

## Arquitetura de Data Warehouse e Data Marts

### Aula 2: Projeto de Data Warehouse: Especificação de Requisitos

#### Apresentação

---

Nesta aula serão apresentados: o ciclo de vida de um projeto de Data Warehouse (DW) definido por Ralph Kimball; a importância do levantamento de requisitos para projetos de DW; os documentos que podem ser elaborados para apoiar os analistas de Business Intelligence (BI), os usuários e todo o ciclo de manutenção e crescimento do ambiente. Nesta aula serão apresentados: o ciclo de vida de um projeto de Data Warehouse (DW) definido por Ralph Kimball; a importância do levantamento de requisitos para projetos de DW; os documentos que podem ser elaborados para apoiar os analistas de Business Intelligence (BI), os usuários e todo o ciclo de manutenção e crescimento do ambiente.

# Objetivos

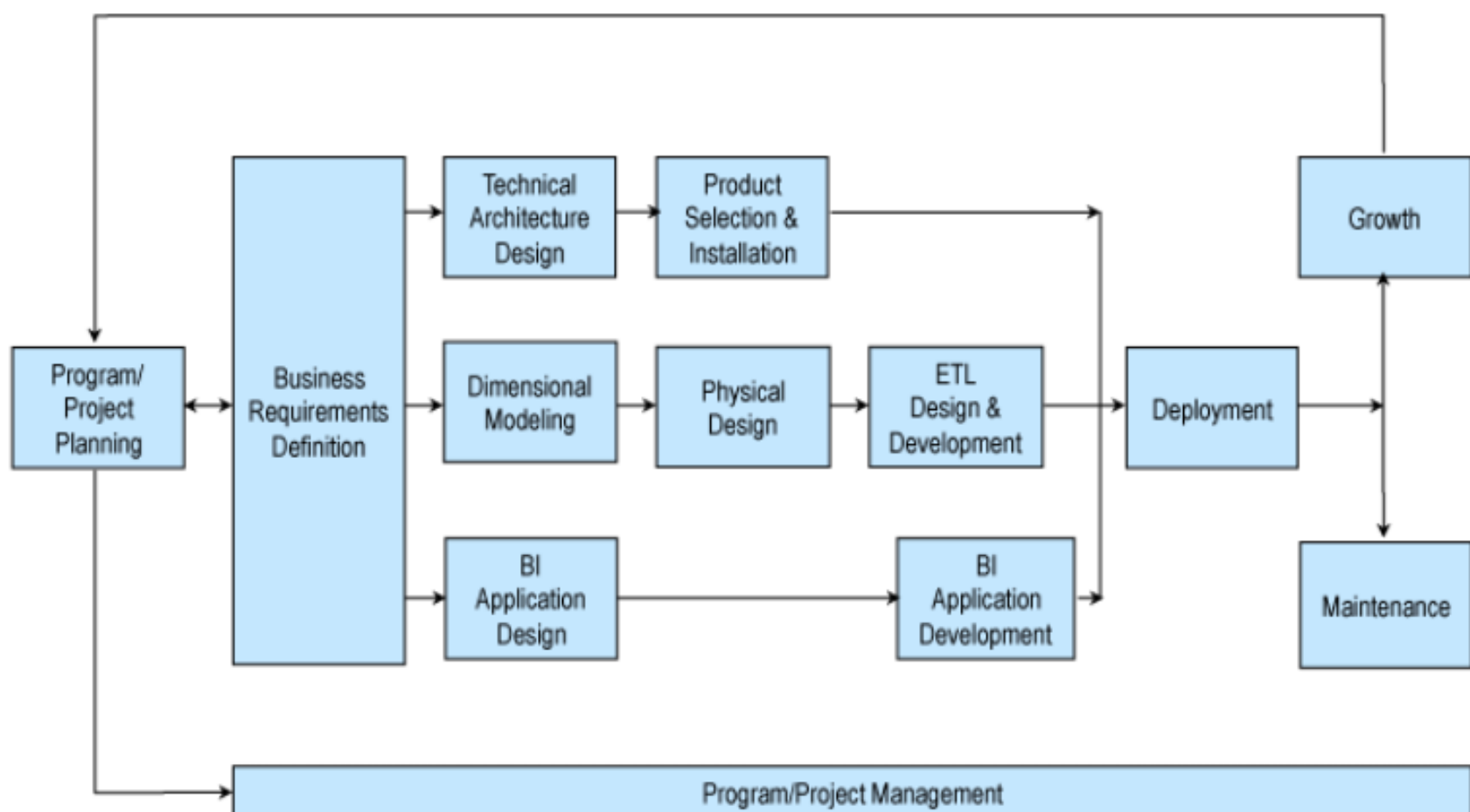
- Descrever o ciclo de vida de um projeto de DW;
- Explicar o processo de levantamento de requisitos para o DW;
- Esclarecer como a documentação para um projeto de DW pode ser construída, os principais pontos a serem documentados nessa fase e o início da construção do Metadados.

