

Análise de modelos dimensionais: um estudo de caso com dados do sistema de processo seletivo do IFET

Leandro Torres Santana, Evaldo de Oliveira da Silva

Bacharelado em Sistemas de Informação - Centro de Ensino Superior de Juiz de Fora
(CES/JF) – Juiz de Fora – MG – Brasil

leandrotorres.santana@gmail.com, evaldo.oliveira@gmail.com

***Resumo.** O objetivo deste trabalho é mostrar como os dados armazenados em sistemas transacionais de uma organização podem ser modelados e analisados através de técnicas de Business Intelligence, com o intuito de apresentar aos gestores informações para tomada de decisões em nível gerencial. Além disso, visando enriquecer o acesso às informações gerenciais, este trabalho também aborda os conceitos e a implementação de soluções para Business Intelligence usando cloud computing. Finalmente, é apresentado um estudo de caso real com dados do Sistema de Processo Seletivo do Instituto Federal de Educação Ciência e Tecnologia (IFET), onde serão extraídos indicadores e disponibilizados em nuvem com o auxílio de um software, analisado e escolhido previamente.*

***Abstract.** The objective of this paper is to show how the data stored in an organization's transactional systems can be modeled and analyzed using technology of Business Intelligence in order to provide information to managers for decision-making at management level. In addition, in order to enrich access to management information, this paper also discusses the concepts and implementation of solutions for Business Intelligence using cloud computing. Finally, is presented a real case study with data from the Selection Process System of Instituto Federal de Educação Ciência e Tecnologia (IFET), which will be extracted indicators and available in the cloud with the help of a software, analyzed and previously chosen.*

1. Introdução

As organizações necessitam cada vez mais de informações que as auxiliem na tomada de decisões estratégicas, e em muitos casos, esses dados precisam estar disponíveis em qualquer tempo e lugar. As empresas com suas respectivas competências, fazem a geração de seus modelos de conhecimento e acumulam essa nova forma de informações em bases de dados acessadas e divulgadas [Barbiere 2011].

Para que essas informações possam ser geradas, as organizações criam processos empresariais, onde através de toda a cadeia do negócio são coletados dados através de sistemas transacionais, que são armazenados em bancos de dados. Uma vez esses dados armazenados, eles podem ser modelados e transformados para posteriormente serem analisados através de regras previamente definidas, gerando informações concisas que auxiliam os gestores a tomar decisões voltadas a estratégia do negócio.

Atualmente com a grande expansão da internet, o acesso a informação ficou muito mais simples. Hoje, no mercado, com a vasta gama de PCs, *notebooks*, *tablets*, *smartphones* e muitos outros dispositivos que se conectam a grande rede é possível acessar os mais variados sistemas e bases de informações. Muitos destes sistemas são acessados usando tecnologias baseadas em computação em nuvem, ou *cloud computing*.

Segundo [Prado 2010], a característica principal do *cloud computing* é ter todas as informações e aplicativos salvos e sendo executados a partir de um lugar na internet (nuvem). Com esse grande acesso a informação, um gestor pode ter informações sobre a organização mesmo que esteja em viagem para outro país, concedendo-o mais flexibilidade para tomar decisões sem limitações geográficas.

O texto deste artigo encontra-se organizado em 5 seções, onde: a seção 1 faz a introdução ao assunto. A seção 2 detalha os aspectos teóricos sobre *Business Intelligence* (BI) e *cloud computing*, que fundamentam a solução proposta; a seção 3 mostra o exemplo de 3 ferramentas existentes no mercado que realizam a análise de dados dimensionais e trabalham com o conceito de *cloud computing*; a seção 4 apresenta os resultados obtidos com a aplicação do trabalho em um estudo de caso real utilizando a base de dados do IFET (Instituto Federal de Educação Ciência e Tecnologia), onde através de dados reais do Sistema de Processo Seletivo, um conjunto de informações gerenciais é modelado com base em regras específicas e apresentado utilizando uma ferramenta específica; e a seção 5 apresenta a conclusão do trabalho.

2. Referencial Teórico

Esta seção apresenta os principais elementos teóricos que foram utilizados para elaborar a solução proposta. Nele serão abordados os conceitos das áreas de BI, *cloud computing* e *Cloud BI*.

2.1. Business Intelligence

A boa gestão da informação é fundamental para o sucesso das empresas. Através dos processos operacionais das organizações, são gerados uma grande quantidade de dados que por sua vez ficam armazenados em bases de dados transacionais, e em muitos casos, esses dados se encontram fragmentados em diversas dessas bases, o que dificulta a obtenção de informações a partir dessas fontes. Ao examinarem os dados, situações e desempenhos históricos e atuais, os tomadores de decisão conseguem uma melhor compreensão dos dados, que podem servir como base para decisões melhores e mais informadas. O processo de BI baseia-se na transformação de dados em informações [TURBAN et. al 2009].

BI, ou Inteligência de Negócios, é um termo criado pela *Gartner Group* na década de 1980. De acordo com [Oliveira 2009], BI é um conceito empregado a ferramentas, tecnologias e metodologias, que tem como objetivo fornecer informações estratégicas, que apoiam na tomada de decisões. Através desse conceito, são descritas técnicas que as corporações podem utilizar para acessar dados e modelar informações importantes, analisando-as e desenvolvendo percepções sobre a gestão organizacional, o que as permitem tornar mais eficiente a tomada de decisão. Com o registro de todas as informações de negócios, a empresa terá através do BI uma fonte de informação. Na Figura 1 é possível ver o *workflow* do processo de BI.

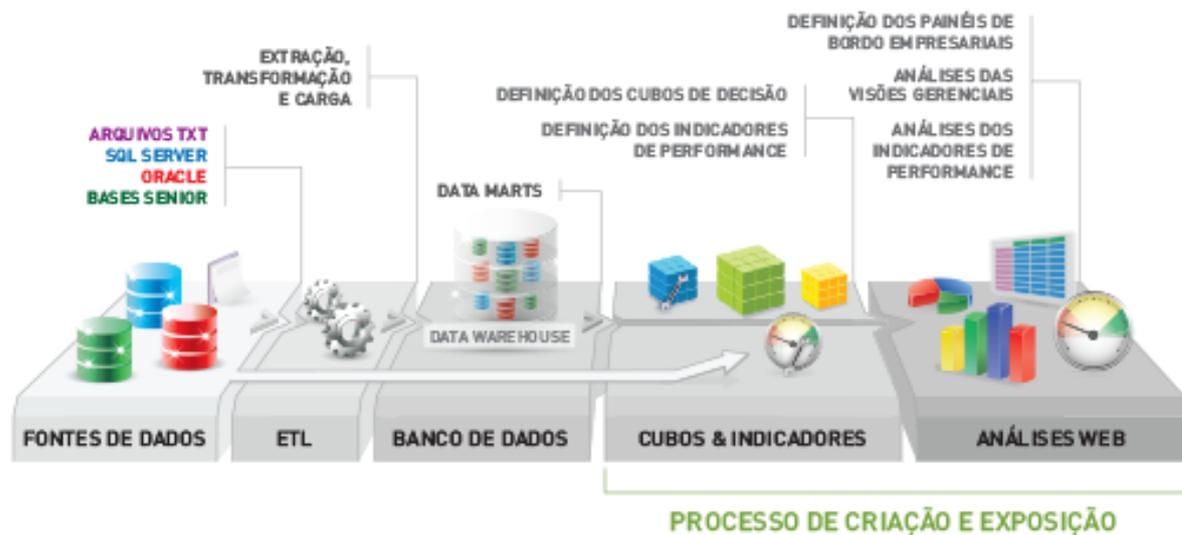


Figura 1: Workflow do processo de BI.

