

**INSTITUTO DE EDUCAÇÃO SUPERIOR DA PARAÍBA /
FACULDADE DE TECNOLOGIA DA PARAÍBA - FATECPB
TECNOLOGO EM SISTEMAS PARA INTERNET
JEILSON SOARES DE ALBUQUERQUE**

***BUSINESS INTELLIGENCE PARA ANÁLISE DE PROCESSO HOSPITALAR EM
HOME CARE UTILIZANDO A FERRAMENTA QLIK SENSE E TALEND OPEN
STUDIO FOR DATA INTEGRATION***

**CABEDELO-PB
2020**

JEILSON SOARES DE ALBUQUERQUE

***BUSINESS INTELLIGENCE PARA ANÁLISE DE PROCESSO HOSPITALAR EM
HOME CARE UTILIZANDO A FERRAMENTA QLIK SENSE E TALEND OPEN
STUDIO FOR DATA INTEGRATION***

Relatório Técnico Científico apresentado ao Curso de Sistemas para Internet da Faculdade de Tecnologia da Paraíba – FATECPB como requisito para obtenção do título de tecnólogo em Sistemas para Internet.

ORIENTADOR: Prof. Me. Hercílio de Medeiros Sousa

**CABEDELO-PB
2020**

JEILSON SOARES DE ALBUQUERQUE

BUSINESS INTELLIGENCE PARA ANÁLISE DE PROCESSO HOSPITALAR EM HOME CARE UTILIZANDO A FERRAMENTA QLIK SENSE E TALEND OPEN STUDIO FOR DATA INTEGRATION

Relatório Técnico Científico apresentado ao Curso de Sistemas para Internet da Faculdade de Tecnologia da Paraíba – FATECPB como requisito para obtenção do título de tecnólogo em Sistemas para Internet.

Aprovada em: ____ de _____ de 2020.

BANCA EXAMINADORA

Prof. Me. Hercílio de Medeiros Sousa (orientador)
de Educação Superior da Paraíba

Prof. XXXXXX
Instituto de Educação Superior da Paraíba

Prof. XXXXXX
Instituto de Educação Superior da Paraíba

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Padre Joaquim Colaço Dourado

A3345b

Albuquerque, Jeilson Soares de.

Business intelligence para análise de processos hospitalar em home care utilizando a ferramenta qlik sense e talend open studio for data integration [recurso eletrônico] / Jeilson Soares de Albuquerque. – Cabedelo, PB: [s.n.], 2020.

47 p.

Orientador: Prof. Me. Hercilio de Medeiros Sousa.
Monografia (Graduação em Sistema para Internet) – Faculdade de Tecnologia da Paraíba – FATEC-PB.

1. Business intelligence. 2. Software – Processo hospitalar. 2. Ferramenta qlik sense. 3. Tecnologia da informação. 4. I. Título.

CDU: 005.94

À minha família e amigos

AGRADECIMENTOS

Agradeço primeiramente a Deus, pois é de sua vontade me conceder capacidade de conquistar mais um sonho. Aos meu pais por acreditarem e pelo suporte e constante apoio em todos os momentos. A minha noiva que teve paciência nos finais de semana quando eu ficava estudando de madrugada na hora de dormir. Ao corpo docente do instituto de ensino pelo compartilhar conhecimento e estímulos nesta jornada até a conclusão. Aos meus professores Hercílio e Humberto por ter me orientado e incentivado a realizar este trabalho de conclusão, sempre dedicados e atenciosos obrigado por toda atenção e paciente. E aos demais professoras pois sem vocês de alguma forma contribuíram para que eu pudesse chegar até aqui. Aos meus amigos da faculdade, que sempre me apoiaram e me ajudaram nas horas difíceis. o meu muito obrigado a todos.

RESUMO

É iminente avanço da tecnologia nos últimos anos, esse avanço tem gerado uma quantidade exaustiva de dados e informações a todo o momento. As empresas podem gerar as mesmas informações em diversos ambientes a fim de atingir o mesmo objetivo, Com base nisso, as organizações necessitam de um bom planejamento estratégico a fim de esboçar objetivos e oferecer ações para atingi-los, a forma como os dados estão sendo disponibilizados as vezes são de difícil compreensão para as pessoas, tornando-se inviável sua utilização. Contudo as informações produzidas em sistema de ERP (*Enterprise Resource Planning*) contidas nessa fonte de dados não estão explícitas, sendo necessário o uso de ferramentas de análises de dados para que se possa extrair algum conhecimento. O BI (*Business Intelligence*) oferece técnicas e métodos que permitem a extração e transformação de dados em conhecimento a partir de grandes distintas bases de dados. A proposta deste trabalho mostrar um estudo de caso para evidenciar os números dos setores de faturamento e farmácia de um hospital com sistema de *home care* a fim de reduzir erros de cadastros e procedimentos, buscando reduzir custo operacional através das análises criadas a partir das técnicas de BI. A metodologia adotada no trabalho fez uso de referências bibliografias, publicações e ferramentas que auxiliaram na aplicação das técnicas e seus métodos, possibilitando com o uso da ferramenta *Talend Open Studio for Data Integration* a realização do processo de ETL (Extração, Transformação) e com a ferramenta *Qlik sense* o *Load* que é carga dos dados. Portanto, permitindo a visualização de forma rápida, precisa e dinâmica das informações acerca análise de rentabilidade por paciente diante do uso de materiais, dietas, e medicamentos, verificar o quantitativo do uso dos gases medicinais alocados em residências e Investigar o aumento/redução de pacientes por cidades, bairros e operadora de saúde.

Palavras-Chaves: *Business Intelligence. Qlik Sense. Tecnologia.*

ABSTRACT

It's imminent the advance in technologies these years, this advance generates an exhausting amount of data and information all the time. The companies can generate the same information in different environments in order to reach the same goal. Based on this, organizations need a good strategic plan in order to define objectives and offer actions to reach them, the way that these data are being made available sometimes makes difficult for people's understanding, making these data unfeasible to use. However, the information produced in ERP systems (Enterprise Resource Planning) that is contained in this data source are not explicit, what makes necessary the use of data analysis tools to make possible to extract some knowledge. The BI (Business Intelligence) offers techniques and methods that allows the extraction and transformation of data into knowledge from a large distinct data basis. The proposal of this work is to show a case study to evidence the numbers of the invoicing and pharmacy departments of a hospital that provides home care system in order to reduce errors in registrations and in procedures, seeking to reduce operational costs through the analysis created from the BI techniques. The methodology of this work used bibliographic references, and tools that assisted in the application of the techniques and in its methods, making possible with the use of Talend Open Studio for Data Integration tool the performing of the ETL process (extraction, transformation) with the Qlik sense tool the data load. Therefore, allowing the visualization in a quick, accurate and dynamic form, of all the information about the profitability analysis per patient against the use of materials, diet and medicine, to verify the quantitative of the use of medicinal gases that are allocated in residences, and to investigate the increase/decrease of patients per city, neighborhood and health operators.

Keywords: *Business Intelligence. Qlik Sense. technology.*

LISTA DE ILUSTRAÇÕES

Figura 1: Pirâmide dos Dados	19
Figura 2: Etapas do ETL	21
Figura 3: Ilustração da granularidade	25
Figura 4: Modelo Estrela	28
Figura 5: Modelo Floco de Neve	28
Figura 6: Mapa Mental 5W2H	31
Figura 7: Interface de desenvolvimento da Ferramenta Talend Open	33
Figura 8: Processo de Extração no Talend Studio	37
Figura 9 – Transformação de Dados usando o componente TMAP da ferramenta Talend Open Studio	38
Figura 10 – Exemplo de notificação após rodar/executar o JOB Carga_Paciente	38
Figura 11: Job de Junção da Ferramenta Talend Open Studio for Data Integration	39
Figura 12: Tmap da Junção da Ferramenta Talend Open Studio	39
Figura 13 – Painel Senso dos Pacientes, análise quantitativo de pacientes por Plano, Faixa Etária, Programa de Internação, Por Cidades e Sexo	40
Figura 14 - Painel com geral dos pacientes,	41
Figura 15 - Painel para Análise de Itens Prescritos	41
Figura 15 - Painel de Análise de item a item prescrito.	42
Figura 16 - Painel de Análise de item a item prescrito.	42
Figura 17 - Painel de Análise do uso de gases medicinais	44

LISTA DE SIGLAS

BI – Business Intelligence

DM - Data mart

DW - Data Warehouse

ERP - Enterprise Resource Planning

ETL - Extract, Transform and Load

ETL – Extract, Transform and Load

KDD - Knowledge Discovery in Databases

KPI - Key Performance Indicator

OLAP – Online Analytical Processing

OLTP – Online Transaction Processing

OLTP - Online Transaction Processing

SGBD – Sistema de Gerenciamento de Banco de Dados

XML - Extensible Markup Language

SUMÁRIO

1. INTRODUÇÃO	12
1.2 OBJETIVOS	14
1.2.1 GERAL	14
1.2.2 ESPECIFICOS	14
1.3 JUSTIFICATIVA	14
2. METODOLOGIA	15
3 FUNDAMENTAÇÃO TEORICA	16
3.1 HOSPITAL	16
3.2 GESTÃO DE FARMÁCIA	16
3.3 BUSSINESS INTELLIGENCE	16
3.4 DIFERENÇA DE DADO, INFORMAÇÃO E CONHECIMENTO	18
3.5 ETL: <i>EXTRACT, TRANSFORM AND LOAD</i>	19
3.6 <i>DATA WAREHOUSE</i>	21
3.7 <i>DATA MART</i>	22
3.8 <i>DATA MINING</i>	23
3.9 GRANULARIDADE	24
3.10 METADADOS	25
3.11 FATO	25
3.12 DIMENSÃO	26
3.13 MODELAGEM MULTIDIMENSIONAL	26
3.14 5W2H	28
3.15 FERRAMENTA <i>QLIK SENSE</i>	31
3.16 FERRAMENTA TALENT DATA INTEGRATION	31
3.17 SISTEMA IW CARE	33
3.18 SISTEMA WFS SOFTWARE	33
3.19 POSTGRES	34
4 ESTUDO DE CASO	35
4.1 MATERIAIS E METODOS	35
4.2 BASE DE DADOS	35
4.3 EXTRAÇÃO	35
4.4 CRIAR O JOB	35
4.5 DIMESSÃO TEMPO	36
4.6 TRANSFORMAÇÃO	37

5 RESULTADOS E DISCUSSÕES	40
5.1ANÁLISE SENSO DOS PACIENTE	40
5.2ANÁLISE ITENS PRESCRITOS	41
6 CONSIDERAÇÕES FINAIS	44
REFERÊNCIA	46

1. INTRODUÇÃO

Nos últimos anos, as empresas vêm tentando melhorar seus processos/resultados de forma eficiente, utilizando sistemas de ERP (*Enterprise Resource Planning*) que atende as necessidades específicas de determinado setor, como gestão de Pacientes, gestão de gases medicinais, sistema financeiro etc. Com isso o sistema de BI (*Business Intelligence*) é utilizado para analisar os processos de negócios como apoio a decisões organizacionais.

ANS (Agência Nacional de saúde suplementar), um setor específico do ministério da Saúde, Nasceu em 28 de janeiro de 2000. Segundo a ANS (2003), o ano de 1923 é o marco da previdência social no Brasil, a Lei Eloy Chaves "(...) em cada uma das estradas de ferro existentes no país, uma caixa de aposentadoria e pensões para os respectivos empregados" Essas caixas além de garantir aposentadoria e pensões financiável também serviços médico-hospitalar. E o plano de saúde mais antigo do Brasil e ainda em operações é a Cassi, constituído em 1944 pelo Banco do Brasil.

A ANS passou por uma reforma, em meados de 2003, pelo fato de não existir uniformidade para o envio de contas médicas por parte das instituições privadas, e era necessário uma grande demanda de tempo e esforço com a finalidade de extrair o resultado dos procedimentos que os pacientes haviam utilizados em suas contas mensais. Cada convenio usava um padrão de formulário, e desse modo, as informações ficavam inexatas e podiam ser facilmente adulteradas, pelo fato de ser apresentada em meios analógicos.