## O ciclo de vida do Data Warehouse (DW)

Como vimos na aula anterior, o Data Warehouse é um sistema de informação gerencial focado no apoio à tomada de decisão, que normalmente é realizada pelos gestores da organização. Vimos também que a construção do DW/DM precisa estar alinhada com as necessidades da organização, com a disponibilidade de recursos e dos dados.

A construção do DW deve considerar esses pontos e, além disso, deve ter um plano de desenvolvimento para que os objetivos sejam alcançados.

O desenvolvimento de um projeto é dividido em fases e possui início e fim. Para iniciar qualquer atividade que envolve várias fases, você precisa planejar a execução delas. O **Planejamento** (Program/Project Planning) é a primeira fase do ciclo de vida de um projeto de DW, conforme definição de Kimball, apresentada na figura a seguir. Nessa fase, são definidos o escopo do projeto, a viabilidade de recursos, as tarefas a serem desenvolvidas no projeto e o encadeamento delas.




 Ciclo de Vida de um Projeto de Data Warehouse por Kimball. Fonte: (Kimball 2013).

O gerenciamento do projeto mantém as tarefas a serem desenvolvidas organizadas para que sejam executadas conforme o planejamento. O gerente de projetos é o responsável por acompanhar todo o ciclo de vida do projeto, atuando como um facilitador na resolução de impedimentos durante o desenvolvimento.

A segunda fase do ciclo de vida é a **Definição de Requisitos de Negócios** (Business Requirements Definition) e está diretamente ligada à primeira fase, devido à necessidade do conhecimento dos requisitos, pois o escopo do projeto é definido pelos requisitos do usuário. A relação entre essas duas fases é representada na figura pela seta de mão dupla.

Kimball afirma que, um bom planejamento e a definição bem elaborada dos requisitos aumentam a probabilidade de sucesso de um projeto de DW, pois seu desenvolvimento é baseado nas necessidades dos usuários do negócio e isso apoia a importância dessas duas fases para o desenvolvimento do DW.

Observe que o ciclo de vida é dividido em três trilhas após a definição dos requisitos do negócio.

 Clique nos botões para ver as informações.

### [Trilha 1](#)



A primeira trilha dedica-se às tecnologias que serão utilizadas no desenvolvimento do DW. A fase **Arquitetura do DW** (Technical Architecture) preocupa-se com a definição estrutural e compreende os componentes necessários à implementação de um DW. Esses componentes estão relacionados à arquitetura de dados, à infraestrutura utilizada e às tecnologias necessárias para a construção e a utilização de um DW. Essa fase é seguida de **Seleção e Instalação de Produtos** (Product Seletion & Installation), que define as ferramentas que serão utilizadas na construção, realiza a instalação, faz o teste de integração e sua execução.

### [Trilha 2](#)



A segunda trilha dedica-se ao tratamento dos dados e encadeia as fases Modelagem Dimensional, definição do projeto físico e especificação de ETL. A fase de **Modelagem Dimensional** (Dimensional Modeling) estuda as análises que serão desenvolvidas no ambiente analítico e une o conhecimento dos requisitos definidos para criar uma estrutura capaz de acomodar os dados dimensionalmente. Nessa fase, é definido o Modelo de Dados Dimensional do DW/DM. Na fase seguinte, **Projeto Físico** (Physical Design), é definida a estrutura física para a construção do modelo de dados dimensional, como a definição do padrão de nomenclatura utilizada e a configuração do ambiente do banco de dados. Após a definição da estrutura física da base de dados, é o momento de definir e construir os processos que irão extrair os dados dos sistemas origens, transformar e carregar os dados nas tabelas definitivas do DW. Esta é a fase **Especificação e Desenvolvimento de ETL** (Extract, Transform and Load) (ETL Design & Development). É importante destacar que o tamanho das caixas de cada fase não representa o esforço realizado em cada uma delas. A construção do ETL é uma tarefa muito custosa, que demanda aproximadamente 70% do esforço empregado na trilha de dados.

