

Leonardo Contezini

**PROJETO DE AUTOMATIZAÇÃO DA AQUISIÇÃO E  
MODELAGEM DE DADOS PARA GERAÇÃO E  
DISTRIBUIÇÃO DE INDICADORES DE EFICIÊNCIA  
APLICADOS NA INDÚSTRIA**

Trabalho de Conclusão de Curso (TCC)  
submetido ao Curso de Graduação em  
Engenharia Elétrica da Universidade  
Federal de Santa Catarina como parte  
dos requisitos para obtenção do grau de  
Bacharel em Engenharia Elétrica

Orientador: Prof. Dr. Mauricio  
Valencia Ferreira da Luz.

Florianópolis  
2019

Ficha de identificação da obra elaborada pelo autor  
através do Programa de Geração Automática da Biblioteca Universitária  
da UFSC.

Contezini, Leonardo  
PROJETO DE AUTOMATIZAÇÃO DA AQUISIÇÃO E  
MODELAGEM DE DADOS PARA GERAÇÃO E DISTRIBUIÇÃO DE  
INDICADORES DE EFICIÊNCIA APLICADOS NA INDÚSTRIA /  
Leonardo Contezini ; orientador, Mauricio Valencia  
Ferreira da Luz, 2019.

71 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Centro  
Tecnológico, Graduação em Engenharia Elétrica,  
Florianópolis, 2019.

Inclui referências.

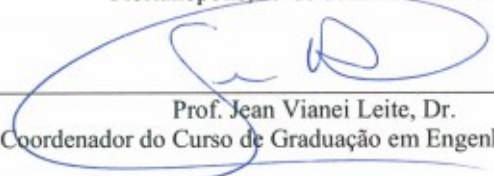
1. Engenharia Elétrica. 2. Business  
Intelligence. 3. KPI. 4. ETL. 5. Distribuição  
Automática de Relatórios. I. Luz, Mauricio Valencia  
Ferreira da . II. Universidade Federal de Santa  
Catarina. Graduação em Engenharia Elétrica. III.  
Título.

Leonardo Contezini

**PROJETO DE AUTOMATIZAÇÃO DA AQUISIÇÃO E  
MODELAGEM DE DADOS PARA GERAÇÃO E  
DISTRIBUIÇÃO DE INDICADORES DE EFICIÊNCIA  
APLICADOS NA INDÚSTRIA**

Este Trabalho foi julgado adequado como parte dos requisitos para  
obtenção do Título de Bacharel em Engenharia Elétrica e aprovado, em  
sua forma final, pela Banca Examinadora

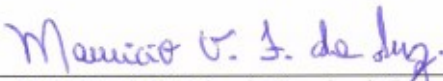
Florianópolis, 09 de dezembro de 2019.



---

Prof. Jean Viane Leite, Dr.  
Coordenador do Curso de Graduação em Engenharia Elétrica

**Banca Examinadora:**



---

Prof. Mauricio Valencia Ferreira da Luz, Dr.  
Orientador  
Universidade Federal de Santa Catarina



---

Prof. Cristhian Becker Cares, M.Sc.  
Universidad de Santiago de Chile / Doutorando PPGEEL-UFSC



---

Eng. Pedro Carvalho Silva Barcelos.  
Mestrando PPGEEL-UFSC



Este trabalho é dedicado a todos que  
cruzaram meu caminho até aqui.





## AGRADECIMENTOS

Agradeço primeiramente à vida e aos momentos que não aconteceram por acaso.

Aos familiares, muito obrigado pelo amor, bons exemplos e todo o suporte necessário. Vocês foram a base para a realização deste sonho.

Natália, sou extremamente feliz por ter você ao meu lado. Obrigado por todo amor, carinho e planos traçados comigo.

Agradeço a todos os amigos que esta jornada me trouxe. Definitivamente tudo acontece por um motivo. Vocês são amizades que quero levar para a vida inteira.

Professor Mauricio, você foi uma referência desde a primeira aula. Obrigado por, além de ensinar o conteúdo com qualidade, preocupar-se em nos mostrar a realidade e abrir nossos olhos para o futuro. É um orgulho ter você como orientador deste trabalho.

Aos clientes, obrigado pelas produtivas horas que passamos juntos e por confiarem no meu trabalho.





## RESUMO

Indústrias que trabalham com uma alta capacidade produtiva precisam se adequar rapidamente às novas tendências do mercado e adotar ferramentas de gestão modernas voltadas ao aumento da eficiência. Neste cenário a tomada de decisão baseada em dados é um ponto chave para a assertividade destas decisões, identificação dos gargalos da indústria, minimização das perdas e aumento da produção. O presente trabalho relata as etapas de implementação de uma ferramenta de Business Intelligence com o objetivo de automatizar tarefas de extração e modelagem de dados, assim como geração e distribuição de indicadores de eficiência. Por fim os tomadores de decisão da empresa podem usar a tecnologia implantada para monitoramentos e melhorias contínuas dos processos da indústria.

**Palavras-chave:** Business Intelligence. KPI. ETL. Distribuição Automática de Relatórios.





## ABSTRACT

Industries with high production capacity need to adapt quickly to changing market trends and adopt modern management tools aimed at increasing efficiency. In this scenario, data-based decision making is a key aspect in order to make the correct decisions, identify the causes of efficiency losses and increase production. This paper describes the implementation steps of a Business Intelligence project aiming to automate tasks of extraction and modeling of data in order to generate and distribute performance indicators. In the end company decision makers can use the project to control and improve industry processes.

**Keywords:** Business Intelligence. KPI. ETL. Automatic Report Distribution.



## LISTA DE FIGURAS

Figura 1 - Exemplo de dashboard desenvolvida no Qlik Sense..	35
Figura 2 - Elementos de um sistema de banco de dados.....	36
Figura 3 - Exemplificação do conector ODBC.....	39
Figura 4 - Exemplo de chaves de ligação entre tabelas de produção, linhas e produtos.....	40
Figura 5 - Hierarquia da empresa em termos de produção. ....	44
Figura 6 - Modelagem Floco de Neve. ....	46
Figura 7 - Exemplo de modelagem estrela. ....	47
Figura 8 - Comparação das modelagens apresentadas.....	48
Figura 9 - Trecho do código de extração. ....	51
Figura 10 - Trecho do código de transformação 1.....	53
Figura 11 - Trecho do código de transformação 2.....	53
Figura 12 - Trecho do script de carga.....	55
Figura 13 - Fluxograma de atualização dos dados.....	56
Figura 14 - Ciclo de tomada de decisões induzido pela metodologia DAR. ....	57
Figura 15 - Exemplo de visualização de dashboard seguindo a metodologia DAR. ....	58
Figura 16 - Exemplo de visualização de analysis seguindo a metodologia DAR. ....	59
Figura 17 - Exemplo de visualização de reporting seguindo a metodologia DAR. ....	60
Figura 18 - Agregação da coluna “Quantidade Produzida” através da função soma.....	61
Figura 19 - Como montar visualizações com itens mestres.....	62









## LISTA DE ABREVIATURAS E SIGLAS

BI – Business Intelligence  
ETL – Extract, Transform and Load  
HD – Horas Desempenho  
HDT – Horas Disponíveis para Trabalho  
HE – Horas Efetivas  
HPNP – Horas Paradas Não Planejadas  
HPP – Horas Paradas Planejadas  
HQ – Horas Qualidade  
HT – Horas Trabalhadas  
KPI – Key Performance Indicator  
ODBC – Open Database Connectivity  
OLE – Overall Labor Effectiveness  
QIP – Quantidade de Itens Produzidos  
QIR – Quantidade de Itens para Reprocesso  
QIS – Quantidade de Itens para Sucata  
QMC – Qlik Management Console  
QVD – QlikView Data  
RAM – Random Access Memory  
SQL – Structured Query Language  
TI – Tecnologia da Informação  
TXPP – Taxa de Produção Planejada  
TXPR – Taxa de Produção Realizada







## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>27</b>
1.1	DESCRIÇÃO GERAL DO PROBLEMA .....	28
1.2	OBJETIVOS .....	29
<b>2</b>	<b>REVISAO BIBLIOGRÁFICA.....</b>	<b>33</b>
2.1	BUSINESS INTELLIGENCE .....	33
2.2	QLIK SENSE.....	33
2.3	BANCOS DE DADOS.....	35
<b>2.3.1</b>	<b>Sistema MES.....</b>	<b>36</b>
2.3.1.1	MES (banco de dados) .....	37
2.3.1.2	MPL.....	37
<b>2.3.2</b>	<b>SQL.....</b>	<b>37</b>
<b>2.3.3</b>	<b>Conexão ODBC .....</b>	<b>38</b>
<b>2.3.4</b>	<b>Chaves de ligação .....</b>	<b>39</b>
2.4	MODELAGEM DE DADOS.....	40
<b>2.4.1</b>	<b>Tabela fato .....</b>	<b>41</b>
<b>2.4.2</b>	<b>Tabela dimensão.....</b>	<b>41</b>
<b>3</b>	<b>MATERIAIS E MÉTODOS.....</b>	<b>42</b>
3.1	ETL .....	42
<b>3.1.1</b>	<b>Especificação da solução.....</b>	<b>43</b>
<b>3.1.2</b>	<b>Mapeamento das fontes de dados.....</b>	<b>44</b>
<b>3.1.3</b>	<b>Análise de modelagem.....</b>	<b>45</b>
3.1.3.1	<i>Flat Table</i> .....	45
3.1.3.2	<i>Snowflake</i> .....	45
3.1.3.3	<i>Star Schema</i> .....	46
3.1.3.4	Comparação entre modelos .....	47
3.1.3.5	Definição do modelo .....	49
<b>3.1.4</b>	<b>Extração .....</b>	<b>50</b>
<b>3.1.5</b>	<b>Transformação .....</b>	<b>51</b>

<b>3.1.6</b>	<b>Carga.....</b>	<b>54</b>
<b>3.1.7</b>	<b>Automatização da ETL.....</b>	<b>55</b>
<b>3.2</b>	<b>VISUALIZAÇÕES E METODOLOGIA DAR.....</b>	<b>56</b>
<b>3.2.1</b>	<b>Dashboard.....</b>	<b>58</b>
<b>3.2.2</b>	<b>Analysis.....</b>	<b>58</b>
<b>3.2.3</b>	<b>Reporting.....</b>	<b>59</b>
<b>3.2.4</b>	<b>Programação dos indicadores.....</b>	<b>60</b>
<b>3.3</b>	<b>ITENS MESTRES.....</b>	<b>62</b>
<b>3.4</b>	<b>VALIDAÇÃO.....</b>	<b>63</b>
<b>3.5</b>	<b>LANÇAMENTO.....</b>	<b>63</b>
<b>4</b>	<b>RESULTADOS.....</b>	<b>65</b>
<b>4.1</b>	<b>TEMPO DE CONFECÇÃO DO RELATÓRIO.....</b>	<b>65</b>
<b>4.2</b>	<b>GESTÃO DO CONHECIMENTO.....</b>	<b>65</b>
<b>4.3</b>	<b>ACURACIDADE DOS APONTAMENTOS.....</b>	<b>66</b>
<b>4.4</b>	<b>IMPACTO NA PRODUÇÃO.....</b>	<b>66</b>
<b>5</b>	<b>CONCLUSÃO.....</b>	<b>67</b>
<b>5.1</b>	<b>SUGESTÕES PARA TRABALHOS FUTUROS.....</b>	<b>67</b>
<b>6</b>	<b>REFERÊNCIAS.....</b>	<b>69</b>



# 1 INTRODUÇÃO

A informação é base fundamental para a evolução da sociedade. A percepção desta necessidade ao longo dos anos fez com que o meio corporativo, incluindo grandes indústrias, percebesse a importância do armazenamento de seus dados de modo a posteriormente extrair valor e perpetuar a evolução dos negócios de maneira acelerada [1].

