

Sistema de Apoio à Decisão para Avaliação de Indicadores para Bacharelados Interdisciplinares

Karen Braga Enes e Eduardo Barrére

Resumo—O processo de tomada de decisão em grandes instituições exige o auxílio de ferramentas capazes de gerir informações consistentes e objetivas em um pequeno intervalo de tempo. Nesse contexto, este trabalho apresenta o processo de desenvolvimento e implantação de um sistema de apoio à decisão, baseado na plataforma Pentaho, para gerenciamento das informações referentes aos cursos de Bacharelados Interdisciplinares (BIs) na Universidade Federal de Juiz de Fora (UFJF). Com intuito de auxiliar os coordenadores dos cursos de BIs nas atividades de gestão de alunos e cursos de segundo ciclo, um repositório de dados foi implementado. Esse repositório é responsável por agrupar e coordenar elementos do banco de dados da UFJF em uma interface simples e objetiva acessível aos coordenadores. A interface disponibiliza dados através de tabelas e gráficos gerados automaticamente pelo servidor analítico da ferramenta, sem que seja necessário um especialista para realização das consultas.

Palavras-Chave—Sistemas de apoio à decisão, repositório de dados, mineração de dados.

I. INTRODUÇÃO

AS instituições de grande porte, diariamente, geram um número cada vez maior de dados que devem ser armazenados referentes a cadastro de pessoal, processos e transações internas e externas. Os Institutos Federais de Ensino Superior (IFES), da mesma forma, contam com grandes bancos de dados para armazenamento de informações do corpo discente, docente e técnicos administrativos. Na Universidade Federal de Juiz de Fora (UFJF), o armazenamento é feito por intermédio do Sistema Integrado de Gestão Acadêmica (SIGA).

Em 2007, o Ministério da Educação e Cultura (MEC) juntamente ao Governo Federal, propuseram um plano de estruturação da educação superior com o Plano de Apoio a Reestruturação e Expansão das Universidades Federais (REUNI). Segundo [1], o governo federal adotou uma série de medidas para retomar o crescimento do ensino superior público, criando condições para a expansão física, acadêmica e pedagógica da rede federal de educação superior.

Diversas IFES aderiram a proposta de expansão do Governo Federal, entre elas a Universidade Federal de Minas Gerais,

Karen Braga Enes é mestranda em Ciência da Computação pela Universidade Federal de Juiz de Fora (PGCC/UFJF). Possui graduação em Ciência da Computação pela Universidade Federal de Juiz de Fora. Email: karenenes@ice.ufjf.br

Eduardo Barrére possui graduação em Bacharelado em Ciência da Computação - UFSCar (1996), mestrado em Ciência da Computação - UFSCar (1998) e Doutorado em Engenharia de Sistemas e Computação - COPPE/UFRJ (2007). Professor adjunto II da Universidade Federal de Juiz de Fora. Ocupa atualmente o cargo de Coordenador dos Bacharelados Interdisciplinares - PROGRAD.

Universidade Federal de São Carlos e a Universidade Federal da Paraíba [1]. A UFJF aderiu ao REUNI em Outubro de 2007 [2]. Entre as diversas medidas adotadas, instituiu-se a criação de três cursos de Bacharelados Interdisciplinares (BIs), em Ciências Exatas, Ciências Humanas e Artes e Design. Essa proposta determinou um aumento significativo no número de ingressantes na UFJF.

Os BIs são formados por dois ciclos. O primeiro, cuja duração varia entre dois anos e meio a três anos, é voltado para uma formação geral e multidisciplinar na área; O segundo, por sua vez, com duração mínima de um a três anos, no qual o aluno cursa disciplinas específicas do curso escolhido ao término do primeiro ciclo.

Nesse contexto, os coordenadores de cursos dos BIs que são responsáveis por acompanhar e avaliar as atividades acadêmicas e o registro acadêmico dos alunos matriculados no curso, passam a lidar com uma demanda maior e mais complexa de informações para gerenciamento. Os coordenadores controlam alunos de primeiro ciclo com enfoque em quaisquer dos cursos de segundo ciclo ofertados.

Controlar e avaliar o rendimento de todos os alunos matriculados no curso, um a um, tornou-se uma tarefa inviável. Outras questões, tais como, avaliar o desempenho de um determinada turma, a evolução dos alunos na grade do primeiro ou segundo ciclo, estudar a migração de ciclos e a forma com que alunos de diferentes formas de ingresso progredem no curso, somente seria possível fazendo uma consulta especializada ao base do SIGA, com auxílio de um especialista em banco de dados.

Assim, esse artigo apresenta a implantação de um sistema de apoio à decisão (SAD) para auxiliar os coordenadores de BIs, em relação as questões previamente apresentadas. O sistema, baseado na plataforma de código aberto Pentaho Business Intelligence [4], agrupa dados de primeiro e segundo ciclo de todos os alunos que ingressaram desde a criação do curso. Com uma interface amigável e simples, o sistema está disponível para os coordenadores na rede interna da UFJF. A plataforma apresenta ainda um servidor analítico responsável por organizar as respostas as consultas além de prover dados e gráficos diretamente ao usuário final.

O primeiro curso a contar com essa ferramenta foi o BI em Ciências Exatas que anualmente conta com cerca de 400 ingressantes, desde o primeiro semestre de 2009. Além disso, o curso de Ciências Exatas apresenta 12 possibilidades de ingresso de cursos de segundo ciclo dentro da área de Exatas e Engenharias. Gerenciar todas essas informações é claramente trabalhoso. Ainda, obter resultados sobre os dados dos cerca de 2500 alunos em questão é ainda mais difícil. As próximas seções são destinadas ao detalhamento do processo

de implementação do SAD até a entrega para os coordenadores do BI de Ciências Exatas.

