionada ao Programa de Melhoramento Genético da Cana-de-açúcar (PMGCA), que continuou a utilizar a sigla RB para identificar seus cultivares. Atualmente, os cultivares de sigla RB, estão sendo cultivados em mais de 50% da área cultivada com cana-de-açúcar no país, chegando em algumas regiões a representar até 70%. Isto significa que a RIDESA apresentou grande eficiência nestes últimos anos, dando incalculável retorno aos investimentos aplicados em pesquisa (BARBOSA, 2008).

Atualmente, a RIDESA é composta por dez Universidades Federais: UFAL, UFRPE, UFS, UFV-MG, UFRRJ, UFSCar-SP, UFG, UFMT e UFPI. A RIDESA é responsável por desenvolver as variedades da sigla RB (Republica do Brasil) que representam uma expressiva porcentagem de área cultivada com cana-de-açúcar no Brasil. Com quase quatro décadas de existência a RIDESA já liberou mais de 65 variedades RB, oriundas de hibridações na Serra do Ouro, que atualmente representam mais de 50% das canas moídas pelas agroindústrias brasileiras, chegando a 70% em algumas regiões (BARBOSA, 2008).

Com os objetivos da RIDESA alcançados, nesta nova era a proposta é de fortalecimento, com a agregação da Universidade Federal de Goiás-UFG a partir de 2004 e da Universidade Federal de Mato Grosso em 2007. Desta forma surgem três novas estações experimentais para o CERRADO, uma localizada em Goiânia-GO, da UFG e outra em Cuiabá-MT da UFMT e uma terceira em Capinópolis - MG da UFV, sendo que nesta última os trabalhos se iniciaram em 2003 (RIDESA, 2011).

### ***3.4 O banco de Germoplasma da RIDESA***

Os cruzamentos entre genótipos de cana-de-açúcar da RIDESA são feitos no banco de Germoplasma de cana da Serra do Ouro, em Murici, AL, sob a responsabilidade da equipe de fitomelhoristas da cana-de-açúcar da UFAL, com a cotização de recursos e trabalho das demais universidades.

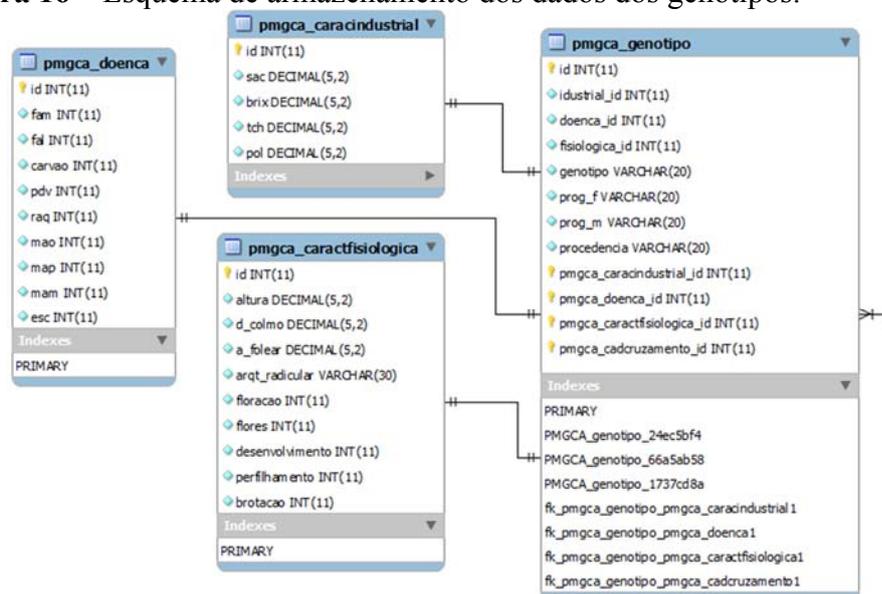
Um dos fatores que dificultam a realização de certos cruzamentos é a dificuldade de sincronização do florescimento dos acessos do banco de germoplasma. O florescimento na Serra do Ouro geralmente inicia no mês de janeiro e se estende até final de junho, e os diversos

genótipos florescem em épocas diferentes, impossibilitando a realização de alguns cruzamentos desejáveis (BARBOSA, 2008). Avaliação do período de florescimento de duzentos genótipos na Serra do Ouro, entre os anos de 1998 a 2004, indicou que 7% florescem entre 8 e 30 de abril, 56,5% florescem entre os dias 1 e 12 de maio, 34,5 % florescem entre os dias 13 e 26 de maio e 1,5 % florescem após 26 de maio. As hibridações na Serra do Ouro começam em abril e se estende até a terceira semana de junho (BARBOSA, 2002) (BARBOSA, 2008) (RAMOS, 2008).

Atualmente o banco de Germoplasma da Serra do Ouro possui mais de 2600 acessos de genótipos de cana-de-açúcar, entre cultivares, clones, outras espécies relacionadas ao gênero *Saccharum* e cultivares importadas das diferentes regiões canavieiras do mundo.

Os dados das variedades de cana-de-açúcar estão dispostos em diversas tabelas no SG-Can, porem as principais tabelas de armazenamento de dados dos genótipos é a tabela *pmgca\_genotipo*, que é dependente das tabelas *pmgca\_caracindustrial*, *pmgca\_caractfisiologica* e a *pmgca\_doenca*, que fornecem os dados básicos para os genótipos armazenados, cujas tabelas estão visíveis na Figura 16.

**Figura 16** – Esquema de armazenamento dos dados dos genótipos.



**Fonte:** Autor

A tabela *pmgca\_genotipo* também fornece todas as informações para a realização dos cruzamentos, pois todos os procedimentos de produção e seleção de uma nova variedade necessita das informações das progênies (Mãe e Pai).

Para melhor explicar a utilização de um banco de germoplasma, bem como a obtenção de uma nova variedade, será descrito todo o processo de obtenção, pois assim ficará claro os

dados que são gerados ao longo dos anos de pesquisa, bem como ideias para futuras interações no SG-can.

### **3.5 Processos para obtenção de novas Variedades de Cana-de-Açúcar.**

#### **3.5.1 Quarentena**

Nos programas de melhoramento genético, a variabilidade genética é indispensável para o seu sucesso. Para a obtenção dessa vasta gama de genótipos, é necessária a introdução de variedades estrangeiras, não só de outras regiões do país, mas também de outros países. Essa introdução não pode ser de imediato nos canaviais, visto que essas variedades podem representar um alto risco para o setor canavieiro se, junto com elas, trouxerem pragas e doenças (MARTINS, 2005).

Para evitar esse problema, o Programa de Melhoramento Genético da Cana-de-açúcar da UFAL possui uma Estação de Quarentena, localizada no bairro de Bebedouro, no município de Maceió, onde as plantas oriundas de outros locais ali permanecem por um determinado período para se verificar a presença ou não de pragas e doenças nas mesmas. Não havendo esse tipo de ocorrência, os genótipos são liberados para o início do processo de melhoramento genético.

A etapa de Hibridação, que marca o início das atividades, compreende os meses de abril a junho e ocorre na Serra do Ouro onde lá são feitos cruzamentos denominados de Biparentais e Múltiplos.

#### **3.5.2 Cruzamentos**

Na Serra do Ouro, os cruzamentos iniciam-se em abril e vão até meados de junho. O processo de hibridação tem início com os cruzamentos manuais, que consistem na realização de hibridações cadastradas ou não ou com diferentes números de panículas (BARBOSA, 2002) (BARBOSA, 2008).

Os dados provenientes das hibridações são processados no Jornada de Cruzamento, que foi desenvolvido no Access um aplicativo do pacote Microsoft Office. Desta forma, os pesquisadores das nove bases, todos os anos devem vir para a Serra do Ouro, para acompanhar os cruzamentos.

As hibridações praticadas na Serra do Ouro consistem praticamente de três tipos de cruzamentos: a) biparental (BP); b) policruzamento ou múltiplo (MP) e c) múltiplo específico (MPE), porém existem ainda as Autofecundações e o Teste de Progênie.

#### a) Cruzamentos Biparentais (BP)

Essa técnica consiste no acasalamento de duas variedades de cana-de-açúcar (A x B) geralmente escolhidas previamente pelo conhecimento de suas características; a variedade “A” é feminina, não produtora de grãos de pólen férteis e a variedade “B” é masculina, produtora de pólen fértil. Cada planta é nutrida por uma solução ácida nutritiva, que consiste em uma mistura de água pura com dióxido de enxofre ( $\text{SO}_2 - 150 \text{ mg.kg}^{-1}$ ), ácido fosfórico ( $\text{H}_3\text{PO}_4 - 75 \text{ mg.kg}^{-1}$ ), ácido nítrico ( $\text{HNO}_3 - 37 \text{ mg.kg}^{-1}$ ) e ácido sulfúrico ( $\text{H}_2\text{SO}_4 - 37 \text{ mg.kg}^{-1}$ ) (MARTINS, 2005).

O cruzamento BP é realizado em campânulas de polietileno para se obter o devido isolamento, evitando a penetração de grãos de pólen estranhos que podem ser levados por insetos ou pelo vento (figura 17). Nesta técnica, são agrupadas seis panículas, três para cada um dos genótipos, no interior de campânulas posicionadas na mata ou em estruturas de concreto em ambiente aberto onde cada genótipo recebe uma etiqueta identificadora. Este tipo de cruzamento é o mais utilizado pelos melhoristas nas hibridações atuais, e permite conhecer ambos os progenitores (masculino e feminino).

**Figura 17** – Exemplo de Cruzamento BP.



Fonte: PMGCA

### b) Cruzamentos Múltiplos (MP)

O policruzamento de cana-de-açúcar foi desenvolvido no Haváí com o objetivo de avaliar grande número de progenitores ao menor custo (BARBOSA, 2002). O método consiste em reunir um grande número de panículas de diferentes progenitores, para que se inter cruzem, sendo reconhecido apenas o progenitor feminino (figura 18) (HEINZ, 1987).