Para tentar resolver esse problema, foi criado um padrão para a troca de informações conhecido como TISS (Troca de informações em saúde Suplementar) que seu principal objetivo foi unificar os métodos administrativos, auxiliar as ações de avaliação de acompanhamento econômico, financeiro e assistencial das operadoras de planos privados de assistência à saúde. Desde então passou a ser eletronicamente no formato de XML (*Extensible Markup Language*), arquivo que contém todas as informações necessárias para verificação do que foi realizado durante o tempo de internação do paciente no leito. Como exames, procedimentos,

consultas, medicações etc. O formato de envio eletrônico além de seguro para quem está enviando é mais rápido e não precisa de tanta mão de obra.

Com a exigência da ANS que as informações trafegassem em formato eletrônico no tipo de arquivo XML, para que os convênios remetessem os seus relatórios de pagamento no formato supracitado.

Com isso, o grande volume de dados trafegados entre os planos privados de Saúde, ofertasse uma demanda de análise de dados em ferramentas de BI (*Business Intelligence*) para a tomada de decisões de negócio rapidamente. Facilitando a comunicação e assertividade. Os gestores necessitam das informações precisas, no tempo exato e no local determinado, este é o perfil do *business Intelligence*, (TURBAN, et al, 2009).

A ferramenta de BI a ser utilizada é o *Qlik Sense*, na qual os dados brutos serão extraídos, transformados e carregados pela ferramenta *Talent Studio Data Integration* para se ter um dado de qualidade e ajudar na procura de padrões coerentes para identificar relacionamentos e subconjuntos de dados a serem mapeados e gerar informações com mais precisão, permitindo que os usuários analisem os dados, evidenciando novas concepções que auxiliem na solução dos problemas de negócios, (VISÃO..., 2011).

Com a utilização da ferramenta de BI, seus processos podem ser otimizados e reduzidos em uma boa porcentagem de tempo, já que apenas normalizando as informações de cadastro de pacientes fornecido pelo sistema cruzando com movimentação de materiais de consumo dos pacientes como medicamentos, materiais descartáveis e dietas industrializadas, além das recargas de gases de usos medicinais, teremos a possibilidade da diminuição de falhas de processos, redução de tempo e até mesmo conflitos entre setores surgindo na falha de uma má solicitação.

Com a implantação da ferramenta de *Business Intelligence* na empresa Hospitalar privada, foi percebido falha no processo administrativo desde o cadastro do paciente, informações em duplicidade e em setores diferentes, devido ao excesso de preenchimento de planilhas eletrônicas e fichas de cadastro com finalidades semelhantes, causando prejuízo além de viagens desnecessárias dos setores de transportes, entrega de medicamentos que não serão pagas pelas operadoras de saúde, tendo em vista o tempo que se perde junto com o prejuízo

gerado pelo acúmulo de dados de baixa qualidade que muitas vezes não são analisados. Com a ferramenta de BI implantada espera-se que os processos possam ser centralizados para que ocorra otimização nas análises custos/despesas por paciente, rápida identificação de falhas de cadastro, além da redução dos prejuízos gerados.

Para isso será necessário à autorização da diretoria da empresa, com finalidade de acessar os dados do sistema de gestão para analisá-los na ferramenta *Qlik sense*.

1.2 OBJETIVOS

1.2.1 GERAL

O objetivo geral do trabalho é apresentar um estudo de caso, no qual será demonstrado os principais benefícios do uso do *Business Intelligence* para analisar a visão geral da movimentação dos insumos, e quantitativos de pacientes para auxiliar os gestores nas tomadas de decisão dentro da empresa.

1.2.2 ESPECIFICOS

Com isso podemos acessar com rapidez os indicadores de:

- Analisar a rentabilidade por pacientes diante do uso de materiais hospitalares, dietas e medicamentos;
- Verificar o quantitativo do uso dos gases medicinais alocados em residências;
- Investigar o aumento/redução de pacientes por cidades, bairros e operadora de saúde.

1.3 JUSTIFICATIVA

A temática em questão é extremamente relevante, por se tratar de um estudo que possibilitará o uso de ferramentas para análises de bancos de dados para um processo de decisões internas.

O estudo com BI diante de um banco de dados na área hospitalar propiciará análises e minerações de informações que permitirão tomadas de decisões assertivas em menor tempo, produzindo assim alta qualidade da gestão.

2. METODOLOGIA

A metodologia aplicada nesse trabalho será de natureza descritiva e bibliográfica. Descritiva por ter a finalidade de observar, registrar e analisar os sistemas técnicos e bibliográfica por ser desenvolvida por meio de livros, artigos e etc. Será realizado um estudo de caso, onde os dados serão coletados no hospital de rede privada através da ferramenta *Talend Open Studio for Data Integration* para que a ferramenta *Qlik Sense* possa fazer as análises. Sua abordagem será quantitativa, pelo fato de analisar muitas informações e qualitativa por visar uma pesquisa mais detalhada dos dados.

3 FUNDAMENTAÇÃO TEORICA

Neste capítulo, temáticas pertinentes ao presente trabalho serão explicadas.

3.1 HOSPITAL

Gonçalves (1983, p. 20) define hospital como:"(...) uma organização completa. Ele incorpora o avanço constante dos conhecimentos, de aptidões, da tecnologia médica e dos aspectos finais desta tecnologia representados pelas instalações e equipamentos". De acordo com Mirshawka (1994, p. 22): "De todas as empresas modernas, nenhuma é mais complexa do que o hospital." Com isso o Hospital é uma instituição em constante evolução em seus complexos processos.

3.2 GESTÃO DE FARMÁCIA

A Assistência Farmacêutica representa hoje um setor de grande impacto financeiro, é constituída por três componentes; Componente Básico, Estratégico e Especializado da Assistência Farmacêutica. "Esta abrange as atividades de seleção, programação, aquisição, armazenamento, distribuição e acompanhamento da utilização de medicamentos. Além de ser uma atividade relevante nas ações de saúde" (DALARMI, 2010, p 84)

3.3 BUSSINESS INTELLIGENCE

É um termo do *Gartner Group* no português significa inteligência de negócio. Conceito que surgiu na década de 80, é um sistema de apoio a decisão e gestão de negócios, é um processo que faz referência à coleta, organização, análise, compartilhamento e acompanhamento de informação e indicadores de desempenho *Key Performance Indicator (KPI)*, oferecendo suporte aos gestores, desenvolvendo e ampliando o entendimento, permitindo a tomada de decisão mais racional.

Segundo Loh (2014), BI é um processo que envolve métodos, técnicas, tecnologia, pessoas, informações, fontes de informações, métricas e ferramentas. E tem por objetivo encontrar causas ou explicações para eventos e resultados. A busca por padrões também é objetivo do BI, é preciso encontrar uma ordem para

que os dados possam fazer sentido e serem úteis. O BI também procura encontrar explicações para eventos com fundamentações em dados.

TURBAN et al (2009, p. 27) afirma que “Os principais objetivos do BI são permitir o acesso interativo aos dados (às vezes, em tempo real), proporcionar a manipulação desses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar a análise adequada”.

Para Loh (2014), para explicar o que é BI, um dos melhores exemplo é o caso da GM (General Motors) e o sorvete de baunilha. Segundo a lenda, um cliente que havia comprado um carro, enviou uma carta reclamando de um eventual problema: quando ele ia a uma sorveteria e comprava um sorvete de baunilha ao voltar para o carro ele tinha dificuldades para dar partida, porém este problema não acontecia se ele comprasse outro sabor de sorvete.

Ninguém conseguia imaginar a ligação que o sorvete de baunilha poderia ter com o tal problema. Um engenheiro pegou o caso para investigar e procurou o cliente. Eles foram juntos à sorveteria testar a teoria, compraram um sorvete de baunilha e ao voltarem ao carro para surpresa do engenheiro, realmente tiveram dificuldades ao dar partida. Em seguida fizeram uma segunda tentativa comprando um sorvete de sabor diferente, e ao voltar para o carro, funcionou perfeitamente.

É claro que o engenheiro sabia que o sabor do sorvete não seria o causador do problema, mas que havia uma relação do problema associada ao sabor escolhido. Então ele percebeu que para comprar o sorvete de baunilha levava menos tempo que os demais sabores, uma vez que o sorvete de baunilha ficava na entrada da loja, enquanto outros sabores no fundo. Ele também notou que havia uma peça no carro que precisava resfriar para que pudesse funcionar. Dessa forma o engenheiro conseguiu fazer uma associação ao tempo que era gasto para comprar o sorvete com o tempo que era necessário para que a peça pudesse ser resfriada, desvendando desta forma, o mistério.

Essa analogia mostra como funciona um processo de um sistema BI, mesmo que tenha sido feita manualmente, sem o auxílio computacional. A análise das informações disponíveis junto com o conhecimento do engenheiro possibilitou a descoberta do problema. Um sistema BI irá fazer exatamente isso, encontrar nos dados disponíveis informações que não estejam em evidencia e tenham relevância para determinada situação. Porém, diferente da analogia, o processo BI tende a ser muito mais eficiente, uma vez que as informações são expostas de forma imediata.

Business Intelligence (BI) é um termo “guarda-chuva” que inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias. Os principais objetivos do BI são permitir o acesso interativo dos dados (às vezes, em tempo real), proporcionar a manipulação desses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar a análise adequada. O Processo do BI baseia-se na transformação de dados em informações, depois em decisões e finalmente em ações (TURBAN, et al, 2009, p. 27).

Como isso o BI não é apenas uma ferramenta que alimenta indicadores e coleciona dados, é uma ferramenta que junto com seu senso crítico é capaz de apontar com precisão o caminho a ser tomado com maior porcentagem de assertividade.

3.4 DIFERENÇA DE DADO, INFORMAÇÃO E CONHECIMENTO

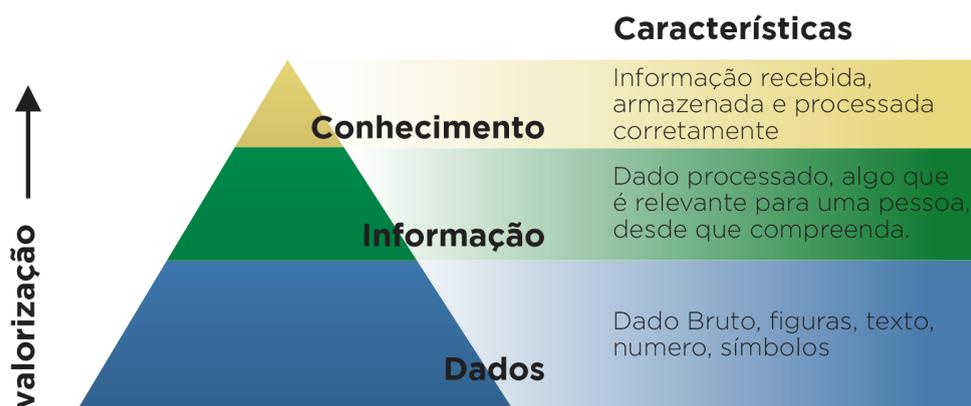
Dado são componentes brutos sem significado e vínculos com a realidade, podemos citar como exemplos: fotos, figuras, texto etc. “Defino dado como uma sequência de símbolos quantificados ou quantificáveis” e “um dado é necessariamente uma entidade matemática e, desta forma, é puramente sintático” (SETZER, 2014, p. 1).

Informação são os dados processados, ou seja, é algo que é representado como relevância para uma pessoa. “A informação pode ser propriedade interior de uma pessoa ou ser recebida por ela”, isto é, “Ao ler um texto, uma pessoa pode absorvê-lo como informação, desde que o compreenda” (SETZER, 2014, p. 2).

Conhecimento é um item inestimável, quer dizer, é a informação recebida, armazenada e processada corretamente. “Caracterizo conhecimento como uma abstração interior, pessoal, de algo que foi experimentado, vivenciado, por alguém” (SETZER, 2014, p. 3).