Segundo Falconi [2], “sem medição não existe gestão”. É preciso ter informações armazenadas a fim de compor métricas e indicadores concretos para avaliar desempenhos e propor melhorias contínuas. Estas melhorias resultam em uma das formas mais eficientes de aumentar a competitividade de uma empresa [3].

No ambiente industrial informações relacionadas à produção apresentam um volume elevado de dados que, armazenados e bem geridos, podem tornar este processo cada vez mais inteligente. A tendência para o futuro é que todos os setores – não apenas da indústria – passem por essa transformação tecnológica [4].

Este grande volume de informações disponível possui uma riqueza escondida. No entanto grande parte delas pode ser irrelevante e ter um efeito contrário ao esperado. Por isso é imprescindível que os dados sejam utilizados de maneira otimizada. Segundo Clay Johnson [5], é essencial que se faça uma dieta das informações disponíveis de modo a trabalhar apenas com as que agregam inteligência às análises. Desta forma tem-se um ambiente enxuto e a tarefa de transformar dados brutos em informação de valor torna-se mais fácil.

Este trabalho descreve um projeto realizado em uma empresa do ramo industrial com a necessidade de extrair valor das informações existentes em suas linhas de produção. Esta indústria tem passado por etapa de modernização de suas fábricas, adicionando sensores ao longo das linhas para captar os status de cada etapa da produção. A empresa é responsável pela fabricação e comercialização de bens a nível mundial e conta com um vastíssimo volume de dados armazenados.

O autor inicia o texto com uma base teórica necessária para a compreensão dos termos técnicos. Em seguida apresenta aspectos sobre modelagem de dados, métodos para levantar as necessidades do negócio e metodologias utilizadas para desenvolvimento do projeto. Posteriormente aborda tópicos referentes ao desenvolvimento prático dos relatórios e indicadores. Por fim, realiza a análise dos resultados e conclusão sobre o trabalho.

## 1.1 DESCRIÇÃO GERAL DO PROBLEMA

O trabalho descreve a solução encontrada para resolver um problema complexo de automatização das informações de uma indústria multinacional com o uso inteligente de seus dados. A empresa em questão comercializa seus produtos em mais de 80 países, possui capacidade produtiva de 40 milhões de unidades ao ano através de plantas em 7 países. O faturamento atual ultrapassa 4 bilhões de reais. A oportunidade de lidar com esta empresa surgiu durante o período de consultoria em *Business Intelligence* na BIX Tecnologia.

A complexidade do problema se caracteriza pela dificuldade da empresa em automatizar a obtenção das informações e suas respectivas transformações referentes às linhas de produção, provida de diferentes fontes, de forma a oferecer dados unificados e confiáveis sobre a eficiência das linhas.

A solução busca aumentar o taxa de produção da empresa através da diminuição do tempo desperdiçado nas linhas de produção. Este problema será resolvido através da automatização dos processos de obtenção e organização dos dados para calcular os indicadores. O primeiro passo é entender o que causa esta ineficiência.

O desperdício de tempo nas linhas provém de diferentes fatores como paradas não planejadas, produtos com falha, ou simplesmente performance reduzida. Produtos com falha, por exemplo, podem virar sucata ou retornar ao início da linha, tomando o lugar de um item sem falhas que poderia ser comercializado. Identificando os principais gargalos e causas destes problemas, os gestores podem tomar decisões embasadas para solucioná-los ou evitá-los.

Relatórios complexos usualmente demandam tempo e esforço dos funcionários da empresa. Isto que gera custos, diminui a eficiência das equipes de trabalho e pode apresentar erros humanos na inserção de dados. Com este projeto de automatização visa-se transformar o tempo de confecção dos relatórios em tempo de gestão efetiva dos processos.

Além da parte técnica a solução conta com a gestão das pessoas envolvidas no projeto para que a entrega atinja seu objetivo. É preciso estar em contato direto com os usuários finais para compreender as regras utilizadas nos cálculos dos indicadores, a origem das informações, o formato das bases de dados e a forma como a eficiência era medida em um período anterior ao projeto.

## 1.2 OBJETIVOS

O objetivo final do projeto é aumentar a taxa de produção da empresa através do gerenciamento eficiente das linhas. Em termos utilizados no ambiente profissional, alavancas são pontos do negócio que podem ser alterados para gerar (ou “alavancar”) algum tipo de resultado. As três alavancas identificadas para atingir o objetivo do projeto são:

1. Disponibilidade: diminuir os tempos de parada.
2. Qualidade: diminuir o número de produtos com falha.
3. Performance: aumentar o desempenho das peças envolvidas no processo.

Para permitir o gerenciamento das linhas e facilitar a tomada de decisões, este projeto providenciará informações relacionadas a estas alavancas. Cada alavanca, portanto, representa um indicador e possui particularidades que precisam ser analisadas e previstas durante a modelagem dos dados.

A disponibilidade - primeira causa da perda de eficiência - reflete o percentual de tempo perdido através das paradas. Estas paradas nas linhas de produção podem ser separadas em duas categorias:

1. Paradas planejadas: paradas previstas. Podem ser para manutenção das máquinas, alimentação, descanso ou ginástica laboral dos operadores, entre outras.
2. Paradas não planejadas: paradas não previstas. Podem ser causadas por quebra nas máquinas ou falta de peças, por exemplo.

O indicador de qualidade representa o segundo agente de ineficiência. Refere-se ao percentual de tempo utilizado para produzir itens com falha. Produtos com falha impactam a eficiência pois o tempo consumido para produzi-los poderia ter sido utilizado para produzir um item sem falhas. Produtos com falha são separados em duas categorias:

1. Reprocesso: falhas pequenas, mas que fazem com que o item precise ser desmontado para retornar ao início da linha de produção.
2. Sucata: devido a falhas mais graves, alguns itens precisam ser descartados e viram sucata.

Ainda assim, mesmo se não existissem paradas ou produtos com falha, as linhas não operam em 100% de sua velocidade. Isto acontece por fatores básicos de desempenho. Um dos motivos pode ser para adicionar margem de segurança aos funcionários da produção, por exemplo. Esta é a terceira causa de ineficiência das linhas: performance.

Após descontar as parcelas de disponibilidade, qualidade e performance do tempo de trabalho disponível, tem-se o tempo utilizado efetivamente para produção de itens prontos para o comércio. Este percentual de tempo é classificado em um quarto indicador, conhecido por OLE (*Overall Labor Effectiveness*).

Os indicadores mencionados representam parcelas de tempo utilizadas na produção. A soma destes indicadores, portanto, deve representar o total de tempo de trabalho disponível. Cada indicador (ou KPI – *Key Performance Indicator*) é apresentado em formato de percentual de tempo e a relação entre eles é dada pela equação (1):

$$[OLE] + [Disp.] + [Qual.] + [Perf.] = 100 \% \quad (1)$$

Onde:

OLE = Percentual de tempo eficaz de trabalho.

Disp. = Percentual de tempo perdido por disponibilidade.

Qual. = Percentual de tempo perdido por qualidade.

Perf. = Percentual de tempo perdido por performance.

O cálculo de cada um destes indicadores é dado pelas equações a seguir:

$$OLE = HE / HT \quad (2)$$

$$Disp. = HPNP / HT \quad (3)$$

$$Qual. = HQ / HT \quad (4)$$

$$Perf. = HD / HT \quad (5)$$

Onde:

HE = Horas Efetivas. São horas de produção eficaz de itens que serão comercializados.

HT = Horas Trabalhadas. É o total de horas disponíveis para trabalho, menos as horas ocupadas por paradas planejadas.

HPNP = Horas utilizadas por Paradas Não Planejadas.

HQ = Horas Qualidade, isto é, utilizadas pela produção de itens de reprocesso ou sucata.

HD = Horas Desempenho. Horas de produção perdidas por questões de performance das linhas.

Para viabilizar a entrega final do projeto os principais objetivos técnicos definidos foram:

1. Mapear fontes de dados disponíveis.
2. Extrair os dados das diferentes fontes.
3. Modelar os dados e integrar todas as fontes.
4. Desenvolver as telas de visualização dos indicadores.
5. Propor melhor maneira de disponibilizar informações para os usuários de forma a criar uma solução *self service*.
6. Montar um processo 100% automatizado de extração, processamento e atualização dos indicadores.
7. Implantar o projeto em até dois meses.



## 2 REVISAO BIBLIOGRÁFICA

Este capítulo apresenta a descrição teórica dos conceitos e métodos utilizados durante a realização deste trabalho. Serão apresentados temas como *Business Intelligence*, noções sobre bancos e transações de dados entre sistemas, além de questões essenciais sobre modelagem de dados. Os tópicos de cada tema são apresentados de forma a auxiliar no entendimento das etapas de desenvolvimento do projeto que serão apresentadas no próximo capítulo.

### 2.1 BUSINESS INTELLIGENCE

O conceito de *Business Intelligence* (BI) surgiu na década de 90 para definir habilidades corporativas de coleta, organização, análise e compartilhamento de dados para futura exploração e tomada de decisões [6]. A tomada de decisão baseada em dados trata-se de uma ação originada principalmente de percepções e cognições providas pelos dados intencionalmente organizados.

Por Business Intelligence entende-se um conjunto de técnicas e ferramentas através das quais é possível coletar dados brutos, organizá-los e extrair informações úteis para, a partir de metodologias específicas, mostrar aos tomadores de decisão a situação atual de seu negócio ou organização e induzi-lo à ação.

Soluções de business intelligence estão presentes praticamente em todos os setores atualmente. Podem servir para tomar decisões embasadas e assertivas, identificar gargalos e causas raízes, monitorar informações da concorrência, alimentar pesquisas de marketing e campanhas, ou auxiliar qualquer outro pilar empresarial que necessite de bases técnicas de informações ou relatórios automatizados.

### 2.2 QLIK SENSE

Qlik Sense é um software destinado ao desenvolvimento de soluções de Business Intelligence. Atualmente é uma ferramenta líder de mercado, sendo uma pioneira no ambiente de BI e desenvolvida pela empresa sueca QlikTech. Desde sua primeira versão, o QlikView, a chave do negócio foi integrar diferentes fontes de dados para a geração automática de relatórios. Segundo o relatório *Magic Quadrant* do Gartner [7], a Qlik tem sido ferramenta líder no desenvolvimento de BI pelos últimos 9 anos.

O Qlik Sense é o produto mais atualizado da desenvolvedora e apresenta uma gama completa de soluções, como: extração e integração de dados, transformação e modelagem das informações, distribuição automatizada de relatórios, controle do ambiente de acesso e interface dinâmica com usuários. Por estes motivos a empresa onde o projeto foi desenvolvido já havia adquirido o licenciamento do produto, e o projeto seria desenvolvido através dele.