**Figura 18** – Exemplo de Cruzamento MP



**Fonte:** PMGCA

Este cruzamento, também conhecido como *Melting pot* ou policruzamento. É realizado na Serra do Ouro através do agrupamento de cinco ou mais diferentes progenitores dispostos num estaleiro de 0,60 m de largura por 5 m de comprimento, oferecendo uma área para cruzamento que comportará 24 baldes, 12 de cada lado. Os estaleiros estão situados em galpões, que possuem capacidade suficiente para comportar 100 ou mais baldes, cada um deles contendo de 5 a 6 inflorescências e identificados por etiquetas, conhecendo-se assim apenas o progenitor feminino ( $A \times ?$ ) que dá origem à semente. As inflorescências são colocadas ao acaso, cada uma atuando como doadora e receptora de pólen. Periodicamente, são trocadas de posição para se aumentar a casualidade e a diversidade nos policruzamentos. Neste tipo de hibridação são coletadas todas as panículas após o amadurecimento. Este cruzamento é realizado no mesmo ambiente de maturação das sementes. Através deste método pode-se avaliar mais rapidamente a capacidade geral de combinação dos progenitores (MATSUOKA, 1999).

### c) Cruzamentos Múltiplos Específicos (MPE)

O cruzamento múltiplo pode ser também realizado reunindo uma série de variedades que apresentam uma característica específica tais como: resistência ao mosaico, carvão, ferrugem, etc., recebendo o nome de cruzamento múltiplo específico.

O policruzamento específico, desenvolvido pelos havaianos, consiste em polinizar várias panículas femininas, comprovadamente macho-estéreis, com apenas um progenitor masculino. É realizado em áreas isoladas, geralmente em caminhos no interior de mata ou em grandes lanternas (BARBOSA, 2002). Na Serra do Ouro, este tipo de cruzamento é realizado isolado na mata por meio de abrigos sobre as panículas também identificadas com etiquetas (figura 19). O progenitor masculino é marcado com tinta e sua panícula é descartada após o período de hibridação, sendo apenas coletadas as panículas dos progenitores femininos.

**Figura 19** – Exemplo de Cruzamento MPE

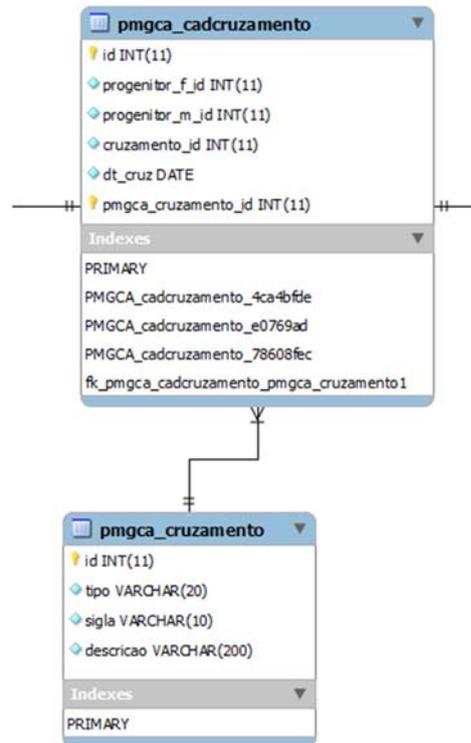


**Fonte:** PMGCA

Terminados os cruzamentos, e obtidas às sementes de cana-de-açúcar ainda aderidas às panículas, segue-se a obtenção de plântulas e, posteriormente, o processo seletivo para se chegar a uma ou mais variedades comerciais. Particularmente, neste programa, as sementes obtidas em cruzamentos na EFCSO, são levadas ao Centro de Ciências Agrárias da UFAL em Rio Largo – Alagoas, para sofrerem um processo de beneficiamento (secagem, armazenamento, etc.), registro e envio às demais universidades que fazem parte do programa.

A figura 20 representa as tabelas onde serão armazenados todos os tipos cruzamento que foram vistos nesta seção, bem como os dados referentes a estes cruzamentos. Na tabela *pmgca\_cruzamento* são armazenados os tipos de cruzamento (tipos de hibridações) possíveis, na tabela *pmgca\_cadcruzamento* são armazenados os dados dos cruzamentos realizados e seus participantes, indexando também a estes cruzamentos os dados já conhecidos no sistema, como é possível verificar nos indexes da tabela *pmgca\_cadcruzamento*.

**Figura 20** – Esquema das tabelas de cruzamento



**Fonte:** Autor

### 3.5.3 Secagem e Deslintamento

A secagem é feita em uma estufa com uma temperatura de 34°C e uma Umidade Relativa (UR) de 55%, por um período de 12 a 24 horas. Porém, se a semente vier com umidade excessiva, o período pode se estender para 48 horas. Dessa forma, perdem o excesso de umidade, havendo assim, condições propícias para um armazenamento prolongado, sem perdas acentuadas de seu vigor e porcentagem de germinação.

O deslintamento é o ato de separar a semente dos materiais inertes (palha, estigma, entre outros) que a envolvem para evitar que, num semeio, os mesmos prejudiquem a germinação da

semente ou mascarem o teste de germinação, importante para que se determine a porcentagem de plântulas por grama utilizadas no semeio.

### 3.5.4 Semeio

Esta operação é executada espalhando as sementes em toda a extensão da caixa o mais uniformemente possível, logo após, a superfície da caixa é molhada com um jato fino de água, batendo de leve com a mão sobre as sementes para promover um maior contato das mesmas com o solo conforme é demonstrado na figura 21. Após o semeio, as caixas são cobertas com um lençol de plástico a fim de melhorar as condições de temperatura e umidade, necessárias à germinação. Normalmente utilizam-se 3g de semente por caixa, podendo essa quantidade variar, dependendo da porcentagem de germinação, pois se sabe que fatores como fertilidade da variedade mãe, amadurecimento e fitossanidade das panículas, secagem e armazenamento das sementes, influem muito na germinação.

**Figura 21** – Exemplo de Semeio

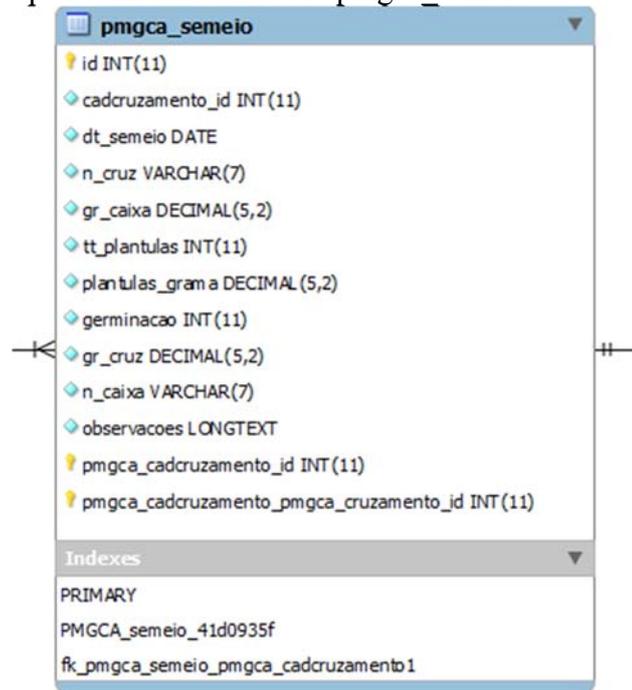


**Fonte:** PMGCA

A ação do semeio no SG-Can só será possível após o cadastramento dos cruzamentos realizados no ano corrente. Após esta digitação (importação) dos dados o sistema liberará a inserção de novos dados na tabela de semeio.

A figura 22 demonstra como a tabela *pmgca\_semeio* está organizada, os dados e os tipos de atributo que são armazenados. Nesta figura também é possível visualizar os relacionamentos entre esta tabela e as demais tabelas do sistema.

**Figura 22** – Esquema interno da tabela pmgca\_semeio



**Fonte:** PMGCA

Os dados que serão armazenados nesta tabela estão ligados diretamente a tabela de cruzamento, como é possível visualizar nos indexes da tabela a chave estrangeira FK\_pmgca\_semeio\_pmgca\_cadcruzamento1 que liga esta tabela a tabela de cruzamento, onde estarão armazenados os dados dos cruzamentos da Serra do Ouro.

### 3.5.5 Repicagem

Ao atingirem 5-6 cm, as plântulas são repicadas individualmente (com etiquetas correspondentes a seus progenitores) para outras caixas e levadas à área de adaptação, até serem transplantadas para o campo, o que ocorre com 45 a 60 dias após a repicagem (figura 23).

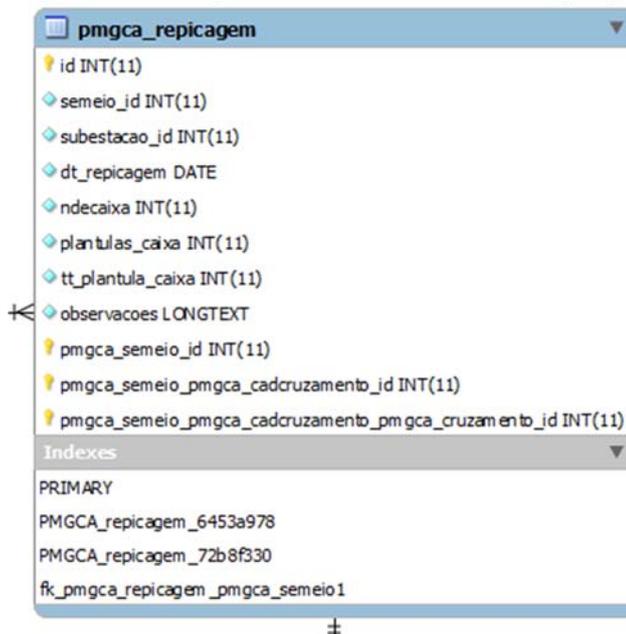
**Figura 23** – Exemplo de repicagem e aclimação das plântulas.



**Fonte:** PMGCA

Após a liberação do semeio as plântulas serão distribuídas em copos, tubetes ou até caixas, para um melhor desenvolvimento e uma futura separação para envio. Nesta repicagem será informado no SG-Can a data da repicagem, a subestação que irá receber esta (s) plântula (s), a quantidade de caixas, a quantidade de plântulas que irão em cada caixa e se possui alguma observação referente as plântulas, como demonstra a figura 24. Estes dados serão armazenados na tabela pmgca\_repicagem.