A próxima seção apresenta os conceitos fundamentais sobre SADs. A seção III descreve as ferramentas da plataforma empregadas no trabalho. Os cubos criados e os resultados obtidos para o sistema implementado são apresentados na seção IV. Considerações finais e trabalhos futuros estão destacados na última seção.

II. CONCEITOS FUNDAMENTAIS SOBRE SAD

Os sistemas de apoio à decisão estão representados por uma gama de sistemas de informação, ferramentas e tecnologias que visam apoiar os decisores, de maneira interativa, em questões relacionadas a identificação e resolução de problemas. O processo decisório está envolto em incertezas e imprevistos sobre complexidade da ação a ser tomada, bem como suas consequências a curto e longo prazo.

De acordo com [5], a capacidade de um decisor em tomar decisões bem como a incerteza relacionada a estas decisões varia de acordo com vários fatores, entre eles: a disponibilidade de conhecimentos e habilidades, o entendimento e comunicação entre os tomadores de decisão e o desejo de cooperação entre os decisores segundo os recursos disponíveis.

As ferramentas proporcionadas pela informática são úteis para otimizar o tempo de resposta e auxiliar o decisor na seleção das informações relevantes, na proposição de cenários a serem estudados, na validação do modelo e na análise dos dados [6].

Os SADs possuem características marcantes e exclusivas visando a satisfação do usuário final que independe de um especialista para buscar resultados. Esses sistemas permitem a abordagem de problemas semi e não estruturados através de um processo colaborativo homem-máquina. São abordadas decisões independentes ou sequências, abrangendo todas as fases do processo de tomada de decisão.

Além disso, em sua grande maioria são adaptativos em função do tempo e recebem a designação de sistemas *Online Analytical Processing (OLAP)*, embora não seja essa uma característica necessária. São sistemas de fácil utilização, com fortes capacidades gráficas e interface agradável ao usuário.

Lupatini [5], define SAD como a atividade que permite através de modelos claramente explicitados, mas não necessariamente completamente formalizados, ajudar na obtenção de respostas às questões que são colocadas a um interventor num processo de decisão.

A Figura 1 representa, esquematicamente, as principais características de um SAD.

A arquitetura de um SAD é definida por [7] e consiste em três principais subsistemas componentes. O gerenciador de banco de dados responsável por armazenar informações relevantes advindas da instituição na qual o sistema é implantado. O gerenciador de modelagem básica cuja função é transformar as informações obtidas da base em fatos ou situações de interesse a serem analisadas e, por fim, o gerador de diálogo e gerenciamento que coordena a interface com o usuário que interage com o sistema.



FIGURA 1. Características de um SAD

Definida a arquitetura, é necessário tratar dos tipos básicos de análise de informações obtidas pelo SAD. Em [8] foi estabelecido quatro tipos básicos de análise de situação.

- Análise do tipo “e se” - Na qual o usuário final pode produzir alterações nas relações estabelecidas entre as variáveis ou até mesmo nas próprias variáveis. Essa análise culmina com a observação das mudanças resultantes e o impacto causado nos valores de outras variáveis.
- Análise de sensibilidade - Considerado por [8] um tipo particular de análise “e se”, envolve alterações sucessivas no valor de uma única variável. O resultado dessa análise é obtido em relação a observação das alterações geradas em outras variáveis. Esse modelo é normalmente empregado quando há indecisão em relação a premissas assumidas nas estimativas de valores de variáveis-chaves.
- Análise de busca de metas ou “como se pode” - Ocorre uma inversão na direção de análise. Ao invés de variar valores e observar as mudanças geradas, uma variável-alvo tem seu valor fixado enquanto outras variáveis dependentes tem seus valores alterados até que se atinja o alvo determinado.
- Análise de otimização - Apresenta uma extensão complexa da análise de busca de metas. Nesse paradigma, a meta é encontrar um valor ótimo para a variável-alvo, que não necessariamente é única, dados as devidas restrições. Atinge-se o objetivo variando incógnitas dependentes repetidas vezes até que sejam descobertos os melhores valores para as variáveis-alvo.

Estabelecidos os tipos de análise, a próxima seção aborda a metodologia de desenvolvimento do SAD. Para a construção e implantação, foi utilizada uma plataforma gratuita e de código aberto chamada Pentaho Business Intelligence.

III. PENTHAO BUSINESS INTELLIGENCE

A metodologia de desenvolvimento do SAD foi apoiada em todas as etapas pelo Pentaho. Segundo [4], o Pentaho é uma suite de código aberto e distribuição gratuita implementada em Java, que oferece um conjunto de ferramentas para integração de dados, servidores de processos analíticos online, mineração de dados e processos de extração transformação e carregamento de dados (ETL). A construção de um SAD é dividida em quatro etapas e apresentada na Figura 2.



FIGURA 2. Fases de construção do SAD

A. Etapa 1 - Fonte de dados

O primeiro procedimento é investigar a fonte de dados que vai suprir o sistema. No caso da UFJF, o SIGA armazena informações acadêmicas e pessoais de todos os alunos da universidade em um banco de dados ORACLE [9].

B. Etapa 2 - ETL

A fase ETL corresponde a transformação dos dados do ORACLE para o banco MySQL compatível com o Pentaho. Essa etapa promove o carregamento efetivo das informações para o repositório de dados do SAD. A ferramenta responsável por esse procedimento é o Pentaho Data Integration (PDI). De acordo com [10][11], é utilizado para integrar dados dispersos de diferentes fontes e fazer a informação integrada possível e viável ao usuário final.

Nesse momento, informações do modelo transacional são carregados no modelo dimensional onde são consolidados. O dados do ORACLE são transformados no chamado modelo estrela [12]. O modelo estrela representa os dados através de tabelas de dimensões e fatos. A tabela de fatos ocupa a posição central do modelo ligando todas as tabelas de dimensões geradas.