A seguir serão apresentados conceitos importantes para o entendimento do processo de BI:

Data Warehouse (DW): Os dados das bases de dados dos sistemas transacionais de uma empresa não apresentam boas características para a realização de consultas analíticas. Para resolver essa limitação são criados os chamados DW. Um DW consiste em um banco de dados contendo dados extraídos dos diversos sistemas do ambiente de produção de uma organização. Esses dados são selecionados e depurados, e então otimizados para processamento de consultas analíticas, deixando de ter características para o processamento de transações.

Processo de Extração, Transformação e Carga de Dados - *Extract, Transform and Load (ETL)*: Em uma base DW, os dados inicialmente são extraídos de fontes externas, geralmente de sistemas transacionais, antes de serem carregados no DW. Após a extração e a carga desses dados, os usuários acessam o DW através de ferramentas analíticas, de modo a obterem informações que permitam a tomada de decisões. Um DW contém dados sumarizados, históricos e detalhados para suportar a tomada de decisões táticas e estratégicas. A seguir são detalhadas as etapas desse processo:

- **Extração:** A primeira etapa na obtenção de dados para o ambiente do DW é a extração. Essa etapa significa ler e entender as fontes de dados e copiar as partes necessárias para a área de transformação de dados, a fim de serem trabalhadas posteriormente. Na grande maioria dos DW, os dados provêm de várias fontes diferentes e independentes. A extração pode ser conduzida através da construção de programas cujo código é executado sobre um sistema fonte de modo a gerar arquivos com os dados desejados. Outra opção é utilizar ferramentas de extração específicas que geram código próprio, interno à ferramenta, executado sobre o sistema fonte, de forma a obter os dados necessários.

- **Transformação dos Dados:** Após a extração dos dados dos sistemas fontes, passa-se à realização de uma série de atividades sobre esses dados, com o intuito de convertê-los em formato adequado para carga no DW e valioso para a organização. Na transformação dos dados estão envolvidos um ou vários processos, dependendo da necessidade e situação. Esses processos são: limpeza, eliminação, combinação, desnormalização, cálculos, derivação e alocação.
- **Carga de Dados:** Na última etapa do processo ETL, logo após os dados serem transformados, eles são carregados no DW. A parte de carga dos dados possui uma enorme complexidade no processo, e os seguintes fatores devem ser levados em conta ao realizar essa etapa: integridade dos dados, tipos de carga a serem realizadas, otimização do processo de carga, suporte completo ao processo de carga.

Modelagem de Dados (Modelo Relacional X Modelo Dimensional): A elaboração do modelo de dados deve se concentrar na observação dos fatos relevantes que ocorrem na realidade das organizações, com a finalidade de construir um sistema que possa atender as necessidades de informação da mesma.

Ao se retratar as necessidades de informação que uma organização possui, é preciso alcançar os objetivos desta mesma realidade. Ao se coletar e selecionar os fatos relevantes, é necessário identificar os elementos geradores de informação (dados). É preciso também fazer uso de um modelo, ou seja, algo que nos mostre como as informações estão relacionadas (fatos). E, com base no modelo criado, é possível interagir com os usuários validando seus objetivos e metas, permitindo a construção de um sistema de informações próximo da realidade.

Os modelos de dados podem ser classificados segundo a arquitetura que utilizam. O modelo relacional surgiu para atender os sistemas transacionais *On-line Transaction Processing* (OLTP), já o modelo dimensional surgiu para atender os sistemas analíticos *On-line Analytical Processing* (OLAP). As diferenças entre esses dois tipos de sistemas podem ser vistas na Tabela 1.

Para definir um projeto de DW é necessário ter em mente os conceitos de modelagem dimensional, ou multidimensional, que compreende um sistema baseado em recuperação de dados que suporta acessos com alto volume de consultas. A representação dos dados é estruturada como um cubo, transmitindo a ideia de múltiplas dimensões. A maior vantagem da multidimensionalidade é que ela permite que os dados sejam organizados de acordo com a preferência de cada usuário dentro de uma organização, dando liberdade as áreas de negócio de fazerem as consultas conforme suas reais necessidades.

O modelo dimensional requer a utilização de ferramentas de consultas analíticas, desenvolvidas especialmente para consultar esse tipo de modelo, o que permite aos usuários a exploração de todos os dados disponíveis durante a elaboração das consultas.

Tabela 1: Tabela comparativa entre sistemas OLAP e OLTP.

	OLAP	OLTP
Foco	Foco no nível estratégico da organização. Visa a análise empresarial e tomada de decisão.	Foco no nível operacional da organização. Visa a execução operacional do negócio.
Performance	Otimização para a leitura e geração de análises e relatórios gerenciais.	Alta velocidade na manipulação de dados operacionais, porém ineficiente para geração de análises gerenciais.
Estrutura dos dados	Os dados estão estruturados na modelagem dimensional. Os dados normalmente possuem alto nível de sumarização.	Os dados são normalmente estruturados em um modelo relacional normalizado, otimizado para a utilização transacional. Os dados possuem alto nível de detalhes.
Armazenamento	O armazenamento é feito em estruturas de <i>Data Warehouse</i> com otimização no desempenho em grandes volumes de dados.	O armazenamento é feito em sistemas convencionais de banco de dados através dos sistemas de informações da organização.
Abrangência	É utilizado pelos gestores e analistas para a tomada de decisão.	É utilizado por técnicos e analistas e engloba vários usuários da organização.
Frequência de atualização	A atualização das informações é feita no processo de carga dos dados. Frequência baixa, podendo ser diária, semanal, mensal ou anual (ou critério específico).	A atualização dos dados é feita no momento da transação. Frequência muito alta de atualizações.
Volatilidade	Dados históricos e não voláteis. Os dados não sofrem alterações, salvo necessidades específicas (por motivos de erros ou inconsistências de informações).	Dados voláteis, passíveis de modificação e exclusão.
Tipos de permissões nos dados	É permitido apenas a inserção e leitura. Sendo que para o usuário está apenas disponível a leitura.	Podem ser feito leitura, inserção, modificação e exclusão dos dados.

O modelo multidimensional é baseado, segundo [Reis et. al 2009], em três tipos de estruturas:

- fatos – a tabela de fatos é a tabela central do modelo e contém os valores (numéricos) que se deseja analisar, geralmente, contendo um grande volume de dados. A tabela fato possui chaves externas, que se relacionam com suas tabelas de dimensões, e campos numéricos que são os valores (medidas) que serão analisados;
- dimensões – as tabelas de dimensões representam um aspecto do negócio que está sendo analisado. Sua chave primária serve para manter a integridade referencial na tabela fato à qual está relacionada. Uma dimensão oferece ao usuário um grande número de combinações e intersecções para analisar os dados, possibilitando diversas formas de visualizar os dados;
- medidas – são atributos numéricos armazenados na tabela de fatos, que representam o desempenho de um indicador em relação às dimensões que participam desse fato.

Com relação aos tipos de modelos dimensionais, [Ballard et. al 2006], descreve três:

- esquema estrela – cada dimensão é formada por apenas uma tabela não padronizada. O fato de suas tabelas não estarem normalizadas, resultará em uma menor quantidade de tabelas no modelo do DW, mas como consequência poderá causar redundância dos dados, fato característico neste tipo de solução;
- esquema multi-estrela – este esquema é baseado no esquema estrela, com uma tabela para cada dimensão, porém existem múltiplas tabelas de fatos, unidas através das dimensões;
- esquema floco de neve – neste esquema as tabelas das dimensões estão padronizadas para eliminar a redundância dos dados. Diferente do esquema estrela, neste esquema os dados das dimensões estão distribuídos em múltiplas tabelas.