O Qlik Sense possui uma linguagem de programação própria denominada Qlik. Através da linguagem Qlik é possível programar tanto rotinas de extração e modelagem de dados, quanto de indicadores. As rotinas de extração e modelagem de dados são denominadas de códigos de ETL, termo que será explicado detalhadamente nas próximas seções deste documento. A programação dos indicadores é chamada de código de expressões, ou código de medidas.

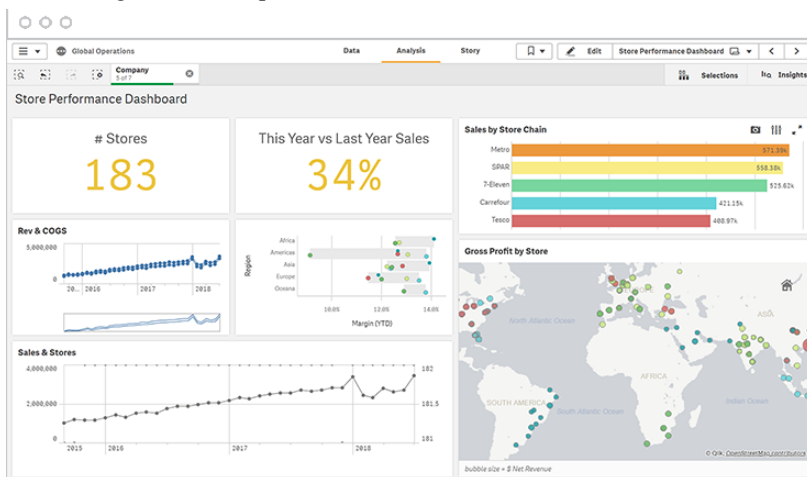
A visualização final dos indicadores no Qlik Sense é organizada em pastas dentro do que é conhecido como aplicação. A aplicação é um arquivo no formato QVF, e contém todos os dados utilizados na visualização. Estas pastas são simplesmente páginas que contém objetos gráficos e KPIs (*Key Performance Indicators*). Por fim, estes objetos contém as medidas e dimensões inicialmente programadas durante o processo de ETL e código de expressões. Assim o usuário final da ferramenta tem acesso às informações - inicialmente extraídas em sua forma bruta – de forma resumida e organizada, através de objetos gráficos pensados especificamente para cada necessidade.

As necessidades variam para cada empresa ou setor, porém é boa prática a primeira pasta apresentar uma visão geral do negócio, funcionando assim como um termômetro da situação atual da empresa. As pastas subsequentes podem apresentar, através de métricas previamente definidas, a evolução do negócio ao longo do tempo, a comparação entre produtos, o ranking de principais clientes, entre inúmeras outras possibilidades.

A figura 1 demonstra um exemplo genérico de dashboard desenvolvido em uma das pastas. Conforme mencionado anteriormente, cada gráfico ou indicador mostrado na imagem contém expressões desenvolvidas durante o script. Todos os gráficos são interativos, o que permite ao usuário interagir com as informações, filtrar as dimensões desejadas, gerar perguntas sobre o negócio e respondê-las através da análise detalhada das informações.



Figura 1 - Exemplo de dashboard desenvolvida no Qlik Sense.



Fonte: Retirado de [8].

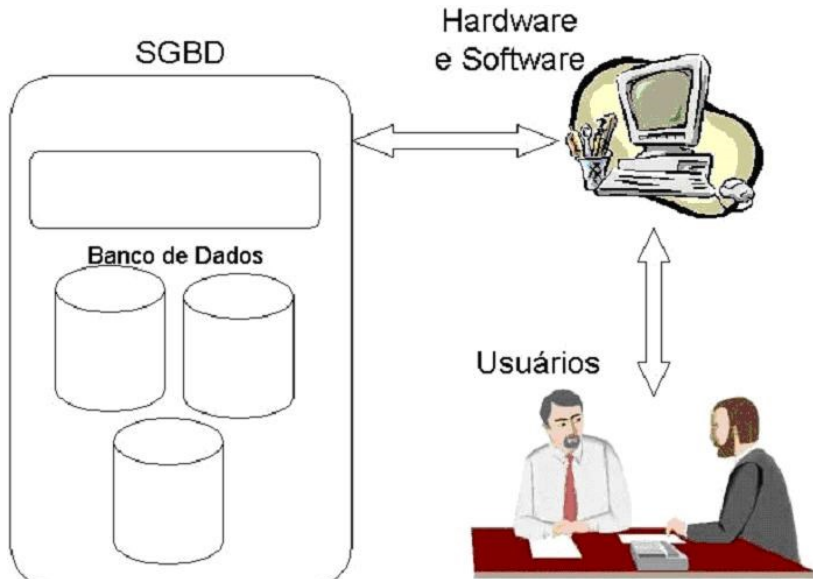
## 2.3 BANCOS DE DADOS

O surgimento dos bancos de dados deu-se pela importância de organizar informações corporativas em locais seguros e organizados, de forma que fosse posteriormente possível acessar estas informações de maneira otimizada. O conceito de banco de dados pode ser resumido de forma mais simplificada através de um conjunto de informações armazenadas digitalmente.

Estas informações devem ser armazenadas em forma de tabela, onde são criadas colunas, e cada coluna representa uma dimensão ou medida do negócio. Nesta estrutura, portanto, cada grupo de informações estará contido em uma tabela, e as tabelas são interligadas entre si através de colunas específicas. Estas colunas são denominadas chaves, e serão explicadas em detalhes posteriormente.

Todo banco de dados, para ser acessado pelo usuário, precisa de um Sistema de Gerenciamento de Banco de Dados [9]. O SGDB, como é chamado, é um software com recursos capazes de manipular as informações armazenadas no banco para disponibilizá-las ao usuário. A figura 2 mostra interação entre banco de dados, SGDB e usuário.

Figura 2 - Elementos de um sistema de banco de dados.



Fonte: Retirado de [9].

### 2.3.1 Sistema MES

MES (*Manufacturing Execution Systems*) é a sigla que contempla sistemas computadorizados utilizados na manufatura para rastrear e documentar cada etapa do processo de transformação de matérias primas em produtos. O sistema MES funciona em tempo real armazenando informações referentes às diferentes etapas que ocorrem nas linhas de produção [10].

A motivação de uma empresa para adquirir o sistema é obter a capacidade de captar informações para que os tomadores de decisão possam analisar a situação atual da manufatura, identificar possíveis áreas de otimização e aumentar a produção final. As informações captadas pelo MES são salvas em um banco de dados específico para posterior utilização em ferramentas de BI.

Neste projeto existem duas bases de dados separadas que seguem o sistema MES. Elas estão explicadas nas próximas seções. Detalhes sobre estas bases serão ocultados por questões de confidencialidade de projeto.

### 2.3.1.1 MES (banco de dados)

As informações que existem em qualquer banco de dados tiveram que ser ao menos uma vez gravadas nestes bancos por programas, algoritmos ou usuários. No caso das informações referentes a cada etapa da linha de produção, estes processos ainda não foram automatizados em todas as plantas e linhas da empresa onde o projeto foi desenvolvido.

Em algumas linhas o apontamento das quantidades produzidas, número e duração das paradas, ou quantidade de itens com falhas ainda é um processo manual. Portanto cada acontecimento é apontado manualmente e registrado em um banco de dados. Este banco foi apelidado de MES pela equipe responsável na empresa.

### 2.3.1.2 MPI

Nas plantas e linhas onde o processo de apontamento foi automatizado, a empresa faz uso do software MPI – Manufacturing Performance Index, ou Monitoramento de Produção Industrial [11]. Alguns dos principais objetivos do software são:

- Coletar dados do chão de fábrica.
- Fornecer em tempo real os status dos setores produtivos.
- Consultar apontamento de produção de fábrica por máquina, funcionário e período.
- Prover relatórios de horas trabalhadas dos funcionários por máquina e período.

O software MPI pode ser classificado também como um sistema MES, uma vez que os objetivos listados atendem aos requisitos definidos pela MESA (*Manufacturing System Association International – MESA, 1997a*) para tais sistemas [11].

### 2.3.2 SQL

A linguagem de programação SQL (*Structured Query Language*) representa a base para utilização de bancos de dados relacionais [12]. A linguagem é implementada através de comandos pré-definidos, divididos em:

- DCL (*Data Control Language*): linguagem de controle de dados.
- DDL (*Data Definition Language*): linguagem de definição de dados.
- DML (*Data Manipulation Language*): linguagem de manipulação de dados.
- DQL (*Data Query Language*): linguagem de recuperação de dados.

Os principais comandos desta linguagem (*insert*, *delete*, *update* e *select*) permitem realizar tarefas como:

- Criação de estruturas básicas de armazenamento (novos campos e tabelas).
- Alteração das estruturas existentes.
- Extração de informações.

A linguagem SQL é complexa e composta por diversos outros comandos. No entanto, para o projeto em questão, ela é utilizada em ambos os bancos exclusivamente para a extração dos dados. Toda a transformação dos dados, incluindo cálculos e agrupamentos, é feita através da linguagem Qlik após a etapa de extração.

### 2.3.3 Conexão ODBC

As conexões *Open Database Connectivity*, ou simplesmente ODBC, representam uma das maneiras mais utilizadas para fazer a integração de ferramentas de BI diretamente a bancos de dados. No desenvolvimento deste projeto, ambos os bancos de dados MES e MPI foram integrados ao Qlik Sense através deste tipo de conexão.

Pode-se entender por conexão uma ponte entre o banco de dados e a aplicação final. Esta ponte pode ser descrita através de duas etapas:

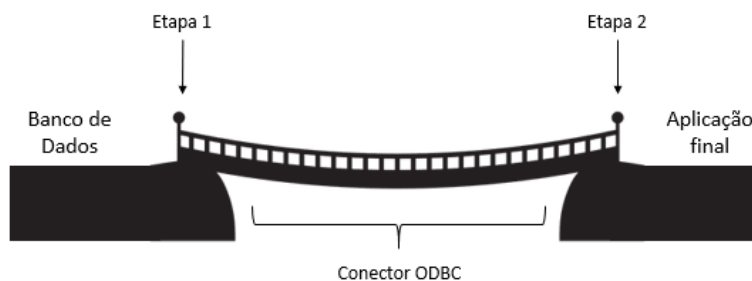
- Etapa 1: conexão entre a ponte (conector ODBC) e o banco de dados.
- Etapa 2: conexão entre a ponte (conector ODBC) e a aplicação final (Qlik Sense).

A primeira etapa é responsável por integrar o conector com as tabelas existentes no banco de dados. Assim tem-se acesso às informações

presentes nestas tabelas. Para utilizar estas informações na aplicação final, deve-se seguir a etapa 2, responsável por transferir estes dados à plataforma de BI.

Após a extração toda a informação fica armazenada no arquivo da aplicação. Desta forma todos os cálculos e transformações são feitos no próprio Qlik Sense através da linguagem Qlik, evitando sobrecarga dos bancos de dados. A figura 3 exemplifica a função do conector ODBC.

Figura 3 - Exemplificação do conector ODBC.



Fonte: Elaborado pelo autor (2019).

### 2.3.4 Chaves de ligação

Conforme já explicado anteriormente, cada conjunto de informações fica organizado em tabelas separadas em um banco de dados. De acordo com boas práticas de desenvolvimento de banco de dados, cada tabela refere-se a determinado assunto. Como consequência, diferentes assuntos não se misturam na mesma tabela.

