

UNIVERSIDADE FEEVALE

SAMUEL AUGUSTO SCHMIDT

CONSTRUINDO UM *DATA WAREHOUSE* PARA UMA
COOPERATIVA DE CRÉDITO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2016

SAMUEL AUGUSTO SCHMIDT

CONSTRUINDO UM *DATA WAREHOUSE* PARA UMA
COOPERATIVA DE CRÉDITO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito parcial
à obtenção do grau de Bacharel em
Ciência da Computação pela
Universidade Feevale

Orientadora: Marta Rosecler Bez

Coorientador: Juliano Varella de Carvalho

Novo Hamburgo
2016

RESUMO

“A nova fonte de poder não é o dinheiro nas mãos de poucos, mas informação nas mãos de muitos.”. A frase de John Naisbitt pode ser interpretada de várias maneiras. Estamos na era conhecida como da informação, o período pós-revolução industrial. Nos dias atuais as contas bancárias, o patrimônio e até mesmo os relacionamentos estão armazenados em bancos de dados. Mas a grande questão que assombra a maioria das organizações é como e o que fazer com toda essa informação. Nos últimos anos têm surgido diversos novos termos criados na computação e utilizados para a denominação de técnicas de tratamento de dados, mas conceitualmente o que toda organização necessita, é extrair o máximo de conhecimento em suas bases de dados. Mais que isso, deixar esse conhecimento como um legado para a empresa e não apenas na memória das pessoas que estão no comando da organização naquele determinado momento. Na grande maioria das organizações a informação ainda é descentralizada, ficando ela alojada em bases de dados distintas, nos discos rígidos das estações de trabalho dos colaboradores, e até mesmo na memória das pessoas. Sendo assim, esse trabalho tem como objetivo tratar da centralização das informações em uma Cooperativa de Crédito que tem enfrentado esse tipo de problema. Uma das tecnologias surgidas na computação com o intuito de organizar a informação foi o *Data Warehouse*. Juntamente com esta tecnologia surgiram várias ferramentas para auxiliar na construção de *Data Warehouses*, muitas delas gratuitas e até mesmo de código aberto. Este trabalho visa o estudo, seleção e uso de ferramentas capazes de organizar as informações de uma Cooperativa de Crédito.

SUMÁRIO

MOTIVAÇÃO	5
OBJETIVOS	9
METODOLOGIA	10
CRONOGRAMA	11
BIBLIOGRAFIA	12

MOTIVAÇÃO

A informação pode ser considerada como o bem intangível mais importante para qualquer organização, pois é a partir dela que são tomadas as decisões estratégicas e até mesmo decisões de cunho operacional. Porém, tão importante quanto ter a informação, é saber onde buscá-la, quando utilizá-la, e mais, como disponibilizar às pessoas certas a informação correta.

A empresa que será estudada nesse trabalho trata-se de uma Cooperativa de Crédito que atualmente conta com 16 unidades de atendimento, 150 colaboradores e mais de 27 mil associados. A mesma já está no mercado há 92 anos e atualmente administra mais de 250 milhões em recursos (SICREDI NORDESTE RS, 2016).

A empresa em questão possui diversos sistemas transacionais, sendo que as informações são geradas a todo momento por pessoas, em todos os níveis da organização. A empresa, por fazer parte de um sistema de Cooperativas de Crédito, não possui acesso direto as bases de dados que gera, pois os sistemas transacionais são desenvolvidos por uma entidade centralizadora que atende a todo o país. Esta tem em sua responsabilidade a custódia e o gerenciamento das informações geradas em todas as Cooperativas ligadas a esse sistema. Porém, muitas dessas Cooperativas têm criado suas próprias áreas de TI, sendo que essas desenvolvem internamente ferramentas que são identificadas como necessárias para o cumprimento pleno de suas atividades.

Algumas dessas são ferramentas operacionais, construídas para auxiliar os colaboradores que estão diretamente no atendimento aos associados. Muitas são ferramentas de apoio à tomada de decisão, utilizadas pelos gestores da empresa no âmbito estratégico.