Para o trabalho será dado ênfase no esquema estrela, que será utilizado para modelar os dados do estudo de caso. Por conta disso, a seguir será feito um maior detalhamento deste esquema.

Esquema Estrela – Este esquema recebe o nome de "estrela", devido a disposição das tabelas no modelo, que consiste de uma tabela central, a tabela de fatos, que se relaciona com diversas outras tabelas, chamadas de dimensão. De acordo com [Poe et. al 1998], o esquema estrela é uma estrutura simples, com poucas tabelas e ligações (relacionamentos) bem definidas.

Este esquema se assemelha ao modelo de negócio das organizações, facilitando assim sua leitura e entendimento, não apenas pelos especialistas, como também pelos usuários finais não familiarizados com estruturas de banco de dados. Permite a criação de um banco de dados que facilita a execução de consultas complexas, podendo ser realizadas de modo eficiente e intuitivo pelo usuário.

O esquema estrela pode representar tanto o modelo lógico, como o modelo físico do banco de dados. A representação mais simples de um modelo dimensional contém um esquema estrela com uma tabela de fatos relacionada com tabelas de dimensão. Uma representação do esquema estrela pode ser visto na Figura 2 [Machado 2004].

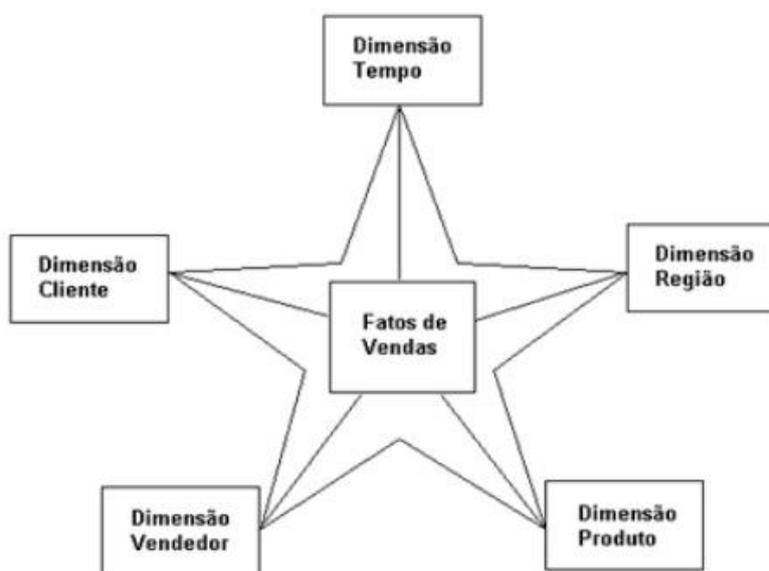


Figura 2: Modelo do Esquema Estrela.

2.2. Cloud Computing e Cloud BI

Cloud computing, ou computação em nuvem em português, é uma expressão relativamente nova, começou a ser muito utilizada a partir de 2008, apesar das ideias por trás da denominação existirem a muito mais tempo. O *National Institute of Standards and Technology* (NIST) define a computação em nuvem como um modelo que possibilita acesso, de modo conveniente e sob demanda, a um conjunto de recursos computacionais configuráveis (por exemplo, redes, servidores, armazenamento, aplicações e serviços) que podem ser rapidamente adquiridos e liberados com mínimo esforço gerencial ou interação com o provedor de serviços [NIST 2009].

Quando se fala em computação em nuvem, fala-se na possibilidade de acessar arquivos e executar diferentes tarefas pela internet. Através desse conceito, é dispensada a instalação de aplicativos nos computadores ou dispositivos móveis, pois é possível acessar diferentes serviços online para fazer o que precisa, já que os dados não se encontram em um computador específico, mas sim em uma rede.

Diversos tipos de aplicações baseadas em computação em nuvem estão disponíveis atualmente. Serviços de *webmail*, *sites*, *softwares* como serviços em geral são alguns exemplos. Outras aplicações que estão sendo muito utilizadas com o conceito de *cloud computing* são as ferramentas de BI. Estas ferramentas necessitam de grande poder de processamento, material humano muito qualificado para a implantação, segurança das informações, alta disponibilidade dos dados, dentre outras características. À junção das tecnologias de BI e *cloud computing* damos o nome de *cloud BI* [Walczak 2014].

Cloud BI é a disponibilização das ferramentas de BI na nuvem. Com isso, grandes empresas de computação como IBM, Microsoft, SAP, dentre outras, investem no desenvolvimento desses aplicativos e na manutenção de grandes *Data Centers* para comportarem grande volume de dados e com alta disponibilidade. A partir disso, essas empresas vendem essas ferramentas como serviço para outras organizações que não possuem condições de manter esse tipo de infraestrutura [Walczak 2014]. Na próxima seção, serão apresentadas ferramentas que oferecem esse tipo de serviço.

3. Ferramentas para análise de dados em nuvem

Para o desenvolvimento do trabalho foram pesquisadas ferramentas existentes no mercado que trabalham com os conceitos de BI e *cloud computing*. A partir dessa análise, uma comparação entre essas ferramentas foi realizada a fim de chegar a uma conclusão sobre qual delas será utilizada no estudo de caso que será proposto no trabalho.

O intuito desta comparação é de mostrar que existem ferramentas no mercado que realizam a tarefa de analisar os dados e disponibilizar as informações na nuvem para acesso a qualquer tempo e em qualquer lugar, e de descrever como é o funcionamento de uma dessas ferramentas através da aplicação dela no estudo de caso. Foram encontradas diversas ferramentas, mas, para o artigo, foram escolhidas 3 das mais importantes citadas em pesquisas na internet. A seguir é apresentado uma breve descrição dessas ferramentas e, na Tabela 2, é feita uma comparação entre elas.

3.1. Microsoft Power BI

A ferramenta *Microsoft Power BI* é um serviço de análise de negócios baseados em nuvem, projetado para atender a demanda de informações por parte de usuários corporativos. Os usuários dessa ferramenta contam com vários recursos que permitem a eles manter o controle de seus negócios por meio de painéis de indicadores operacionais em tempo real através da *web* ou do aplicativo móvel *Microsoft Power BI*. Além desses indicadores, os usuários desta ferramenta contam com relatórios visuais interativos, os quais podem ser expandidos através da conexão com fontes de dados adicionais. O *Microsoft Power BI* também conta com uma ferramenta de criação de relatórios que pode ser explorada livremente pelos usuários [Microsoft 2015].

3.2. SAP Lumira Cloud

O software *SAP Lumira Cloud* utiliza o conceito de visualização de dados para permitir às organizações tirar o máximo de informações das enormes quantidades de dados disponíveis em seus sistemas.

A ferramenta permite visualizar rapidamente grandes volumes de dados sem ter que sacrificar o desempenho ou a segurança. Ela é projetada para acelerar o tempo de análise dos dados, faz com que os usuários possam descobrir prognósticos e extrair informações através do navegador ou dispositivos móveis. A versão gratuita oferece 1GB de armazenamento. A carga de dados é feita através de planilhas, e após essa carga, é possível utilizar recursos de colaboração com outros usuários, através do compartilhamento de conjuntos de dados e análise de informações [SAP 2015].