Esses tópicos são de extrema importância para uma tomada de decisão precisa.

Figura 1: Pirâmide dos Dados



Fonte: Próprio Autor (2020)

A Figura 1 representa esse conceito hierárquico (pirâmide). A base da pirâmide representa os dados brutos, esses têm o menor valor em uma tomada de decisão e possuem os maiores volumes. Ao passo que subimos de nível aumentamos a complexidade e o valor ao passo em que diminuímos o volume.

Os dados representam apenas valores isolados. A análise e combinação dos dados resultam nas informações, que são mensagens com significado e contextualização. Por sua vez, uma combinação de informações tem o poder de gerar conhecimentos. O conhecimento está diretamente relacionado com a capacidade humana de refletir e absorver ideias. A experiência humana está diretamente ligada ao conhecimento. Em uma tomada de decisão será o conhecimento que irá desempenhar a função mais importante, auxiliando o gestor nesse processo.

3.5 ETL: *EXTRACT, TRANSFORM AND LOAD*

O *ETL* (*Extract, Transform and Load*) que quer dizer do inglês Extração, Transformação e Carga é a parte do processo mais desafiadora para a gestão de TI, por envolver todos os dados, que costuma envolver mais de 70% do projeto.

No coração da parte técnica do processo de *data warehouses* estão extração, transformação e carga (ETL). As tecnologias de ETL, que já existem há algum tempo, são providenciais para o processo e uso de *data warehouses*. O processo de ETL é um componente integral de qualquer projeto centrado em dados. Os gerentes de TI constantemente enfrentam desafios, pois os processos de ETL costumam consumir 70% do tempo em um projeto centrado em dados. (TURBAN, et al, 2009, p.72)

Turban et al (2009), descreve que seu principal objetivo é carregar os dados de forma limpa no *Data Warehouse* e integrados. A origem desses dados podem ser um ou mais bancos de dados, além de fontes e tipo diversificadas que tenha a sua importância para o devido fim. Durante esse processo também é feito a transformação para que sejam armazenados na forma que precisa ser analisado no momento em que os dados é movimento entra a origem e destino.

O processo de ETL consiste em extração (leitura dos dados de um ou mais bancos de dados), transformação (conversão dos dados extraídos de sua forma anterior na forma em que precisam estar, para que sejam colocados em um *Data Warehouse* ou apenas em outro banco de dados) e carga (colocação dos dados no *Data Warehouse*) (TURBAN, et al, 2009, p. 72).

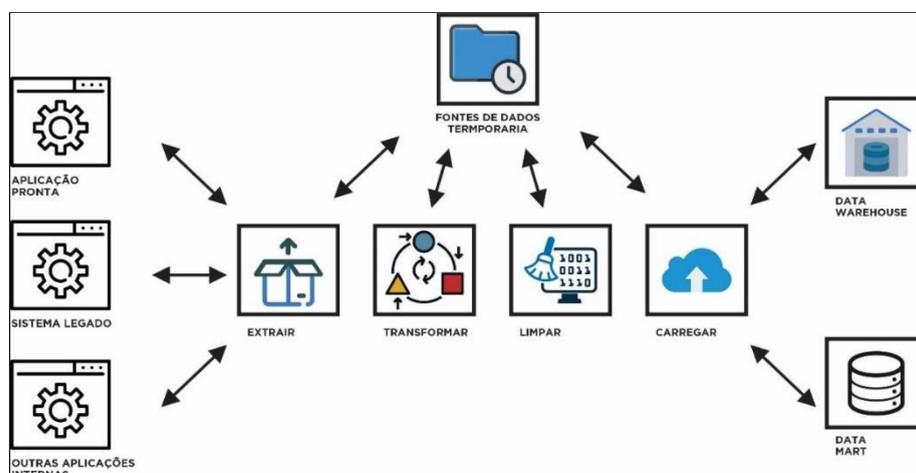
As etapas de ETL fundamentam-se em Primark (2008):

- Definir as fontes de dados: Esses dados podem ser de fontes internas (ERP, bancos de dados etc.) e externas (sistemas fora da empresa, planilhas ou arquivos externos).
- Extrair os dados: A extração é a leitura dos dados com base na fonte de dados definida, no qual o sistema de ETL é capaz de executar independentemente dos sistemas operacionais.
- Transformar e limpar: A transformação é realizada com base nos dados extraídos, diante disso é realizada a limpeza nas imperfeições e uma combinação entre os dados normalizando-os. Nessa etapa, várias regras são aplicadas para afinar os dados que serão carregados. Nem todas as fontes de dados carecem de manipulação, porém é necessário realizar a manipulação em algumas fontes de dados, tendo como exemplo: Tradução de campos codificados (Fonte de dados de origem armazena 1 para o sexo masculino e 2 para o feminino, contudo o DW armazena M para o masculino e F para o feminino), derivação de um novo valor calculado, junção de dados de diversas fontes, etc.

Carga dos dados ou análise dos resultados: A carga é colocar os dados transformados em um *Data Warehouse*. Os DW podem adicionar dados a cada hora, como também repor semanalmente informações existentes, com dados

cumulativos e atualizados. Sistemas complexos mantêm um histórico de todas as alterações sofridas pelos dados.

Figura 2: Etapas do ETL



Fonte: Ilustração próprio autor (2020)

A figura 2 representa a arquitetura de BI da esquerda para direita, o primeiro nível refere-se aos sistemas legados que possuem várias bases de dados diferentes, são neles que estão armazenados todos os dados transacionais da empresa, além de sistemas secundário e planilhas. A camada do meio é do ETL é responsável pela extração, transformação e carga dessas diferentes fontes de dados para uma única fonte, chamada *Data Warehouse*. Os *Data Mart* são subconjuntos do *Data Warehouse*, que determinam demandas específicas de uma área de negócio.

3.6 DATA WAREHOUSE

Data Warehouse ou Armazém de dados é um repositório de dados desenvolvido para apoiar a tomada de decisão. Ele é um banco de dados que possui como principal característica a capacidade de extrair dados que geralmente são inseridos pelos sistemas tradicionais baseados em OLTP (*Online Transaction Processing* ou Processamento de Transações em Tempo Real).

Também é organizado por assunto, ou seja, agrupa as informações mais importantes da empresa, quem em comparação aos sistemas operacionais que são orientados a processos. Os dados de interesse dos dois são diferentes também, visto que no *Data Warehouse* os dados são importantes para a tomada de decisão e nos sistemas transacionais os dados existem para controle operacional.

Nesse contexto, o *Data Warehouse* proporciona uma sólida e concisa integração dos dados da empresa, para a realização de análises gerenciais estratégicas de seus principais processos de negócios. Ele se preocupa em integrar e consolidar as informações de fontes internas, na maioria das vezes heterogêneas, e fontes externas, resumindo, filtrando e limpando esses dados, preparando-os para análise e suporte à decisão (MACHADO, 2007, p.27).

A característica mais importante em um *Data Warehouse* é a integração dos seus dados, uma vez que as informações carecem de unicidade. “Por exemplo, as convenções de nomes, valores de variáveis, tais como sexo masculino e feminino, e outros atributos físicos de dados como data types são formalmente unificados e integrados nessa base única” (MACHADO, 2007, p.31)

A integração de dados compreende três grandes processos que, quando implementados corretamente, permitem que os dados sejam acessados por e disponibilizados a uma gama de ferramentas de ETL e análise e ao ambiente de data warehousing. Os processos são: acesso aos dados (a capacidade de acessar e extrair dados de qualquer fonte), federação de dados (a integração das visualizações de negócios em diversos data stores) e captura de alterações (com base na identificação, captura e entrega das alterações feitas nas fontes de dados da empresa). (TURBAN, et al, 2009, p. 72).

O *Data Warehouse* possui dados não volátil, consiste que uma vez armazenado, os dados não são perdidos e não sofrem alteração; diferentemente dos bancos de dados transacionais que tem as operações: incluir, excluir, alterar e acessar. “Um *Data Warehouse* tem duas operações básicas: a carga dos dados (inicial e incremental) e o acesso a esses dados em modo de leitura” (MACHADO, 2007, p.30).

3.7 DATA MART

O *Data Warehouse* contém o conjunto de dados de toda a organização, o *Data Mart* foca apenas uma parte dos dados, geralmente focado em assuntos ou departamentos. Na visão de Kimball (2000) o *Data Mart* é um subconjunto lógico e físico do *Data Warehouse*. Em um primeiro momento os *Data Marts* eram classificados como um subconjunto altamente agregado de dados, que normalmente eram selecionados para uma questão específica do negócio, mas essa definição não funcionou.

Data Mart (DM) é um subconjunto de dados do *Data Warehouse* que são direcionados a uma determinada unidade de negócios, sendo executados separadamente por departamentos. “Uma das principais vantagens de seu emprego é a possibilidade de retorno rápido, garantindo um maior envolvimento do usuário final, capaz de avaliar os benefícios extraídos de seu investimento” (MACHADO, 2007, p.44). Logo um conjunto de Data mart forma um Data warehouse, os DM são espelhos de determinado setor/assunto da empresa que facilita o acesso a informações com alta performance

Essa arquitetura permite muitas outras funções e capacidade de informação em relação à arquitetura independente. Por meio dessa integração, conseqüentemente, aumenta sensivelmente o nível de complexidade de requisitos (MACHADO, 2007, p.51).

Arquitetura de um *Data Mart* integrada, ao contrário da independente, os *Data Mart* são interconectados com as outras áreas de negócios, onde um pode acessar e utilizar os dados do outro.

3.8 DATA MINING

A Mineração de Dados (*Data Mining*) é a etapa de maior notoriedade no processo de Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases* - KDD). Ela utiliza um processo de descoberta de informações através de uma combinação de métodos tradicionais de análise de dados com algoritmos sofisticados (TAN, STEINBACH, KUMAR, 2009).

Data Mining é a exploração e análise, de forma automática ou semiautomática, de grandes bases de dados com objetivo de descobrir padrões e regras. O objetivo principal do processo de data mining é fornecer as corporações informações que a possibilitem montar melhores estratégias de marketing, vendas e suporte, melhorando assim os seus negócios (BERRY, LINOFF, 1997, p.7).

As técnicas de mineração de dados tornam possíveis traçar perfis dos mais diversos tipos de clientes, podendo dessa forma direcionar vendas, detectar fraudes e oferecer os melhores serviços para aquele determinado cliente. Também torna possível obter informações acerca de situações que poderão ocorrer, por exemplo, uma grande parte de um determinado grupo de cliente que comprou o produto x, tem grande tendência de comprar o produto y nos próximos 90 dias. Vale lembrar que a mineração de dados não está limitada apenas ao uso de tomadas de

decisões, seu uso vai muito além e atinge as mais diversas situações (TAN, STEINBACH, KUMAR, 2009).

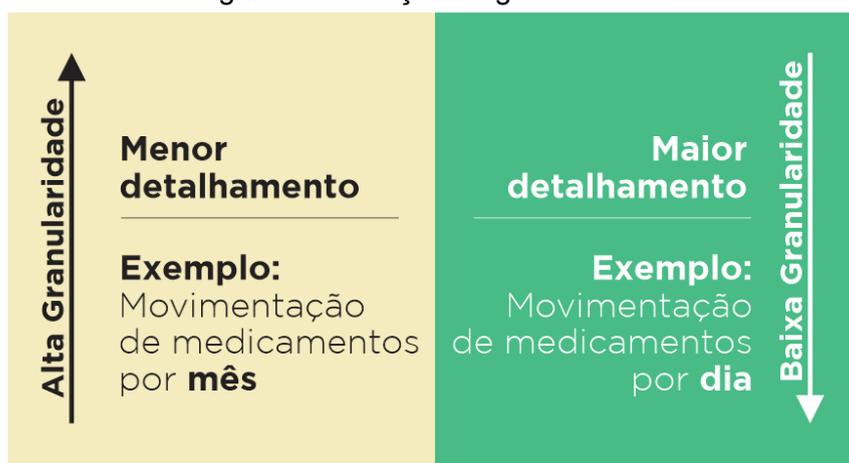
3.9 GRANULARIDADE

A Granularidade é responsável pela definição do nível de detalhamento dos dados, esse é um dos aspectos mais importantes no projeto de um *Data Warehouse*. Ela irá definir se os dados serão resumidos ou se apresentarão um alto nível de detalhamento. Um nível muito baixo de granularidade irá resultar em uma grande quantidade de dados armazenados no *Data Warehouse*. O baixo nível também irá afetar o desempenho, uma vez que terá uma quantidade maior de dados para serem analisados, dessa forma podendo fazer com que as consultas sejam mais demoradas (INMON, 1997).