Exemplo:

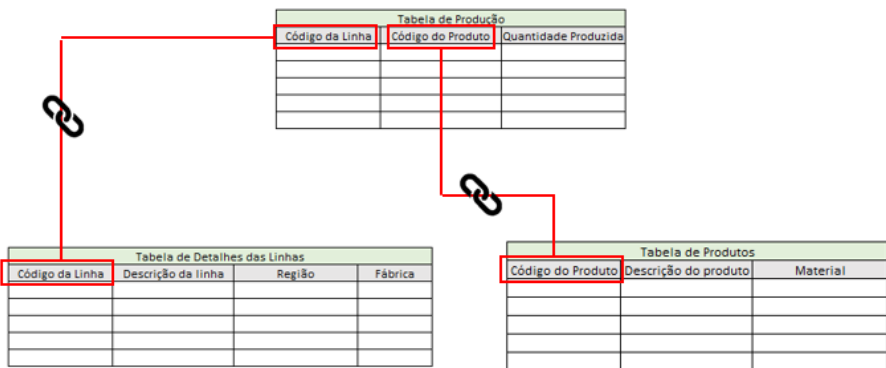
O banco de dados do cliente deste projeto apresenta informações referentes a produção, detalhes das linhas, detalhes dos produtos, entre outras. Cada conjunto de informação encontra-se em uma tabela diferente.

Na tabela de produção pode-se ter código da linha, código do item produzido e quantidade de itens produzidos. Na tabela de detalhes das

linhas pode-se ter campos como: código da linha, descrição, região e fábrica. Já a tabela de produtos pode ter código de produtos, descrição e materiais utilizados. É necessário unir estas informações para se ter um relatório detalhado das operações. Para tal, as tabelas precisam ser conectadas de alguma forma.

Chaves de ligação são utilizadas com objetivo de correlacionar estas tabelas através de campos em comum. Estes campos precisam ter o mesmo conteúdo para que a conexão entre as tabelas seja estabelecida. A figura 4 mostra uma das maneiras de ligar as tabelas do exemplo através de chaves de ligação.

Figura 4 - Exemplo de chaves de ligação entre tabelas de produção, linhas e produtos.



Fonte: Elaborado pelo autor (2019).

## 2.4 MODELAGEM DE DADOS

A modelagem de dados é uma etapa de extrema importância em um projeto de BI. Durante o período de estágio na BIX Tecnologia, utilizava-se a seguinte frase de Abraham Lincoln para descrever a importância de uma boa modelagem: “se eu tivesse oito horas para cortar uma árvore, passaria seis afiando meu machado”.

A escolha errada na modelagem de dados pode resultar em aumentos consideráveis na complexidade dos cálculos, resultando em maior esforço computacional e humano, muitas horas a mais de desenvolvimento, maiores custos de desenvolvimento, e possivelmente o fracasso do projeto e insatisfação do cliente.

No momento da definição da modelagem é preciso compreender os cálculos e regras de negócio, a maneira como diferentes informações se correlacionam, qual o intervalo de tempo analisado, entre outros fatores. Os subcapítulos a seguir servem como uma base teórica sobre aspectos básicos da modelagem de dados, como tabelas fato e dimensões. Os métodos utilizados para compreender as necessidades de negócio serão explicados no capítulo 3.

#### **2.4.1 Tabela fato**

A tabela fato é a tabela principal do modelo de negócios. Nela armazenam-se os dados centrais do negócio, aquilo que realmente ocorreu – o fato, e por isso a denominação da tabela. É com base nos dados presentes nesta tabela que todos os demais indicadores serão gerados. Existem dois tipos de coluna nas tabelas fato:

- Métricas: campos contendo dados utilizados para medição, quantificação de informações. Logo, são números originados de transações do negócio, seja de vendas ou produção, por exemplo.
- Chaves: campos geralmente compostos por códigos que servirão de chave para outras tabelas (tabelas de dimensão).

Por estas características a fato tende a ser a tabela com maior número de linhas no modelo de negócios. À vista disso tende-se a manter somente as colunas necessárias para as análises.

#### **2.4.2 Tabela dimensão**

Tabelas de dimensão tem o objetivo de fornecer detalhes descritivos sobre as cada registro armazenado nas tabelas fato. Ligam-se a tabela fato através da chave em comum para completar os indicadores com descrições e detalhes.

No exemplo mostrado na Figura 4, a tabela de produção seria a fato com todas as métricas enquanto as tabelas de produtos e detalhes das linhas seriam as tabelas dimensão, com os detalhes de cada registro de produção.

### 3 MATERIAIS E MÉTODOS

Ao longo deste capítulo serão explicadas as técnicas utilizadas no desenvolvimento do projeto. São elas:

- ETL (extract, transform, load).
- Metodologia DAR (dashboard, analysis and reporting).
- Itens Mestres.
- Validação.
- Lançamento.

#### 3.1 ETL

ETL é a sigla usada para representar os três passos do fluxo de dados durante desenvolvimento de uma aplicação de BI: Extração, Transformação e Carga (*Extract, Transform, Load*). Trata-se do fluxo de dados por completo, desde sua extração diretamente do banco de dados até sua utilização nas aplicações finais. São etapas que acontecem separadamente, porém são dependentes entre si. A transformação depende dos dados extraídos na etapa de extração, assim como a carga depende dos dados transformados na etapa de transformação.

As configurações da automatização da aquisição e transformação dos dados são feitas dentro das etapas da ETL, assim como modelagem de dados e cálculo dos indicadores. Isto posto, o conjunto de eventos contemplados pelo processo de ETL faz com que esta seja uma etapa fundamental para o sucesso do projeto.

No início de um projeto de BI são realizados testes com bases limitadas para as etapas de extração, transformação e cálculo de indicadores. Através dos testes começa-se a mapear fontes de informações, comparar possíveis modelos de dados e planejar a estrutura da ETL.

O objetivo dos testes é diminuir ao máximo as perdas de performance durante as etapas. Por perdas entende-se maior tempo de carga e esforço computacional. Qualquer perda em um ambiente de testes (com bases menores) pode ser maximizada e convertida em perdas gigantescas ao migrar para o ambiente de produção.

O próximo passo é escalar o projeto e expandir a base de testes para a base de dados em produção. O volume de informações do banco de dados em produção é disparadamente maior, e o sucesso no desenvolvimento da ETL deve resultar em uma solução rápida,



performática e compatível com o volume de dados real da solução proposta.

Pode-se dizer que o planejamento correto da ETL é o que vai garantir a migração de um protótipo para o ambiente de produção., e oferecer ao usuário a informação necessária em tempo hábil. As etapas de desenvolvimento da ETL estão descritas nas próximas seções.

### 3.1.1 Especificação da solução

O início de qualquer solução vem do bom entendimento do problema. Na etapa de especificação o consultor de BI deve reunir-se com gestores e futuros usuários da solução para entrevistá-los e levantar as principais necessidades do negócio. Estes usuários não necessariamente conhecem os bancos de dados, mas conhecem os principais problemas que a organização encontra.

Algumas das principais dificuldades ao iniciar novos projetos são enfrentadas logo nesta etapa, como:

- Entender a hierarquia do negócio.
- Entender as principais necessidades.
- Identificar soluções realizáveis no tempo disponível.
- Identificar oportunidades de soluções não previstas pelos usuários.
- Entender como os usuários, de gerentes até analistas, costumam interpretar os dados para encontrar soluções. A mudança muitas vezes precisa ser suave para atrair estes usuários ao projeto.

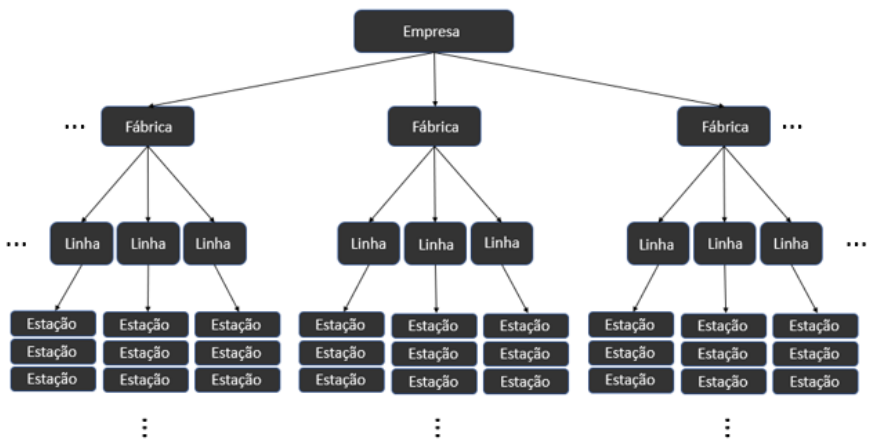
O sucesso desta etapa resulta em um documento de especificação contendo o detalhamento da solução proposta. Neste documento estão descritos os indicadores (KPIs) e métricas que serão utilizados, a frequência de atualização dos dados, o período histórico que será analisado, entre outros. As especificações levantadas pelos usuários para este projeto estão listadas abaixo.

- Indicador de % OLE (*Overall Labor Effectiveness*).
- Indicador de % Disponibilidade.
- Indicador de % Qualidade.
- Indicador de % Performance.
- Visualizações intuitivas.

- Implementação de BI self service (os usuários podem criar suas próprias visualizações).
- Atualização dos dados cinco vezes ao dia.

Os cálculos dos KPIs foram explicados na seção 1.2 deste documento. A Figura 5 apresenta um organograma simplificado da estrutura da empresa, o que compreende o entendimento da hierarquia do negócio.

Figura 5 - Hierarquia da empresa em termos de produção.



Fonte: Elaborado pelo autor (2019).

### 3.1.2 Mapeamento das fontes de dados

Depois de especificar os indicadores desejados pelo cliente, deve-se conseguir as informações necessárias para calculá-los. Para tanto os funcionários responsáveis por levantar as necessidades na primeira reunião acionam profissionais da Tecnologia da Informação (TI). Estes possuem total conhecimento dos bancos de dados e podem fornecer a origem dos dados que vão compor a solução.

Em uma segunda reunião, portanto, deve ser confeccionado um documento de extração contendo as tabelas de onde os dados serão extraídos. Este documento contém as credenciais de acesso ao banco de dados, o nome das tabelas, descrição das colunas e especificação das chaves de ligação.

### 3.1.3 Análise de modelagem

Nesta etapa é preciso definir a estratégia de modelagem de dados que melhor se adequa ao caso do cliente. Existem diversos tipos de modelagem de dados, e cada ambiente ou necessidade precisa ser avaliado individualmente. Para fins comparativos o autor decidiu separar três tipos de modelagem. São abordagens performáticas, bem aceitas e de uso conhecido no mercado.