3.3. IBM Watson Analytics

A ferramenta *IBM Watson Analytics* nasceu de um projeto da IBM para um programa de perguntas e respostas da TV americana chamado *Jeopardy*. Trata-se de um supercomputador capaz de processar e analisar linguagem natural.

A ferramenta possui uma versão gratuita do seu serviço baseado em nuvem, projetada para rodar na *web* e em dispositivos móveis. O *software* pode se conectar a diversas fontes de dados de uma organização permitindo refinar esses dados, descobrir ideias, prever os resultados, visualizar os resultados, criar relatórios ou indicadores operacionais, e colaborar com outras pessoas, por exemplo, de um mesmo setor [IBM 2015].

Tabela 2: Tabela comparativa entre as ferramentas de análise de dados.

Ferramenta	Microsoft Power BI	SAP Lumira Cloud	IBM Watson Analytics
Característica			
Possui Versão Gratuita	SIM	SIM	SIM
Espaço Disponível em Nuvem	1GB	1GB	500MB
Visualização em Dispositivos Móveis	SIM	SIM	SIM
Facilidade de Obtenção	Cadastro e Requisição da Ferramenta	Cadastro e Requisição da Ferramenta	Cadastro Simples

Através da comparação realizada, foi escolhida a ferramenta *IBM Watson Analytics* para ser utilizada no estudo de caso a ser desenvolvido, pois foi a ferramenta que apresentou mais facilidade de obtenção e utilização.

4. Estudo de caso (IFET)

O Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais (IF Sudeste MG) é uma autarquia ligada ao Ministério da Educação (MEC). É uma instituição que articula as educações superior, básica e profissional, pluricurricular e multicampi, especializada na oferta de educação profissional e tecnológica nas diferentes modalidades de ensino, com base na conjugação de conhecimentos técnicos e tecnológicos.

4.1. Problema

Um dos sistemas de maior importância para o IFET atualmente é o sistema de processo seletivo. Neste sistema é realizada grande parte da gestão do processo de seleção de novos alunos para os diversos cursos existentes na instituição, do ensino médio à pós-graduação. Desse sistema são solicitados, principalmente pela Pró-Reitoria de ensino, informações gerenciais sobre cada processo seletivo que ocorre no instituto. Atualmente ocorrem dois processos seletivos por ano, e ter informações gerenciais sobre eles é de fundamental importância para o IFET.

O IFET não conta hoje com um DW de onde poderia se extrair informações de forma mais fácil e rápida, devido ao melhor desempenho desse tipo de base de dados para pesquisas analíticas. Em razão dessa limitação, as informações, indicadores e relatórios importantes para a instituição são extraídos diretamente da base de dados transacional, o que faz com que as consultas fiquem lentas, dificultando o acesso aos dados.

Em virtude dessa limitação, é proposto neste trabalho um estudo de caso onde serão realizados os procedimentos de modelagem de uma base de dados analítica. Para o estudo, primeiramente serão definidos dois indicadores considerados importantes pela administração do IFET. Após a definição dos dois indicadores, será mostrado como modelar os dados para extraí-los da base de dados transacional para o modelo dimensional. Por fim, após a modelagem dimensional dos dados, será mostrado como esses dados podem ser analisados, com a finalidade de gerar informação gerencial, através do auxílio de uma ferramenta específica para essa finalidade, que como foi visto na seção anterior, a ferramenta escolhida para o estudo de caso é a *IBM Watson Analytics* devido a maior facilidade de utilização e obtenção da versão gratuita.

Para a realização desse estudo de caso, foi obtida uma autorização de uso dos dados do sistema de processo seletivo do IFET, exceto dados de informações pessoais e de caráter sigiloso da instituição. Essa autorização foi assinada pelo diretor de tecnologia da informação e pelo coordenador de sistemas de informação da instituição.

4.2. Definição dos Indicadores

Para o trabalho foram definidos dois indicadores provenientes dos dados do sistema de processo seletivo considerados importantes para o IFET. Esses indicadores são os seguintes:

- Relação Candidato / Vaga (RCV) – Trata-se da relação de candidatos inscritos dividido pelo número de vagas ofertadas.
- Conversão Aprovados Matriculados (CAM) – É a porcentagem de matrículas efetivadas sobre o número de candidatos aprovados.

4.3. Modelagem dos Dados

Nessa etapa do trabalho é feita a modelagem dos dados com a finalidade de transformar a estrutura da base de dados do modelo relacional para o modelo dimensional.

Relação Candidato / Vaga (RCV)

Primeiramente é identificado no banco de dados relacional os campos e as tabelas necessárias para gerar o indicador em questão. A seguir são apresentadas essas informações:

- Ano (ps_candidato);
- Semestre (ps_candidato);
- Câmpus (ps_candidato, ps_processo_seletivo_has_curso, glo_curso, glo_campus);
- Curso (ps_candidato, ps_processo_seletivo_has_curso, glo_curso);
- Inscrições Realizadas;
- Vagas Ofertadas (ps_processo_seletivo_has_curso);
- Relação Candidato / Vaga (Inscrições Realizadas / Vagas Ofertadas)

Na Figura 4 é mostrado o modelo relacional com as tabelas utilizadas para a geração do indicador:



Figura 4: Modelo relacional para a geração do indicador Relação Candidato / Vaga.

Definido, no modelo relacional, de onde os dados serão extraídos, então é desenhado o modelo dimensional. No trabalho, para a modelagem dimensional, foi utilizado o esquema estrela. O desenho do modelo dimensional, com a tabela fatos e suas dimensões, pode ser visto na Figura 5:

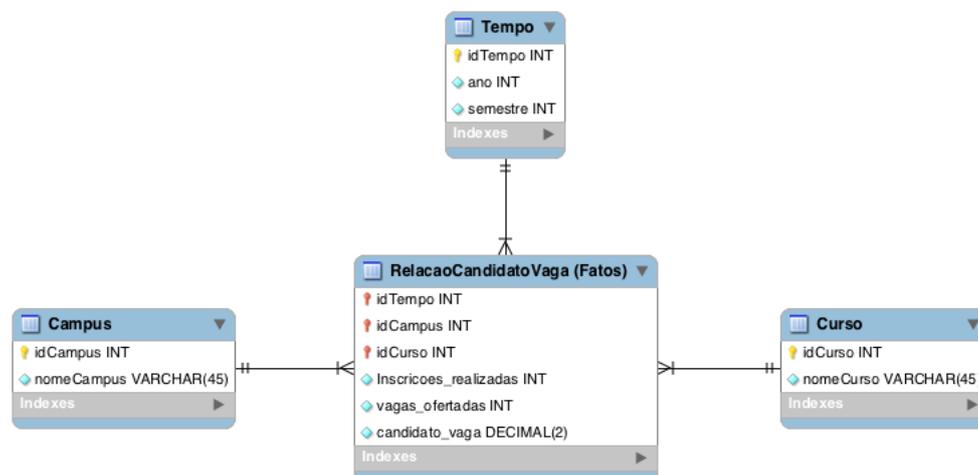


Figura 5: Modelo dimensional para a geração do indicador Relação Candidato / Vaga. Esquema estrela.

Conversão Aprovados Matriculados (CAM)

Assim como no indicador (RCV), é identificado no banco de dados relacional as tabelas e dos campos necessários para gerar o indicador em questão.