Para Inmon (1997, p.45): O mais importante aspecto do projeto de um *Data Warehouse* é a questão da granularidade. A granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no *Data Warehouse*. Quanto mais de detalhe, mais baixo o nível de granularidade. Quanto menos detalhe, mais alto o nível de granularidade.

Já o inverso, um alto nível de granularidade pode afetar a qualidade das consultas, resultado do alto resumo dos dados, gerando em consequência consultas tendenciosas.

Figura 3: Ilustração da granularidade



Fonte: Próprio Autor (2020)

É fundamental analisar o tipo de métrica aplicada na tabela fato, “Explicando de outro modo, com um nível muito baixo de granularidade, é possível responder a

qualquer consulta. Todavia, com um alto nível de granularidade, o número de questões a que os dados podem satisfazer é limitado.” (INMON, 1997, p. 47).

3.10 METADADOS

O metadados é um repositório único de dados, ou seja, dados sobre dados. “Metadado é uma abstração do dado, capaz, por exemplo, de indicar se uma determinada base de dados existe, quais são os atributos de uma tabela e suas características, tais como: tamanho e/ou formato.” (CERQUEIRA, 2015, §8).

Inmon (1997, p. 163) destaca alguns componentes do metadados:

- Estruturas das tabelas do *Data Warehouse*;
- Atributos das tabelas do *Data Warehouse*;
- Fonte de dados do *Data Warehouse* (o sistema de registro);
- Correspondência entre o sistema de registro e o *Data Warehouse*;
- Especificação do modelo de dados;
- *Logging* de extrações;
- Rotinas comuns de acesso a dados.

Um importante aspecto do ambiente de *Data Warehouse* diz respeito aos metadados. Metadados são dados sobre dados. Os metadados fazem parte do meio de processamento de informações há tanto tempo quanto os programas e os dados. Contudo, é no mundo dos *Data Warehouses* que os metadados assumem um novo nível de importância. É por meio dos metadados que a utilização mais produtiva do *Data Warehouse* é alcançada (INMON, 1997, p. 108).

Outro aspecto muito importante dos metadados descrito por Inmon (1997), remete-se ao gerenciamento do mapeamento entre o ambiente operacional e o ambiente de *Data Warehouse*. Os dados que são carregados no *Data Warehouse* passaram por mudanças distintas durante todo o processo de ETL, entre essas transformações ocorrem: conversões, filtrações, resumos e alterações estruturais. É no metadados onde ficam armazenadas todas as informações sobre as transformações que ocorreram com os dados. Essas informações contidas no metadados também possibilita chegar à fonte de dados operacional (origem) a partir do *Data Warehouse*.

3.11 FATO

O Fato corresponde a uma tabela de fatos central que contém dados de medidas e de contexto ligada a diversas tabelas de dimensão. “Fato é tudo aquilo que pode ser representado por um valor aditivo, ou melhor, sem academicismos, por meio de valores numéricos. Esse conjunto de valores numéricos é denominado de métricas ou medidas simplesmente.” (MACHADO, 2007, p. 100).

Um fato pode ser a venda de um produto no varejo, o preço de uma ação em um determinado momento no tempo, a quantidade de cobertura de seguro incluída em uma nova apólice sendo criada, o balanço de uma conta ou a alteração em seu salário como resultado de sua promoção. Todos esses fatos são numéricos. Alguns deles são verdadeiros em um instante específico no tempo, e outros representam uma medida acumulada sobre um período de tempo (KIMBALL, MERZ, 2000, p. 122).

Nessa tabela não é para conter informações descritivas e sim as medidas do negócio, onde será determinada de forma quantitativa. Essas medidas são definidas em três tipos, Elias (2014):

- Aditivas: Podem ser sumarizadas, cruzando qualquer uma de suas dimensões.
- Semi-aditivas: Podem ser sumarizadas, cruzando apenas uma parte de suas dimensões.
- Não aditivas: Não podem ser sumarizadas por nenhuma das suas dimensões.

3.12 DIMENSÃO

As **Dimensões** são tabelas que possuem atributos que descrevem os dados contidos na tabela de fatos, normalmente elas não possuem atributos numéricos, uma vez que são apenas descritivas. As dimensões também irão ser as responsáveis pela forma de como os dados serão analisados (MACHADO, 2007).

Os atributos numéricos que representam um fato são denominados de **Medidas**. “Uma medida é determinada pela combinação das dimensões que participam de um fato, e estão localizadas como atributos de um fato.” (MACHADO, 2007, p. 81).

3.13 MODELAGEM MULTIDIMENSIONAL

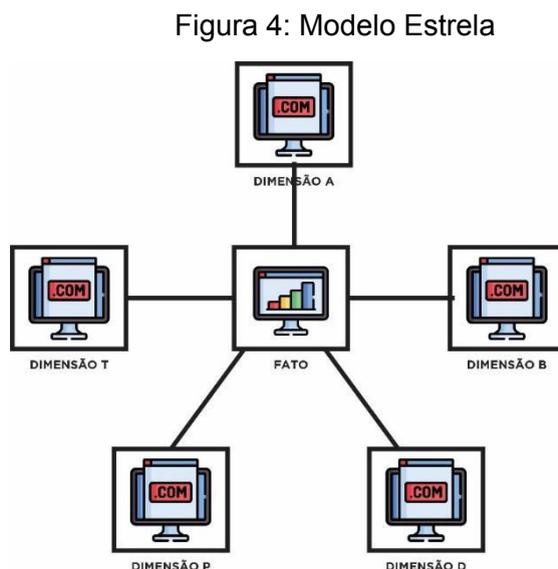
O modelo dimensional é aprimorado para consultas e promove a compreensão e melhor desempenho dos usuários, para Kimball (2002), “modelagem dimensional é uma técnica de design de banco de dados projetada para suportar consultas dos usuários finais de um *data warehouse*”.

Modelagem dimensional é uma disciplina que busca modelar dados para aprimorar o entendimento e o desempenho. Modelos dimensionais são o formato preferido para a apresentação dos dados no *Data Warehouse*. A modelagem dimensional é uma alternativa à disciplina de modelagem tradicional de entidade-relação (E/R), com a qual a maioria de nós crescemos (KIMBALL, MERZ, 2000, p. 121).

Em sua estrutura as tabelas de um modelo dimensional são denominadas como tabela Fato, sendo está a principal e a central da formatação, e as demais tabelas são as de Dimensão. A tabela de Fatos armazena a referência temporal, as chaves estrangeiras das dimensões e os valores quantitativos das medidas. A tabela Dimensão caracteriza e institui atributos ao fato, melhorando o desempenho da consulta.

Os modelos mais conhecidos desse em que as tabelas podem ser predispostas são o modelo Estrela ou o modelo floco de neve.

O modelo estrela é caracterizado por todas as tabelas de dimensão se relacionar diretamente com a tabela fato, lembrando a estrutura de uma estrela, conforme ilustra a figura 4:

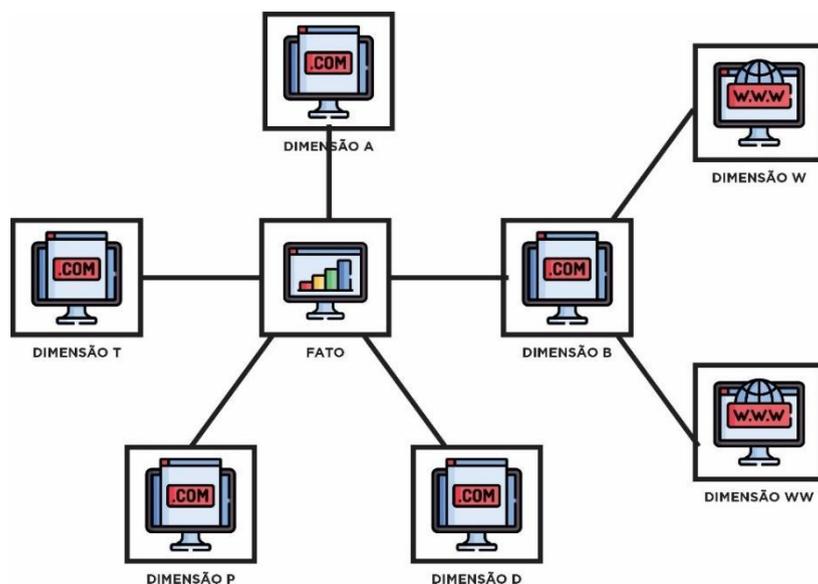


Fonte: Próprio Autor (2020)

Moreira (2006) afirma que o acesso mais rápido e simples aos dados e uma navegação mais inteligível favorece o uso do modelo estrela ao floco de neve.

Já o modelo floco de neve é uma variação do modelo estrela, pois as tabelas de dimensão se relacionam também com a tabela fato, mas algumas dimensões auxiliares se relacionam apenas entre elas, conforme retratado na figura 5:

Figura 5: Modelo Floco de Neve



Fonte: Próprio Autor (2020)

O modelo Floco de Neve reduz o espaço de armazenamento dos dados, mas insere várias tabelas, tornando a modelagem dimensional mais complexa. Outro fator é que o desempenho das consultas é afetado por conta da necessidade de mais JOINS de instruções SQL. Já no modelo Estrela os relacionamentos são mais simples e conseqüentemente possui melhor desempenho no retorno das consultas do que o modelo anterior.

3.14 5W2H

A ferramenta 5W2H foi criada como uma ferramenta auxiliar na utilização do PDCA, mais precisamente na parte de planejamento, por profissionais da indústria automobilística do Japão. A técnica é uma ferramenta prática que permite, a qualquer momento, identificar dados e rotinas mais importantes de um projeto ou de uma unidade de produção (SEBRAE, 2008).

Segundo Sebrae (2012) essa ferramenta consiste em um plano de ação para atividades pré-estabelecidas que tem a necessidade de serem desenvolvidas com a

maior clareza possível, além de funcionar como um mapeamento dessas atividades. Também tem como objetivo principal responder a sete questões e organizá-las.

Segundo o SEBRAE (2008), a ferramenta 5W2H é prática e permite, a qualquer momento, identificar as rotinas mais importantes de um processo, projeto ou até mesmo de uma unidade de produção. Também possibilita identificar quem é quem dentro da organização, o que faz e porque realiza tais atividades. O método é constituído de sete perguntas, utilizadas para implementar soluções conforme a tabela 1:

Tabela 1: Método Do 5W2H

Método do 5W2H			
5W	<i>What</i>	O que?	Que ação será executada?
	<i>Who</i>	Quem?	Quem irá executar/participar da ação?
	<i>When</i>	Onde?	Onde será executada a ação?
	<i>When</i>	Quando?	Quando a ação será executada?
	<i>Why</i>	Por que?	Por que a ação será executada?
2H	<i>How</i>	Como	Como será executada essa ação?
	<i>How much</i>	Quanto Custa?	Quanto custará para executar a ação?