Essa etapa é responsável pela maior parte de tempo de criação do SAD. Cada novo problema avaliado demanda um novo processo ETL. Essa etapa compreende uma análise não trivial da fonte de dados para seleção das informações. Uma transformação incorreta ou uma seleção inadequada de informações compromete todo o processo posterior.

C. Etapa 3 - Repositório de dados e cubos OLAP

Determinadas as bases que agrupam as tabelas de dimensões e fatos. O Pentaho Administration Console (PAC) carrega as bases que são utilizadas no servidor analítico. O repositório de dados é formado pelas bases de dados agrupadas no PAC [13].

A próxima fase corresponde a criação dos cubos OLAP. Essa fase é definida pelo Pentaho Schema Workbench [14]. Como documentado, o workbench fornece uma interface gráfica para criação e testes de cubos OLAP. O esquema é criado e armazenado como um arquivo XML. Esse modelo XML é considerado um cubo, embora não seja estruturado fisicamente, que utiliza estruturas de tabelas para Fatos e Dimensões, geradas pelas transformações do PDI.

O cubo OLAP é responsável por determinar o cruzamento de informações entre as tabelas de dimensões e fatos. Esse cruzamento de informações deve ser estruturado visando o entendimento do usuário final. São determinadas, nessa fase, as medidas de avaliação do problema e as dimensões acessíveis ao usuário. O cruzamento de informações determina os tipos de análise de situação viáveis para cada problema, seguindo a estrutura mencionada na seção anterior.

Cada modelagem feita pelo PDI pode gerar um número indefinido de cubos, conforme a necessidade da instituição. O cubo, depois de construído, deve ser testado dentro do próprio Workbench e depois publicado na ferramenta que provê a interface com o usuário final, o Pentaho User Console (PUC) [15].

O PUC interpreta o arquivo XML após a publicação e fornece o cubo ao usuário através da interface gráfica de interação chamada Saiku [16].

D. Etapa 4 - Exploração dos dados

O Saiku é o servidor analítico da plataforma e trabalha fazendo a interpretação do cubo. Ele mostra as dimensões obtidas e as medidas criadas para filtragem de dados e geração de resultados. Os resultados podem ser apresentados em tabelas ou gráficos. A consulta pode ser salva no PUC de acordo com as permissões do usuário.

Todo o processo de estruturação do SAD é sumarizado na Figura 3.

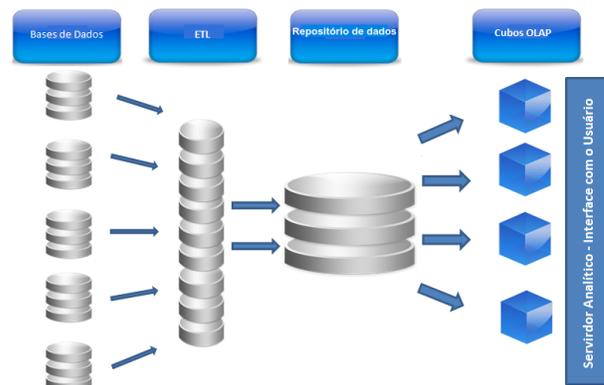


FIGURA 3. Estruturação do SAD

Determinadas todas as etapas de desenvolvimento de SAD e tendo sido esclarecidas as ferramentas utilizadas no processo, a próxima seção apresenta a criação dos cubos implementados no SAD do BI em Ciências Exatas.

IV. RESULTADOS

Os coordenadores do curso de Ciências Exatas tem por objetivo avaliar o desempenho e a evolução dos alunos de primeiro e segundo ciclos. Essa avaliação é feita com base na forma de ingresso, no perfil acadêmico dos alunos, nas disciplinas cursadas e nas disciplinas obrigatórias de primeiro ciclo.

As medidas são avaliadas em relação a movimentação nos cursos de segundo ciclo, nas taxas de avaliação das disciplinas e nos índices de trancamento, e desligamento do curso.

Pretende-se averiguar se a busca pelo curso é genuína ou se dá devido a falta de competitividade no momento do vestibular. Além disso, por se tratar do primeiro curso de BI da UFJF é importante caracterizar o perfil dos ingressantes desde a primeira turma até a mais atual. A principal meta do BI em Ciências Exatas é promover um conhecimento interdisciplinar, com intuito de possibilitar à comunidade discente uma escolha de segundo ciclo com maior embasamento e maturidade. É de fundamental interesse investigar se a meta do curso vem sendo cumprida no decorrer das turmas ingressantes.

Em parceria com os coordenadores de curso e, após uma pesquisa sobre a necessidade dos mesmos, foram determinados os cubos a serem criados. Essa seção aborda detalhadamente as etapas no processo de criação de três cubos implantados no sistema.

A. Sistemas de Ingresso

O primeiro cubo descrito, Sistemas de Ingresso, tem por objetivo avaliar a situação dos alunos no BI provenientes das diferentes formas de ingresso da UFJF.

Pretende-se identificar medidas quantitativas em relação ao número de alunos em primeiro e segundo ciclo, bem como número de alunos jubilados, transferidos e ativos. Além disso, é possível identificar também quantos desses alunos já concluíram o primeiro ciclo.

Uma outra medida interessante relaciona a forma de ingresso dos alunos e o Índice de Rendimento Acadêmico (IRA). O IRA é uma média ponderada do somatório das notas obtidas nas disciplinas multiplicadas pelos créditos correspondentes sobre o somatório dos créditos. Essa medida é de extrema importância por ser requisito para obtenção de bolsas de iniciação científica, participação em grupos de educação tutorial e intercâmbios institucionais. Portanto, é imprescindível observar essa variação.

Estabelecidos as metas do cubo a ser devolvido e as medidas que devem ser abordadas, as etapas de construção do SAD são abordadas a seguir.

1) *Dimensões*: As dimensões do cubo devem ser estabelecidas visando cumprir os objetivos determinados.