- Ano (ps_candidato),
- Semestre (ps_candidato),
- Câmpus (ps_candidato, ps_processo_seletivo_has_curso, glo_curso, glo_campus),
- Curso (ps_candidato, ps_processo_seletivo_has_curso, glo_curso),
- Candidatos Aprovados (ps_matricula),
- Matrículas Efetivas (ps_matricula, ps_matricula_situacao),
- Conversão Candidatos Aprovados / Matrículas Efetivas

Na Figura 6 é mostrado o modelo relacional com as tabelas utilizadas para a geração do indicador.

Definido, no modelo relacional de onde os dados serão extraídos, então é desenhado o modelo dimensional que pode ser visto na Figura 7.

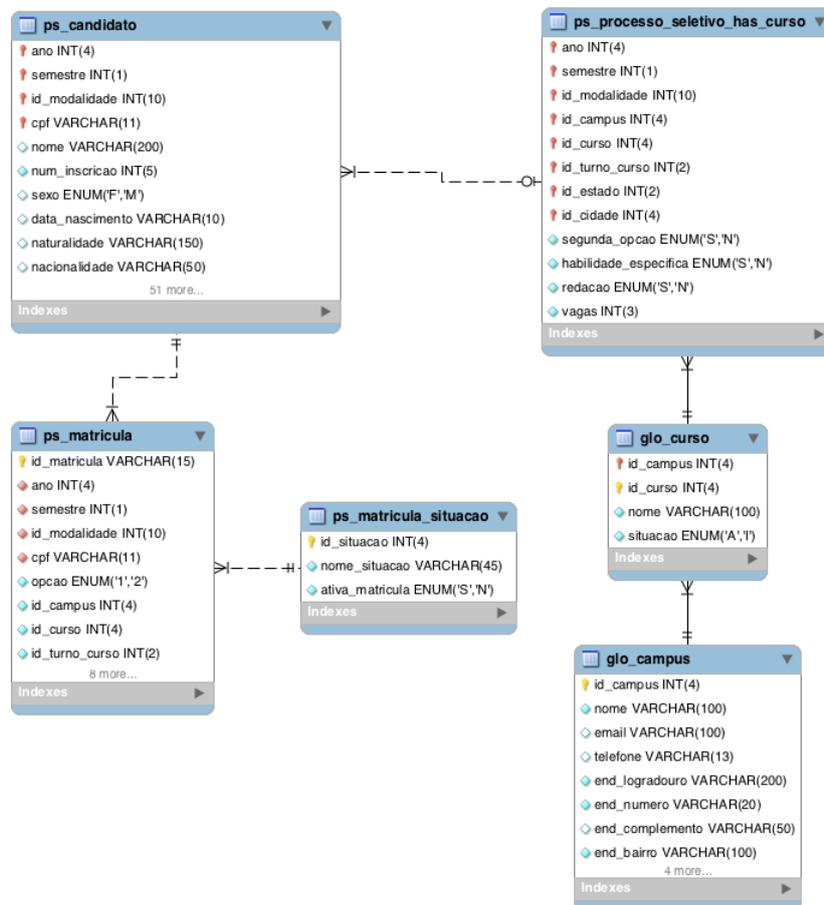


Figura 6: Modelo relacional para a geração do indicador Conversão Candidatos Aprovados / Matrículas Efetivas.

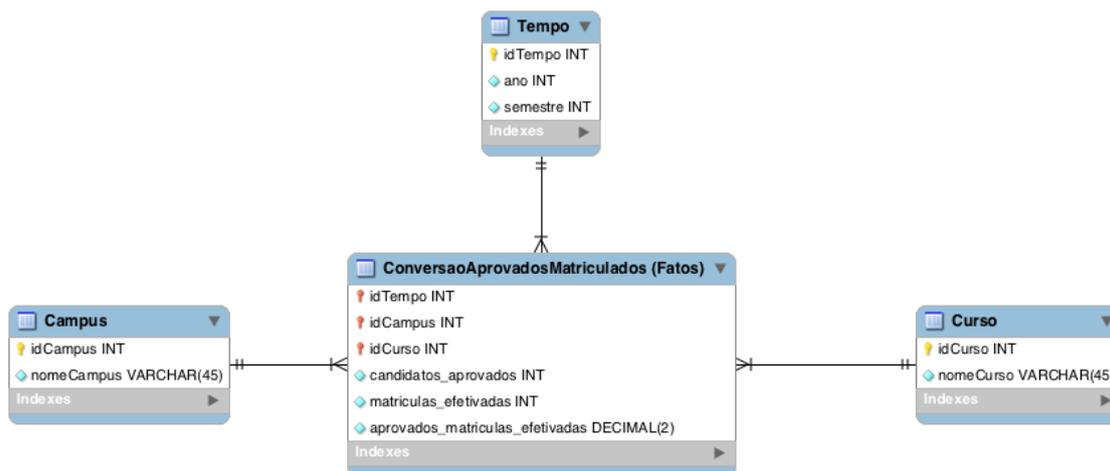


Figura 7: Modelo dimensional para a geração do indicador Conversão Candidatos Aprovados / Matrículas Efetivas. Esquema estrela.

4.4 Importação dos Dados na Ferramenta IBM Watson Analytics

Após o desenho do modelo dimensional, é executado o processo ETL. O processo ETL é realizado através da extração dos dados das bases relacionais, transformando esses

dados de acordo com as regras definidas previamente e logo após dar a carga desses dados na base dimensional onde serão realizadas as pesquisas analíticas pela ferramenta *IBM Watson Analytics*. Na Figura 8 é mostrado o script de extração e transformação dos dados presentes na base de dados relacional para o indicador “Candidato / Vaga”, e na Figura 9 para o indicador “Candidatos Aprovados / Matrículas Efetivas”.

```

SELECT
ps_candidato.ano as ANO,
ps_candidato.semestre as SEMESTRE,
ps_candidato.id_campus as IDCAMPUS,
ps_candidato.id_curso1 as IDCURSO,
glo_campus.nome as CAMPUS,
glo_curso.nome as CURSO,
count(ps_candidato.cpf) as INSCRICOES_REALIZADAS,
(SELECT sum(vagas)
FROM ps_processo_seletivo_has_curso
WHERE semestre = ps_candidato.semestre AND ano = ps_candidato.ano AND id_curso = ps_candidato.id_curso1
AND id_campus = ps_candidato.id_campus) as VAGAS_OFERTADAS
FROM
ps_candidato,
glo_campus,
glo_curso
WHERE
ps_candidato.id_campus = glo_campus.id_campus AND
ps_candidato.id_curso1 = glo_curso.id_curso
GROUP BY ANO, SEMESTRE, IDCAMPUS, IDCURSO, CAMPUS, CURSO
ORDER BY ps_candidato.ano desc, ps_candidato.semestre desc;

```

Figura 8: Script de extração dos dados da base de dados relacional “Indicador Candidado / Vaga”.

```

SELECT
ps_candidato.ano as ANO,
ps_candidato.semestre as SEMESTRE,
ps_candidato.id_campus as IDCAMPUS,
ps_candidato.id_curso1 as IDCURSO,
glo_campus.nome as CAMPUS,
glo_curso.nome as CURSO,
(SELECT count(cpf)
FROM ps_matricula
WHERE semestre = ps_candidato.semestre AND ano = ps_candidato.ano AND id_curso = ps_candidato.id_curso1
AND id_campus = ps_candidato.id_campus) as CANDIDATOS_APROVADOS,
(SELECT count(cpf)
FROM ps_matricula
WHERE semestre = ps_candidato.semestre AND ano = ps_candidato.ano AND id_curso = ps_candidato.id_curso1
AND id_campus = ps_candidato.id_campus AND id_situacao = 2) as MATRICULAS_EFETIVAS
FROM
ps_candidato,
glo_campus,
glo_curso
WHERE
ps_candidato.id_campus = glo_campus.id_campus AND
ps_candidato.id_curso1 = glo_curso.id_curso
GROUP BY ANO, SEMESTRE, IDCAMPUS, IDCURSO, CAMPUS, CURSO
ORDER BY ps_candidato.ano desc, ps_candidato.semestre desc;

```

Figura 9: Script de extração dos dados da base de dados relacional “Conversão Candidatos Aprovados / Matrículas Efetivas”.