Fonte: Próprio Autor (2020)

Dentro de cada uma dessas perguntas, inúmeros outros questionamentos podem ser feitos para melhor entendimento da função do 5W2H:

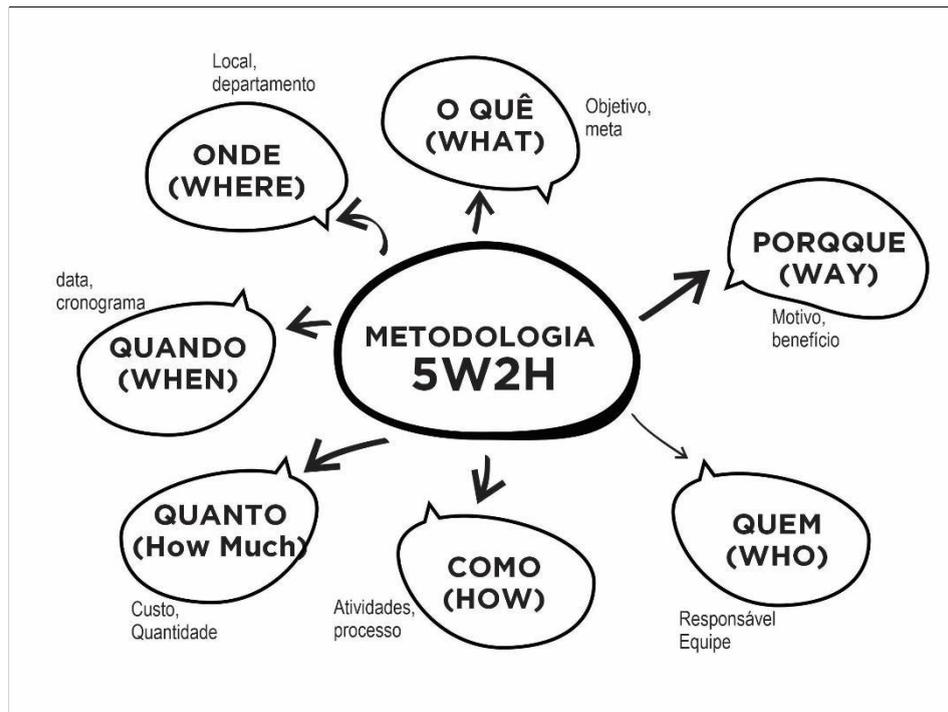
- a) O quê? Qual a atividade? Qual é o assunto? O que deve ser medido? Quais os resultados dessa atividade? Quais atividades são dependentes dela? Quais atividades são necessárias para o início da tarefa? Quais os insumos necessários?
- b) Quem? Quem conduz a operação? Qual a equipe responsável? Quem executará determinada atividade? Quem depende da execução da atividade? A atividade depende de quem para ser iniciada?

- c) Onde? Onde a operação será conduzida? Em que lugar? Onde a atividade será executada? Onde serão feitas as reuniões presenciais da equipe?
- d) Por quê? Por que a operação é necessária? Ela pode ser omitida? Por que a atividade é necessária? Por que a atividade não pode fundir-se com outra atividade? Por que A, B e C foram escolhidos para executar esta atividade?
- e) Quando? Quando será feito? Quando será o início da atividade? Quando será o término? Quando serão as reuniões presenciais?
- f) Como? Como conduzir a operação? De que maneira? Como a atividade será executada? Como acompanhar o desenvolvimento dessa atividade? Como A, B e C vão interagir para executar esta atividade?
- g) Quanto custa realizar a mudança? Quanto custa a operação atual? Qual é a relação custo / benefício? Quanto tempo está previsto para a atividade? Através dessas perguntas é possível direcionar, planejar, definir as responsabilidades e quantificar as ações.

Ainda segundo o SEBRAE (2008), a técnica 5W2H é uma ferramenta simples, porém poderosa, para auxiliar a análise e o conhecimento sobre determinado processo, problema ou ação a serem efetivadas, podendo ser usado em três etapas na solução de problemas:

- a) Diagnóstico: na investigação de um problema ou processo, para aumentar o nível de informações e buscar rapidamente as falhas;
- b) Plano de ação: auxiliar na montagem de um plano de ação sobre o que deve ser feito para eliminar um problema;
- c) Padronização: auxilia na padronização de procedimentos que devem ser seguidos como modelo, para prevenir o reaparecimento de modelos

Figura 6: Mapa Mental 5W2H



Fonte: Próprio Autor (2020)

A figura 6 mostra o uso da ferramenta 5w2h em forma de mapa mental que facilita a aplicação de forma do centro para as extremidades simplificando o uso sem necessariamente seguir uma ordem sequencial para respostas das questões.

3.15 FERRAMENTA QLIK SENSE

A ferramenta *Qlik Sense* é focada para o usuário que prefere autonomia e flexibilidade usando o conceito de auto-service, onde o usuário é capaz de desenhar e modificar vários painéis com seus indicadores, que não dependerá exclusivamente do departamento de TI da Empresa para poder produzir conteúdo informativo. (QLIK, 2020).

O Qlik® Sense define o padrão para uma nova geração de analytics. Com seu motor de análise associativa exclusivo, IA sofisticada e arquitetura multi-cloud escalável, você permite que todos na sua organização tomem melhores decisões diariamente, criando uma empresa verdadeiramente orientada por dados (QLIK, 2020).

Com essa ferramenta facilitar que as perguntas sejam respondidas de forma rápida e dinâmica pelos os gestores. Além de poder formar varias combinações de dados.

O objetivo principal desta tecnologia inovadora e revolucionária é possibilitar que perguntas possam ser feitas pelos gestores sem que

os mesmos estejam presos ou engessados a uma estrutura hierárquica de dados pré-definida, dando em tempo real análise infinita, possibilidade de arranjo e/ou combinação de análises. (SMARTSOLUTIONS, 2020, § 2).

Os dados do aplicativo são selecionados rapidamente conforme as seleções dos usuários. “Um usuário pode fazer uma pergunta no *Qlik Sense* de diversas maneiras diferentes, por exemplo, ao delimitar uma área em diagramas, gráficos e mapas, ao clicar em itens em listas, ao utilizar deslizadores e ao selecionar datas em calendários” (VISÃO..., 2011, p. 5).

3.16 FERRAMENTA TALENT DATA INTEGRATION

De acordo com Altos (2008), a empresa *Talend* é um dos primeiro a fornecer software OpenSource para integração de dados.

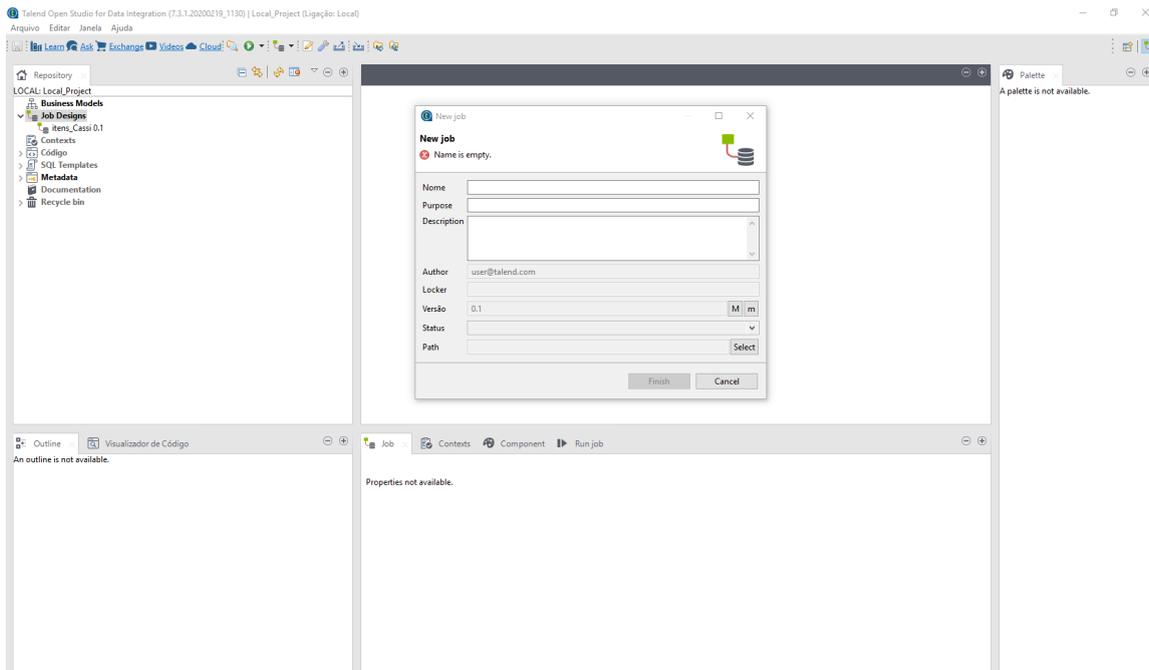
E segundo Bowen (2012), a ferramenta contém mais de 600 conectores pré construídos que facilita a conexão com o banco de dados, transformações de arquivos, carga de dados, copiar e renomear arquivos e conectar componentes individuais a fim de construir um processo de integração complexo.

Uma ferramenta de fácil utilização para desenvolvedores experientes e pessoas não desenvolvedores. A figura a seguir mostra a interface da ferramenta *Talend Open Studio for Data Integration* versão 7.3.1

Um software de integração de dados corporativos para conectar, acessar e transformar qualquer dado na nuvem ou no local (TALEND, 2020).

A Ferramenta foi desenvolvida na linguagem Java e tem como base a plataforma do Eclipse, aproveitando todos os recursos que a IDE disponibiliza na hora de começar um job ou debugar o código. Além de oferecer de ofertar um ambiente de desenvolvimento amigável, baseado na composição de componentes visuais. Isso beneficia os desenvolvedores que não estão acostumados com a linguagem Java a utilizar a ferramenta. A figura 7 mostra a interface gráfica da ferramenta.

Figura 7: Interface de desenvolvimento da Ferramenta *Talend Open Studio for Data Integration*



Fonte: Próprio Autor (2020)

No *Talend* a criação de Jobs é bastante comum e configurar componentes em vez de codificar, os Jobs podem ser executados a partir do ambiente gráfico de desenvolvimento ou executado como scripts independentes (Bowen, 2012). Na execução é o fluxo é visualizado graficamente além de possuir um debug para tratativas de erros.

Funciona em diversas plataformas como Windos, Linux, Mac e na cloud.

Para Bowen (2012), essa ferramenta fornece os seguintes benefícios:

- É OpenSource, livre para download e uso, com acesso a código fonte, permitindo aos utilizadores estender o produto de acordo com as suas necessidades específicas.
- É um grande impulsionador da produtividade. É fácil de aprender e rápido para desenvolvimento. Mesmo os desenvolvedores principiantes poderão construir integrações complexas de uma forma muito rápida.
- Usa componentes pré-construídos para auxiliar nas tarefas.
- Tem uma grande comunidade de utilizadores aberta e ativa.

3.17 SISTEMA IW CARE

De acordo com o IW SOFTWARE o sistema IW CARE é uma ferramenta que é possível monitorar todos os detalhes da execução do plano de atenção domiciliar dos pacientes, desde a captação onde encontrasse no hospital até a sua conclusão na desospitalização.

O IW Care Oferece um ambiente integrado que propicia um ótimo controle online do planejamento a programação e controle de todos os aspectos de atenção domiciliar.

- Automação dos processos e redução dos custos da atenção domiciliar
- Prontuário eletrônico especializado para a atenção domiciliar
- Controle 100% automatizado da dispensação
- Ambiente integrado de apuração e análise dos custos online
- Otimização da prescrição médica e de enfermagem
- Geração automatizada de relatórios clínicos periódicos

3.18 SISTEMA WFS SOFTWARE

A WFS Software de (Wellington Farias de Sousa) entrou no mercado em 1993 quando Wellington decide deixar seu trabalho de programador na UNIPE e ampliar seus conhecimentos com desenvolvimento de softwares, oferecendo esse tipo de serviço para pequenas transportadoras de mercadorias e livrarias de João Pessoa.

A marca WFS Software consolidou-se no mercado em 2005, com o desenvolvimento de um produto chamado SET_PV (marca registrada da Empresa), conquistando um número crescente de usuários/clientes. O atendimento da WFS também precisou passar por modificações inerentes ao fluxo da demanda exigida pela rotatividade e sucesso que o software alcançava, despontando a WFS entre as líderes no mercado de softwares. (WFSSOFTWADE, 2020)

3.19 POSTGRES

De acordo com MILANI O PostgreSQL é um sistema gerenciador de banco de dados (SGBD) relacional, utilizando para armazenar informações de soluções de informática em todas as áreas de negócios existente, bem como administrar o acesso e está informações.