#### 3.1.3.1 *Flat Table*

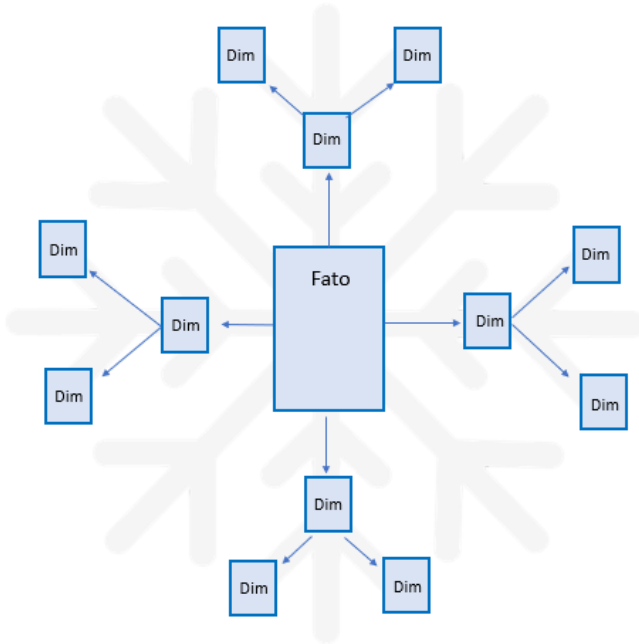
O esquema *flat table* ou tabela simples significa todas as informações relativas ao negócio em uma única tabela. Por consequência esta tabela apresenta tanto dados de métricas quanto dimensões, ou seja, detalhes e informações extras sobre cada métrica.

Neste modelo a pode misturar dados de diferentes assuntos. Por este motivo há um grau elevado de redundância de dados chamado por desnormalização [13]. Relembrando as tabelas apresentadas na seção 2.4, neste modelo as tabelas fatos e dimensões ficam consolidadas em uma única tabela.

#### 3.1.3.2 *Snowflake*

A modelagem *snowflake* (floco de neve) surgiu como uma solução para normalizar os dados através da modelagem. Diferentes assuntos, antes misturados no modelo de tabela simples, são agora decompostos e divididos em tabelas menores. Neste caso, as métricas e informações principais ficam separadas em uma tabela fato que se ligará às dimensões através de chaves de ligação. As dimensões também são decompostas, e ligadas às dimensões anteriores também através de chaves. O formato da estrutura dos dados mostrado na Figura 6 explica a origem do nome desta modelagem.

Figura 6 - Modelagem Floco de Neve.

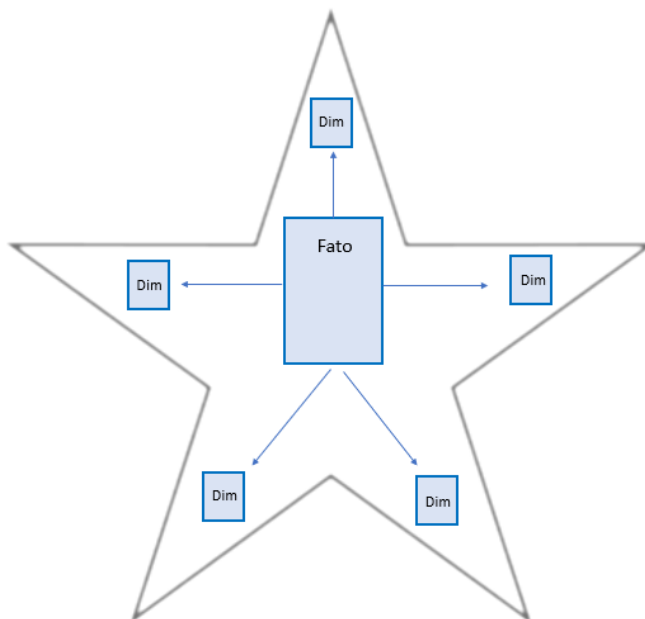


Fonte: Elaborado pelo autor (2019).

### 3.1.3.3 *Star Schema*

Traduzindo para o português, a modelagem estrela é parecida com o modelo floco de neve. A essência do modelo está no fato de manter a tabela fato com as informações principais e apenas uma tabela para cada dimensão. Desta forma cada dimensão liga-se somente à fato. Assim como no modelo floco de neve, tem-se uma tabela fato com dados normalizados. As tabelas de dimensão, no entanto, possuem um grau maior de redundância quando comparadas ao modelo floco de neve pois os são menos decompostos. A Figura 7 apresenta um exemplo de modelagem estrela.

Figura 7 - Exemplo de modelagem estrela.

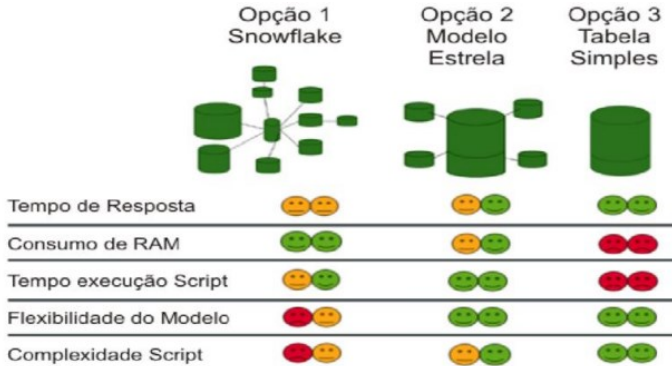


Fonte: Elaborado pelo autor (2019).

#### 3.1.3.4 Comparação entre modelos

Cada um dos modelos apresentados possui características próprias. Eles se diferenciam através da complexidade para modelagem e cálculos no script, performance e esforço computacional, flexibilidade, tempo de carga, entre outros aspectos. A figura 8 mostra uma comparação simples e objetiva dos mesmos.

Figura 8 - Comparação das modelagens apresentadas.



Fonte: Retirado de [14].

Seguindo a ordem apresentada na figura pode-se concluir que o modelo de dados com o melhor tempo de resposta é o modelo de tabela simples. Isto ocorre devido ao fato de que o tempo de processamento é maior onde existem ligações complexas entre as tabelas. Quando todas as informações estão em uma única tabela, os cálculos são efetuados com dados que já estão na memória RAM, dispensando a necessidade de realizar correlações entre tabelas. Por possuir menos ligações complexas que o modelo floco de neve, o modelo estrela assume a segunda colocação neste quesito.

A tabela simples, como um efeito colateral do reduzido tempo de resposta, consome uma quantidade maior de memória RAM devido ao fato de apresentar um grau elevado de redundância em seus dados. O grau de normalização de dados é inversamente proporcional ao consumo de RAM. O melhor modelo neste quesito, portanto, é o floco de neve, seguido pelo modelo estrela.

O terceiro item refere-se ao tempo de execução do script. Para tal, deve-se considerar dois aspectos: volume de informações e quantidade de chaves de ligações. Quanto menor o volume de informações e chaves, melhor (e menor) o tempo de execução do *script*. Conhecendo os três modelos apresentados, pode-se concluir que a modelagem estrela apresenta a melhor performance. Não possui tanta redundância de informações quanto a tabela simples, nem muitas chaves de ligação quanto o modelo floco de neve.

Por flexibilidade do modelo, entende-se facilidade de escalabilidade. Um modelo flexível é aquele que, com o aumento da demanda e do volume de dados, possa ser escalado sem apresentar perdas

na performance. Os modelos estrela e tabela simples apresentam os melhores resultados.

Em relação à complexidade de implementação do script, o nível de dificuldade é diretamente proporcional a quantidade de chaves de ligação. Isto faz com que a complexidade aumente consideravelmente no modelo floco de neve. No entanto o modelo estrela tende a não apresentar um número muito elevado de chaves, o que o deixa em segundo lugar neste aspecto. Ainda, por não possuir chaves de ligação, a tabela simples é superior neste quesito.

### 3.1.3.5 Definição do modelo

Após as etapas de especificação e mapeamento das fontes de dados, detalhadas anteriormente neste capítulo, o projetista tem a capacidade de escolher o melhor modelo de dados. Para isso é preciso relacionar os aspectos apresentados na figura 8 com o ambiente encontrado no cliente. Três aspectos foram decisivos para a decisão do modelo:

1. O servidor onde a aplicação será executada é compartilhado com outras soluções de BI que consomem sua memória RAM. A tabela simples, por apresentar um consumo enorme, pode não atender este requisito.
2. A aplicação será atualizada diversas vezes ao dia para que os gestores possam manter um acompanhamento da produção e das paradas nas linhas os turnos de trabalho. Para que isso seja possível o tempo de execução do script deve ser reduzido. O modelo estrela é a principal opção neste aspecto.
3. A solução será implementada, inicialmente, para algumas das linhas de produção da empresa – aquelas que possuem o sistema MES ou MPI na planta brasileira. Com o passar do tempo todas as plantas terão adquirido o sistema, o que resulta na necessidade de uma modelagem escalável. O modelo floco de neve pode não ser performático neste caso e por isso foi excluído das possibilidades

Por fim, o modelo estrela foi adotado para o projeto. Este apresenta consumo reduzido de memória RAM, é altamente escalável considerando a quantidade reduzida de dimensões, e por este mesmo motivo apresenta o menor tempo de execução do *script* em relação aos demais modelos.

### 3.1.4 Extração

Definido o modelo de dados, é hora de iniciar o primeiro dos três passos da ETL. De acordo a ordem das cargas a primeira etapa é de extração. O objetivo é extrair dados brutos do banco de dados para utilizá-los posteriormente nas etapas de transformação e carga.

Por questões de boas práticas, é criada uma aplicação exclusiva para cada etapa da ETL. É na aplicação de extração que é feita a conexão com o banco de dados através do conector ODBC para ambos os casos: MES e MPI. A linguagem SQL utilizada também é a mesma, e serve para trazer os dados através do conector. Para isto é utilizado o comando *SELECT* na linguagem SQL, seguido de um comando *LOAD* na linguagem Qlik.

Os dados trazidos do banco ficam então contidos no arquivo da própria aplicação de extração, um documento com extensão QVF [15]. Após completar a conexão com o banco e a extração dos dados brutos, estes são armazenados em arquivos com extensão QVD. Nenhuma transformação é feita nesta etapa, portanto cada tabela extraída é armazenada exatamente como vem do banco.

O QVD é um arquivo otimizado de armazenamento de dados desenvolvido pela mesma empresa do Qlik Sense [16]. A otimização reflete no tempo de execução da carga destes arquivos, que é altamente reduzido. Após o comando *STORE* utilizado para armazenar os dados no QVD, os dados carregados na aplicação são retirados da memória do QVF.

Por fim, a aplicação de extração não deve conter nenhum dado. Todas as tabelas extraídas permanecem armazenadas separadamente nos arquivos QVD que serão carregados de maneira otimizada nas próximas etapas da ETL. A partir deste ponto conexões com os bancos de dados não são mais necessárias.

A figura 9 mostra um trecho da extração dos dados. Nela são destacadas, de cima para baixo: conexão com banco, carga dos dados na aplicação, carga dos dados no banco, e armazenamento dos dados no arquivo QVD seguido da retirada dos dados na aplicação.



Figura 9 - Trecho do código de extração.

```
LIB CONNECT TO ;
[DIM_SHIFT]:
LOAD

SELECT

FROM ;
STORE [DIM_SHIFT] INTO [ \DIM_SHIFT.qvd](qvd);
DROP TABLE [DIM_SHIFT];
```

Fonte: Elaborado pelo autor (2019).

### 3.1.5 Transformação

A etapa de transformação costuma representar a maior parcela de tempo utilizada em um projeto de BI devido à sua complexidade. Neste passo os QVDs criados na etapa anterior são carregados para a aplicação de transformação onde são realizados os ajustes necessários.