No caso do cubo Sistemas de Ingresso, é pertinente obter informações sobre a forma de ingresso do aluno, a cota pela qual o aluno ingressou, sexo. Informações pessoais como

número de matrícula devem ser observados para verificar a presença de dados duplicados no sistema. Deve-se portanto, criar uma dimensão para os dados da comunidade discente.

O BI de Ciências Exatas conta com 12 opções de cursos de segundo ciclo. Assim, é preciso gerenciar os dados do curso e determinar a dimensão curso.

Finalmente, a entrada de alunos é anual. Dessa maneira, deve existir uma dimensão temporal.

2) *Fonte de dados*: Inicialmente, é preciso realizar uma busca na fonte de dados e identificar com precisão as tabelas que serão úteis na construção do modelo. É importante ressaltar que os dados necessários devem ser identificados de maneira precisa, evitando assim que a necessidade de reconstruir o modelo após a fase de teste do cubo. Qualquer detalhe é decisivo para a efetividade da solução do problema.

A fase de análise da fonte de dados determina as tabelas necessárias a construção de cada uma das dimensões.

3) *ETL*: A fase ETL do processo de criação do cubo é iniciada após a verificação da conexão com a base fonte de dados. Para agrupar cada uma das transformações feitas para as dimensões, é criado um arquivo auxiliar chamado *job*. Quando é construída a tabela fato, um outro *job* deve ser formado. Finalmente, o agrupamento de todas as dimensões e a tabela fato é feito pelo *job* de criação do repositório de dados. A Figura 4 ilustra a construção da dimensão aluno. A Figura 5 apresenta a transformação da tabela fato.

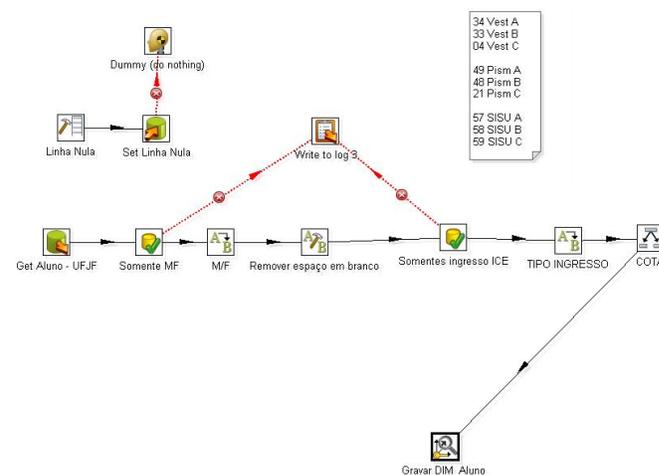


FIGURA 4. Transformação - Dimensão Aluno

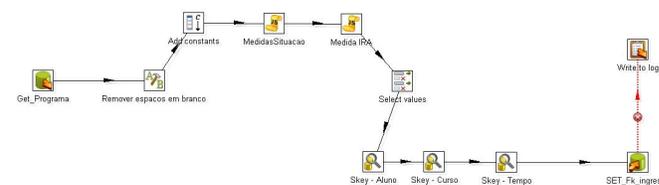


FIGURA 5. Transformação - Tabela Fatos

Durante a criação das dimensões é necessário executá-las algumas vezes afim de evitar erros. Não existe um número limite de dimensões, entretanto, existe apenas uma tabela

fato por modelagem de problema. Ao final do processo, o *job* que cria o repositório de dados deve ser executado. Caso não haja erros, as tabelas de dimensões e fatos são automaticamente transportadas para base de dados MySQL. O dados transportados são mostrados na Figura 6.

ID	NUM_CURSO	NUM_ALUNO	NUM_CICLO	STATUS	DATA_MATRICULA	DATA_CANCELAMENTO	DATA_CONCLUSAO	DATA_TRANSFERENCIA	DATA_JUBILACAO	DATA_MEDIACALCULO
1	1	1	1	1	2009-01-01					
2	1	2	1	1	2009-01-01					
3	1	3	1	1	2009-01-01					
4	1	4	1	1	2009-01-01					
5	1	5	1	1	2009-01-01					
6	1	6	1	1	2009-01-01					
7	1	7	1	1	2009-01-01					
8	1	8	1	1	2009-01-01					
9	1	9	1	1	2009-01-01					
10	1	10	1	1	2009-01-01					
11	1	11	1	1	2009-01-01					
12	1	12	1	1	2009-01-01					
13	1	13	1	1	2009-01-01					
14	1	14	1	1	2009-01-01					
15	1	15	1	1	2009-01-01					
16	1	16	1	1	2009-01-01					
17	1	17	1	1	2009-01-01					
18	1	18	1	1	2009-01-01					
19	1	19	1	1	2009-01-01					
20	1	20	1	1	2009-01-01					
21	1	21	1	1	2009-01-01					
22	1	22	1	1	2009-01-01					
23	1	23	1	1	2009-01-01					
24	1	24	1	1	2009-01-01					
25	1	25	1	1	2009-01-01					
26	1	26	1	1	2009-01-01					
27	1	27	1	1	2009-01-01					
28	1	28	1	1	2009-01-01					
29	1	29	1	1	2009-01-01					
30	1	30	1	1	2009-01-01					
31	1	31	1	1	2009-01-01					
32	1	32	1	1	2009-01-01					
33	1	33	1	1	2009-01-01					
34	1	34	1	1	2009-01-01					
35	1	35	1	1	2009-01-01					
36	1	36	1	1	2009-01-01					
37	1	37	1	1	2009-01-01					
38	1	38	1	1	2009-01-01					
39	1	39	1	1	2009-01-01					
40	1	40	1	1	2009-01-01					
41	1	41	1	1	2009-01-01					
42	1	42	1	1	2009-01-01					
43	1	43	1	1	2009-01-01					
44	1	44	1	1	2009-01-01					
45	1	45	1	1	2009-01-01					
46	1	46	1	1	2009-01-01					
47	1	47	1	1	2009-01-01					
48	1	48	1	1	2009-01-01					
49	1	49	1	1	2009-01-01					
50	1	50	1	1	2009-01-01					
51	1	51	1	1	2009-01-01					
52	1	52	1	1	2009-01-01					
53	1	53	1	1	2009-01-01					
54	1	54	1	1	2009-01-01					
55	1	55	1	1	2009-01-01					
56	1	56	1	1	2009-01-01					
57	1	57	1	1	2009-01-01					
58	1	58	1	1	2009-01-01					
59	1	59	1	1	2009-01-01					
60	1	60	1	1	2009-01-01					