### [Trilha 3](#)



A terceira trilha do ciclo de vida está concentrada na definição e construção da camada de visualização dos dados. O desenho das consultas desejadas pelos usuários é um artefato muito interessante, que contribui no alinhamento das expectativas dos usuários que acessarão o DW por meio de análises predefinidas. Essa definição é realizada na fase de **Especificação de Aplicação de BI** (BI Application Design).

# Título e texto

---

## Somente título e texto

Seguindo a tarefa de especificação, a fase **Desenvolvimento das Aplicações de BI** (BI Application Development) constrói as consultas na ferramenta de relatórios analíticos definida para o projeto.

Saiba mais

A fase de **Implementação** (Deployment) é a união das tarefas desenvolvidas em cada trilha do ciclo, devendo ocorrer quando todas as fases estiverem concluídas. Novas necessidades surgirão após a implementação do ambiente analítico, o que faz parte do processo de desenvolvimento e crescimento do DW de uma organização. O **Crescimento** (Growth) é representado pela fase que tem início com o planejamento de um novo projeto, mas, nesse caso, será um projeto de complemento. Por fim, a **Manutenção** (Maintenance) é representada no ciclo de vida de um projeto de DW.

Nesta aula vamos aprofundar o conhecimento sobre as tarefas a serem realizadas na Definição de Requisitos de Negócio.

**Atenção!** Aqui existe uma videoaula, acesso pelo conteúdo online

## Levantamento de Requisitos para a construção do Data Warehouse

O planejamento do projeto e o levantamento de requisitos produzem o entendimento sobre a necessidade da organização e o conhecimento do objetivo para a construção do DW, que deve estar bem definido e que justifique essa necessidade. Sem essas definições, o sucesso do projeto está comprometido, pois, se não há um objetivo para tal solução, o ambiente não será utilizado ou sua construção pode não ser finalizada.

**Atenção!** Aqui existe uma videoaula, acesso pelo conteúdo online

### Imagine o seguinte cenário:

Marcos é gerente de vendas em uma grande rede de fast-food. Todos os dias, às 16 horas, ele precisa verificar se é necessário fazer a reposição de algum item utilizado na confecção dos lanches da lanchonete. Se o item estiver com a disponibilidade comprometida, ele precisa enviar a solicitação de reposição ao setor de reabastecimento para que o item seja entregue na manhã seguinte.

Para fazer o controle dos itens, Marcos imprime a lista dos pedidos, calcula a quantidade de lanches servidos em cada pedido e faz o cálculo de kits utilizados para saber se é necessário repor ou não algum item. Esse processo é tão custoso para Marcos que há dias em que ele não consegue terminar a análise em tempo de solicitar os itens para o dia seguinte.

Qual é a solução mais adequada para ajudar Marcos?

Vamos analisar o problema do Marcos?

- Qual é o problema do Marcos?

A dificuldade é saber se é necessário, ou não, solicitar a reposição de itens, até as 17 horas, todos os dias da semana.

- Qual é o objetivo do Marcos?

Saber se há necessidade de solicitar a reposição de algum item diariamente e fazer a solicitação dentro do prazo, de forma mais rápida.

- De que forma o Marcos faz a análise sobre os itens?

Marcos verifica todos os pedidos e calcula a média, manualmente, dos itens utilizados com o objetivo de saber se há algum item que precisa ser repostado.

O que podemos oferecer para resolver o problema do Marcos?

De acordo com a necessidade descrita, um relatório no sistema de vendas irá fornecer a informação sobre os itens que precisam ser repostos.

A investigação sobre o problema do Marcos permitiu verificar a solução mais adequada para resolver seu problema. Imagine se, para a solução do problema do Marcos, fosse projetado um DM que levaria dois meses para ser construído e, além disso, fossem construídas consultas em que o menor nível de análise estivesse em Mês. Por exemplo, Consulta de quantidade de itens por Mês.

Comentário

Essa solução resolveria o problema do Marcos? Não resolveria! Essa consulta pode ser útil para outro tipo de tomada de decisão, inclusive para a melhoria do processo do Marcos, mas não para a necessidade atual dele.