Com a modelagem e a criação do script de extração dos dados, já é possível fazer a carga de dados na ferramenta *IBM Watson Analytics*. A ferramenta permite a importação de dados através de diversas fontes pela conexão que ela realiza em diversas bases de dados, porém na versão gratuita que foi utilizada nesse estudo de caso, só é possível fazer a carga de dados através de planilhas ou arquivos “.csv”. Na Figura 10 é possível ver a interface da ferramenta antes da carga de dados.

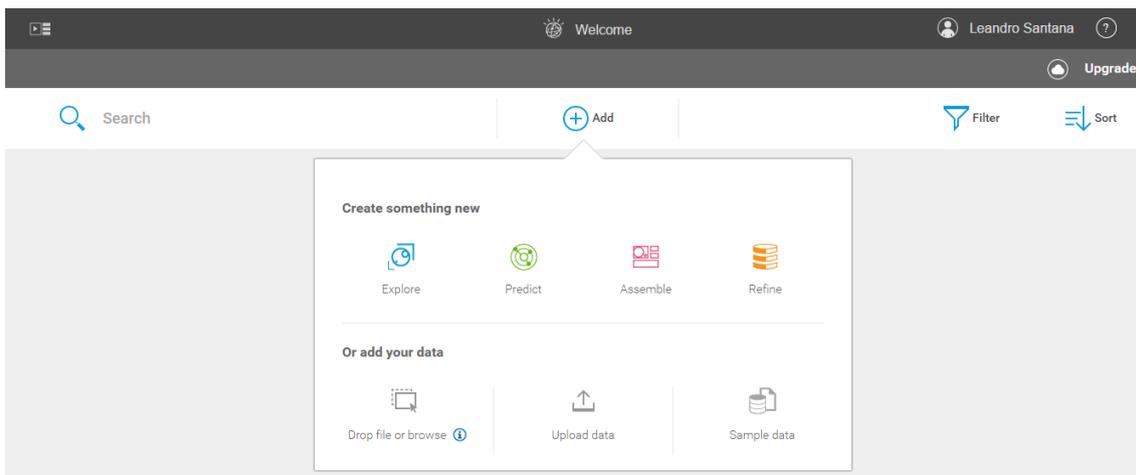


Figura 10: Interface da ferramenta IBW Watson Analytics (carga de dados).

Depois dos dados serem modelados e extraídos como um arquivo “.csv”, é possível fazer sua carga na ferramenta *IBW Watson Analytics*. Conforme mostra a Figura 10, é preciso escolher a opção “*Upload data*” e uma caixa de seleção de arquivo será aberta. Na Figura 11 é mostrado os arquivos “.csv”, de cada indicador, gerados para dar a carga de dados na ferramenta.

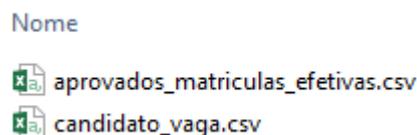


Figura 11: Arquivos “.csv” gerados para a carga de dados na ferramenta IBW Watson Analytics.

Após a carga dos dados, a ferramenta mostrará uma opção com o nome do arquivo importado, que no caso desses indicadores foram “*candidato_vaga*” e “*aprovados_matriculas_efetivas*”. Essa opção pode ser vista na Figura 12. Com esses dados importados, já é possível aos usuários realizarem suas análises nesse conjunto de dados modelados.

Uma das opções que a ferramenta apresenta é a possibilidade de refinar os dados após ser feita a carga. Para o indicador “*Candidato Vaga*” o arquivo de carga de dados apresentou 616 registros, com dados dos campi, polos de ensino a distância e de todos os cursos do IFET desde o ano de 2011. Para o indicador “*Relação Aprovados e Matriculados*” a base de dados possui dados a partir do ano de 2015 e conta com 163 registros. Na Figura 13 é apresentada uma tabela com os dados que foram carregados na ferramenta *IBM Watson Analytics*.

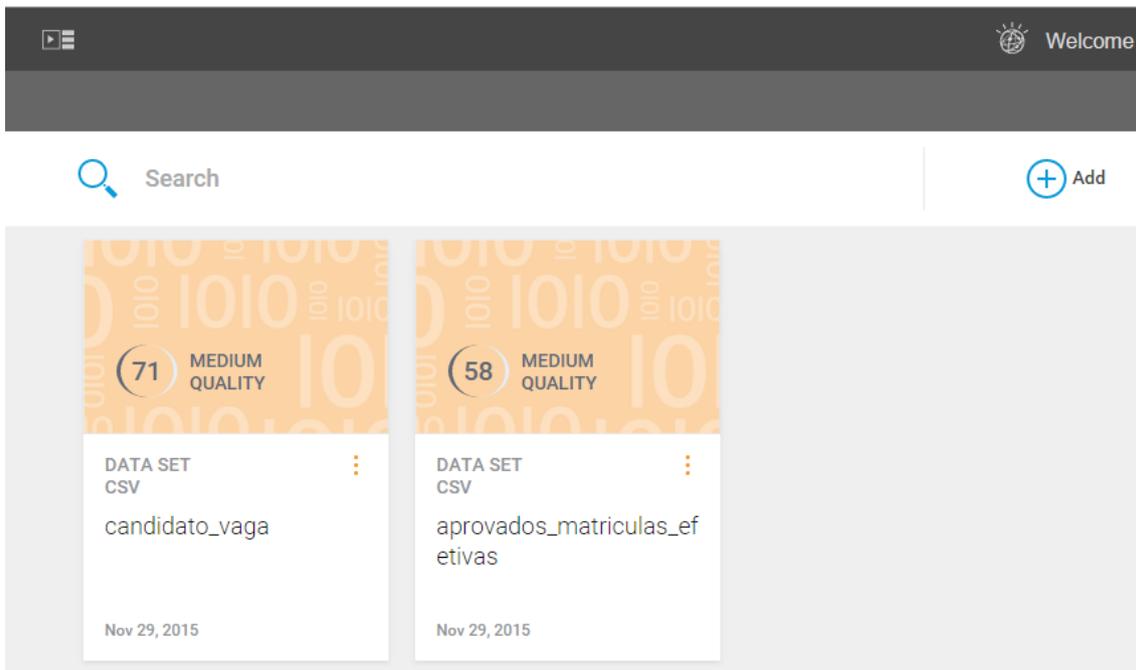


Figura 12: Interface da ferramenta IBW Watson Analytics com a carga de dados completa.

The screenshot shows the IBM Watson Analytics interface displaying a data table. The table has the following columns: ANO, CAMPUS, CURSO, SEMESTRE, CANDIDATOS..., MATRICULAS..., and APR_MAT_EFE. The data is filtered for the year 2015. The table shows enrollment and approval percentages for various campuses and courses. The table is titled 'aprovados_matriculas_efetivas Refinement'.