FIGURA 6. Dados transferidos para o MySQL

Após a transferência dos dados, a nova base é adicionada ao PAC. O teste determina se os dados de entrada estão corretos. Só assim é possível iniciar a fase de criação do cubo. A Figura 7 mostra a interface da ferramenta de administração do SAD.

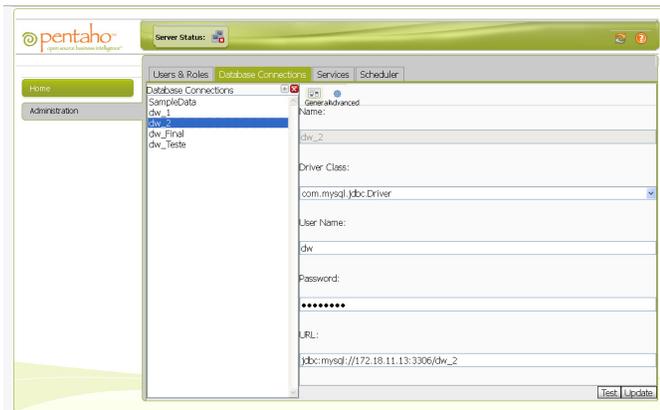


FIGURA 7. PAC

4) *Workbench - Criação do cubo*: Novamente, antes de iniciar o processo é verificado a conexão com o repositório de dados. A construção do cubo respeita os componentes das dimensões, bem como a tabela fato e a hierarquia existente nessas relações.

Essa construção permite avaliar quais tipos de perguntas podem ser respondidas pelo SAD de acordo com o problema. Embora o cubo não seja físico e não possua um número limitado de dimensões, trata-se o modelo como tal devido ao cruzamento de informações proporcionado. Esquemáticamente, esse cruzamento de informação é mostrado pela Figura 8

A tabela fato é o ponto central do cubo. Toda a construção é baseada nas características hierárquicas do cubo, respeitando sempre das dimensões estabelecidas. As dimensões devem ser

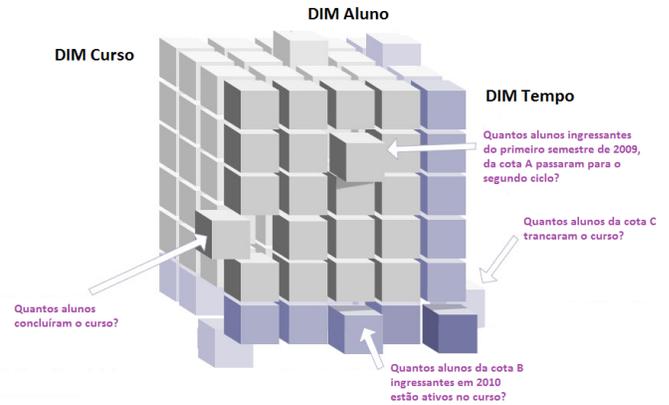


FIGURA 8. Cubo esquemático

ajustadas e conectadas a cada umas das hierarquias estabelecidas.

Configuradas as hierarquias do cubo, são estabelecidas as métricas de avaliação do problema. Nesse caso específico, as métricas envolvem avaliações quantitativas e cálculo da média do IRA. A ferramenta fornece atributos matemáticos capazes de ajudar nos cálculos de taxas, médias, somatórios, entre outros.

O modelo gerado para o cubo sistemas de ingresso é definido conforme a Figura 9.

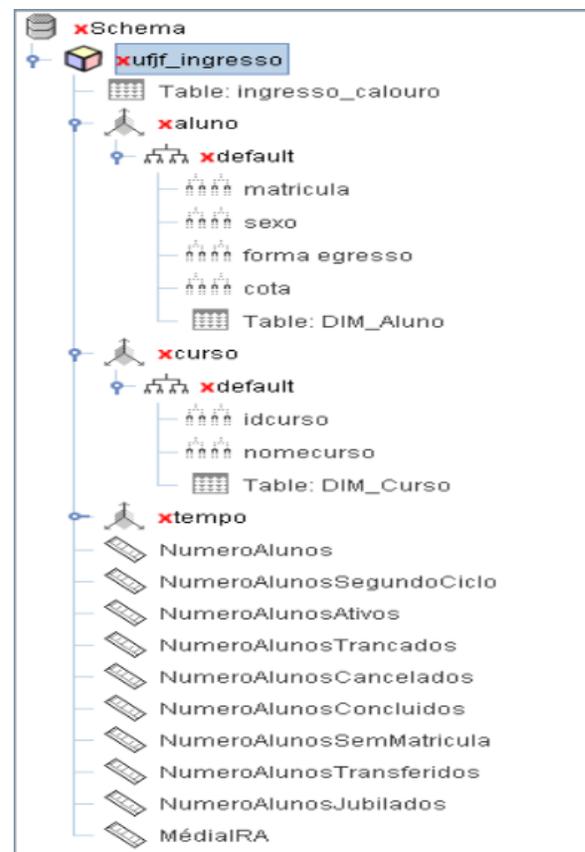


FIGURA 9. Cubo Sistemas de Ingresso

Esse modelo deve ser publicado no ambiente de usuário do servidor Pentaho, o PUC.