Com a impossibilidade de acesso as bases de dados, tanto as ferramentas de apoio à operação quanto as de apoio à decisão, têm suas bases de dados abastecidas a partir de relatórios gerados pelos sistemas legados da empresa e importados para as bases locais a partir de *scripts* de importação desenvolvidos pela área de TI interna. Esses relatórios são gerados pelas diversas áreas de empresa, e muitas dessas informações são redundantes. Desta forma, determinados relatórios são gerados por áreas distintas em momentos diferentes,

porém objetivando a mesma informação, o que muitas vezes acaba gerando divergência e colocando em criticidade a autenticidade e integridade da informação.

Além de descentralizada, a recuperação da informação não é dinâmica, pois os relatórios dos sistemas internos são estáticos, uma vez que foram construídos para fins específicos. Isto gera a necessidade de demandar a área de TI para buscar informações que não se encontram nesses relatórios. Além disso, quando identificada a necessidade da criação de novos relatórios, a área de TI é demandada para reprogramar os sistemas já desenvolvidos de tal forma que esse relatório seja fornecido, ocasionando uma demora maior na entrega dessas solicitações.

Atuando em um mercado dinâmico como é o financeiro e tendo como concorrentes as maiores empresas do país, o volume de informação necessário para a operação da Cooperativa tem aumentado exponencialmente, gerando uma demanda de trabalho bastante considerável da área de TI na confecção de relatórios customizados. Esse tempo poderia estar sendo investido na construção de novas ferramentas ou até mesmo na melhoria das ferramentas existentes.

A partir dos problemas relatados e das necessidades da empresa, identificou-se indispensável a criação de um repositório de dados (*Data Warehouse*), onde todas as informações compiladas pelas diferentes áreas do negócio ficassem disponíveis e de fácil acesso. Além disso, também se identificou a necessidade da criação de uma política de acesso a informação, restringindo o acesso da mesma nos diferentes níveis da empresa. E também a implantação de uma ferramenta para que essa informação pudesse ser recuperada de forma rápida e dinâmica.

Um processo de BI (*Business Intelligence*), segundo a definição de Han e Kamber (2001), trata-se de um processo que objetiva a elaboração normativa de sistemas computacionais responsáveis pela organização dos grandes volumes de dados. E foi a partir dessa necessidade que surgiram os *Data Warehouses* (armazém de dados). Kimball (1998) define um *Data Warehouse* como "Uma cópia dos dados de transação especificamente estruturados para consulta e análise". Segundo Kimball (2002), existem 6 requisitos que são fundamentais na criação de um *Data Warehouse*, que são eles:

- 1 – O *Data Warehouse* deve fornecer acesso aos dados corporativos ou organizacionais: o *Data Warehouse* deve ter conexão a partir dos computadores pessoais de

gerentes e analistas, sendo essa conexão imediata e de alto desempenho, ou seja, as menores consultas devem ser executadas em milésimos de segundo. Além disso, as ferramentas disponíveis aos gerentes e analistas devem ser de fácil manuseio.

2 – Os dados devem ser consistentes: quando duas pessoas solicitam um determinado relatório contendo a mesma informação, os resultados devem ser idênticos. Além disso, o usuário deve ser informado caso uma carga de dados ainda estiver parcialmente carregada.

3 – Os dados podem ser separados e combinados usando qualquer medição possível do negócio: esse requisito está diretamente ligado a abordagem dimensional. Existe uma definição mais operacional para esse termo que são *cabeçalhos de linha e restrições*, sendo esses os alicerces de qualquer *Data Warehouse*.

4 – O *Data Warehouse* não consiste apenas em dados, mas também em um conjunto de ferramentas no qual se possa consultar, analisar e apresentar as informações: componentes como o servidor central, o Sistema de Gerenciamento de Banco de Dados (SGBD) e os dados propriamente ditos, consistem em apenas 60% do *Data Warehouse*. Os demais 40% consistem em um conjunto de ferramentas de *front-end*.

5 – Os dados devem ser confiáveis: o sucesso do *Data Warehouse* não consiste apenas em centralizar os dados em um repositório central e depois liberá-los. Os dados devem ser cuidadosamente coletados, limpos, e apenas após terem sua autenticidade comprovada é que devem ser liberados.