Você já deve ter escutado comentários sobre um projeto que não deu certo e o desenvolvimento foi cancelado, ou que o desenvolvimento foi finalizado, mas os usuários não utilizaram o produto que foi entregue. O entendimento sobre o problema a ser resolvido deve ser a primeira tarefa realizada para o desenvolvimento de um projeto, pois a investigação permite conhecer o cenário, os stakeholders ou partes interessadas, o problema e as possíveis soluções a serem adotadas. Essa primeira etapa é o levantamento de requisitos e aplica-se a qualquer tipo de projeto, inclusive ao projeto de DW.

O levantamento de requisitos para o DW possui características particulares em relação ao levantamento de requisitos para os sistemas de apoio operacional. Essas características estão presentes em Sistemas de Apoio à Decisão (SAD) e vamos conhecê-las a seguir.



1. Entender as necessidades do Negócio (Stakeholders).
2. Elaborar documento com definições e conceitos das perspectivas de análises.
3. Elaborar documento com as medidas que serão analisadas
4. Elaborar documento que descreva as análises desejadas
5. Elaborar documento com apontamento das origens dos dados

Características do Levantamento de Requisitos. Fonte: O autor.

Características do Levantamento de Requisitos

Clique no botão acima.

## Características do Levantamento de Requisitos

### Passo 1 - Entender as necessidades do Negócio

O entendimento da necessidade é realizado pelo Analista de Negócios. Ele é responsável por investigar a necessidade, entender as dores dos usuários e traduzir o entendimento em requisitos para o projeto. Kimball aborda o levantamento de requisitos focado na necessidade do negócio e coloca que os requisitos determinam quais dados devem estar disponíveis no DW, como são organizados e com que frequência são atualizados.

O primeiro passo é entrevistar os usuários e entender quais são as atividades realizadas por eles. Conhecer a atividade realizada pelo usuário auxilia o entendimento do fluxo dos dados que será analisado. Você pode realizar reuniões mais específicas com usuários individuais, pequenos grupos ou grupos que reúnem todos os interessados no desenvolvimento do DW. A estratégia pode ser traçada conforme a necessidade.

O levantamento de requisitos é apoiado por técnicas que auxiliam a condução das entrevistas. O BABOK (*Business Analysis Body of Knowledge*) é um guia que reúne os principais conceitos e técnicas que apoiam a Análise de Negócios.

Durante o levantamento dos requisitos, as informações coletadas devem ser anotadas. O resultado do levantamento conterá a descrição de cenário do negócio com as dores, os objetivos, as análises desejadas, entre outros. Nas análises desejadas podem ser identificadas as possíveis perspectivas de análise e os indicadores. As perspectivas de análise descrevem os fatos que ocorreram em determinado assunto e os indicadores são as medidas que podem ser descritas pelas perspectivas de análise.

### **Periodicidade de carga dos dados no Data Warehouse**

Uma importante informação que deve ser verificada no levantamento de requisitos para o DW é a periodicidade com que os dados serão carregados no ambiente. A periodicidade pode ser diária, semanal ou mensal, ou ainda quase em tempo real. Essa decisão depende da necessidade da organização.

Quando a carga dos dados ocorre diariamente, o processo de ETL acessa a base de dados do sistema transacional todos os dias, obedecendo uma janela temporal para a extração dos dados. Normalmente, a extração ocorre no período em que as transações dos sistemas de origem são desaceleradas, por exemplo à noite. Essa estratégia é usada para que a extração dos dados não concorra com as operações transacionais prejudicando o andamento das operações na organização.

Quando a carga é realizada mensalmente, o processo de ETL acessa a base de dados do sistema transacional após o fechamento mensal do negócio, populando a base do DW apenas uma vez ao mês.

Essa informação deve estar registrada no documento principal de especificação do projeto.

## **Passo 2 - Elaborar documento de Perspectivas de Análises (Visões)**

Todo entendimento deve ser documentado para que os demais analistas tenham acesso às informações do projeto. Normalmente, cada organização usa a metodologia que melhor se encaixa às suas necessidades. No entanto, independentemente da metodologia adotada, as perspectivas de análise precisam ser definidas e descritas. Elas são representadas pelas tabelas Dimensões do modelo de dados do DW e contêm os dados que descrevem os fatos. Vamos entender com um exemplo!

Você se lembra do cenário de análise do supermercado, comentado na aula anterior? Vamos relembrá-lo.