As tabelas criadas são carregadas na ordem de execução do código através da linguagem Qlik. Entre os ajustes realizados na transformação estão:

- Adaptação das tabelas e campos para o modelo escolhido: o modelo estrela. Para isso são definidas as informações que ficarão na tabela fato e dimensão. É preciso definir chaves de ligação e unir ou separar as informações de cada

tabela. As tabelas de diferentes bancos de dados que tratam do mesmo assunto, por exemplo, são unidas nesta etapa.

- Criação de campos calculados: nem todos os campos presentes nos QVDs extraídos estão no formato como serão utilizados na etapa de transformação. É comum criar campos novos através de cálculos, chamados de campos calculados.

Estes resumem informações presentes em múltiplas tabelas, por exemplo: através de um campo com quantidade vendida e outro com valor unitário, pode-se criar um terceiro campo calculado de valor vendido.

Também podem ser criadas flags, isto é, campos binários úteis para otimizar o apontamento de determinadas características. A performance computacional de comparações no Qlik Sense é melhor quando são utilizados valores binários em relação à valores de texto, além de facilitar o cálculo de indicadores.

- Alteração de nomes dos campos: dados brutos extraídos do banco nem sempre apresentam nomes intuitivos. É comum encontrar títulos de coluna representados por códigos, o que o força o desenvolvedor a consultar o manual do banco para saber do que a coluna se trata. Na etapa de transformação, o nome das colunas pode ser renomeado para facilitar o entendimento dos cálculos realizados.

As figuras 10 e 11 mostram alguns trechos da etapa de transformação, explicados logo em seguida.

Figura 10 - Trecho do código de transformação 1.

```
[FACT_MANUFACTURING]:
NoConcatenate
LOAD
  ID_LINHAIID_TURNO AS CHAVE, 1
FROM
WHERE
CONCATENATE([FACT_MANUFACTURING]) 2
LOAD
```

Fonte: Elaborado pelo autor (2019).

Figura 11 - Trecho do código de transformação 2.

```
[DEFECT_TABLE_QUALITY_TOTAL]:
NoConcatenate
LOAD
  sum(NoOfDefects) as QTD_QUALIDADE 3
  RESUME [DEFECT_TABLE_TMP]
Group by 4
Drop Table [DEFECT_TABLE_TMP];
[RESOURCE]:
LOAD
FROM [.] \DIM_RESOURCE.qvd
(qvd);
Left Join([RESOURCE]) 5
LOAD
FROM [.] \DIM_PRODUCTION_LINE.qvd
(qvd);
```

Fonte: Elaborado pelo autor (2019).

As etapas destacadas na figura representam os seguintes passos:

1. Criação de chave, composta por dois campos distintos;
2. Ajuste de dados na tabela fato. Neste caso, duas tabelas estão sendo concatenadas.
3. Criação de campo calculado. Neste caso, é feita uma soma a partir das dimensões agrupadas no item 4.
4. Função de agregação de dimensões para a realização de operações.
5. Ajuste de dados para o modelo estrela. Neste caso, dados de duas tabelas de dimensão estão sendo inseridos em uma única tabela.

As tabelas transformadas são armazenadas em QVD, assim como na extração. Os dados também são retirados da aplicação QVF e permanecem apenas nos QVDs para serem utilizados na etapa de carga.

### 3.1.6 Carga

Esta é a última etapa da ETL e por este motivo as tabelas carregadas não precisam ser armazenadas em arquivos fora da aplicação. Esta etapa tende a ser a mais performática pois os dados são carregados diretamente dos QVDs criados anteriormente, já modelados, e nenhuma transformação é realizada. Além de carregar as tabelas para aplicação final, pequenos ajustes são realizados nesta etapa.

No *script* de carga os campos são renomeados para ficarem mais intuitivos aos usuários finais. Afinal é na aplicação de carga que os indicadores, gráficos e relatórios serão visualizados pelos usuários. O código contém também a definição de variáveis que podem ser mostradas dinamicamente na aplicação.

Em termos de consumo de memória, após a etapa carga todos os dados ficam armazenados no documento Qlik Sense (arquivo QVF). O documento, quando aberto pelo usuário, fica armazenado na memória RAM onde são realizados cálculos e operações. A figura 12 mostra um trecho do *script* com criação de variáveis e renomeação de campos da tabela fato.

Figura 12 - Trecho do script de carga.

```
Derive Fields From Fields [Data] Using [Calendário];  
SET vLocalExpressaoUltimaAtualizacao = 'Última atualização em '&reloadtime();  
[Fato]:  
LOAD  
  ID_TURNO as [Turno],  
  ID_LINHA as [Linha],  
  ID_SETOR as [Setor],
```

Fonte: Elaborado pelo autor (2019).

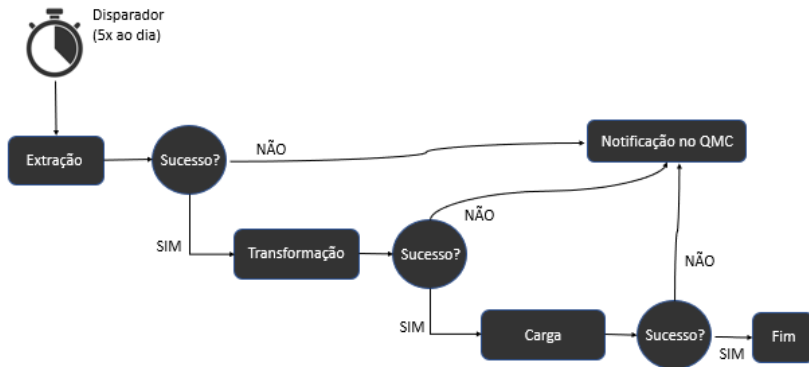
### 3.1.7 Automação da ETL

É de extrema importância que soluções de BI apresentem informações atualizadas aos tomadores de decisão. Para que isto aconteça sem a necessidade de esforço humano, todas as etapas da ETL foram automatizadas. O Qlik Sense possui uma plataforma própria para gerenciamento de cargas e aplicações, conhecida como QMC (*Qlik Management Console*).

O QMC permite agendar as etapas e criar dependências entre elas. Desta forma cada etapa da ETL será iniciada somente se respeitar condições pré-determinadas. No projeto em questão, as etapas são totalmente dependentes entre si. A etapa de transformação só deve ser iniciada após o sucesso da etapa de extração, e a etapa de carga apenas após o sucesso da transformação. Isto faz com que a etapa de extração seja a única com horários agendados para iniciar.

Para este projeto era desejável que a aplicação fosse atualizada cinco vezes ao dia. O processo de extração foi configurado para iniciar 05h30min, 11h30min, 13h30min, 22h30min e 09h30min. O sucesso de cada tarefa de extração resulta no disparo das demais tarefas de transformação e carga. A figura 13 representa este fluxograma de cargas.

Figura 13 - Fluxograma de atualização dos dados.



Fonte: Elaborado pelo autor (2019).

### 3.2 VISUALIZAÇÕES E METODOLOGIA DAR

O contato dos usuários com a solução implementada acontece após o desenvolvimento das pastas de visualização. Através das pastas os usuários podem interagir com indicadores, gráficos e tabelas. A visualização precisa ser intuitiva, de forma que o usuário consiga entender rapidamente os dados e retirar *insights* sem o auxílio de um consultor especializado.

Para isso o desenvolvimento da visualização segue a metodologia DAR (*Dashboard, Analysis and Reporting*). A metodologia DAR surgiu através de uma análise da interação humano-computador a partir da década de 1940, quando computadores eram utilizados apenas para fins militares e científicos [17]. Devido à complexidade dos comandos naquela época, os usuários precisavam de treinamentos específicos para lidar com as máquinas.

Com o lançamento dos desktops a partir do ano de 1980, os computadores passaram a apresentar uma interface mais amigável para que usuários comuns pudessem interagir e realizar atividades ordinárias no dia a dia. Profissionais HCI (*Human-Computer Interaction*) passaram então a estudar como funciona esta interação, e este conhecimento hoje é aplicado inclusive em aplicações de BI através da metodologia DAR.

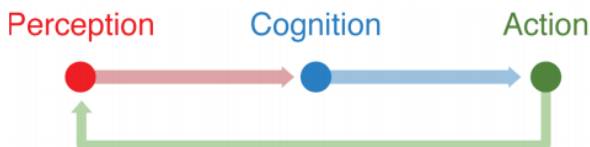
A base teórica da metodologia segue a premissa de que o fluxo de dados em um computador acontece em três estágios para que qualquer ação seja realizada: entrada, processamento e resultado (ação).

Para que o usuário de uma aplicação tome uma decisão baseada em dados, existe uma transferência similar de informações entre humano e aplicação. Esta transferência também ocorre em três estágios, sendo:

1. Percepção (*perception*): o usuário recebe informações através de indicadores, e cria questionamentos sobre as mesmas;
2. Cognição (*cognition*): processamento da informação, através de interação com filtros e gráficos. Nesta etapa o usuário procura entender as causas e criar questionamentos mais aprofundados;
3. Ação (*action*): após análises detalhadas sobre as perguntas geradas nas etapas anteriores, o usuário encontra respostas embasadas e é induzido a tomar uma ação.

Após o resultado o usuário tende a retornar ao estágio de percepção para receber novos *inputs* sobre as decisões tomadas anteriormente. A figura 14 mostra o ciclo causado pela metodologia DAR.

Figura 14 - Ciclo de tomada de decisões induzido pela metodologia DAR.



Fonte: Retirado de [17].

Para induzir o usuário a este ciclo a metodologia DAR sugere o desenvolvimento das visualizações em 3 partes na seguinte ordem:

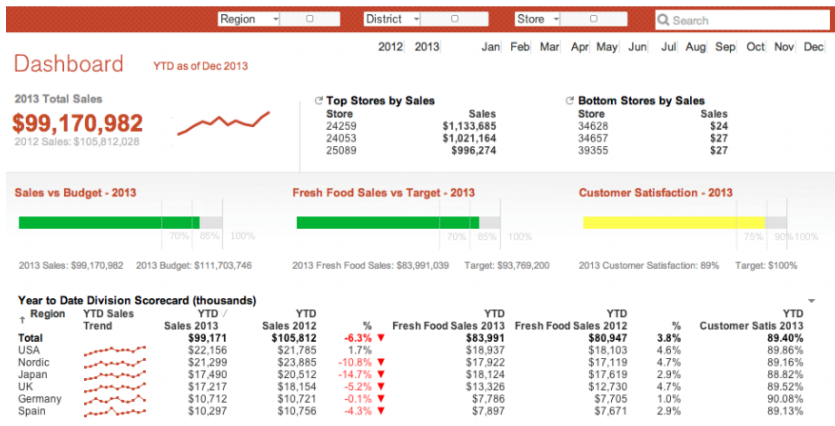
1. Dashboard
2. Analysis
3. Reporting

As próximas seções referem-se aos objetivos e funcionalidades de cada parte.

### 3.2.1 Dashboard

Indicadores abstraídos a um nível superior de visualização, sem muitos detalhes. O objetivo desta tela é servir como um termômetro geral do negócio gerando percepção através de números, ícones e cores indicando onde o usuário precisa prestar mais atenção. Na figura abaixo é possível identificar estas características.