5) *PUC - Saiku*: A última etapa compreende a adequação do modelo ao servidor analítico, Saiku. Os passos de publicação são estabelecidos entre o Schema Workbench e o PUC conjuntamente.

Uma nova consulta analítica é iniciada a fim de identificar se os objetivos estabelecidos para o cubo foram alcançados. A partir daí, sucessivos testes são realizados para identificar o cumprimento da meta. A interface gráfica com o usuário é apresentado pela Figura 10.

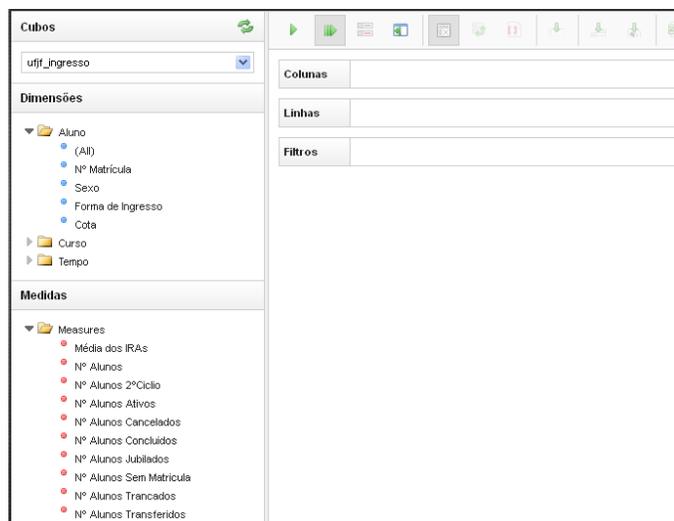


FIGURA 10. Cubo publicado

Toda consulta realizada pode ser salva no próprio servidor analítico e retomada posteriormente. Os resultados obtidos são apresentados por meio de gráficos, tabelas ou ambos, conforme a necessidade do usuário.

O funcionamento do PUC e, conseqüentemente, do Saiku depende diretamente do servidor do Pentaho, controlado pelo PAC. Por isso é necessário verificar frequentemente o status de conexão dessas ferramentas.

Estabelecidos os passos mencionados e realizados os respectivos testes, o cubo está pronto para ser disponibilizado aos usuários finais. A Figura 11 mostra uma consulta-exemplo ao cubo sistemas de ingresso.

B. Disciplinas Obrigatórias

O curso de BI em Ciências Exatas possui 17 disciplinas obrigatórias distribuídas em 4 semestres que devem ser necessariamente cumpridas para que o aluno possa concluir a formação de primeiro ciclo. Essas matérias não são suficientes para a conclusão do ciclo inicial, mas sem que o aluno tenha sido aprovado em todas elas, ele não poderá colar grau.

A escolha de segundo ciclo é outro fator dependente de aprovação em disciplinas. Nesse caso, o aluno em questão precisa ter concluído as primeiras 14 cadeiras, ou seja, ter cursado e ter sido aprovado nas matérias obrigatórias dos três primeiros períodos da graduação.

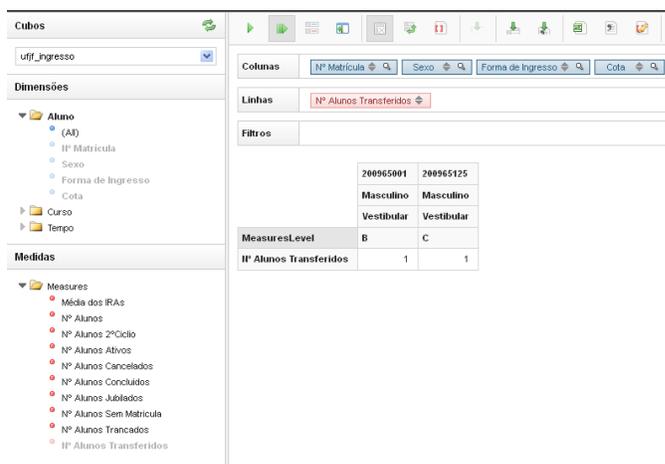


FIGURA 11. Consulta-exemplo

Nesse sentido, é de particular interesse para os coordenadores de curso avaliar o fluxo de alunos nas disciplinas básicas. É importante obter um levantamento dos discentes que eventualmente estarão aptos a fazer a escolha de segundo ciclo ao final do semestre. Além disso, o modelo possibilita identificar possíveis formandos a partir do cumprimento das cadeiras.

Devido ao elevado número de ingressantes existe mais de uma turma de cada uma das disciplinas obrigatórias. Os índices qualitativos, como taxa de aprovação e trancamento são também interessantes para o modelo determinado.

Em relação a distribuição das cadeira, o primeiro semestre da graduação possui 7 dos assuntos obrigatórios. Por sua vez, o segundo período conta com 4 e os terceiro e quarto semestres com 3 matérias de cunho obrigatório cada.

Determinadas as metas que devem ser atingidas, o próximo passo, novamente, é estabelecer as dimensões.

1) *Dimensões*: A dimensão temporal é de fundamental importância para a criação da maioria dos modelos. Nesse esquema, as disciplinas do BI são semestrais e a entrada de alunos anual, portanto, a dimensão tempo é indispensável.

Como a principal meta é avaliar os alunos em relação as disciplinas cursadas, está claro que as dimensões aluno e disciplina devem ser instituídas. Entretanto, nesse modelo as informações pertinentes ao aluno envolvem, além do número de matrícula, apenas o currículo relaciona e o curso atual e seu respectivo código.

Em relação a dimensão disciplina, devem ser restauradas dados sobre a turma, disciplina e docente responsável.

2) *Fonte de Dados*: A fonte de dados para esse cubo permanece a mesma, o SIGA. Entretanto, as tabelas exploradas respeitam as novas transformações abordadas. O mesmo cuidado deve ser tomado, a fim de evitar erros nas próximas etapas.