**Figura 24** – Esquema interno da tabela de pmgca\_repicagem.



**Fonte:** Autor

### 3.5.6 Plantio da fase T1

A primeira fase de seleção é denominada T1. Normalmente são emitidas, em torno de, 400.000 plântulas por ano. Essa quantidade pode variar, para mais ou para menos, de um ano para outro, essa variação dependerá da quantidade de cruzamentos realizados na Serra do Ouro. O plantio tem sido feito com plantas individuais, transplantadas em espaçamento de 0,50 m entre plantas, em sulcos convencionais distanciados entre si em 1,30 m e com 11,50 m de comprimento (figura 25). Em cada sulco tem-se, no máximo, 24 plantas. Atualmente esses campos de T1 são instalados em subestações que fazem parte do programa de melhoramento, que estão espalhadas por todo Estado de Alagoas, e ainda em algumas subestações em outros estados. Após este procedimento a plântula recebe o nome de clone.

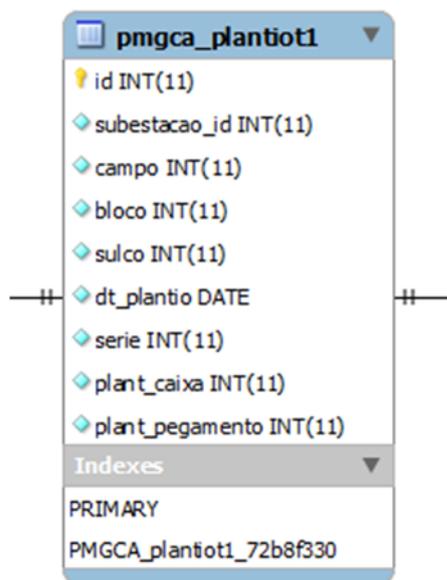
**Figura 25** – Exemplo de plantio de T1.



**Fonte:** PMGCA.

Após a liberação da remessa no SG-Can o sistema libera a ação de plantio de T1, esta liberação será a responsável por conectar as informações das subestações (dados de campo) aos dados do genótipo que já estão armazenados no SG-Can. A formulação desta tabela é demonstrada na Figura 26 e esta tabela recebe nome de `pmgca_plantiot1`.

**Figura 26** – Esquema interno da tabela pmgca\_plantiot1.



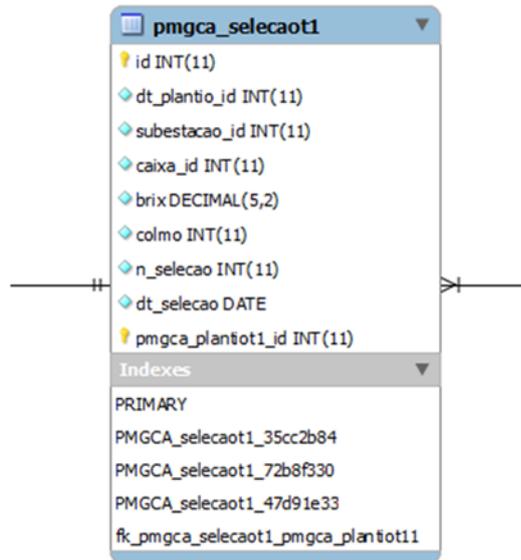
**Fonte:** Autor.

Após o plantio, estes campos de T1 serão avaliados de duas formas:

1. Avaliação de T1 Cana-planta – Após um ano do plantio é feita uma avaliação onde são observadas características como o desenvolvimento, morfologia, ocorrência de florescimento e incidência de doenças, ocamento, enraizamento aéreo, rachadura, tamanho das gemas. Nos anos de verão seco, é feito o controle da lagarta *elasmo* (*Elasmopalpus lignosellus*). E ainda dados como: BRIX, quantidade de colmo, porém atualmente, está sendo analisado, em alguns campos, apenas a quantidade de colmos, ou seja, o BRIX não é obrigatório. Após este procedimento de seleção o clone recebe uma numeração que o identificara durante os estágios restantes.
2. Avaliação de T1 em Soca – São feitos os mesmos procedimentos da seleção em cana-planta, a diferença é que esta é feita em no segundo ano do plantio, onde são avaliadas várias características como o chochamento, incidência de pragas, Brix, morfologia, incidência de doenças, ocamento, enraizamento aéreo, diâmetro do colmo, rachadura e tamanho das gemas. Sendo obrigatório o BRIX dos clones e ainda a quantidade de colmos.

A seleção de T1 só será liberada após a inserção dos dados de plantio de T1. A tabela que armazena os dados desta seleção é a tabela pmgca\_selecaot1 que está ilustrada de acordo com a Figura 27, que ilustra o esquema interno da tabela pmgca\_selecaot1. Esta seleção pode ser por meio do BRIX ou do número de colmos que um clone possui.

**Figura 27** – Esquema interno da tabela de pmgca\_selecaot1.



**Fonte:** Autor.

### 3.5.7 *Plantio da fase T2*

Os clones selecionados em T1 (cana-planta ou soca) são plantados no esquema 2 x 3,5 m (dois sulcos de três metros e meio cada) utilizando o espaçamento da unidade produtora correspondente ao campo de instalação. Cada clone recebe uma numeração de ordem crescente para identificar os progenitores do material. São plantadas também variedades padrões (variedades já conhecidas) cujas características se deseja superar (figura 28).

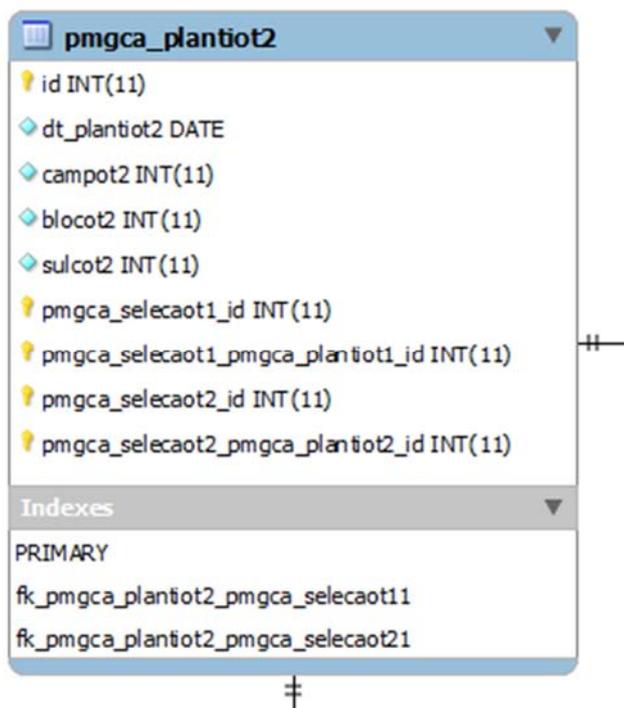
**Figura 28** – Exemplo de plantio de T2



**Fonte:** PMGCA.

Como todas as informações no SG-Can estão relacionadas, o sistema só liberará a área de plantio de T2 quando forem informados os dados de seleção de T1 do referido local e/ou campo, estes dados são armazenados na tabela pmca\_plantiot2. A Figura 29 demonstra como os dados estão organizados e a dependência das ligações com as tabelas antecessoras.

**Figura 29** – Esquema interno da tabela de pmgca\_plantiot2.



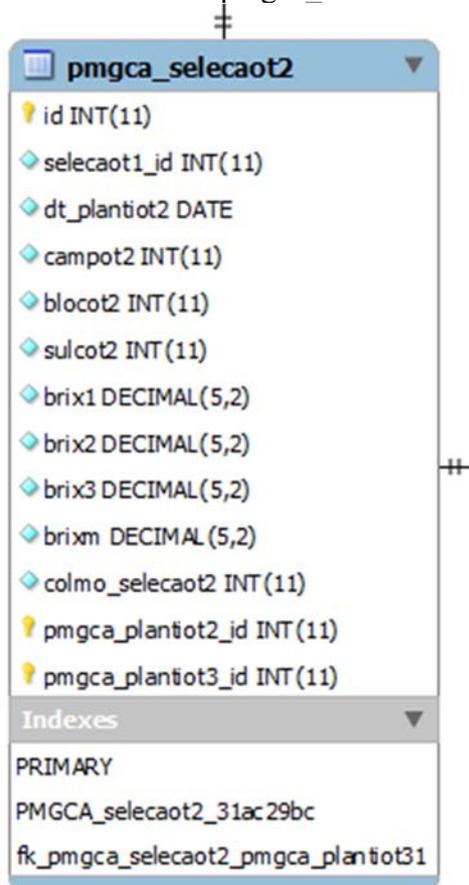
**Fonte:** Autor.

### 3.5.8 Seleção de T2

Na seleção avaliam-se morfologia, incidência de pragas e doenças, chochamento, diâmetro do colmo, observando também se há isoporização. Aos 60 dias, é feita a avaliação de brotação; aos 120 dias, são observados o perfilho e o desenvolvimento e, entre 210 e 240 dias avaliam-se o número de colmos e porcentagem de florescimento além de características como tamanho das gemas e ocamento. Nessa fase verifica-se o kg de Brix por parcela, onde faz-se a pesagem de 30 colmos adultos, extrapolando-se o valor obtido para o total de colmos da parcela. Tomando-se o resultado de Brix, é finalmente calculado o valor de kg de Brix por parcela. Como a produção por área é resultado da conjugação do número de colmos numa área determinada com o peso unitário por colmo, a produção em kg de Brix por parcela, por conseguinte, é o produto do Brix por aquela produção por área.

A área de seleção de T2 estará disponível assim que o plantio de T2 for realizado, porém como é explicado melhor acima esta seleção normalmente só se faz após dois anos do plantio do T1 deste campo. A Figura 30 ilustra os tipos de dados que são armazenados nessa seleção, conseqüentemente armazenados na tabela pmgca\_plantiot2.

**Figura 30** – Esquema interno da tabela pmgca\_selecaot2.



**Fonte:** Autor.

### 3.5.9 Plantio da fase T3

Na fase T2, os materiais que apresentarem os melhores resultados nas avaliações são plantados na fase T3 (figura 31). São utilizadas no plantio dessa fase 15 gemas/m (acompanhadas dos padrões), distribuídas em parcelas de 5 sulcos de 4 m de comprimento, com 2 repetições, utilizando o espaçamento da unidade produtora, com 3 m de sulcos de bordadura (frente e fundo do campo) e 2 sulcos de bordadura nas laterais para que haja proteção do material. Faz-se também a multiplicação da sobra desse material, no mesmo período. Nessa fase, cada clone recebe um número de ordem que será sua identidade até uma possível liberação. Essa numeração é composta pelas letras RB – República do Brasil: variedades obtidas pelas

Universidades ligadas à RIDESA e por números, onde os dois primeiros representam o ano do cruzamento. Os demais identificam o número de ordem de seleção daquele cruzamento.