O PostgreSQL é um poderoso sistema de banco de dados relacional de objeto de código aberto que usa e estende a linguagem SQL combinada com muitos recursos que armazenam e escalam com segurança as cargas de trabalho de dados mais complicadas. As origens do PostgreSQL remontam a 1986 como parte do projeto POSTGRES da Universidade da Califórnia em Berkeley e tem mais de 30 anos de desenvolvimento ativo na plataforma principal. (MEDIUM, on-line)2020

Portanto o PostgreSQL é um banco de dados de código aberto e arquitetura confiável. Podendo integrar com todos os principais sistemas operacionais.

O PostgreSQL ganhou uma forte reputação por sua arquitetura comprovada, confiabilidade, integridade de dados, conjunto robusto de recursos, extensibilidade e dedicação da comunidade de código aberto por trás do software para fornecer consistentemente soluções inovadoras e de alto desempenho. O PostgreSQL é executado em todos os principais sistemas operacionais, é compatível com ACID desde 2001 e possui complementos poderosos, como o popular extensor de banco de dados geoespacial PostGIS. Não é surpresa que o PostgreSQL tenha se tornado o banco de dados relacional de código aberto de escolha para muitas pessoas e organizações. (POSTGRESQL, on-line)2020

Então com sua licença liberação, ele pode ser modificado e distribuído por desde que seja livre de encargos.

4 ESTUDO DE CASO

4.1 MATERIAIS E METODOS

A seguinte infraestrutura e *hardware* foi utilizada para realizar o trabalho:

Notebook, processador core i5 8gb de memória *ram*, 240 hd, para pesquisa bibliográfica, construção de testes , análise dos resultados e escrita da dissertação;

Ambiente de desenvolvimento Ferramenta *Talend Open Studio for Data Integration* para processo de ETL

Bando de Dados PostGres versão 12.1 para armazenas as informação do Datawarehouse

Ambiente de Analise, ferramenta *Qlik Sense* para construção do Dashboard

Servidor com processador core i5, 4gb de memoria ram, armazenamento de 240GB SSD,

4.2 BASE DE DADOS

A base de dados são de fontes distintas para análise, uma das bases para ser analisado é o cadastro de paciente, movimentações de insumos e os dados de uso de gases medicinais encontra se no sistema IW care em um banco de dados oracle. Já informações complementares foram extraídas de planilhas localizadas em setores distintos. Já informações referentes a qualidade da satisfação dos pacientes utiliza outra fonte de dados do sistema WFS.

A missão é extrair o máximo de informações de qualidade de todas as bases, limpar os dados de transformar em uma única base para poder ser analisado.

4.3 EXTRAÇÃO

Primeira Etapa é Criar a *Stage Area* ou área auxiliar, serve como ponto único para a carga efetiva no *data warehouse*, a cada carga seu conteúdo é limpo, evita a produção em caso de recarga durante o dia.

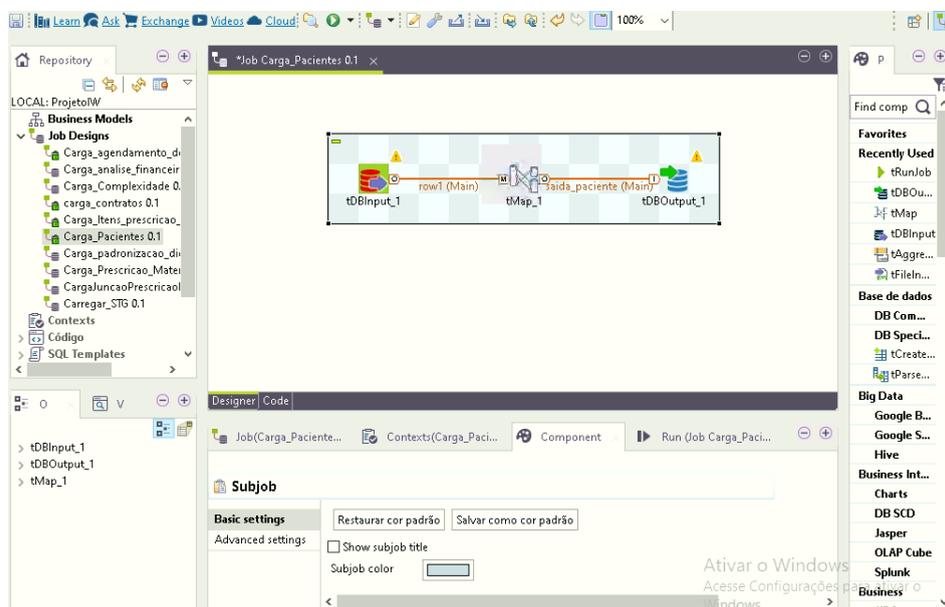
4.4 CRIAR O JOB

A área de Staging ou área de transição é o armazenamento de dados intermediário entre uma fonte de dados original e o repositório de dados centralizado. Toda purificação, reconciliação, transformação e relacionamentos necessários acontecem nessa área. Essa implementação é muito bem definida e progressiva. Ter uma área de Staging é uma estratégia comum e simples de obter informações de um sistema transacional sem a necessidade de realizar transformações complexas.

É comum colocar um registro chamado de sem informação (código = -1), isso devido à inconsistência de dados, pois podem ocorrer em projetos de DW, onde os dados do operacional não são íntegros. Por exemplo, CPF de um determinado paciente na tabela PACIENTE do banco de dados, cujo a informação não existe, ou não foi cadastrado, a colocação de uma identificação de inconsistência permitirá que a gestão da empresa veja que há dados inconsistentes na base. Colocado em

cada dimensão o código=-1, possibilitará que quando exibirmos os dados em uma ferramenta OLAP, Data Discovery ou Report qualquer, este seja identificado.

Figura 8: Processo de Extração no *Talend Studio*



Fonte: Próprio Autor (2020)

A figura 9 mostra como acontece essa carga usando alguns componentes nativos da ferramenta *Talend Open Studio for Data Integration*.

4.5 DIMENSÃO TEMPO

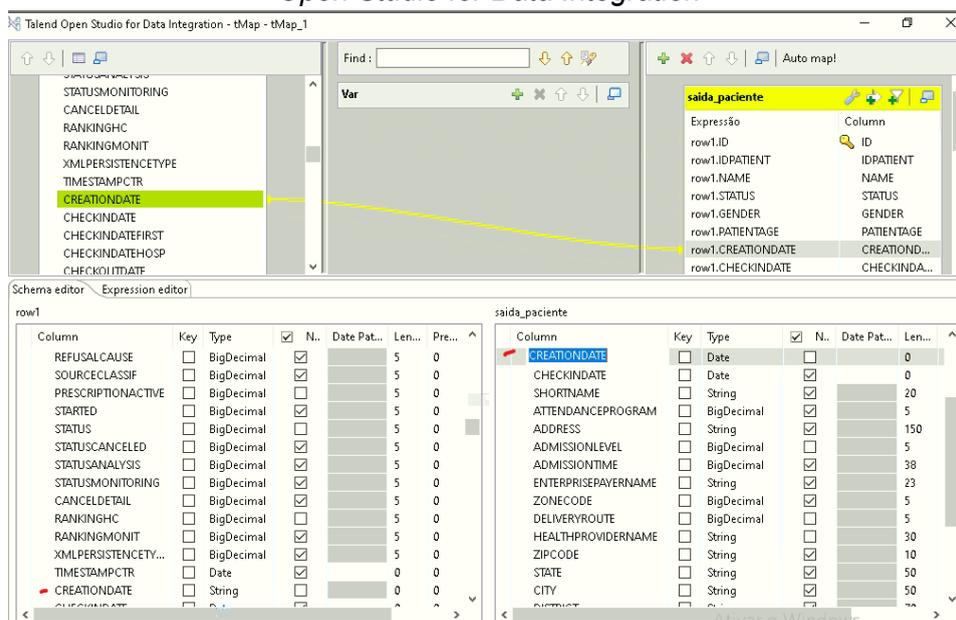
A dimensão TEMPO é uma das mais importantes que existem em um projeto de *Data Warehouse*, sem ela, você não consegue estabelecer análises temporais. Com os dados da dimensão TEMPO você identifica quando uma determinada métrica foi inserida.

4.6 TRANSFORMAÇÃO

O processo de transformação é feito pela carga da tabela Stage Area para o Data Warehouse. É nessa etapa que os dados são transformados e limpos, eliminando as informações que não são necessárias. Foi necessário dividir a transformação dois bancos prenominaos em STG_ e DW para identificar as tabelas dos grupos que serão carregados os dados puros da base e onde serão limpos para serem analisados

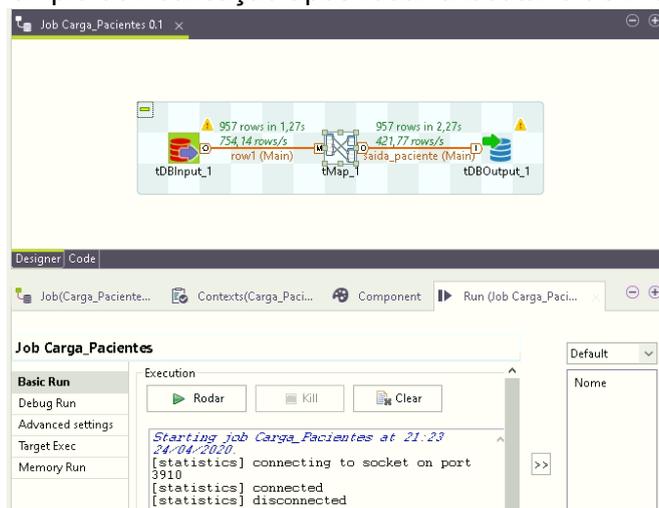
STG_PACIENTE foi realizado uma consulta no banco de dados Oracle em 17 tabelas para obter informações e serem gravadas em uma única tabela na Stage Area de pacientes os dados como por exemplo, Numero de atendimento, Operadora, Nome do Paciente, Data de Início de Atendimento, Motivo de Internação. Endereço e outros dados considerados importantes para ser gerar uma chave forte, alguns campos sofreram alteração de tipo como de STRING para DATE, BIGDECIMAL para INT

Figura 9 – Transformação de Dados usando o componente TMAP da ferramenta *Talend Open Studio for Data Integration*



Fonte: Próprio Autor (2020)

Figura 10 – Exemplo de notificação após rodar/executar o JOB Carga_Paciente



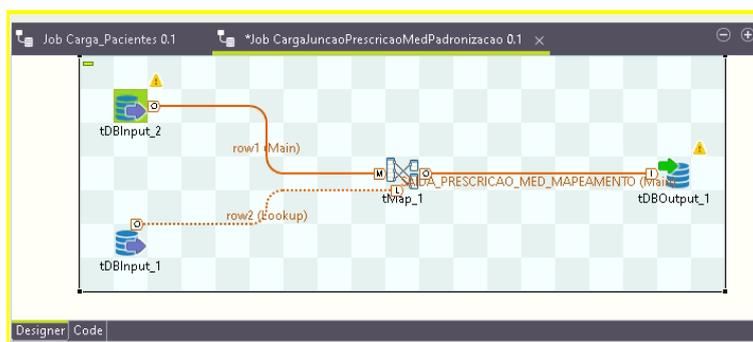
Fonte: Próprio Autor (2020)

a base de ANALISE FINANCEIRA, USO_DE_GASE, ANALISE_PRESCRICAO e PESQUISA_DE_SATISFAÇÃO foram realizados os

mesmos procedimento, porem a ANALISE_PRESCRICAO, além desse procedimento foi associada a uma tabela do Excel para realizar as padronização não existente no sistema nativo.