6 – A qualidade dos dados em um *Data Warehouse* impulsiona a reengenharia dos negócios: se determinados dados são considerados de alta relevância por uma organização, a entrada dessa informação deve ser garantida, não sendo sua entrada opcional nos sistemas de *input*. Portanto, é possível afirmar que um *Data Warehouse* pode exercer um papel importante nos esforços de reengenharia de uma organização, uma vez que forçaria a obrigatoriedade na entrada nas informações que são consideradas relevantes para o negócio.

Com o surgimento de técnicas como o *Data Warehouse*, surgem ferramentas para auxiliar o processo de carregamento e análise dos dados. E duas das principais ferramentas são: ETL (*Extract, Transform and Load*) e OLAP (*Online Analytical Processing*). A primeira delas consiste em um processo de extrair dados de fontes externas, limpá-los e, caso necessário, ajustá-los, deixando apenas o que seja necessário e importante para o negócio, e feito isso, carregá-los dentro do *Data Warehouse*. Para Machado (2004), o propósito de um

processo de ETL é garantir a integridade e consistência dos dados, construindo uma base de informações que retrate a realidade da empresa de uma forma fiel.

A maior parte do legado de sistemas de uma organização é composta pelos sistemas que apoiam a operação da empresa de fato, esses sistemas são conhecidos como sistemas transacionais ou OLTP (*Online Transaction Processing*) (KIMBALL e MERZ, 2000). Porém, com o considerável aumento de informações geradas pelos sistemas OLTP, entraram no mercado sistemas construídos com o intuito de apoiar análises e consultas dos dados gerados pela organização, e também de sintetizar informações para comparações, visões históricas e visões personalizadas. Esses sistemas são conhecidos como sistemas OLAP (*Online Analytical Processing*). Segundo Fortulan e Filho (2005), os sistemas OLAP caracterizam-se por permitirem uma visão multidimensional dos dados, possibilitando uma análise a partir de múltiplas perspectivas de forma fácil e intuitiva.

Dado este contexto, este trabalho criará um repositório de dados único, conhecido como *Data Warehouse*. Desta forma, a informação gerada em uma cooperativa estará acessível de forma rápida e dinâmica, permitindo aos colaboradores da organização gerarem a informação necessária para a tomada de decisões.

OBJETIVOS

Objetivo geral

Criar um repositório de dados único, *Data Warehouse*, onde a informação gerada em uma cooperativa fique acessível de forma rápida e dinâmica, permitindo aos colaboradores da organização a geração da informação necessária.

Objetivos específicos

- Estudar as metodologias disponíveis para a implementação de um *Data Warehouse*;
- Estudar as ferramentas disponíveis para construção de *Data Warehouses*, e sistemas de OLAP;
- Modelar um banco de dados multidimensional;
- Estabelecer uma política de acesso a informação;
- Desenvolver os *scripts* de importação de relatórios;
- Definir os *Dashboards* necessários para as decisões estratégicas da cooperativa;
- Criar um sistema único contendo todos os *Dashboards* necessários para as decisões estratégicas da cooperativa;

METODOLOGIA

Todo trabalho de pesquisa, exige uma questão norteadora. Esta é apresentada na sequência: “É possível, através de um *Data Warehouse*, dar autonomia a um usuário sem conhecimento em TI, para buscar nos dados gerados pelo legado de sistemas de uma organização, a informação que necessitar, no período que quiser, e visualizá-la da forma que achar mais apropriada?”.

Esse trabalho caracteriza-se como uma pesquisa de natureza aplicada, visto que objetiva a solução de um problema prático com um escopo definido (organização da base de dados em uma Cooperativa de Crédito).