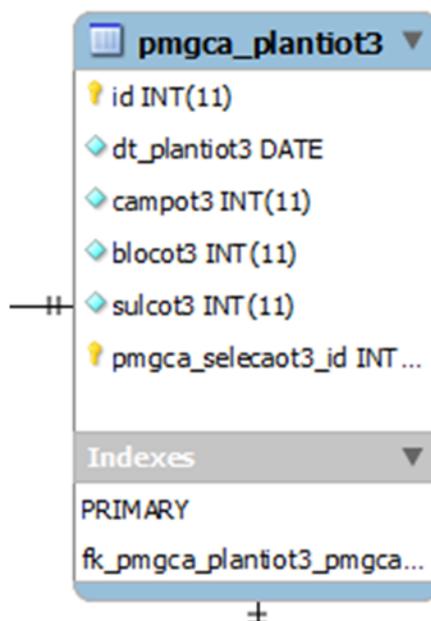
**Figura 31** – Exemplo de plantio de T3.



**Fonte:** Autor

Os dados do plantio do T3 são armazenados na tabela `pmgca_plantiot3`. Esta tabela anexa todos os dados das fases iniciais da liberação de uma variedade com os testes agroindustriais que começam a ser realizados a partir do T3. Esta tabela será a responsável por vincular as informações das fases subsequentes com os dados anteriores dos clones, desta forma ela vincula todos os novos dados aos progenitores das novas variedades. A Figura 32 ilustra a organização da tabela `pmgca_plantiot3`.

**Figura 32** – Esquema interno da tabela pmgca\_plantiot3.



**Fonte:** Autor.

### 3.5.10 Avaliação de T3

Na avaliação, em cana-planta, observam-se, aos 60 dias, a brotação e a presença de ferrugem (*Puccinia melanocephala*); aos 120 dias, são avaliados o perfilhamento, desenvolvimento e incidência de doenças; aos 210-240 dias, verifica-se o número de colmos e a porcentagem de florescimento; são observados, também, nessa avaliação: morfologia, ocamento, enraizamento aéreo, diâmetro do colmo, rachadura e tamanho das gemas.

Em seguida, colhe-se, no primeiro mês de moagem da unidade produtora, pesando-se a parcela utilizando uma carregadeira e um dinamômetro. Nessa pesagem obter-se-á o TCH, t cana.ha-1. Fazem-se feixes de 10 colmos, enviando o material à unidade produtora, onde é feita toda a análise tecnológica (determinação de Brix, POL da cana, PC, POL, % cana, Fibra, Pureza e ATR – Açúcares Totais Recuperáveis); O TPH, t pol.ha-1 é calculado no laboratório do PMGCA/CECA/UFAL.

A tabela desta fase é responsável por conectar os dados das fases FE e CM, pois essas duas fases dependem diretamente dos dados armazenados na tabela pmgca\_selecaot3. A ilustração da forma de armazenamento da tabela está disponível na Figura 33.

**Figura 33** – Esquema interno da tabela pmgca\_selecaot3.

Column Name	Data Type	Key Type
id	INT(11)	PRIMARY
selecaot1_id	INT(11)	Foreign Key
corte	INT(11)	
n_gen	INT(11)	
fibra	DECIMAL(3,2)	
pureza	DECIMAL(3,2)	
pcc	DECIMAL(3,2)	
atr	DECIMAL(3,2)	
tch	DECIMAL(3,2)	
tph	DECIMAL(3,2)	
mcuss_ha	DECIMAL(4,2)	
pmgca_cmaturacao_id	INT(11)	Foreign Key
pmgca_fexperimental_id	INT(11)	Foreign Key

Index Name	Index Type
PRIMARY	PRIMARY
PMGCA_selecaot3_31ac29bc	PRIMARY
fk_pmgca_selecaot3_pmgca_cmaturacao1	Foreign Key
fk_pmgca_selecaot3_pmgca_fexperimental1	Foreign Key

**Fonte:** Autor.

### 3.5.11 Avaliação da Fase Experimental – FE

Após os sucessivos testes, vem a fase experimental dos clones eleitos em T3. O delineamento utilizado é o de blocos ao acaso com parcelas de 6 sulcos de 6 m de comprimento com 4 repetições onde são utilizadas 15 gemas por metro no plantio. Colhem-se pelo menos três ciclos (planta, soca e ressoca) e o experimento é repetido três vezes em locais diferentes, procedendo-se avaliações em três épocas de corte, o que normalmente se faz em anos subsequentes ao do experimento de FE, com materiais pré comerciais. O objetivo de estudar diferentes épocas é adequar cada futura variedade ao seu período ideal de cultivo, o que é essencial, dada a grande extensão do período de safra.

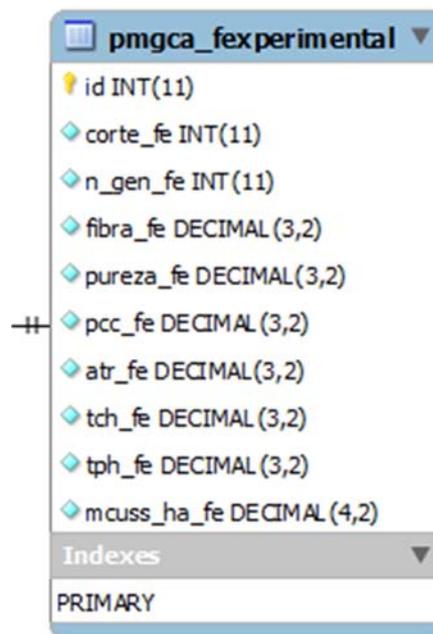
Nessa fase, os materiais são intercambiados entre todas as subestações, com o objetivo de analisar o comportamento desses clones em diversos ambientes e presumir um ambiente ideal de cultivo. A multiplicação do material de FE é feita em 2 sulcos de 10 metros de

comprimento. Na avaliação, são observadas as seguintes características: morfologia, incidência de pragas e doenças, chochamento, observando também se há isoporização. Aos 60 dias, é feita a avaliação de brotação; aos 120 dias, são observados o perfilho e o desenvolvimento e, entre 210 e 240 dias avaliam-se o número de colmos e porcentagem de florescimento, além de características como diâmetro do colmo, enraizamento aéreo, tamanho das gemas e ocamento.

Nos meses de setembro a fevereiro é realizada a colheita de FE na qual o material é enviado para a subestação correspondente ao campo onde o experimento foi instalado, que passará por uma análise tecnológica, analisando-se Brix, PC, Fibra, Pureza e ATR. Na fase experimental, é feito o peso total da parcela. Normalmente estes campos são cortados diversas vezes, no mínimo por 3 anos consecutivos, e os resultados são utilizados para se fazer um estudo de longevidade do canavial.

A tabela *pmgca\_fexperimental* é responsável por armazenar os dados da fase FE, estes dados serão confrontados diretamente com os dados das outras subestações. Esse tipo de interação somente ocorrerá nesta fase. A figura 34 mostra a estrutura interna da tabela *pmgca\_fexperimental*.

**Figura 34** – Estrutura interna da tabela *pmgca\_fexperimental*.



**Fonte:** Autor.

### ***3.5.12 Avaliação da Curva de Maturação – CM***

A curva de maturação dos clones correspondentes à FE é obtida a partir de plantas cultivadas normalmente numa área contígua ao experimento, com parcelas de 3 sulcos de 4 m, com 6 repetições ou seis épocas de amostragem. É plantada no final de agosto ou início de setembro e é nessa fase onde vai se conhecer a maturação da planta, podendo classificá-las em precoce, mediana ou tardia, o que possibilitará definir a melhor época de colheita do genótipo. Na avaliação, observam-se: morfologia, incidência de pragas e doenças, chochamento e isoporização. Aos 60 dias, é feita a avaliação de brotação; aos 120 dias, são observados o perfilho e o desenvolvimento e, entre 210 e 240 dias avaliam-se o número de colmos e porcentagem de florescimento. São avaliadas também características como diâmetro do colmo, enraizamento aéreo, tamanho das gemas e ocamento. São três amostras (compostas por feixes de 10 canas por genótipo/época) que são colhidas aos 12 meses (14 meses se não for à época de corte), fazendo-se análises laboratoriais completas, obtendo-se o pol cana, fibra a pureza e os açúcares recuperáveis. São seis os pontos de avaliação, com os mesmos sendo colhidos de setembro a fevereiro.

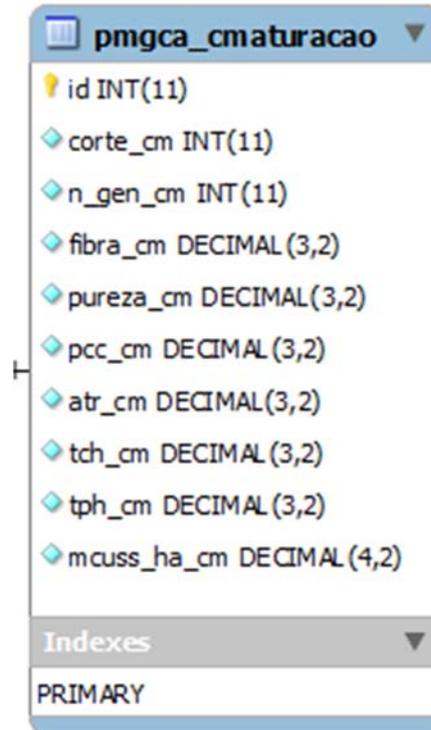
A avaliação da maturação é realizada nos meses de setembro a fevereiro, compreendendo as seis épocas da curva de maturação. É feita uma análise mensal da curva de PC, que consiste em observar o período em que a curva de PC atinge o seu pico possibilitando a classificação da maturação do genótipo em: precoce (até 12 meses), mediana (12-14 meses) e tardia (acima de 14 meses). É importante saber a época da maturação para se definir o melhor período de colheita da planta.