ANO	CAMPUS	CURSO	SEMESTRE	CANDIDATOS...	MATRICULAS...	APR_MAT_EFE
2015	Campus Juiz...	Engenharia M...	2	19	17	89%
2015	Campus Juiz ...	Licenciatura e...	2	22	19	86%
2015	Campus Juiz ...	Técnico em E...	2	47	28	60%
2015	Campus Juiz ...	Técnico em El...	2	46	34	74%
2015	Campus Juiz ...	Técnico em D...	2	32	17	53%
2015	Campus Juiz ...	Técnico em M...	2	39	35	90%
2015	Campus Juiz ...	Técnico em In...	2	39	24	62%
2015	Campus Juiz ...	Técnico em Tr...	2	45	27	60%
2015	Campus Rio P...	Mestrado Prof...	2	0	0	0%
2015	Campus Sant...	Técnico em M...	2	30	24	80%
2015	Campus Sant...	Técnico em El...	2	34	24	71%
2015	Campus Sant...	Técnico em Tr...	2	49	34	69%
2015	Campus Sant...	Técnico em M...	2	47	35	74%
2015	Polo Além Par...	Técnico em Al...	2	40	40	100%
2015	Polo Além Par...	Técnico em R...	2	39	39	100%
2015	Polo Alfenas	Técnico em Z...	2	50	50	100%
2015	Polo Boa Espe...	Técnico em Z...	2	53	50	94%
2015	Polo Carangola	Técnico em M...	2	48	48	100%
2015	Polo Catagua...	Técnico em L...	2	51	50	98%

Figura 13: Interface da ferramenta IBW Watson Analytics apresentando os dados que foram carregados.

4.5. Análise dos Dados na Ferramenta IBM Watson Analytics

A ferramenta *IBM Watson Analytics* permite analisar os dados importados, gerando tabelas e gráficos gerenciais, através de processamento de linguagem natural, ou seja, utiliza conceitos de inteligência artificial e linguística para facilitar a análise dos dados

por parte dos interessados [IBM 2015]. Na Figura 14 é possível ver a interface da ferramenta onde ocorrem as análises dos dados.

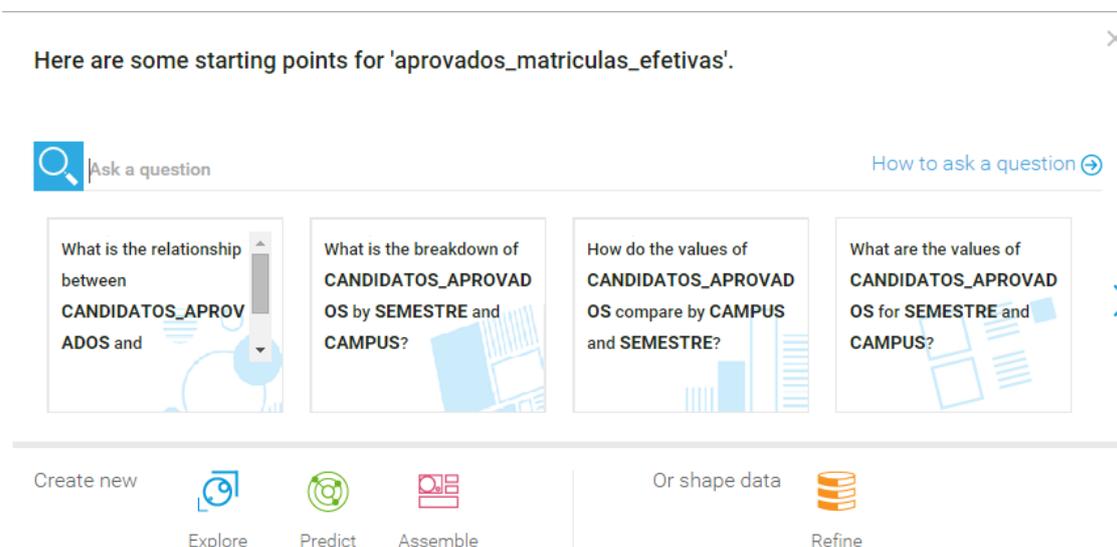


Figura 14: Interface da ferramenta IBW Watson Analytics onde o usuário pode analisar os dados fazendo perguntas através de linguagem natural.

Após a execução da consulta através de linguagem natural, a ferramenta exibirá as informações em forma de gráficos e tabelas que podem ser personalizados conforme a necessidade do gestor. Para este indicador foi solicitado à ferramenta exibir a relação de “candidato/vaga” por Ano e por Curso, filtrando os dados apenas do campus São João del-Rei para os anos de 2012 e 2013, conforme visto na Figura 15, no gráfico de barras.

Através da consulta realizada, o gestor terá a informação do aumento ou da diminuição da concorrência no processo seletivo nos anos 2012 e 2013 para cada curso. Analisando gráfico é possível notar que apenas um dos cursos do campus teve diminuição na concorrência entre estes dois anos, um curso não possui dados referente ao ano de 2012 e dois cursos mantiveram a concorrência.

Da mesma forma, é gerado um gráfico com as informações para o indicador “Candidatos Aprovados / Matrículas Efetivas”, conforme visto na Figura 16. Nessa visualização, o gestor pode perceber a variação de candidatos aprovados que efetivamente se matricularam, por curso, de um semestre para o outro. Nesse indicador foram filtrados os dados do ano de 2015, do primeiro e segundo semestre, dos cursos do campus Juiz de Fora. Alguns dos cursos não apresentam dados de um dos semestres. É possível constatar que a grande maioria dos cursos apresenta um número de matrículas efetivas maior que 50%.

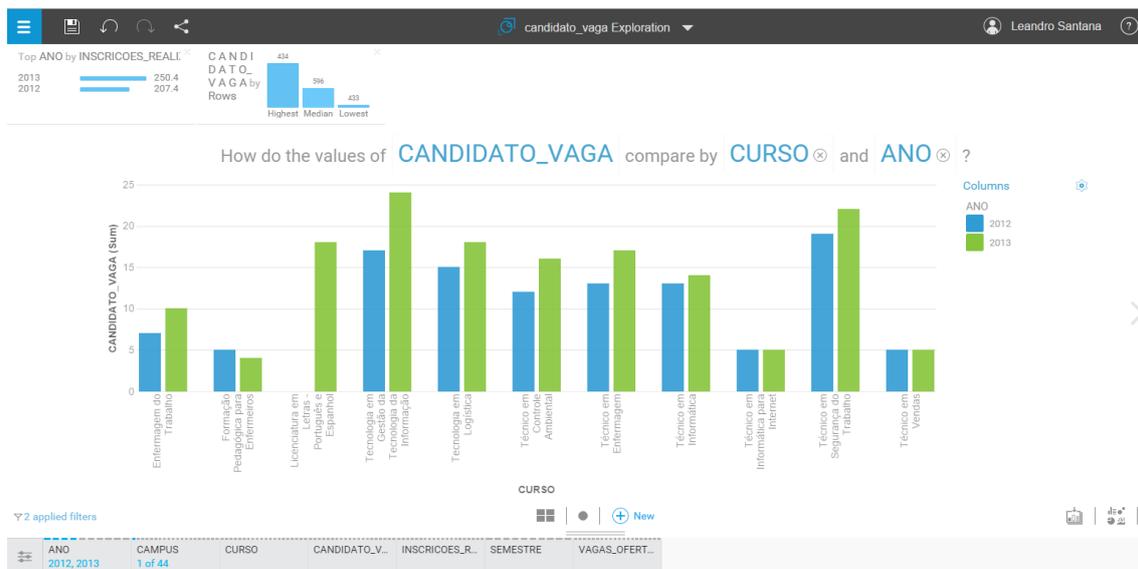


Figura 15: Interface da ferramenta IBW Watson Analytics com o gráfico solicitado pelo usuário.