3) *ETL*: A conexão com a base de dados deve ser verificada novamente. A partir daí, passa-se para a construção das transformações e dos jobs.

Esse modelo conta com 6 dimensões além da tabela fato. Tem-se a dimensão aluno, tempo e uma dimensão para cada

semestre de disciplina obrigatória, ou seja, 4.

A transformação para as matérias do primeiro período da graduação e a tabela de fatos criação para esse modelo, são apresentadas, respectivamente, pela Figura 12 e Figura 13.

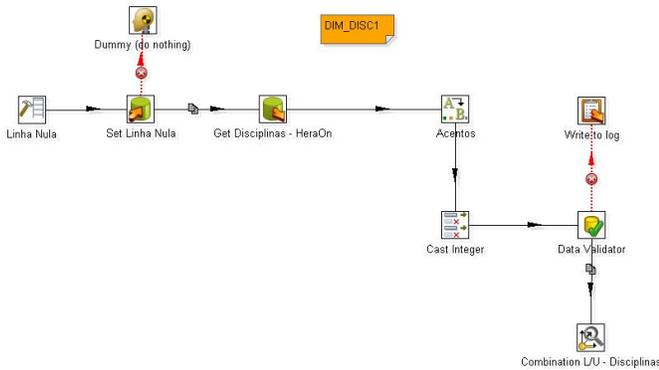


FIGURA 12. Transformação - Dimensão Disc1

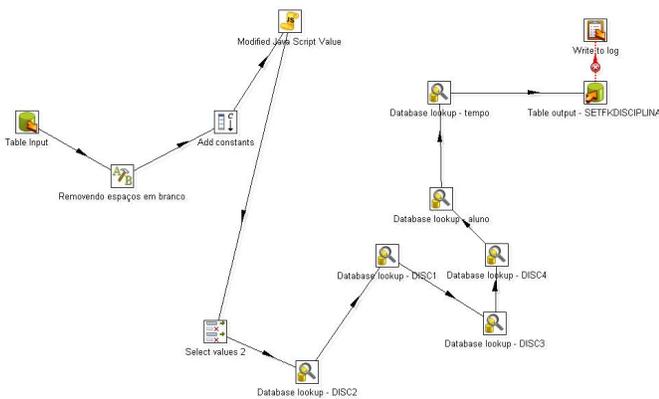


FIGURA 13. Transformação - Tabela Fatos

Determinadas as transformações e jobs para o modelo. Testadas as transformações, as informações são transportadas para a base de dados MySQL.

A nova base de dados é adicionada ao PAC e o processo de criação do cubo com o Workbench é iniciado.

4) *Workbench - Criação do cubo*: O passo inicial permanece após a verificação da base de dados.

As hierarquias são estabelecidas e conectadas com base das 6 dimensões e a tabela fatos.

As métricas para esse problema envolvem fatores quantitativos e taxas relacionadas a aprovação, reprovação por nota e por frequência, trancamentos, dispensas, entre outros.

A Figura 14 apresenta o modelo final gerado, pronto para publicação.

5) *PUC - Saiku*: O cubo disciplinas obrigatórias é então publicado no PUC e anexado ao Saiku. O servidor do Pentaho é reiniciado e posteriormente disponibilizado para os coordenadores de curso.

As Figuras 15 e Figura 16 apresentam, respectivamente, o cubo publicado e uma consulta-exemplo.

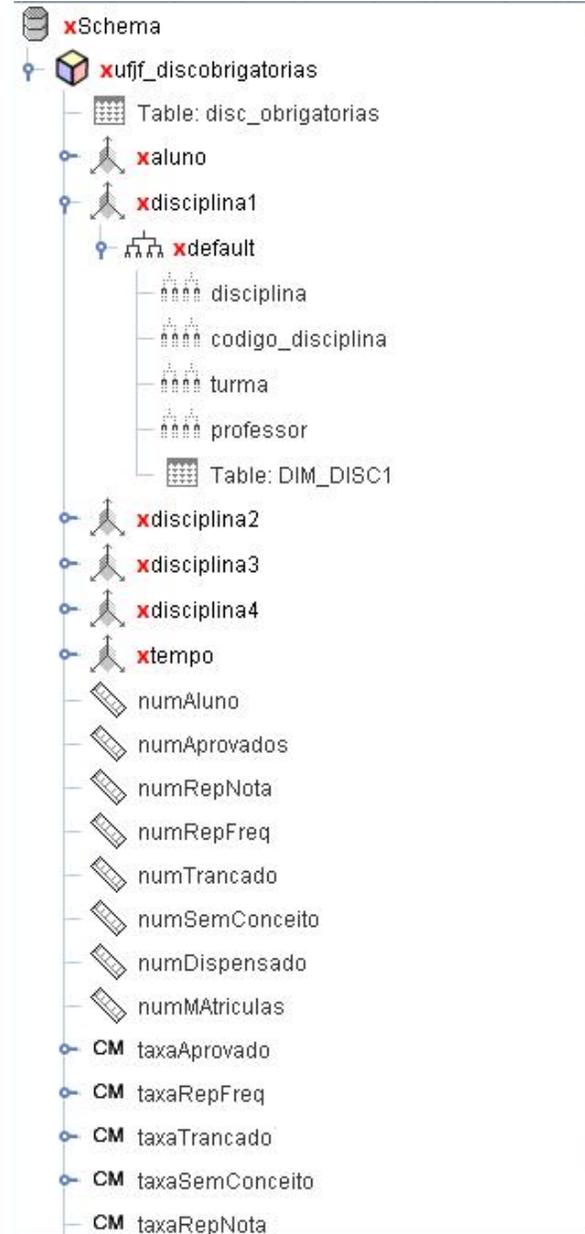


FIGURA 14. Cubo Disciplinas Obrigatórias

V. CONSIDERAÇÕES FINAIS

Através do presente trabalho, foi possível incorporar ações de mineração de dados na rotina dos coordenadores de curso da UFJF, apresentando informações nem sempre triviais para o controle acadêmico dos BIs. Ao todo, cinco cubos foram implementados e disponibilizados para auxiliar os coordenadores do BI de Ciências Exatas. O sistema está disponível para acesso a partir de login e senha.