Após passar pelas fases de experimentação os clones que mais se sobressaírem serão registrados, protegidos e, posteriormente, liberados para o plantio comercial, como novas variedades. É importante registrar e proteger as variedades, pois isso gerará recursos ao programa através dos royalties oriundos da utilização das mesmas pelas unidades produtoras e fornecedores de cana-de-açúcar.

Desta forma, o sistema de gerenciamento de banco de dados para o germoplasma da cana-de-açúcar, que será descrito no próximo capítulo, tem como maior finalidade, auxiliar os melhoristas em otimizar a produção de novas variedades, bem como, possibilitar a automação das fases que foram descritas e ainda a conexão com outros programas computacionais que são utilizados nos estudos de produção varietal da RIDESA.

A tabela `pmgca_c maturacao` é a tabela responsável por armazenar os dados que são colhidos nas seis épocas de colheitas de um FE. Cada colheita será armazenada separadamente, e estará conectada ao respectivo clone do T3. A figura 35, mostra como está organizada a tabela.

**Figura 35** – Estrutura da tabela `pmgca_c maturacao`.



**Fonte:** Autor.

### 3.6 *Desenvolvimento do SG-Can*

Após conhecer toda a estrutura interna do PMGCA e de todo o seu funcionamento, é importante saber como o mesmo foi agrupado em um banco de dados. Este agrupamento foi feito de acordo com todas as fases utilizadas, porem utilizando todos os princípios básicos que regem o modelo de banco de dados E-R.

Desta forma, o banco de dados do SG-Can ficou organizado da seguinte maneira:

1. `auth_group`
2. `auth_group_permissions`
3. `auth_message`
4. `auth_permission`
5. `auth_user`
6. `auth_user_groups`

7. auth\_user\_user\_permissions
8. django\_admin\_log
9. django\_content\_type
10. django\_session
11. django\_site
12. pmgca\_cadcruzamento
13. pmgca\_campo
14. pmgca\_caracindustrial
15. pmgca\_caractfisiologica
16. pmgca\_cmaturacao
17. pmgca\_cruzamento
18. pmgca\_doenca
19. pmgca\_fexperimental
20. pmgca\_genotipo
21. pmgca\_parametrorelatorio
22. pmgca\_pessoa
23. pmgca\_pessoa\_subestacao
24. pmgca\_plantiot1
25. pmgca\_plantiot1\_caixa
26. pmgca\_plantiot1\_dt\_remessas
27. pmgca\_plantiot2
28. pmgca\_plantiot2\_nselecao1
29. pmgca\_relatorio
30. pmgca\_remessas
31. pmgca\_replicagem
32. pmgca\_selecao1
33. pmgca\_selecao2
34. pmgca\_selecao3
35. pmgca\_semeio
36. pmgca\_subestacao

Destas tabelas, as onze primeiras, são geradas automaticamente pelo *Django*. Estas tabelas fazem parte do pacote de segurança, que estão embutidos no *framework*, *Django*, onde as tabelas iniciadas por *auth* controlam todo o cadastro de usuários e grupos, bem como as permissões que estes usuários e grupos terão no sistema. As tabelas iniciadas por *Django* fazem

parte do sistema administrativo do banco de dados. Nestas tabelas serão armazenados os dados rotineiros de utilização do sistema, assim o administrador do sistema terá um histórico de utilização do sistema, o qual o ajudará a resolver possíveis problemas causados por erros de usuários. Já as tabelas iniciadas por pmgca fazem parte do conjunto de tabelas de armazenamento dos dados do germoplasma da cana-de-açúcar.

A maioria das tabelas iniciadas com pmgca já foram descritas nas seções anteriores.

### **3.6.1 O SGBD SG-Can**

Inicialmente para melhor compreender todas as rotinas e relacionamentos do SG-Can se faz necessário apresentar os diagramas de entidade-relacionamento (ER) do SG-Can.

A figura 36 demonstra o diagrama ER do setor administrativo do SGBD SG-can. Neste diagrama é possível identificar duas áreas de atuação de controle: a área de controle de grupos e usuários e a área administrativa do SG-Can.

Na área de controle de Grupos e Usuários o sistema armazena todos os dados referentes permissões que os usuários e/ou grupos terão dentro do SG-Can. Este controle poderá ser feito diretamente por usuário ou pelos respectivos grupos que o usuário estiver ligado.

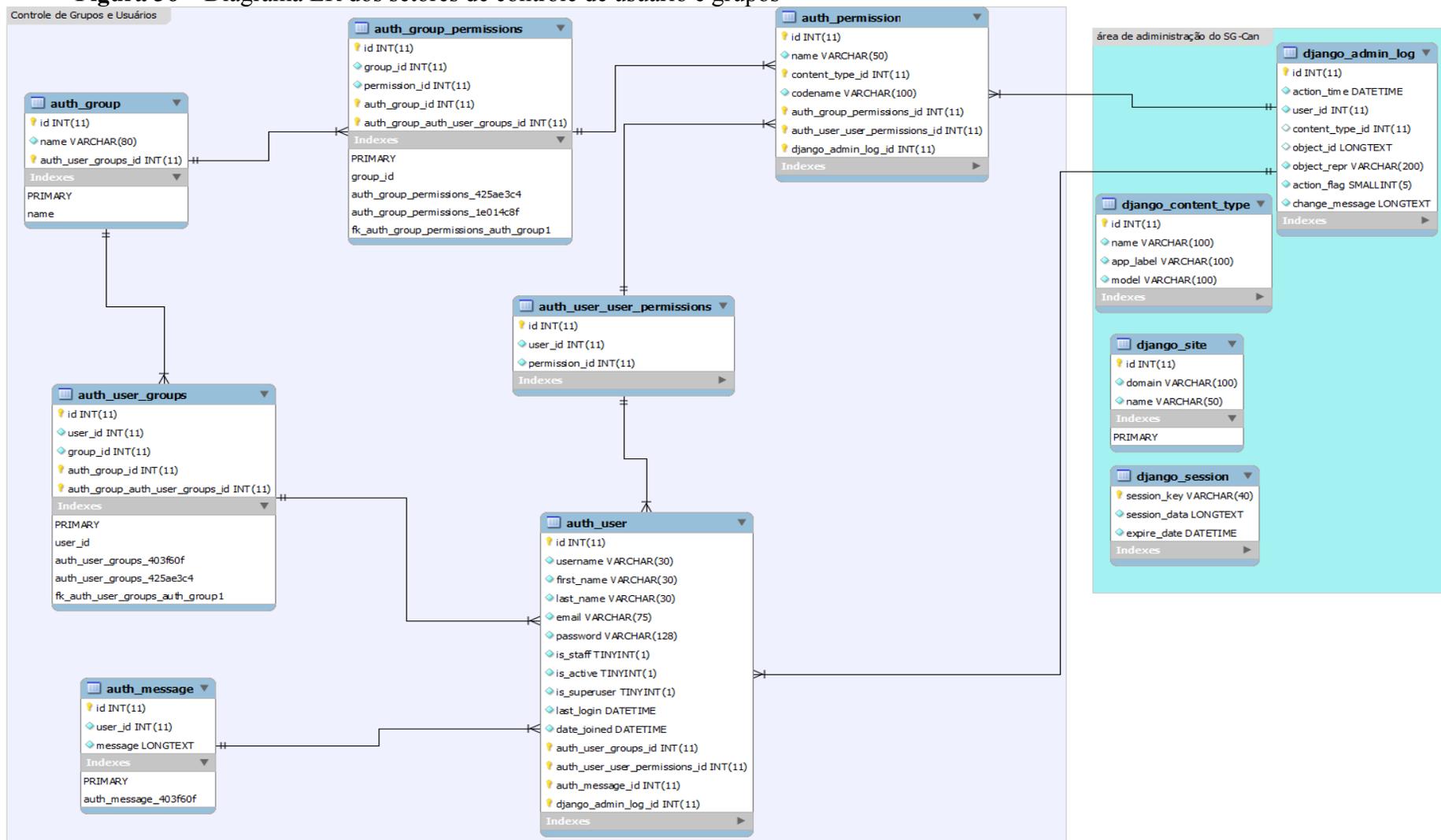
Desta forma, o sistema é possível controlar desde a área que o usuário ou grupo pode visualizar, bem como controlar todas as funções que serão liberadas para os mesmos. Isto é um fator primordial para controle e correção de possíveis erros operacionais dentro do sistema.

Já na área de administração do SG-Can, é possível controlar o domínio a que o usuário ou grupo está ligado. Esta identificação se faz necessária para que o SG-Can identifique, processe e armazene toda a operação realizada por um usuário. Delimitando o acesso a determinadas áreas do SG-Can.

A união entre as duas áreas de controle descritas acima, permite ao administrador do sistema um maior controle sobre o acesso aos dados restritos da cana-de-açúcar, bem como facilita a liberação de dados comuns os quais são disponibilizados para qualquer usuário.

O diagrama ER da Figura 36 representa todas as entidades e atributos que o SG-Can possui. Desta forma é possível identificar as formas de relacionamento entre as tabelas, bem como toda a padronização das tabelas do SG-Can nas três formas normais de banco de dados. Acompanhando todo o processo de armazenamento de dados proposto pelo diagrama ER da figura 16 é possível identificar todo o processo de criação de uma nova variedade, bem como os dados estão interligados entre si facilitando pesquisas nos dados armazenados.

Figura 36 – Diagrama ER dos setores de controle de usuário e grupos



Fonte: Autor

Sendo assim, o SG-Can se utiliza dos dados gerados através dos processos de “produção” de uma variedade para identificar as características primordiais para se obter uma nova variedade, ou das características que devem ser “eliminadas” no processo de produção varietal de cana-de-açúcar.

### 3.6.2 Exemplo de utilização do SG-Can.

Para melhor demonstrar o manuseio do SG-Can, será demonstrado como gerar o relatório de seleção T1.

Quando um usuário for identificado pelo SG-Can, o mesmo será redirecionado para a sua área de atuação, desta forma, apenas os usuários que estão “setados” para área de T1, poderão visualizar este Relatório. Estando reconhecido pelo SG-Can, o usuário deverá ir para a área de geração de relatório, onde poderá identificar a subárea de fase de testes/Seleção/T1, como ilustra a Figura 37.