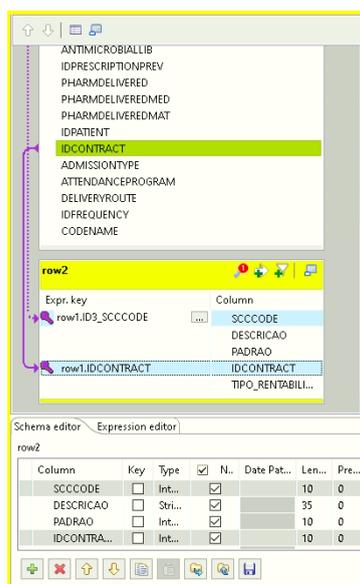
Realizado a inserção da tabela do excel para a STG_PADRONIZACAO, para unir com a tabela STG_ITENS_PRESCRICAO e ser transformado em uma única tabela CHAMADA DW_ITENS_PRESCRICAO conforme figura 12.

Figura 11: Job de Junção da Ferramenta *Talend Open Studio for Data Integration*



Fonte: Próprio Autor (2020)

Figura 12: Tmap da Junção da Ferramenta *Talend Open Studio for Data Integration*



Fonte: Próprio Autor (2020)

A figura 12 representa a Propriedades do componente TMAP da ferramenta *Talend Studio Integration* para executar a Junção das duas tabelas associando a padronização de uma tabela do Excel a uma Tabela já existente no postgres.

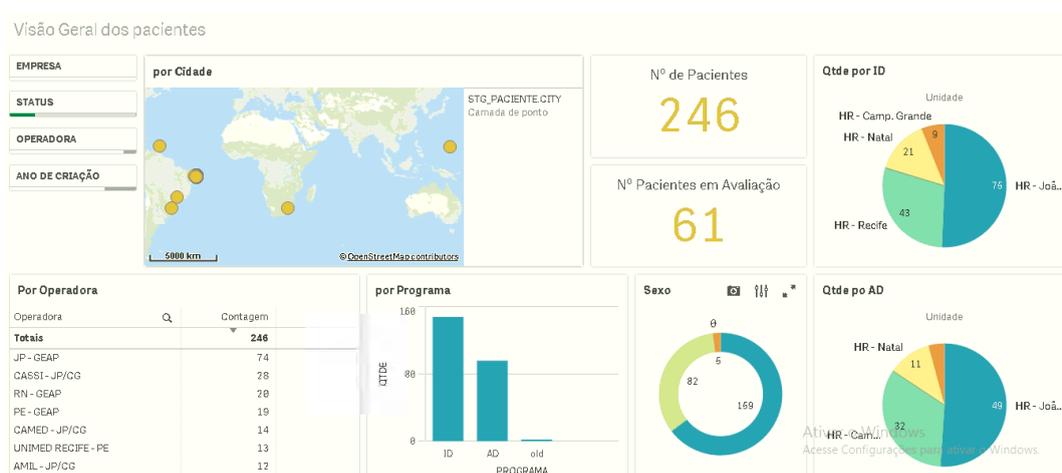
5 RESULTADOS E DISCUSSÕES

A partir de agora vamos usar a ferramenta a ferramenta *Qlik sense* para demonstrar todos os resultados em dashboard dos dados extraídos pela ferramenta *talend studio*.

5.1 ANÁLISE SENSO DOS PACIENTE

Para as análises propostas foram criados painéis ou dashboard com instrumentos visuais empregados gráficos do tipo barra, pizza para facilitar a compreensão dos líderes dos setores.

Figura 13 – Painel Senso dos Pacientes, análise quantitativo de pacientes por Plano, Faixa Etária, Programa de Internação, Por Cidades e Sexto



Fonte: Próprio Autor (2020)

O primeiro painel apresentado na figura 14 demonstra os quantitativos de pacientes por programa de internação, análise quantitativo de pacientes por Plano, Faixa Etária, Programa de Internação, Por Cidades e Sexto

Após realizar os filtros ou iterações no painel, é possível passar para o próximo painel, respeitando os filtros aplicados no painel anterior, assim é possível ver o nome de cada paciente e seu código de internação.

Figura 14 - Painel com geral dos pacientes,

Pasta Exibição dos pacientes

ID	Paciente	Sexo	Programa	Filial	Operadora
429	ADALCINA DA SILVA ARAUJO	Feminino	AD	HR - João Pessoa	CASSI - JP/CG
826	ADELMO FELIX DE LIMA	Masculino	AD	HR - Natal	RN - GEAP
86	ADEVAL CHAGAS DA SILVA	Masculino	AD	HR - Camp. Grande	CAMED - JP/CG
6	ADRIANA FERREIRA LINS ALBUQUERQUE	Feminino	ID	HR - João Pessoa	JP - GEAP
773	ALBA ELITA FONSECA DOS SANTOS LIMA	Feminino	AD	HR - Camp. Grande	CASSI - JP/CG
882	ALCIDES PAULO DE ALBUQUERQUE	Masculino	ID	HR - Recife	PE - GEAP
714	ALICIA PITT SALES SERENO	Feminino	AD	HR - João Pessoa	HOSPITAL RESIDENCIAL
878	ALZIRA CAMARA DE ANDRADE	Feminino	ID	HR - Natal	RN - GEAP
485	ALZIRA MARIA SILVA FREIRE	Feminino	AD	HR - Camp. Grande	JP - GEAP
812	ALZIRA MENDES DE MEDEIROS	Feminino	ID	HR - João Pessoa	FUSEX - JP
588	AMBROSINA RIBEIRO PESSOA	Feminino	ID	HR - Recife	UNIMED RECIFE - PE
548	ANA ALVES DE MELO	Feminino	AD	HR - Camp. Grande	JP - GEAP
789	ANA DE CARVALHO OLIVEIRA	Feminino	ID	HR - Natal	RN - GEAP
83	ANA FLAVIA CARLOS DE MELO	Feminino	ID	HR - João Pessoa	AMIL - JP/CG
783	ANA IVANETE DE LIMA FERNANDES	Feminino	ID	HR - Natal	RN - GEAP
392	ANTONIETA MARIA ARAUJO BARBOSA	Feminino	AD	HR - Camp. Grande	JP - GEAP
581	ANTONIO ALVES DA COSTA	Masculino	ID	HR - Recife	PE - GEAP
882	ANTONIO BERNARDINO DE SENA	Feminino	AD	HR - Camp. Grande	CASSI - JP/CG
881	ANTONIO CARLOS DA CUNHA LIMA	Masculino	AD	HR - João Pessoa	CAMED - JP/CG
887	ANTONIO FERNANDO NEGRIMONTE DA SILVA	Masculino	ID	HR - Recife	PE - GEAP

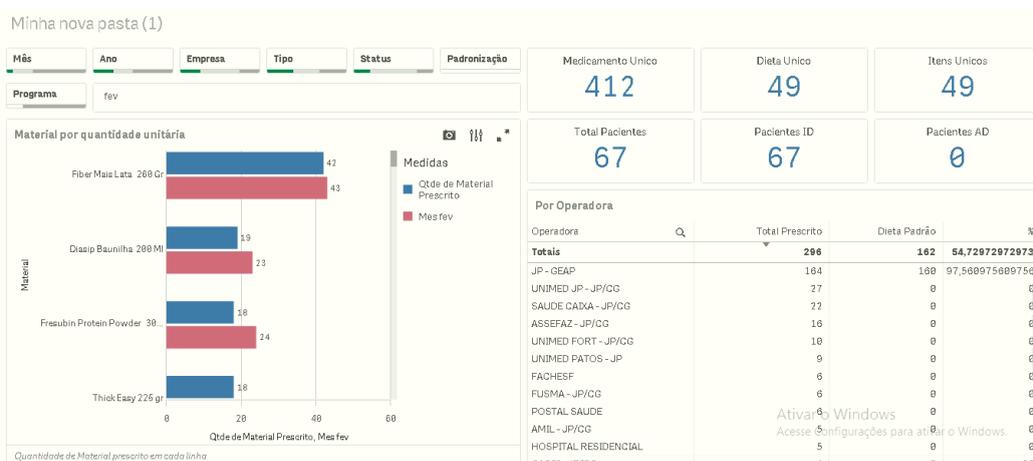
Fonte: Próprio Autor (2020)

A figura 14 mostra o resultado do filtro aplicado no painel anterior e ainda podendo executar um procedimento de download em forma de planilha eletrônica.

5.2ANÁLISE ITENS PRESCRITOS

Na análise de itens prescritos é possível verificar toda a movimentação de materiais, dietas e medicamentos alocados em casa paciente. Além de identificar se o item alocado é padronizado pela empresa.

Figura 15 - Painel para Análise de Itens Prescritos



Fonte: Próprio Autor (2020)

Como mostrado no painel da figura 14, foi possível também fazer uma exportação do detalhe do filtro aplicado no painel da figura 15.

Figura 15 - Painel de Análise de item a item prescrito.

Minha nova pasta

STARTDATE... STARTDATE... HEALTHPR... IDTABLE_TL... STATUS PADRAO Programa

ID	Q	PRES...	PACIENTE	Q	SCC...	Q	PRINC. AT	Q	NOME COMERCIAL	Q	QTDE	Q	MEDI
Totais													
520		3403	MARIA MADALENA PEREIRA DOS ANJOS		590150		Fresubin Protein Energy Drink Baunilha 200 MI		Fresubin Protein Energy Drink Baunilha 200 MI		200	ml	
520		3403	MARIA MADALENA PEREIRA DOS ANJOS		514119		Nutri Renal 2.0 Kcal/ml 200 MI		Nutri Renal 2.0 Kcal/ml 200 MI		200	ml	
520		3572	MARIA MADALENA PEREIRA DOS ANJOS		514119		Nutri Renal 2.0 Kcal/ml 200 MI		Nutri Renal 2.0 Kcal/ml 200 MI		200	ml	
134		3411	ERIKA RODRIGUES VASCONCELOS		514111		Glutamax (Glutamina) Saché 10 Gr		Glutamina Saché 10 Gr		1	envel	
134		3604	ERIKA RODRIGUES VASCONCELOS		514111		Glutamax (Glutamina) Saché 10 Gr		Glutamina Saché 10 Gr		1	envel	
301		3488	ISABEL PINTO DE MENEZES		514111		Glutamax (Glutamina) Saché 10 Gr		Glutamina Saché 10 Gr		1	envel	
301		3652	ISABEL PINTO DE MENEZES		514111		Glutamax (Glutamina) Saché 10 Gr		Glutamina Saché 10 Gr		1	envel	
316		3525	MARIA DAS NEVES DE SOUSA ARAÚJO		514111		Glutamax (Glutamina) Saché 10 Gr		Glutamina Saché 10 Gr		1	envel	
316		3701	MARIA DAS NEVES DE SOUSA ARAÚJO		514111		Glutamax (Glutamina) Saché 10 Gr		Glutamina Saché 10 Gr		1	envel	
38		3412	TEREZINHA FERNANDES VITORINO		514089		Probiotop Saché 1 Gr		Probiotop Saché 1 Gr		1	gram	
38		3598	TEREZINHA FERNANDES VITORINO		514089		Probiotop Saché 1 Gr		Probiotop Saché 1 Gr		1	gram	
47		3374	STELLA MARIA VASCONCELOS VELLOSO FREIRE		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		1	gram	
47		3586	STELLA MARIA VASCONCELOS VELLOSO FREIRE		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		1	gram	
53		3515	EUFROZINA RODRIGUES DA SILVA		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		6	gram	
53		3691	EUFROZINA RODRIGUES DA SILVA		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		6	gram	
62		3499	IRACEMA SOUZA BRITO		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		6	gram	
62		3675	IRACEMA SOUZA BRITO		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		6	gram	
334		3500	JAIRO DIAS PAREDES		514097		Simbioflora Saché 6 Gr		Simbioflora Saché 6 Gr		6	gram	

Fonte: Próprio Autor (2020)

A figura 15 exibe uma planilha com todos os pacientes e itens alocados por período.

Através da pesquisa de satisfação mensal é possível identificar através de alguns indicadores saber a satisfação do usuário /ou responsável pelo paciente.