O processo de criação de cubos, embora trabalhoso, não inviabiliza a implantação do sistema, levando cerca de 7 dias para implementação de um novo cubo. Os cubos são criados conforme a necessidade dos coordenadores.

Os dados quando transferidos, são armazenados na base MySQL. Essa, por sua vez, não é atualizada de maneira

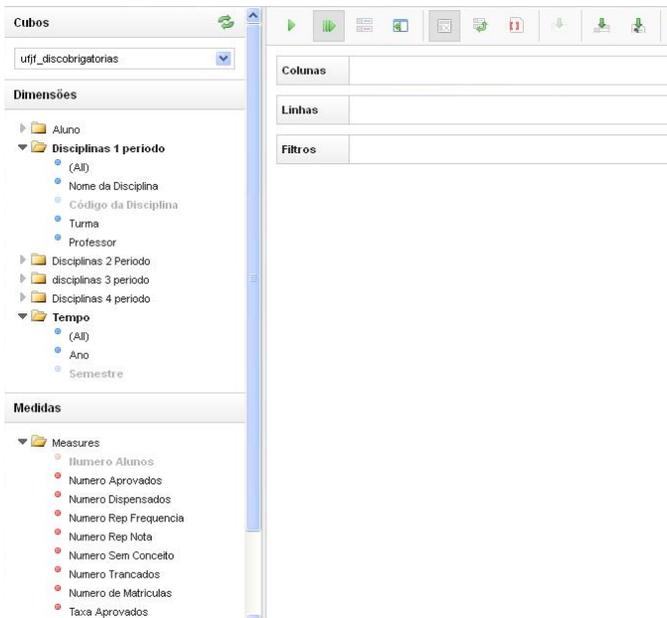


FIGURA 15. Cubo publicado

Medida est. Local	ALGORITMOS	CALCULO I	GEOMETRIA ANALITICA E SISTEMAS LINEARES	LABORATORIO DE PROGRAMACAO	QUIMICA FUNDAMENTAL
Numero Alunos	1.587	1.639	1.844	1.584	1.270

FIGURA 16. Consulta-exemplo

automática. Assim, é necessário uma atualização a cada seis meses, conforme os dados da base ORACLE são atualizados.

Dessa forma, sugere-se como passo futuro a criação de uma rotina de atualização automática da base ORACLE para a MySQL. Os dados atualizados são transmitidos aos cubos quando atualizadas as conexões de base de dados do PAC. Além disso, e reinicialização do servidor Pentaho, atualiza os cubos para o PUC.

Outro projeto futuro é promover uma análise quantitativa do sistema. Essa análise será baseada na medição do tempo necessário para a realização de cada ação e consulta. Se necessário, possíveis reestruturações na base de dados serão administradas.

Atualmente, somente a coordenação do curso de Ciências Exatas tem acesso pleno ao sistema criado. O projeto deverá ser estendido aos outros cursos de Bacharelados Interdisciplinares da UFJF.

REFERÊNCIAS

- [1] REUNI, Reestruturação e Expansão das Universidades Federais. [Online; Acessado em 9 de Maio de 2014]. <http://reuni.mec.gov.br>
- [2] Pró-reitoria de Graduação, Universidade Federal de Juiz de Fora. [Online; Acessado em 9 de Maio de 2014]. <http://www.ufjf.br/arquivos/planoUfjf.pdf>
- [3] Secretaria de Comunicação, Universidade Federal de Juiz de Fora. [Online; Acessado em 9 de Maio de 2014]. <http://www.ufjf.br/secom/2013/01/10/entenda-como-funcionam-os-bacharelados-interdisciplinares-da-ufjf/>
- [4] R. Bouman and J.V. Dongen, *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL*. Wiley Publishing, 2009.
- [5] LUPATINI, G. *Desenvolvimento de um sistema de apoio à decisão em escolha de áreas para aterro sanitário*. Dissertação de Mestrado em Engenharia Ambiental. Programa de Pós Graduação em Engenharia Ambiental, Universidade Federal de Santa Catarina, 2002.
- [6] G. Lachtermacher. *Pesquisa Operacional na Tomada de Decisões*. Editora Campus, 2002.
- [7] R. H. Sprague and E. D. Carlson, *Building Effective Decision Support Systems*. Prentice-Hall, 1982.
- [8] J. A. O'Brien, *Sistemas de informação e as decisões gerenciais na era da Internet*. Editora Saraiva, 2001.
- [9] Oracle. [Online; Acessado em 9 de Maio de 2014]. <http://www.oracle.com/index.html>
- [10] M. Casters, R. Bouman and J. V. Dongen, *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration*. Wiley Publishing, 2010.
- [11] M. C. Roldan, *Pentaho 3.2 Data Integration: Beginner's Guide*. Wiley Publishing, 2010.
- [12] C. G Martins, R. A. Stefan and L. B. Martins, *Data Warehouse*. UFRJ, 1997.
- [13] Pentaho Administration Console. Pentaho Business Intelligence. [Online; Acessado em 9 de Maio de 2014]. <http://wiki.pentaho.com/display/ServerDoc2x/The+Pentaho+Administration+Console>
- [14] Documentação Workbench. Pentaho Business Intelligence. [Online; Acessado em 9 de Maio de 2014]. <http://mondrian.pentaho.com/documentation/workbench.php>
- [15] Pentaho User Console Guide. Pentaho Business Intelligence. <http://infocenter.pentaho.com/help48/topic/puc-user-guide/puc-user-guide.pdf>
- [16] Saiku Analytics. [Online; Acessado em 9 de Maio de 2014]. <http://www.meteorite.bi/saiku>