**Figura 37** – Interface para escolha do relatório de seleção de T1.



**Fonte:** Autor.

Após selecionar o relatório de Seleção de T1, o usuário deverá interagir com quatro filtros básicos para a filtragem dos dados e posterior geração do relatório. O primeiro filtro refere-se ao intervalo de data que o usuário deseja visualizar os dados; o segundo filtro corresponde à seleção da subestação, pois os materiais são plantados e avaliados separadamente em cada subestação; o terceiro filtro corresponde ao número do campo, que identifica uma

determinada área da subestação e por fim o quarto filtro que se refere ao bloco que subdivide um campo. Tudo isto está demonstrado na Figura 38.

**Figura 38** – Relatório de T1

**Fonte:** Autor

Após clicar em gerar relatório o usuário será perguntado sobre a forma de classificação dos dados, sendo assim o usuário deverá selecionar a classificação entre: BRIX ou número de colmo, pois existe uma diferença nos dados dos campos selecionados em planta (não possuem BRIX tendo apenas o número de colmos) dos dados do campo selecionado em socaria (têm os dados de BRIX e número de colmos), esta abordagem é feita na Figura 39.

**Figura 39** – Classificação da visualização dos dados.

**Fonte:** Autor.

Desta forma, após a filtragem e classificação dos dados, os mesmos serão visualizados de forma resumida, contendo apenas os principais dados (Figura 40).

**Figura 40** – Demonstração dos dados filtrados.

Dados de Seleção de T1										Subestação USA	
G.Fem	G.Mas	Cruz.	N de Seleção.	BRIX.	N Colmos.	Caixa	Dt. Semeio	Dt. Plantio			
RB72910	RB855511	463"A	93F	14,6	15	13	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	94F	16	9	13	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	95F	16,4	11	13	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	67F	13,6	12	1490	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	68F	16	11	1490	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	69F	15,6	10	1490	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	70F	15,6	13	1490	17/10/2005	12/12/2005			
RB72910	RB855511	463"A	71F	17,4	19	1490	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	63F	14,4	9	3093	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	64F	14,4	7	3093	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	65F	16	10	3093	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	66F	14,2	10	3093	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	92F	15,8	11	725	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	88F	15,4	10	3184	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	89F	15	7	3184	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	90F	15,8	8	3184	17/10/2005	12/12/2005			
RB855511	RB72910	463"B	91F	15,8	11	3184	17/10/2005	12/12/2005			
RB739735	RB745464	336"A	61F	16,6	10	3789	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	62F	15,6	9	3789	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	15F	15,5	8	2651	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	16F	17	6	2651	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	17F	15,5	7	1978	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	20F	15,5	9	1978	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	84F	14,8	10	3462	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	85F	12,6	8	3462	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	86F	13,8	7	3462	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	87F	14,6	14	3462	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	18F	16	7	3818	18/10/2005	12/12/2005			
RB739735	RB745464	336"A	19F	16,8	13	3818	18/10/2005	12/12/2005			
RB72910	RB863129	185"A	27F	13,6	10	3386	25/10/2005	12/12/2005			
RB72910	RB863129	185"A	28F	13,6	14	3386	25/10/2005	12/12/2005			
RB72910	RB863129	185"A	29F	16,4	10	3386	25/10/2005	12/12/2005			
RB72910	RB863129	185"A	77F	15	11	3181	25/10/2005	12/12/2005			
RB72910	RB863129	185"A	78F	16,8	8	3181	25/10/2005	12/12/2005			
RB72910	RB863129	185"A	8F	16	9	2138	25/10/2005	12/12/2005			
RB72910	RB863129	185"A	9F	15,4	10	2138	25/10/2005	12/12/2005			

Fonte: Autor.

Se o usuário quiser imprimir os dados de forma mais completa e com o modelo de relatório do PMGCA, basta clicar no botão imprimir e o relatório será gerado em pdf (figura 41).

**Figura 41** – Modelo de Relatório de seleção de T1

SUB-ESTAÇÃO: Usina Santo Antônio		Nº DE CLONES: 2005		PMGCA - CECA - UFAL		CAMP 002/																
FASE: T1 (2 x 3)		SÉRIE: 2005		DATA DE SELEÇÃO T2: ___/___/___		ADUBAÇÃO: _____																
PADRÕES:				DATA DE PLANTIO T2: 19/9/2007		BORDADURA: _____																
T1 PLANTADO POR: _____				ESPAÇAMENTO T2: _____ m		1																
Parc	PROGENTOR	Nº SELEÇÃO	T1		GRADUAÇÃO T2 (PLANTA)								SELEÇÃO T2 (SOCA)									
			BRIX	Nº de Colmos	BRIX Médio	Nº de colmos	(%) FLO	Brot.	Des.	Perf.	Doe.	Brix 01	Brix 02	Brix 03	Brix M	Colmos/ Parc.	P 30 cel (g)	Peso unt.	kgBrix/ Parc.	Nº Flor	Nº Sel.	Nº T3
1	SP89-1115	1314	22,8	11																		
2	RB855511	1075	20,6	16																		
3	RB867513	1352	21,2	8																		
4	RB72910	1032	19,8	17																		
6	RB855511	1361	21,4	21																		
7	SP80-3280	1048	20,6	11																		
9	RB92606	1364	21,0	11																		
10	RB72910	1028	19,0	20																		
11	RB906043	1047	21,6	8																		
13	RB863129	1069	20,2	14																		
14	CP88-1540	1062	20,4	10																		
16	SP80-1842	1199	21,0	7																		
17	RB835089	1313	19,4	14																		
18	RB855511	1359	21,8	12																		
20	RB855511	1072	20,2	10																		

Fonte: Autor.

## 4.0 O MODELO DO BANCO

### 4.1 Instalações da Infraestrutura

Inicialmente o sistema foi instalado em um servidor local do PMGCA/CECA/UFAL/RIDESA. Desta forma, foi utilizado um sistema gerenciador de banco de dados que fosse compatível com o Windows Server 2003 Service Pack 2 (SP2).

O SGBD escolhido foi o *MySQL Database Version 5.0.51b*. A instalação foi feita de forma simples utilizando o pacote de aplicativo para desenvolvedores *WEB AppServ Network Version 2.5.10*, que em sua instalação permiti escolher entre os SGBDs, MySQL e *Postgre SQL*, ainda permite instalar e configurar *PHP Script Language Version 5.2.6* e o servidor *WEB Apache Web Server Version 2.2.8*.

Para que se entenda o procedimento, através do qual os métodos foram aplicados, permitindo assim que a informação saia da sua origem até encontrar-se disponível, é necessário saber que as etapas ocorreram de forma mutuamente exclusivas e, sobretudo, na sequência que estão mostradas a seguir.

### 4.2 Implementação do modelo de dados

Antes de gerar os arquivos da criação física do banco de dados, correspondentes ao modelo lógico de dados, é fundamental identificar neste modelo a correspondência entre as regras e os resultados esperados. Então, não se deve utilizar muitos atributos nas chaves das tabelas, apesar de muitas vezes o modelo lógico assim o recomendar. Quando se implementa no banco de dados um modelo com muitos atributos na chave, complica-se desnecessariamente a programação da aplicação que acessará esta tabela e sobretudo, o desempenho das consultas ao banco de dados. Neste estudo as chaves PKs são compostas por no máximo três atributos, que se refere a casos isolados de tabelas, pois a normalidade das tabelas, deste trabalho, possui apenas dois atributos compondo as Pks.

O projeto físico é a implementação do modelo de dados e prima por questões relacionadas ao desempenho das consultas, majoritariamente, além de traduzir a integridade existente no modelo lógico. Para isso, utiliza-se da criação de índices, particionamento de mídias, paralelismo de processamento, ajustes etc. Estes detalhes não pertencem ao modelo lógico de dados. Assim, diversas características do projeto físico estão associadas a

peculiaridades do produto SGBD específico, pois nesta etapa, além de efetivarem-se fisicamente as rígidas regras de integridade ditadas pela normalização do modelo lógico de dados, é necessário também que o desempenho seja satisfatório. A criação de índices para chaves estrangeiras, por exemplo, tende a melhorar o desempenho de consultas complexas ou que envolvam muitas tabelas ou ainda tabelas com muitos registros. No caso deste trabalho, as tabelas possuem um grande e crescente volume de dados, justificando, assim, a existência de índices em colunas do tipo PK, assim como a implementação física das *constraints* que traduzem os relacionamentos do modelo lógico, fazendo valer a integridade referencial ditada pela normalização.

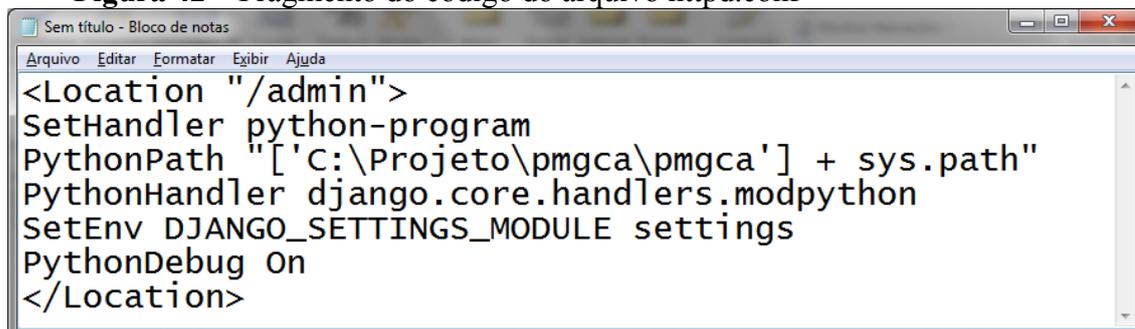
Assim, com a implementação física das tabelas, parâmetros, relacionamentos e índices, temos o banco de dados SG-Can fisicamente gerado.

### 4.3 *Conectividades ao banco de dados*

Parte-se dos seguintes pressupostos para iniciar esta etapa: O sistema operacional Windows Server 2003 Service Pack 2 (SP2) está instalado e o banco de dados SG-Can instalado e customizado neste. É importante salientar que inicialmente a conexão ao SG-Can só estará disponível na intranet do PMGCA/CECA/UFAL/RIDESA, porem o desenho do modelo de segurança utilizado no SG-Can foi desenvolvido partindo do pressuposto de que este banco de dados futuramente seja disponibilizado para todas as universidades integrantes da rede RIDESA.