Do ponto de vista da forma de abordagem, o projeto será abordado na forma quantitativa. Segundo Prodanov e Freitas (2013), a pesquisa quantitativa considera que tudo pode ser quantificável, ou seja, ser traduzido em números, opiniões e informações para que se possa fazer uma análise. O sucesso do presente projeto e o entendimento dele por parte dos colaboradores da empresa se dará quando o número de demandas referente a solicitação de informações feitas a área de TI for reduzido. A análise dos resultados se dará a partir de um sistema de gestão de demandas utilizado pela área de TI, pois atualmente as demandas de solicitação de relatórios pontuais são feitas a partir desse sistema. Também se dará de forma qualitativa através de questionário de satisfação de uso e qualidade da informação a ser aplicado a pelo menos cinco executivos que utilizarão a ferramenta.

A conclusão dos resultados desse trabalho de conclusão será feita da seguinte forma: após o *Data Warehouse* e a ferramenta de OLAP estarem implantados, será observado o histórico das demandas referentes a relatórios pontuais em um período anterior a implantação desses sistemas. Será comparada a quantidade de solicitações de demandas anterior a implantação dos sistemas, com a quantidade posterior aos sistemas implantados em um período a ser determinado.

Segundo Prodanov e Freitas (2013), quando um estudo, em sua fase preliminar, procura buscar maiores informações sobre o fato investigado, possibilitando assim, sua definição e delineamento, considera-se uma pesquisa exploratória. No caso dessa pesquisa, é possível enquadrá-la nesse contexto, visto que será trabalhado com um problema já conhecido, e a partir de estudos de técnicas já existentes, buscar-se-á uma proposta de solução para o mesmo.

Prodanov e Freitas (2013) classificam como pesquisa experimental, do ponto de vista dos procedimentos técnicos, como a determinação de um objeto de estudo, a seleção das variáveis capazes de influenciá-lo, e a definição das formas de controle de observação que a variável produz nesse objeto. É possível enquadrar esse contexto como tal, visto que, tem-se um objeto de estudo (a base de dados de uma Cooperativa de Crédito), uma variável de influência (a forma de organização dos dados e disponibilização deles aos colaboradores) e a observação dos efeitos que essa variável produzirá sobre o objeto (a autonomia das pessoas em gerarem a própria informação e saberem onde buscá-la).

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
Identificação do problema	X			
Busca por conhecimento bibliográfico e de mercado e bibliográfico na área	X	X	X	X
Definição dos requisitos do projeto	X	X		
Avaliação das alternativas existentes no mercado		X	X	
Definição da ferramenta a ser utilizada		X	X	
Escrita do TC I		X	X	X
Entrega do TC I				X

Trabalho de Conclusão II

Etapa	Meses				
	Jan	Fev	Mar	Abr	Mai
Busca por conhecimento de mercado na área	X	X	X		
Definir cubos de dados a serem montados no Data Warehouse		X	X		
Elaborar <i>scripts</i> de ETL para montagem dos cubos de dados		X	X		
Criar política de acesso a informação			X	X	
Montar os cubos de dados sobre a ferramenta escolhida para criação do Data Warehouse			X	X	
Criação dos <i>DashBoards</i>				X	X
Avaliação dos resultados quanto ao histórico de demandas antes e pós implantação				X	X
Escrita TC II	X	X	X	X	X
Entrega TC II					X

BIBLIOGRAFIA

FORTULAN, M. R.; FILHO, E. V.G. Uma proposta de aplicação de Business Intelligence no chão-de-fábrica. *Gestão e Produção*, São Carlos, 2005.

HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufmann, 2001.

KIMBALL, Ralph. *Data Warehouse Toolkit: Lifecycle Toolkit*. John Wiley & Sons, 1998.

KIMBALL, Ralph. *Data Warehouse Toolkit: Técnicas para Construção de Data Warehouses Dimensionais*, São Paulo: Makron Books, 2002.

KIMBALL, R.; MERZ, R. *Data Webhouse: construindo o Data Warehouse para a WEB*. Rio de Janeiro: Editora Campus, 2000.

MACHADO, F. N. R. *Tecnologia e Projeto de Data Warehouse*. São Paulo: Érica, 2004.

PRODANOV, C. C., FREITAS, E. C., *Metodologia Do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico*. Novo Hamburgo: Feevale, 2013.

SICREDI Nordeste RS. *Balanco Semestral*. 2016.