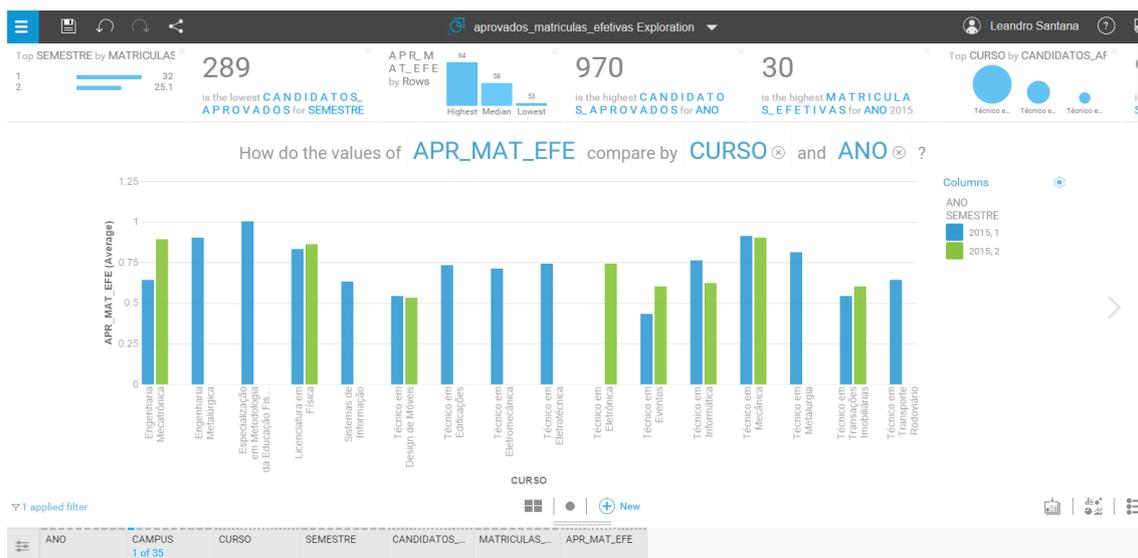


Figura 16: Interface da ferramenta IBW Watson Analytics com o gráfico solicitado pelo usuário.

4.6. Apresentação das Informações Geradas

Realizado todo o processo de modelagem de dados provenientes do banco de dados relacional, para o modelo dimensional, e a importação desses dados na ferramenta *IBM Watson Analytics*, essas informações ficam armazenadas no servidor da ferramenta em nuvem e estão disponíveis para acesso a qualquer pessoa autorizada a acessá-las mediante *login* e senha.

A ferramenta também permite compartilhar as informações que foram processadas e analisadas via e-mail, redes sociais, arquivos e links. Essas opções estão disponíveis na ferramenta conforme mostra a Figura 17.

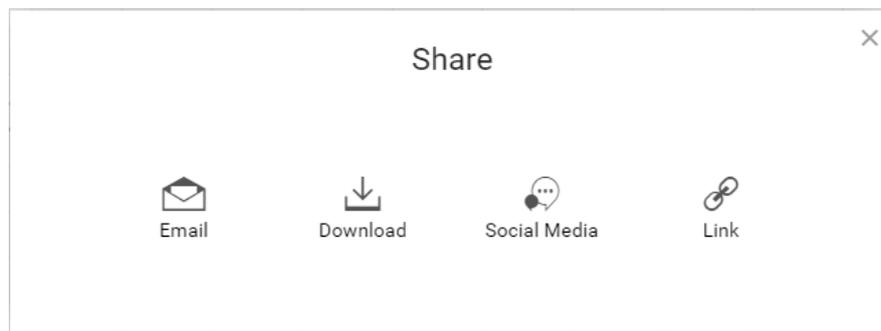


Figura 17: Interface da ferramenta IBM Watson Analytics com as possibilidades existentes de compartilhar as informações geradas.

5. Considerações Finais

Este trabalho mostrou os conceitos que uma organização deve considerar para gerar informação gerencial para seu negócio, desde a concepção de BI, DW, processo ETL, modelagem de dados dimensionais, até como esses dados podem ser analisados em uma ferramenta para esse fim, que no caso foi a *IBM Watson Analytics*, gerando uma grande quantidade de informações através de recursos de visualização como gráficos e indicadores.

A partir desse estudo, espera-se oferecer ao IFET uma alternativa a geração de informação para a tomada de decisões estratégicas, com a adoção de um DW para refinar os dados que se originam em seu sistema de processo seletivo, e pela adoção de uma ferramenta que analise esses dados e forneça visualizações das informações obtidas.

Referências

- Ballard, C., Farrell, D., Gupta, A., Mazuela, C. e Vohnic, S. (2006), “*Dimensional Modeling: In a Business Intelligence Environment*”, IBM, 645 p.
- Barbiere, Carlos. Bi2 - BI – “Modelagem e Qualidade”. Rio de Janeiro: Elsevier, 2011.
- IBM Watson Analytics. Disponível em: <<http://www.ibm.com/analytics/watson-analytics/>>. Acesso em: 22 de Outubro de 2015.
- Machado, F. N. (2004), “Tecnologia e Projeto de Data Warehouse”, Érica, 318 p.
- Microsoft Power BI. Disponível em: <<https://powerbi.microsoft.com/en-us/>>. Acesso em: 20 de Outubro de 2015.
- NIST (National Institute of Standards and Technology), “*Draft Definition of Cloud Computing*”. Disponível em: <<http://csrc.nist.gov/groups/SNS/cloud-computing/>>. Acesso em: 01 de Novembro de 2015.
- Oliveira, Djalma de Pinho Rebouças de. “Planejamento estratégico: conceitos, metodologia e práticas”. 14ª ed. rev. São Paulo: Atlas, 1999.
- Poe, Vidette, Klauer, Patricia, Brobst, Stephen. “*Building a Data Warehouse for Decision Support*”. New Jersey. Prentice-Hall, Inc, 1998. 285 p.

- Prado, Roberto Longhi Rodrigues. “A ampliação do valor de uma empresa através da adoção de um novo modelo: o *Cloud Computing*”. 2010. 51 f. Monografia (Especialista) Puc-sp, São Paulo, 2010. Disponível em: <http://www.mbis.pucsp.br/monografias/Monografia_-_Roberto_Prado.pdf>. Acesso em: 8 out. 2015.
- Reis, E., Teixeira, F. e Araújo, M. A. (2009), “Implementando uma solução de *Business Intelligence* com o Microsoft SQL Server 2005 – Parte 1”, SQL Magazine, Rio de Janeiro, n. 59, p. 52-66.
- SAP Lumira Cloud. Disponível em: <<http://www.saplumira.com/products/sap-lumira-cloud.php>>. Acesso em: 20 de Outubro de 2015.
- Turban, E., Sharda, R., Aronson, J. e King, D. (2009), “*Business Intelligence*: um enfoque gerencial para a inteligência do negócio”, Artmed, 254 p.
- Walczak, Magda. “*What is Cloud BI (Business Intelligence)?*”. 2014. Disponível em: <<http://einsights.com/what-is-cloud-bi-business-intelligence/>>. Acesso em: 15 out. 2015.