Figura 15 - Exemplo de visualização de dashboard seguindo a metodologia DAR.



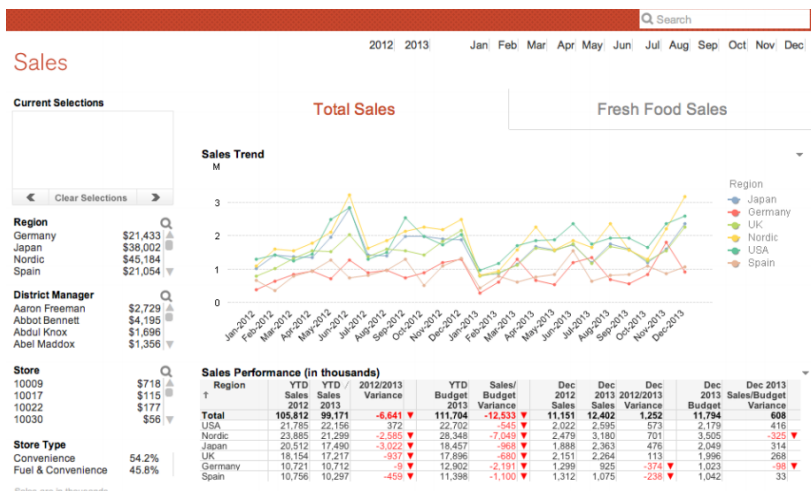
Fonte: Retirado de [17].

### 3.2.2 Analysis

Pastas de análise que induzem à cognição das informações. Para isso incluem gráficos interativos, maior quantidade de filtros e uma granularidade menor das informações. Cada medida apresentada no dashboard contém uma pasta específica de análise para que o indicador seja analisado a partir de diferentes dimensões e pontos de vista. É possível identificar estas características no exemplo da figura 16.



Figura 16 - Exemplo de visualização de analysis seguindo a metodologia DAR.



Fonte: Retirado de [17].

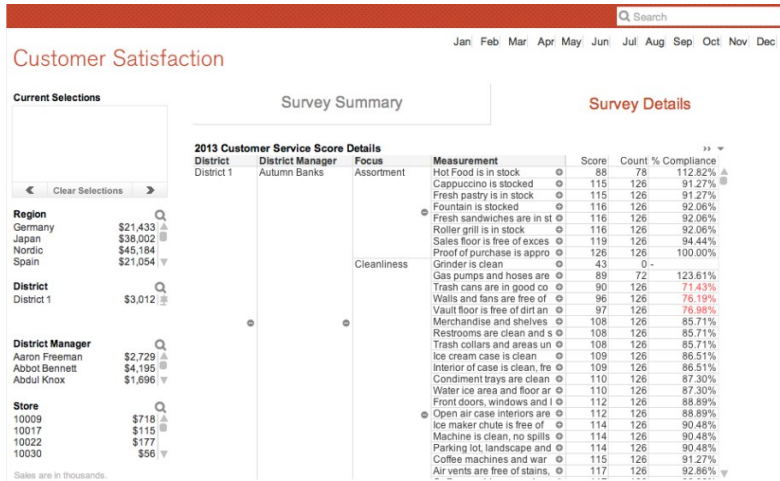
### 3.2.3 Reporting

Após passar pelas pastas de *Dashboard* e *Analysis* o usuário é direcionado ao *Reporting* (relatório, em tradução livre para o português). Através desta visualização é possível entrar no menor nível de granularidade dos dados e identificar as causas raízes dos problemas levantados nas etapas anteriores.

O seguinte exemplo pode ser usado em relação a este projeto: supõe-se que no *Dashboard* certo analista de BI detectou uma baixa eficiência causada pelo indicador de qualidade para determinado mês. Com esta informação, através da pasta de *Analysis*, este usuário pode identificar a fábrica ou linha que causou esta ineficiência. Na etapa de *Reporting*, o analista é capaz de investigar ainda mais e descobrir qual estação dentro daquela linha é responsável por mais falhas e inclusive as causas raízes destas falhas.

A partir do *Reporting* é que as respostas surgem e os usuários sabem exatamente onde precisam agir. Após identificar a falha do exemplo, a próxima ação desta analista será notificar seu gestor para que alguma ação seja tomada.

Figura 17 - Exemplo de visualização de reporting seguindo a metodologia DAR.



Fonte: Retirado de [17].

### 3.2.4 Programação dos indicadores

Este item refere-se a um dos passos realizados para desenvolver as pastas de visualização dos indicadores seguindo a metodologia DAR. Serão apresentados exemplos de códigos e funções utilizadas, exemplos de KPIs e funcionalidades disponíveis no Qlik Sense para assegurar que um BI *self service* fosse entregue ao cliente final.

A programação dos KPIs pode ser feita durante a etapa de carga da ETL ou na própria aplicação, após ter os dados carregados. Como o cliente final levantou a necessidade de ter um BI *self service* foi decidido realizar estes cálculos diretamente na aplicação.

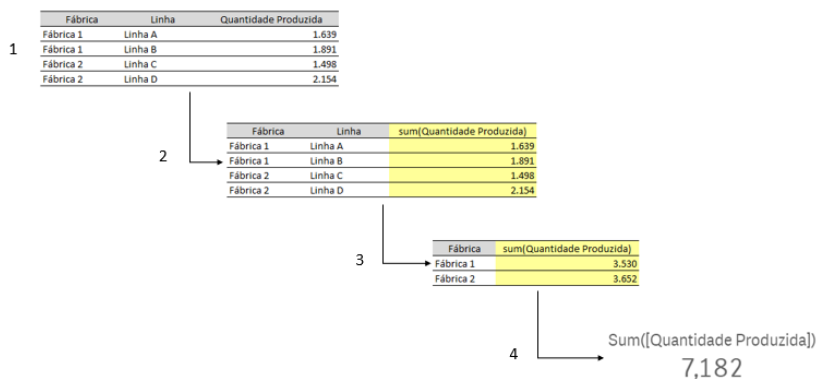
Existem diversas funções disponíveis no Qlik Sense para realizar cálculos e programar indicadores. Para fins de exemplificação e explicação do funcionamento dos indicadores será utilizada a função soma: uma das mais utilizadas no ambiente de BI.

Através da soma são calculados valores acumulados de peças produzidas, vendas, horas trabalhadas, quantidade de funcionários, e uma infinidade de outros indicadores.

A figura 18 mostra exemplos fictícios de aplicações da função sum através de diferentes objetos e dimensões, como:

1. Objeto de tabela sem aplicação da função sum. Neste caso, tem-se a quantidade produzida por Fábrica e Linha, da forma como os dados foram modelados. Ou seja, nenhuma operação é feita neste caso.
2. Aplicação da função sum na coluna “Quantidade Produzida” através das mesmas dimensões: Fábrica e Linha. Neste caso, como a granularidade é a mesma, os valores não se alteram.
3. Mesmo caso do item (2), porém com um nível de abstração maior. Os valores das linhas são acumulados nas linhas de suas respectivas fábricas. A única dimensão utilizada neste caso é Fábrica.
4. Função sum aplicada sem dimensões, diretamente em um objeto de KPI. O resultado é o valor total de “Quantidade Produzida”, em todas as fábricas e linhas.

Figura 18 - Agregação da coluna “Quantidade Produzida” através da função soma.



Fonte: Elaborado pelo autor (2019).

As funções do Qlik Sense funcionam de forma dinâmica para que o usuário possa utilizar filtros e obter resultados de acordo com as granularidades desejadas. No item 4 por exemplo, caso o usuário filtrasse a “Linha D”, o valor do KPI seria alterado para 2.154. Todos os objetos da visualização também serão adaptados para os filtros aplicados.

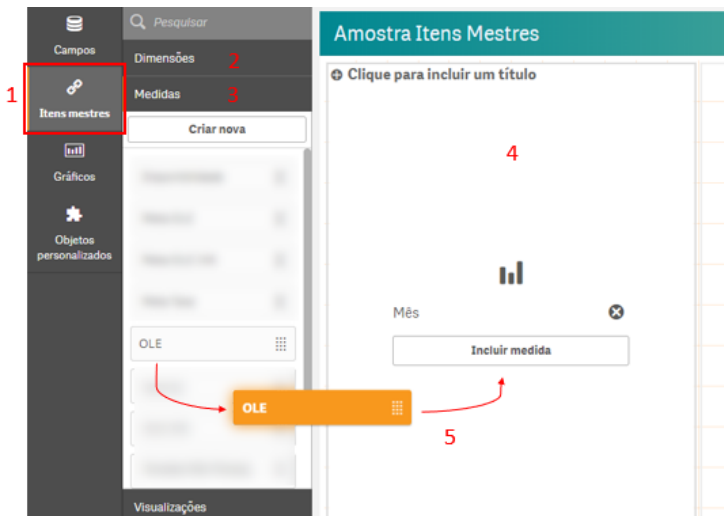
### 3.3 ITENS MESTRES

O Qlik Sense permite ao desenvolvedor entregar aos usuários finais indicadores já programados que podem ser facilmente replicados na aplicação. Através de itens mestres – como são chamadas este conjunto de funcionalidades – é possível encapsular expressões de indicadores, dimensões ou visualizações para o usuário copiar durante a criação de novas telas.

Os itens mestres possuem utilização altamente intuitiva através da funcionalidade *drag-and-drop* (do inglês, arrastar e soltar). Portanto para montar uma visualização específica, não prevista no período do projeto, o usuário apenas escolhe os gráficos que deseja visualizar e arrasta sobre estes os itens mestres de medidas e dimensões aplicáveis.

Para criar estes itens o desenvolvedor precisa implementar as dimensões, medidas ou visualizações necessárias e adicioná-las – também através de *drag-and-drop* – ao espaço dos itens mestres. A figura 19 demonstra a facilidade de montar visualizações através de itens mestres já criados.

Figura 19 - Como montar visualizações com itens mestres.



Fonte: Elaborado pelo autor (2019).

Os destaques da figura 19 fazem referência aos seguintes pontos:

1. Espaço destinado aos itens mestres.
2. Espaço destinado para dimensões adicionadas como itens mestres (dimensões mestre).
3. Espaço destinado para medidas adicionadas como itens mestres (medidas mestre).
4. Gráfico de barras criado pelo usuário. Dimensão mestre “Mês” já foi adicionada.
5. Medida mestre “OLE” sendo adicionada ao gráfico através da funcionalidade *drag-and-drop*.

### 3.4 VALIDAÇÃO

Esta seção refere-se a uma das etapas de finalização do projeto. Após entrevistar os usuários, entender necessidades, implementar a ETL e desenvolver todos os demais passos explicados em capítulos anteriores, é preciso validar a solução encontrada.

A validação consiste em um momento onde determinados usuários finais da aplicação, especialmente aqueles que participaram da etapa de especificação, terão acesso ao BI para usá-lo exaustivamente e determinar se:

- A solução atende às necessidades levantada.
- O projeto segue o que foi definido na especificação.
- Os números apresentados são corretos e confiáveis.

Durante a etapa de validação podem ser feitos pequenos ajustes de cálculos, métricas ou visualizações até que as necessidades iniciais sejam supridas.