Uma forma comum de ataque à segurança do banco de dados é a injeção de SQL, em bancos de dados que façam uso desta linguagem, devido a este tipo de ataque foi pensada uma estratégia de defesa aos dados, que em sua primeira barreira seu utiliza da linguagem de programação *Python* e o *Framework Django*.

Na instalação do pacote de aplicativos *WEB AppServ Network* a configuração da conexão do Banco de dados *MySQL* com o servidor *Web apache* e o *framework Django* e imediata, facilitando todo o processo de conexão que o desenvolvedor teria que fazer manualmente. Porém, ainda resta ao desenvolvedor configurar o arquivo *settings.py* e o *httpd.conf* (Figura 42) do apache para que seja liberada a requisição (*HttpRequest*), como protocolo, domínio, portas, caminhos, parâmetros, etc.

**Figura 42** – Fragmento do código do arquivo httpd.conf

```
Sem título - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda
<Location "/admin">
SetHandler python-program
PythonPath "'C:\Projeto\pmgca\pmgca' + sys.path"
PythonHandler django.core.handlers.modpython
SetEnv DJANGO_SETTINGS_MODULE settings
PythonDebug On
</Location>
```

**Fonte:** Autor.

#### ***4.4 Por que utilizar Python e Django?***

As vantagens de se utilizar Python e Django para desenvolver o SG-Can estão disponíveis na seção 2.4.2 do capítulo 2, porém, para melhor especificar a utilização destas duas ferramentas em conjunto é preciso salientar que o desenvolvimento do SG-Can partiu da decisão de se deixar o sistema funcional em qualquer sistema operacional, bem como disponível em qualquer aparelho que possua conexão com internet ou intranet e que tenha um browser (navegar de internet).

Desta forma, a junção de Django + Python possibilita um ambiente rápido, seguro e confiável. Além do que, como são ferramentas gratuitas que estão despontando no desenvolvimento de sistemas WEB, o desenvolvimento de novas bibliotecas e rotinas é quase que diário, possibilitando um grande suporte para o desenvolvedor e agilizando o processo de desenvolvimento de softwares.

#### ***4.5 A interação do Django e Python com um SGBD?***

Diferente das rotinas incluídas diretamente ao banco de dados pelo SQL, a criação das tabelas, PKs, FKs, consultas, etc., possibilitam ataques diretamente ao código SQL do banco de dados. Já com o desenvolvimento na linguagem de programação Python, a geração completa do banco de dados, bem como suas consultas e rotinas de segurança, são geradas diretamente em Python, ou seja, o Django interpreta o código em Python e compila diretamente no SGBD.

Desta forma, todo o processo de consulta, inserção, remoção e correção de dados têm que passar pelas rotinas e parâmetros de segurança contidos nos códigos em Python.

A rotina desenvolvida para geração de um banco de dados, em Python, serve para a maioria dos SGBDs, sendo necessárias pequenas alterações no arquivo `settings.py` e no servidor Web, que neste caso é o apache.

#### 4.6 *Segurança do Banco de Dados utilizando Django*

A segurança inicial do processamento das informações no Django tem como base o processo MVC (*Model View Controller*) ilustrado na figura 43.

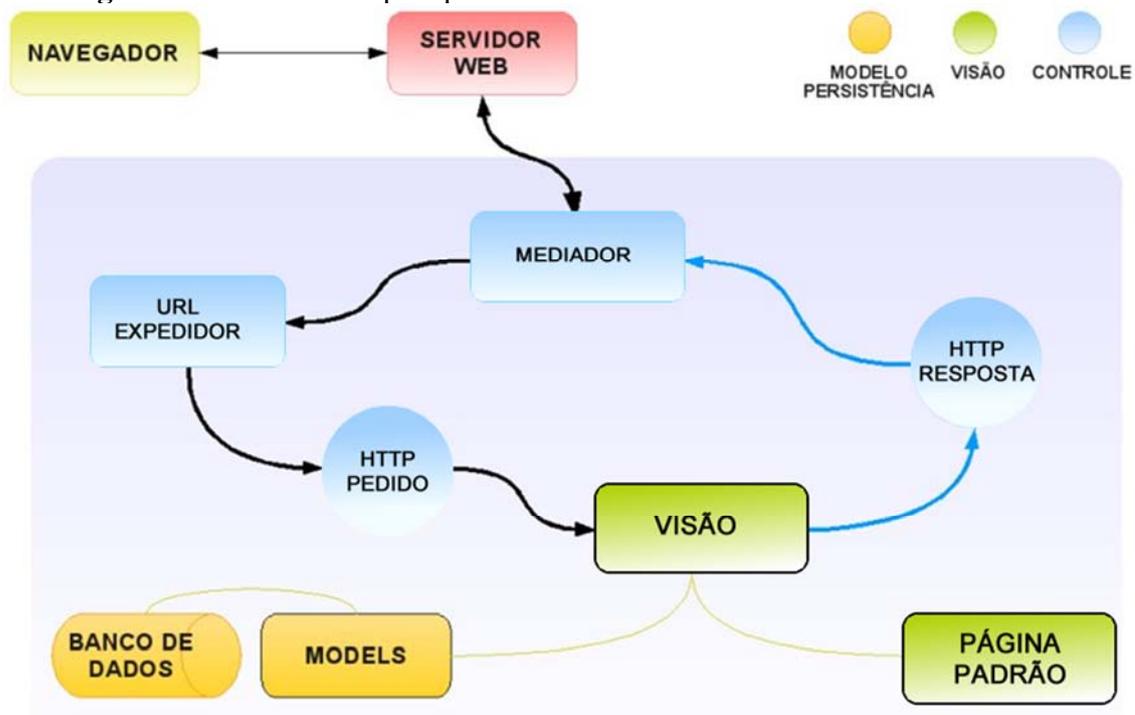
Desta forma, toda e qualquer informação que esteja sendo processado no Banco de dados tem que pôr obrigatoriedade passar por alguns requisitos do Django, que são:

1. *Model* (Modelo) – É onde estão às definições dos dados; como eles devem ser armazenados e tratados. É aqui onde se diz quais campos uma tabela deve ter, seus tipos e valores padrão e se eles são obrigatórios ou não, dentre outras coisas. Para armazenar e resgatar informações do banco de dados, não é necessário ir até ele e conhecer a linguagem dele (SQL). O Django usa uma ferramenta chamada ORM, que interpreta o código, leva aquilo até o banco de dados, e depois devolve as informações desejadas, ou seja, as consultas não são feitas diretamente ao banco de dados. A parte do código que configura os modelos de dados e os tipos de informações que devem armazenar é um arquivo chamado `models.py`.
2. *View* (Visão) – São as funções que recebem requisições e retornam respostas, ao usuário, a outro computador, a uma impressora ou qualquer outra coisa externa. Estas requisições são armazenadas em arquivos chamados `views.py`.
3. *Controller* (Controle) – São todas as coisas que ficam no meio do caminho, como o *handler*, os *middlewares* e o URL *dispatcher*. A maior parte destas requisições é feitas pelo próprio Django, o usuário, a priori, não se preocupa em desenvolver estas requisições, que podem ser:
  - *Handler* – É o primeiro passo antes de a requisição chegar aos *middlewares*, e também é o último passo antes da resposta voltar do Django para o usuário final.
  - *Middleware*s - São pequenos trechos de código que analisam a requisição na entrada e a resposta na saída. A requisição é analisada por eles. É possível determinar quais *middlewares* estarão na fila à mesa para analisar a

requisição e pode até criar os novos middlewares. Um dos middlewares faz com que toda a segurança e autenticação, de usuários, seja feita. Outro adiciona a função de sessão, uma forma de memorizar o que o usuário está fazendo para atendê-lo melhor. Há ainda outro middleware, que memoriza as respostas no Cache, pois se caso a próxima requisição tenha o mesmo caminho e os mesmos parâmetros, ele retorna a resposta memorizada, evitando o trabalho de ser processada novamente pela view. Sendo assim, após passar por todos esses middlewares, a requisição passa a ter outras informações, como qual usuário está autenticado, sessão atual e outras coisas.

- *URL Dispatcher* - Depois de passar pelos middlewares, a requisição é analisada pelo *URL Dispatcher*, um importante componente do Django que verifica o endereço - especialmente a parte do caminho - e verifica o arquivo `urls.py` do projeto para apontar qual *view* será chamada para dar a resposta

**Figura 43** – Modelo HttpRequest de um SGBD



**Fonte:** Autor adaptado da web.

Desta forma os dados solicitados são entregues via *Really Simple Syndication* (RSS), que é somente um formato que segue padrões XML, mas que foi criado para carregar informações como Título, Data de Publicações e conteúdo de artigos. Também é muito útil para

*Podcasts*, exportar e importar esses mesmos tipos de informações. Este recurso já vem pronto no Django, desta forma, não se faz necessário configurá-lo.

## 5.0 CONCLUSÃO

### 5.1 *Gerais*

A “corrida contra o tempo” para desenvolver tecnologias e técnicas para obtenção de uma fonte de energia mais sustentável, tanto pelo fator econômico como pelo fator ambiental, cumulou em trazer as atenções das autoridades mundiais, no campo energético, para o Brasil.

A cana-de-açúcar sai de uma visão de uma monocultura extrativista, que apenas deteriorava as terras brasileiras, para a cultura que pode substituir o petróleo como fonte energética e além do que ainda é uma das maiores fontes de resgate de carbono.

Nesse momento, onde a tecnologia está cada vez mais acessível ao ser humano, não é possível entender como a cultura da cana-de-açúcar que é de grande importância para o Brasil, seja no âmbito econômico como no âmbito socioeconômico e até no âmbito ambiental, ainda está extremamente atrasada quanto a utilização de tecnologias computacionais para o acompanhamento do desenvolvimento e até da produtividade de uma variedade.

Através deste trabalho constatou-se a importância para a utilização de novas tecnologias para o acompanhamento da cultura da cana-de-açúcar, mostrando a importância de se criar novas ferramentas que contribuam para o processo de obtenção de novas variedades, acompanhamento das variedades em campo, armazenamento e processamento dos dados obtidos nos estágios de plantio e colheita e ainda pode auxiliar no desenvolvimento de uma nova variedade de cana-de-açúcar.