### **Cenário de análise: supermercado**

Paulo e Ricardo são gerentes de uma grande rede de supermercados. Eles contrataram o desenvolvimento de uma solução que apoie a tomada de decisão da organização. Para entender as necessidades de Paulo e Ricardo, algumas reuniões de levantamento foram feitas com eles e com alguns usuários que constroem análises gerenciais. Durante as reuniões foram coletadas as seguintes informações:

O supermercado possui um sistema de apoio ao fluxo de vendas que recebe todas as ocorrências de eventos de compras realizadas pelos clientes, em lojas físicas e pelo e-commerce. Todas as operações de inclusão, alteração e deleção de registros ocorrem durante o período do atendimento ao cliente.

Sempre que uma venda ocorre, um serviço informa ao sistema de estoque quais produtos foram vendidos e em que quantidade.

Paulo e Ricardo precisam realizar as seguintes análises:

- Quais são os produtos mais vendidos no verão?
- Quais são os clientes com maior potencial de compras em um determinado grupo de produtos?
- O estoque está zerado?
- Quais são os fabricantes dos produtos que oferecem maior lucro na comercialização de seus itens?



De acordo com o cenário, é possível entender que, para analisar quais são os produtos mais vendidos no verão, precisamos saber a quantidade vendida de cada produto e em que momento ela ocorreu. Aqui temos a visão **Produto**, a visão **Tempo** e a medida Quantidade de Produtos Vendidos. A visão Produto e a visão Tempo descrevem a medida Quantidade de Produtos Vendidos, ou seja, elas informam qual produto foi vendido e em que momento.

Para acompanhar a venda de produtos e o estoque novamente identificamos a visão Produto e a visão Tempo. No entanto, precisamos saber qual a Quantidade do Produto no Estoque. A Quantidade de Produto no Estoque é mais uma medida identificada.

As medidas são os fatos que ocorreram em um determinado momento. Por exemplo, o produto foi vendido. O fato ocorrido é a venda do produto e, nesse caso, além de sabermos que a venda ocorreu, sabemos também a quantidade que foi vendida. Então, o fato nesse caso é: foram vendidas 10 unidades do produto sabonete. Esse conceito será detalhado mais à frente.

Na última análise desejada pelos usuários, além da visão Produto, qual(is) outra(s) visão(ões) ou medida(s) podem ser identificadas?

Resposta: A visão **Fabricante do Produto** e a medida Lucro.

Contudo, durante o levantamento de requisitos, foi informado pelos usuários que o Lucro não está no sistema origem. Para obter Lucro no final do mês, o valor da venda do produto é extraído por meio de um relatório do sistema SisVendas, assim como o preço do produto comprado no Fabricante é extraído do sistema SisEstoque. Com as duas informações em uma planilha o lucro é calculado.

Aqui temos uma medida calculada e que precisa ser documentada com a fórmula de cálculo, para que seja possível apresentar o resultado esperado.

Após identificar as visões de análise, é hora de documentar as informações obtidas sobre elas. Essas informações podem ser verificadas com os gestores e aprofundadas com os analistas responsáveis pelos sistemas de origem (sistemas transacionais).

A Visão (Dimensão) contém os dados referentes ao domínio que está sendo tratado. Por exemplo, a visão Produto contém o código do Produto, que é importante na identificação do produto no sistema origem; a descrição do produto, que permite saber qual é o produto analisado; e o fabricante do produto.

A figura a seguir ilustra a documentação da Visão Produto. A coluna Visão de Análise contém o nome da visão, a coluna Atributo contém os dados referentes ao produto, a coluna Conceito descreve cada um dos atributos. O conceito é extremamente importante para um ambiente analítico, pois o usuário e os analistas saberão o que é o dado, tanto na construção das análises quanto na manutenção do ambiente.

A coluna Exemplos contém alguns exemplos dos dados para auxiliar nas próximas etapas do projeto. O campo Observação é livre para adicionar comentários importantes sobre cada um dos dados, caso tenham, e regras de negócio que deverão ser aplicadas a eles.

Visão de Análise	Atributo	Conceito	Exemplos	Observação
Produto	-	Descreve os produtos do DW Supermercado	-	-
	Código do Produto	Identifica unicamente um produto no sistema SisVendas	1,2,3	Códigos de produtos que deixaram de ser comercializados não podem ser reutilizados em novos produtos
	Descrição do Produto	Nome do produto que está sendo comercializado no SisVendas	Detergente	Pode conter até 100 caracteres
	Fabricante do Produto	Fabricante do produto que está sendo comercializado no SisVendas	Limpa+	Pode conter até 200 caracteres
	Categoria do Produto	Grupamento do produto que está sendo comercializado no SisVendas	Limpeza	Pode conter até 50 caracteres

Vamos praticar?