### 3.5 LANÇAMENTO

A empresa em questão possui plantas em sete países. Cada planta possui dezenas de linhas, cada linha dezenas de estações e assim por diante. O volume de dados gerados por esta estrutura é gigantesco e exige uma solução escalável, isenta de erros e com governança total sobre as informações disponíveis. Por este motivo o lançamento do projeto apresenta alguns pontos de atenção.

Conforme descrito na seção 2.3, no início do projeto nem todas as linhas possuíam sistema MES implantado. Com o passar do tempo o

sistema seria instalado em todas as fábricas, e a plataforma criada deve ser capaz de suportar este crescimento.

Antes de publicar a aplicação e liberar o acesso a todos os usuários, passou-se por um período de testes de carga para garantir que os dados fossem atualizados corretamente. À medida que novas linhas eram adicionadas ao banco de dados, foram feitos acompanhamentos das etapas da ETL e validações dos dados obtidos. Com escalabilidade e confiabilidade garantidas é possível lançar a solução para que seja acessada pelos usuários.

Através do QMC é possível criar ambientes de acesso restritos a determinadas áreas ou usuários. Com isso pode-se garantir que somente pessoas com credenciais habilitadas possam acessar dados confidenciais mostrados na aplicação. A partir deste ponto todos os processos descritos anteriormente estão automatizados e a aplicação está lançada no ambiente de produção da empresa.

## 4 RESULTADOS

Neste capítulo serão detalhados os resultados obtidos através da implementação do projeto na empresa em questão. O resultado será analisado a partir dos seguintes pontos:

- Tempo de confecção do relatório.
- Gestão do conhecimento.
- Acuracidade dos apontamentos.
- Impacto na produção.

### 4.1 TEMPO DE CONFECÇÃO DO RELATÓRIO

Antes do projeto funcionários da área responsável por medir a eficiência das linhas levavam em média 5 horas por dia para gerar indicadores em uma base diária, para apenas uma fábrica. Em outras palavras, diariamente mais da metade do expediente destes funcionários era ocupado para confecção de relatórios que só demonstravam a situação do dia anterior e para uma parcela muito pequena da empresa.

Implantado em dois meses, o projeto gera informações confiáveis, atualizadas cinco vezes ao dia e com dados relativos a uma janela de análise anual. Cada atualização dura em média 13 minutos e o esforço para adicionar novas fábricas e linhas na análise é nulo, uma vez que o processo é automático. O tempo utilizado antes para gerar informações hoje é utilizado para tomar decisões relativas ao negócio. Usuários classificaram o projeto como um divisor de águas para a empresa.

### 4.2 GESTÃO DO CONHECIMENTO

Embora seja necessário que cada funcionário tenha pleno conhecimento dos resultados de suas ações, é comum encontrar em empresas situações onde somente pessoas do alto escalão possuem acesso a estes dados. Segundo Henry Mintzberg [18], considerado um dos melhores especialistas em estratégia, “gestão significa influenciar a ação. (...) É sobre ajudar as organizações e as unidades a fazerem o que tem que ser feito”.

A plataforma implantada neste projeto tornou-se acessível para diferentes públicos, desde funcionários do chão de fábrica quanto da alta diretoria. A solução possui atualmente uma média de 1200 acessos mensais através de aproximadamente 180 usuários. Com a pulverização

destes indicadores todos os funcionários possuem o conhecimento necessário para aumentar a eficiência das linhas, bem como os efeitos causados após cada melhoria efetuada em seu turno.

#### 4.3 ACURACIDADE DOS APONTAMENTOS

Algumas linhas possuem apontamento manual dos eventos. Ou seja, para cada falha ou parada que ocorra na linha, algum funcionário da fábrica deve apontar o ocorrido. Devido aos longos períodos de criação dos relatórios antes do projeto ser publicado, os gestores não tinham as informações em mãos em tempo hábil queda na qualidade destes dados.

Com a automatização destes processos ficou mais fácil monitorar e cobrar acuracidade dos apontamentos. Consequentemente a qualidade dos dados aumenta, e os gestores passam a ter uma base sólida para elaboração de projetos de melhoria.

#### 4.4 IMPACTO NA PRODUÇÃO

A implantação do projeto possibilitou a identificação dos principais agentes agressores da eficiência nas linhas de produção. Através da metodologia DAR os usuários foram possibilitados de encontrar causas raízes destas perdas e assim realizar ações para reverter a situação.

O impacto exato na quantidade de itens produzidos após a implantação do projeto e o valor monetário do impacto não serão mostrados em detalhes por questões de confidencialidade do projeto. Para fins de estimativa, em seis meses após a publicação do projeto houve um aumento médio de 9% na taxa de produção das linhas da empresa onde o sistema foi aplicado. Ao aplicar este incremento na capacidade produtiva da empresa, tem-se um potencial de aproximadamente 3,6 milhões de unidades a mais produzidas anualmente.



## 5 CONCLUSÃO

Empresas de grande porte com alcance mundial na produção e comercialização de seus produtos precisam estar em constante busca por inovação para manterem-se competitivas no mercado. Pequenas melhorias, implantadas localmente nestas empresas, constantemente tendem a atingir um âmbito global e tornam-se grandes avanços no negócio.

Este projeto teve um papel importante neste quesito, pois através de melhorias de gestão iniciadas em linhas de produção locais, tem se expandido para as demais fábricas e causado um aumento considerável na eficiência da organização. Ademais, além de alavancar a taxa de produção, colaborou na gestão do conhecimento entre os funcionários e no impulsionamento da cultura de dados e de inovação.

Uma das maiores alegrias obtidas ao entregar um projeto desta natureza é perceber a satisfação, felicidade e o alívio dos clientes. Por clientes, neste caso, pode-se entender por pessoas que antes utilizavam quase toda sua jornada de trabalho para gerar relatórios através de atividades repetitivas. Para estes colaboradores o projeto certamente traz, além do aumento na eficiência, uma redução do stress.

As tecnologias utilizadas no trabalho mostraram-se excelentes durante todo o período de projeto. O resultado é uma solução escalável que tem atendido cada vez mais linhas de produção e alcançado centenas de usuários de maneira muito performática.

É importante destacar também a necessidade de saber trabalhar em grupo. Todo o período do projeto foi de constante interação com especialistas do negócio para que as necessidades fossem levantadas e os dados validados. Sem um time de pessoas trabalhando com um objetivo em comum o projeto não teria o mesmo resultado.

### 5.1 SUGESTÕES PARA TRABALHOS FUTUROS

Projetos deste porte apresentam melhorias e ajustes constantes para que as necessidades dos usuários sejam atendidas. Por hora, pode-se listar as seguintes melhorias para os próximos trabalhos relacionados ao projeto:

- 1- Incluir das informações de preço de venda dos produtos.
- 2- Incluir das informações relacionadas aos custos dos produtos.
- 3- Implementar envio automático de relatórios atualizados via e-mail.

Os pontos levantados serão capazes de medir o valor em moeda das paradas nas linhas de produção, além do impacto no lucro final da empresa. Através da quantidade de itens que deixaram de ser produzidos (informação já existente no projeto), e do preço de comercialização destes itens, pode-se calcular o custo de cada parada. Isto poderá melhorar a inteligência de priorização na resolução das paradas, ou seja, as paradas mais custosas à empresa ficarão no topo da lista de espera por soluções.

O terceiro item fará com que os usuários tenham as informações atualizadas com mais frequência. Estes relatórios poderão ser enviados na frequência desejada, seja com dados diários ou dados acumulados mensais ou semestrais, por exemplo. O formato também pode ser variado: estes e-mails podem ter indicadores no próprio corpo da mensagem, ou relatórios anexados em diferentes formatos.

O nível de análise também pode ser personalizado para cada usuário que receber este relatório. Podem ser confeccionados relatórios operacionais, com detalhes das linhas e paradas, ou gerenciais, para gestores terem uma visão geral do negócio. De qualquer forma, as informações estarão disponíveis para consulta e compartilhamento sempre que o usuário desejar, fortificando na empresa a cultura de decisão baseada em dados.

## 6 REFERÊNCIAS

- [1] C. Bazzotti e E. Garcia, “A Importância do Sistema de Informação Gerencial na Gestão Empresarial para Tomada de Decisões,” *Ciências Sociais Aplicadas em Revista*, vol. 6, nº 11, 2006.
- [2] Revista Veja, “As lições de gestão do consultor Vicente Falconi,” *Veja*, 14 Novembro 2017. [Online]. Disponível em: <https://veja.abril.com.br/economia/as-lico-es-de-gestao-do-consultor-vice-falconi/>. [Acesso em 11 Outubro 2019].
- [3] L. Attadia e R. Martin, “Medição de desempenho como base para evolução da melhoria contínua,” *Revista Produção*, pp. 33-41, 2003.
- [4] M. Lemos, Interviewee, *Nenhum setor ficará imune à transformação tecnológica*. [Entrevista]. 21 Outubro 2019.
- [5] C. Johnson, *A Dieta da Informação*, 1 ed., Novatec, 2012.
- [6] J. E. Lemos, E. Turban, D. King e R. Sharda, *Business Intelligence - Um Enfoque Gerencial para a Inteligência do Negócio*, 2 ed., Bookman, 2009.
- [7] Gartner, “Magic Quadrant for Analytics and Business Intelligence Platforms,” 2019.
- [8] “Qlik Sense,” Qlik, [Online]. Disponível em: <https://www.qlik.com/pt-br/products/qlik-sense>. [Acesso em 15 Julho 2019].
- [9] R. Rezende, “DEV MEDIA,” 2006. [Online]. Disponível em: <https://www.devmedia.com.br/conceitos-fundamentais-de-banco-de-dados/1649>. [Acesso em 20 Setembro 2019].
- [10] J. Kletti, *Manufacturing Execution Systems - MES*, Springer, 2007.
- [11] H. Belan, J. Palma e R. Lima, “Proposta de um processo de construção da arquitetura de software para aprendiz,” *SIMPEP*, Novembro 2006.
- [12] E. Duarte, *SQL e Programação de Banco de Dados*, 2006.
- [13] P. Rob e C. Coronel, *Sistemas de banco de dados*, 2011.
- [14] F. Tonial, *Apostila de QlikView Avançado*, Toccato, 2014.
- [15] Qlik, “Introduction to loading data,” QlikTech International, November 2017. [Online]. Disponível em: [https://help.qlik.com/pt-BR/qlikview/November2017/Subsystems/Client/Content/Loading\\_data.htm](https://help.qlik.com/pt-BR/qlikview/November2017/Subsystems/Client/Content/Loading_data.htm). [Acesso em 19 Outubro 2019].

- [16] Qlik, “QVD files,” QlikTech International, November 2017. [Online]. Disponível em: [https://help.qlik.com/en-US/qlikview/November2017/Subsystems/Client/Content/QVD\\_files.htm](https://help.qlik.com/en-US/qlikview/November2017/Subsystems/Client/Content/QVD_files.htm). [Acesso em 26 Outubro 2019].
- [17] Qlik, “Dashboard, Analysis, Reporting (DAR),” Novembro, 2013.
- [18] J. Elaina, “Henry Mintzberg - A estratégia não se planeia, contrói-se,” Portal Gestão, 10 Dezembro 2012. [Online]. Disponível em: <https://www.portal-gestao.com/artigos/6855-henry-mintzberg-a-estrat%C3%A9gia-n%C3%A3o-se-planeia>. [Acesso em 12 Novembro 2019].