A proposta deste trabalho foi desenvolver um sistema de gerenciamento de banco de dados que permita processar, analisar e responder questões pertinentes aos dados do germoplasma da cana-de-açúcar. Desta forma, o presente trabalho foi moldado para que os pesquisadores das bases da RIDESA, espalhadas por todo território brasileiro, possam ter uma ferramenta segura, veloz e prática para o desenvolvimento e acompanhamento dos estudos em cana-de-açúcar.

### 5.2 *Específicos*

Com base no conhecimento da cultura, foi possível modelar um sistema de gerenciamento de banco de dados, onde o acompanhamento das variedades de cana-de-açúcar

é feito desde sua concepção na base Serra do Ouro, passando pelos diversos estágios de estudos em campo (fase t1, t2, t3, FE, FM, CM e etc.) e até mesmo os dados agroindustriais que são fornecidos por todas as usinas que estão associadas à RIDESA.

Também foi possível adicionar ao SG-Can, uma modelagem que permita armazenar os dados fenotípicos da cana-de-açúcar. Estes dados irão possibilitar um maior aprofundamento nos estudos de produção de uma variedade.

Desta forma, o SG-Can possibilita o cruzamento dos dados, em campos distintos, dos estudos em cana-de-açúcar. Como é o caso de poder analisar variedades através de seus resultados agroindustriais e fenotípicos de uma só vez, o que normalmente só é analisado de forma separada.

Por fim, o banco de dados do SG-Can permiti a conexão com outros softwares que facilitam os estudos na área, como é o caso da conexão com o SETAX, o que gerou dois trabalhos em congressos no brasil e no exterior.

### **5.3 *Trabalhos Futuros***

É importante salientar que o desenvolvimento do SG-Can, ainda não está completo. O SG-Can, ainda não está apto para armazenar e processar dados do Genoma e do Proteoma da cana-de-açúcar. Desta forma, sugere-se para trabalhos futuros, os seguintes temas:

1. Uma abordagem computacional sobre os dados do Genoma Cana;
2. Estudos de mineração de dados das informações fenotípicas, agroindustriais e genéticas da cana-de-açúcar.
3. Interação entre bancos de dados fenotípicos, genéticos e proteômicos.

## 6.0 REFERÊNCIAS

ALVES, W.P. **Fundamentos de bancos de dados**. São Paulo: ÉRICA, 2004. N.1, p.28-36.

AMARAL, T.M. & NEVES, M.F. **As cadeias Agroalimentares do Açúcar: Um estudo comparativo entre França e Brasil**. In: 2º Congresso do Instituto Franco-Brasileiro de Administração de Empresas IFBAE-2003. Anais do 2º Congresso do Instituto Franco-Brasileiro de Administração de Empresas IFBAE-2003. Franca, 26 e 27 de maio de 2003 – ISBN:85-87406-14-0, CDB: 658.4012.

APACHE. **Introdução ao software**. Disponível na página web< <http://httpd.apache.org/>> visitada em 27/07/2011.

APPSERV. **Appserv Open Project: what is Appserv**. Disponível na página web <http://www.appservnetwork.com/modules.php?name=Content&pa=showpage&pid=11>> visitada em 27/07/2011.

BARBOSA, G.V.S., SOUSA, A.J.R., ROCHA, A.M.C., RIBEIRO, C.A.G., FERREIRA, J.L.C., CRUZ, M.M., SOARES, L. e MOURA FILHO, G. **Relatório Técnico PMGCA: Safra 2001/2002**. Rio Largo, 2002.

BARBOSA, G.V.S., SOUSA, A.J.R., ROCHA, A.M.C., RIBEIRO, C.A.G., FERREIRA, J.L.C., CRUZ, M.M., SOARES, L., SANTOS, J.M., SOARES, L., SILVA, P.P. e VIVEIROS, A.J.A. **Relatório Técnico safra 2006-2007**. Rio Largo, 2007. 9-10p.

BARRETO, E. J. S. **Estação de floração e cruzamento Serra do Ouro – EFCSO: características do banco de germoplasma, comportamento do florescimento e técnicas de hibridação em cana-de-açúcar**. Rio Largo, 2002.

BOOCH, Grady; RUMBAUGH, James; JACOBSON, Ivar. **UML, guia do usuário**. Rio de Janeiro: Campus Ltda, 2000.

CÂNDIDO, F., MVS/VSAM. **Manual técnico - T-Systems**. São Paulo, 2007.

COMCIENCIA. **Cana brasileira: combustível para todo o mundo**. <<http://www.comciencia.br/reportagens/2004/12/11.shtml>> visita: 02/05/11.

COPERSUCAR. 1995. **Censo varietal quantitativo**. Piracicaba: Centro de Tecnologia Copersucar. 1994. 16p.

DATE, C. J. **Introdução a Sistemas de Banco de Dados**. Tradução de Daniel Vieira. 8º ed. Rio de Janeiro: Elsevier, 2003.

MYSQL. **MySQL Documentation: MySQL Reference Manuals**. Disponível na página web< <http://dev.mysql.com/doc/>> visitada em 27/07/2011.

DJANGO. **Aprendendo Django no planeta terra**. Disponível na página web <<http://www.djangobrasil.org/>> visitada em 27/07/2011

ECLIPSE. **Eclipse SDK docs: refence Manuals**. Disponível na página web <[https://www.eclipse.org/docs/ />](https://www.eclipse.org/docs/), visitada em 27/07/2011.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. Revisor Técnico Luis Ricardo de Figueiredo. 4º ed. São Paulo: Addison Wesley, 2005.

HEINZ, D.J.; TEW, T.L. **Hybridization procedures**. In: Heinz, D.J. (ed.). Sugarcane improvement through breeding. Elsevier, Amsterdam, 1987. p.313-342.

IBM. Corp. and Fund. Software. IMS / ESA. **Performance Analyzer**. San Jose, CA.: IBM, 1998.

JUNIOR, A.N.B., 2009. **A importância da cana de açúcar na economia e meio ambiente**. Artigo publicado em 19/03/09 no portal da educação e da informação <<http://www.jomar.pro.br/portal/>>. No link: <[http://www.jomar.pro.br/portal/modules/smartsection/item.php?itemid=254&keywords=imp ortância+cana+açúcar+economia+meio+ambiente](http://www.jomar.pro.br/portal/modules/smartsection/item.php?itemid=254&keywords=imp%20ort%C3%A2ncia+cana+a%C3%A7%C3%BAcar+economia+meio+ambiente)>.

MALAVOLTA, E. et al. **Cultura e adubação da cana-de-açúcar**. São Paulo: Instituto Brasileiro de Potassa. 1964. p368.

MANGELSDORF, A.J. **Um programa de melhoramento da cana-de-açúcar para a agroindústria canavieira do Brasil**. Brasil Açúcar, 1967. p.69, 208-223.

MARTINS, G. O. **Programa de Melhoramento Genético da Cana-de-Açúcar no Estado de Alagoas**. Trabalho de Conclusão de Curso da Universidade Federal de Alagoas, Maceió, 2005.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. **Melhoramento da cana-de-açúcar**. In: BOREM, A. (Org.). Melhoramento de espécies cultivadas. Viçosa; Universidade Federal de Viçosa, 1999. p.204-251.

MICROSOFT. **Informações sobre o Produto**. Disponível na página web <<http://www.microsoft.com/brasil/servidores/windowsserver2003/evaluation/default.msp>> visitada em 27/07/2011.

MINORELLO, D. **PHP/MYSQL: WEB**. Santa Cruz do Rio Pardo: Viena, 2007. n 1.

MIOCQUE, J.Y.J. **O melhoramento da cana-de-açúcar no Brasil**. STAB 11. 1993. p24-28.

MOZAMBANI, A.E; PINTO, A.S.; SEGATO, S.V.; MATTIUZ, C.F.M. **História e morfologia da cana-de-açúcar**. In: Atualização em produção de cana-de-açúcar. Piracicaba, 2006. p 11-18.

OLIVEIRA, W. J. **Data Warehouse**. Florianópolis: Visual Books, 2002.

POLETTO, A. S. R. S. **Um modelo para projeto e implementação de bancos de dados analítico-temporais**. Edição Revisada / A.S.R.S. Poletto. São Paulo, 2008. 167 p.

PRESSMAN, Roger S. - **Engenharia de Software**. Rio de Janeiro: McGraw Hill, 2002.

PYTHON. **Aprendendo Python**. Disponível na página web<<http://www.python.org/>> visitada em 27/07/2011.

RAMOS, R. P. **Influência da Temperatura do Ar no Período de Florescimento de Genótipos de Cana-de-Açúcar na Serra do Ouro, Alagoas**. Trabalho de Conclusão de Curso da Universidade Federal de Alagoas, Maceió, 2008.

ROCHA, A. M. C.; SANTOS, J. M.; BARBOSA, G. V. S.; CRUZ, M. M.; SOUZA, A. J. R.; RIBEIRO, C. A. G.; FERREIRA, J. L. C.; SILVA, P. P.; SANTOS, T. W. T.; NASCIMENTO, B. F. C. **Quatro décadas de contribuição da Serra do Ouro para a agroindústria da cana-de-açúcar do Brasil**. Anais, 9º Congresso Nacional da STAB. Maceió, 2008. p. 532.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistemas de Bancos de Dados**. Tradução de Daniel Vieira. Rio de Janeiro: Elsevier, 5º ed, 2006.

SILVA, P.B. **Aspectos Fisiológicos de Seis Genótipos de Cana-de-Açúcar Submetidos a Stress Hídrico**. Dissertação (Mestrado em Agronomia: Produção Vegetal e Proteção de Plantas) Universidade Federal de Alagoas. Centro de Ciências Agrárias. Rio Largo, 2010 – CDU: 633.61.

SUAPESQUISA. **História do Brasil Colônia - O Período Colonial**. Disponível em: <<http://www.suapesquisa.com/colonia/>> visita 02/05/11.

ÚNICA. **Setor Sucroenergético – Histórico: Cultivo da Cana hoje**. Disponível em: <[www.unica.com.br](http://www.unica.com.br)>. Visita 14/09/11.

## APÊNDICE

### Diagrama E-R do Sistema SG-Can