Além das visões citadas, há mais duas visões importantes para o cenário. Você consegue identificá-las?

Resposta: Visões Cliente e Categoria do Produto.

### Passo 3 - Elaborar documento com as medidas que serão analisadas (Indicadores)

Após a documentação das visões de análise, é hora de documentar as medidas, também conhecidas como Indicadores. Os Indicadores são organizados em tabelas Fato, que registram os fatos ocorridos, como informado anteriormente.

No cenário supermercado foram identificados alguns Indicadores: **Quantidade de Produtos Vendidos, Quantidade de Produto no Estoque, Preço do Produto Vendido, o Preço do Produto Comprado do fabricante e o Lucro do Produto Vendido.**

A figura a seguir ilustra a conceituação dos Indicadores identificados durante o levantamento com os usuários. A coluna Indicador lista o nome dos indicadores, a coluna Conceito lista os conceitos/definições dos indicadores, a coluna Fórmula descreve como os indicadores devem ser calculados e a coluna Observação contém informações adicionais.

Indicador	Conceito	Fórmula de cálculo	Observação
Quantidade de Produtos Vendidos	Quantidade do produto vendido em um pedido	Soma das unidades do produto	
Quantidade de Produto no Estoque	Quantidade de produto disponível no estoque	Soma das unidades do produto	
Preço do Produto Vendido	Preço do produto no momento da venda do produto	Não há	Apresentar o valor com formatação de moeda e com duas casas decimais
Preço do Produto Comprado do fabricante	Preço do produto quando ele foi comprado do fabricante ou distribuidor do produto	Não há	Apresentar o valor com formatação de moeda e com duas casas decimais
Lucro do Produto Vendido	Lucro obtido na venda do produto	Preço do Produto Vendido - Preço do Produto Comprado do fabricante	Apresentar o valor com formatação de moeda e com duas casas decimais


 Conceituação dos Indicadores identificados durante o levantamento com os usuários. Fonte: próprio autor

#### Matriz de granularidade

Para facilitar o entendimento e a compreensão da relação existente entre as Visões e os Indicadores do DW/DM, temos a matriz de granularidade. Em formato de matriz, são organizados as Visões (Atributos) e os Indicadores que estão relacionados a essas visões. A Figura 5 ilustra a relação entre as visões identificadas no levantamento e os indicadores que serão analisados nas consultas predefinidas. No eixo X da matriz estão organizadas as Visões Tempo, Cliente, Fabricante e Produto. No eixo Y estão organizados os Indicadores Quantidade de Produtos Vendidos, Quantidade de Produto no Estoque, Preço do Produto Vendido, Preço do Produto Comprado do fabricante e Lucro do Produto Vendido.

De acordo com a matriz, sabemos que a Quantidade de Produtos Vendidos pode ser analisada pela data de venda do produto ao cliente. Por exemplo, eu posso saber a quantidade de sabonetes vendidos no dia 20/08/2020, no mês 08/2020 ou ainda no ano de 2020.

Indicadores	Visões											
	Venda ao Cliente			Estoque			Cliente		Produto		Fabricante	
	Data da Venda	Mês da Venda	Ano da Venda	Data do Estoque	Mês do Estoque	Ano do Estoque	Nº do Cliente	Nome do Cliente	Cód do Produto	Descrição do Produto	Cód do Fabricante	Descrição do Fabricante
Quantidade de Produtos Vendidos	X	X	X				X	X	X	X	X	X
Quantidade de Produto no Estoque				X	X	X			X	X	X	X
Preço do Produto Vendido	X	X	X				X	X	X	X		
Preço do Produto Comprado do fabricante					X	X			X	X	X	X
Lucro do Produto Vendido					X	X			X	X	X	X

 Relação entre as visões identificadas no levantamento. Fonte: próprio autor

No nosso exemplo há poucas visões e indicadores e isso facilita saber quais são os possíveis cruzamentos entre eles. No entanto, no levantamento de um DW/DM real, há inúmeros cruzamentos e a matriz permite a visualização das análises que serão possíveis no ambiente analítico de forma mais simples e objetiva. Além disso, a matriz de granularidade apoia os analistas que estão atuando no projeto.