Esse indicado foi produzido um *Chatbot* pela empresa WFS soft aonde alimenta um banco, para ser consumido pela ferramenta *Qlik sense*

Figura 16 - Painel de Análise de item a item prescrito.

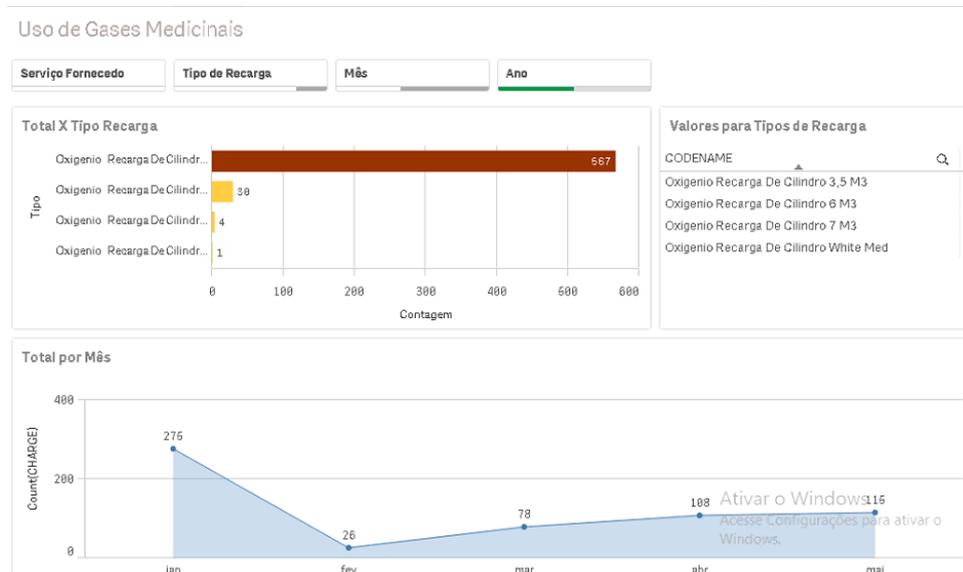


Fonte: Próprio Autor (2020)

A figura 16 apresenta os indicadores de satisfação de técnico de enfermagem, mobília / equipamentos, disponibilidades de materiais, dietas e medicamentos para o paciente, serviço de urgência e emergência, assistência da equipe multiprofissional e atendimento dos profissionais de enfermagem.

O uso de gases medicinais tem papel bastante significativo na conta de cada paciente, sua investigação para saber o total usado e os tipos de cilindros usados em cada paciente é bastante importante para fazer o fechamento das contas.

Figura 17 - Painel de Análise do uso de gases medicinais



Fonte: Próprio Autor (2020)

A figura 17 demonstra no *dashboard* com análise dos usos medicinais com seus totalizadores separados por tipo de recarga para saber qual tipo foi mais consumido no período filtrados, mostra uma linha do tempo para saber qual mês houve mais consumo de gases e tipos de recarga caso queira realizar um filtro.

6 CONSIDERAÇÕES FINAIS

A informação no âmbito empresarial tornou-se um dos ativos mais preciosos, pois com ela os gestores podem tomar decisões fundamentadas na situação real da empresa

Este trabalho apresentou um estudo de caso sobre as ferramentas *Talend Open Studio for Data Integration* para fazer integração ou extração de banco de dados no processo de ETL e *business intelligence Qlik Sense*, uma tecnologia muito usada atualmente para ajudar a tomada de decisões nas empresas, com a finalidade de fornecer suporte aos gestores. Com a intenção de buscar o melhor gerenciamento de cada área.

Inicialmente este trabalho focou em estudar a fundamentação teórica sobre o tema abordado, principalmente nos processos hospitalar, posteriormente foi desenvolvido o projeto com as ferramentas *Talend Open Studio for Data Integration* para fazer a extração e transformação dos dados e ferramenta de BI *qlik sense* para visualização das informações, o que tornou de forma mais simples a manipulação feita pelos usuários com as vantagens de retornar rápidas repostas ao que fora requisitado e de maneira mais clara.

Foi encontrado certa dificuldade em relação ao bando de dados, pois o bando de dados escolhido, não tinha nenhum conhecimento sobre PostgreSQL para poder construir o *data warehouse* além da identificação das tabelas do banco de dados de origem para extração dos dados, não existe nenhum dicionário de dados e mudança de nome de campos no decorrer do processo de desenvolvimento por parte da empresa desenvolvedora IWCARE, dificultando desenvolvimento do DW uma vez já construído. Além de uma melhoria no projeto é a aquisição da licença do *Qlik Sense* já que é um programa comercial e com essa aquisição é possível ter acesso a outras ferramenta como o *Qlik Sense Server* onde posso automatizar processo de carga de dados.

Tendo em vista todos os aspectos observados, pode-se concluir que o *business Intelligence* é uma tecnologia necessária, que auxiliam os gestores nas tomadas de decisões garantindo uma maior assertividade com base nas informações centralizadas e obtidas em diversas fontes de dados, e para que não haja frustração por parte do usuário faz necessário uma definição clara dos objetivos, com esta ferramenta é possível tomar decisões mais precisas, otimizar trabalho,

organizar, reduzir custo, estimativas de crescimento e construir para construção de novas estratégias.

A execução deste trabalho está de acordo com a metodologia apresentada, com pontos positivos na execução dos aplicativos, mostrou se promissor por apresentar inconsistência de vários cadastros e redução de tempo para análises antes feito em um quadro, hoje com a ferramenta totalmente integrada ao sistema. E mostra bastante promissor para análise de vários setores dentro da empresa como setor de enfermagem, setor financeiro e setor de compras. Com os resultados alcançados dar-se-á continuidade e uso da ferramenta dentro da organização.

REFERÊNCIA

ALTOS, L. (22 de 01 de 2008). Talend to Optimize Open Source Data Integration on Windows Platform. Obtido em 18 de 06 de 2013, de Talend:

<http://www.talend.com/about-us/press/talend-optimize-open-source-dataintegration-windows-platform>

AURÉLIO. *Dicionário Aurélio Eletrônico - Século XXI*. [S.l.]: Editora Nova Fronteira, 1999.

BOWEN, J. (2012). *Getting Started with Talend Open Studio for Data Integration*. Birmingham: Packt Publishing.

ELIAS, Diego. **Tipos de métricas existentes no Data Warehouse**. 04 jul. 2014. Disponível em:

<<http://corporate.canaltech.com.br/materia/business-intelligence/Tipos-de-metricas-existente-s-no-Data-Warehouse/>>. Acesso em: 12 maio. 2020.

DALARMI, 2010, p84 *Visão Acadêmica*, Curitiba, v.11, n.1, Jan. - Jun./2010 - ISSN 1518-5192 **GESTÃO DE SUPRIMENTOS NA FARMÁCIA HOSPITALAR PÚBLICA**

GONÇALVES, E. L. *O hospital e a visão administrativa contemporânea*. São Paulo: Pioneira, 1983

GONÇALVES, José Ernesto Lima. **As empresas são grandes coleções de processos**. São Paulo: ERA – Revista de Administração de Empresas, n.1, v. 40, p. 6-19, Jan./Mar. 2000.

INMON, William H.; tradução de Ana Maria Neto Guz. **Como construir o data warehouse**. Rio de Janeiro: Campus, 1997.

IW CARE (org), *Softwares Inteligentes* . Disponível em:

<<http://www.iwsoftware.com.br/iwcare.php>>. Acesso em: 20.abril.2020.

IW CARE, *Softwares Inteligentes* . Disponível em:

<<http://www.iwsoftware.com.br/iwcare.php>>. Acesso em: 20.abril.2020.

KIMBALL, Ralph; MERZ, Richard; tradução Edson Furmankiewicz, Joana Figueiredo. **Data Webhouse: construindo o Data Warehouse para a WEB**. Rio de Janeiro: Campus, 2000.

KIMBALL, Ralph; MERZ, Richard; tradução Edson Furmankiewicz, Joana Figueiredo. **Data Webhouse: construindo o *Data Warehouse* para a WEB**. Rio de Janeiro: Campus, 2000.

KIMBALL, Ralph; ROSS, Margy. **The *data warehouse* toolkit : the complete guide to dimensional modeling. - 2nd Ed.** New York: Wiley Computer Publishing, 2002.
LOH, Stanley. **BI na Era do Big Data para Cientistas de Dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões**. Porto Alegre, 2014.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de *Data Warehouse*: Uma visão multidimensional**. 3ª ed. São Paulo: Érica, 2007.

MEDIUM, PostgreSQL — um pouco deste SGBD muito robusto e Open Source.

Disponível em: <

<https://medium.com/@HexaDC/postgresql-um-pouco-deste-sgbd-muito-robusto-e-open-source-b48d84b5dea7>

>. Acesso em: 12 maio. 2020

MILANI, André; **PostgreSQL: Guia do programador / André Milani**. – São Paulo: Novatec Editora, pg 25, 2008

MIRSHAWKA, V. *Hospital fui bem atendido: a hora e a vez do Brasil*. São Paulo: Makron Books, 1994

MOREIRA, Eduardo. **Modelo Dimensional para *Data Warehouse***. Disponível em <<http://imasters.com.br/artigo/3836/gerencia-de-ti/modelo-dimensional-para-data-warehouse>>. Acesso em: novembro de 2015.

PORTAL, Colunista. **Auditoria de enfermagem na conta hospitalar**. Campo Grande – MS, 23 mar. 2012. Disponível em: <<http://www.portaleducacao.com.br/educacao/artigos/11280/auditoria-de-enfermagem-na-conta-hospitalar>>. Acesso em: 04 out. 2015

PORTAL, Colunista. **Recusas de pagamentos hospitalares**. Campo Grande – MS, 23 mar. 2012. Disponível em: <

<http://www.portaleducacao.com.br/enfermagem/artigos/11299/recusas-de-pagamentos-hospitales>>. Acesso em: 04 out. 2015

POSTGRESQL (org.). **O que é o PostgreSQL?** 2020. Disponível em: <https://www.postgresql.org/about/>. Acesso em: 20 abr. 2020.

QLIK . **Qlik Sense Plataforma de Análise de dados** . João Pessoa – PB – 2020. Disponível em: <https://www.qlik.com/pt-br/products/qlik-sense> >. Acessado em: 21 abril 2020.

SEBRAE. Ferramenta 5W2H. Disponível em: <<https://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/5W2H.pdf>>. Acesso em: 20.abril.2020.

SETZER, Valdemar W. **Dado, Informação, Conhecimento e Competência**. São Paulo: Depto. De Ciência da Computação, Universidade de São Paulo, p. 1-3, 12 set. 2014. Disponível em: <<http://www.joinville.ifsc.edu.br/~heverton.pedri/tcc/dado%20informacao%20conhecimento%20e%20competencia.pdf>>. Acesso em: 22 abr. 2015.

SMARTBISOLUTIONS. **QlikView Análise Qualitativa Avançada**. São Paulo – 2020. Disponível em: <<https://www.smartbisolutions.com.br/qlikview-analise-qualitativa-avancada/> >. Acessado em: 15 abril 2020.

TALEND .Ferramenta Talend Data Integration. Disponível em: <<https://www.talend.com/products/data-integration/>>. Acesso em: 20.abril.2020.

TAN, Pang-Nin; SETEINBACH, Michael; KUMAR, Vipin. **Introdução ao DATA MINING**. Rio de Janeiro: Ciência Moderna LTDA., 2009.

TURBAN, Efrain ...[et al]; tradução Fabiano Bruno Gonçalves. **Business Intelligence: Um enfoque gerencial para a inteligência de negócio**. Porto Alegre: Bookman, 2009.

VISÃO geral da arquitetura do QlikView. **Documento sobre a Tecnologia do QlikView**. São Paulo, 2011. Disponível em: <http://www.nordicasoft.com.br/dv_files/arquivos/20130115170003_dbarquivos.pdf>. Acesso em: 02 mai. 2016.

WFS SOFTWARE. Disponível em: < <http://www.wfssoftware.com.br/sobre.php>>.
Acesso em: 20.abril.2020.