Você observou que essa matriz se chama Matriz de Granularidade? A granularidade é referente ao grão de análise do DW/DM. O grão é o nível de detalhamento dos dados, em que quanto mais granular/menor granularidade, mais detalhada é a informação, quanto mais alta a granularidade, menos detalhada é a informação.

Por exemplo, é possível analisar o Preço do Produto Vendido por data da venda (dia, mês e ano) e o Preço do Produto Comprado do Fabricante só pode ser analisado por mês e pelo ano. Isso significa que a informação sobre a venda dos produtos ao cliente é mais granular do que a informação sobre a compra do produto com o fabricante para o abastecimento do estoque.

## Passo 4 - Elaborar documento que descreva as análises desejadas (Consultas)

O documento das análises predefinidas deve conter o layout de todas as consultas desejadas pelos usuários e identificadas durante o levantamento das necessidades. Pode acontecer de novas análises surgirem ao longo do projeto. Se essa nova análise utilizar as visões e indicadores já mapeados no levantamento, é simples desenhar esse novo layout e entregar a análise ao cliente, deixando-o satisfeito com a entrega, agregando valor à organização.

Contudo, se as visões ou os indicadores não estiverem mapeados, os participantes do projeto, tanto analistas e usuários devem ser reunidos para que seja definida a melhor forma de atendimento da nova necessidade. Para isso, alguns pontos precisam ser considerados no impacto no projeto, como tempo e dinheiro.

A descrição de uma análise deve conter o desenho do relatório ou *dashboard* para que seja possível o alinhamento das expectativas com o cliente. O desenho permite que ele visualize suas futuras análises de forma mais fácil e mais aproximada do produto que será entregue.

Além dos desenhos, devem estar presentes: descrição de cada análise com o objetivo, atributos que estarão na análise, indicadores, filtros obrigatórios e filtros dinâmicos, caso sejam necessários. A figura ilustra um exemplo de uma especificação de consulta.

Vendas de Produtos por Mês				
Mês Venda	Produto		Categoria	Quantidade de Produtos Vendidos
	Código	Descrição	Descrição	
Abril/2020	1	sabonete	Higiene	1523
	2	Pão de forma	Padaria	150
	3	Suco de Uva Integral	Bebida	63
<b>Descrição</b> O Objetivo do relatório é apresentar a quantidade de produtos vendidos por mês.				
<b>Visões</b> - Mês da Venda - Produto (Código e Descrição) - Categoria do Produto				
<b>Indicadores</b> - Quantidade de Produtos Vendidos				
<b>Filtros</b> - O filtro Mês é de preenchimento obrigatório. - O relatório deve permitir filtrar por Categoria de Produtos.				

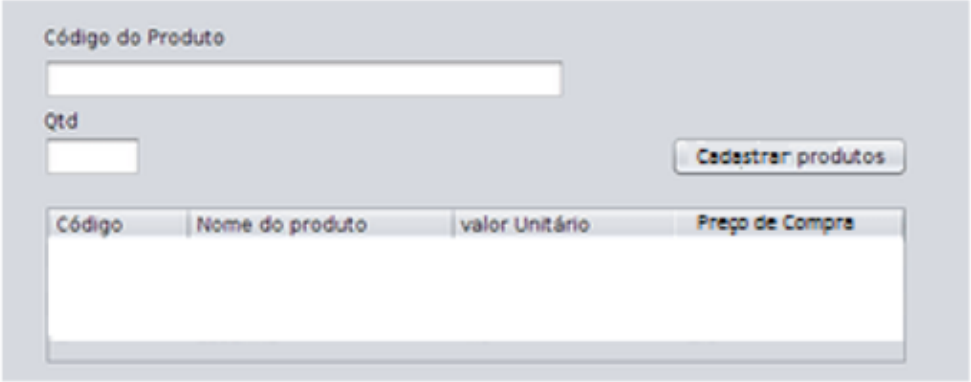
 Exemplo de uma especificação de consulta. Fonte: próprio autor

## Passo 5 - Elaborar documento com apontamento das origens dos dados

Com o mapeamento das visões de análise e dos indicadores é possível verificar a origem dos dados. Essa verificação, normalmente, é feita com os analistas responsáveis pelos sistemas transacionais. A existência de cada uma das visões e dos indicadores no sistema origem deve ser checada.

O apontamento da origem dos dados é muito importante, porque pode ser que o dado não exista no sistema transacional, ou ainda pode não ser possível extraí-lo do sistema origem. Uma vez que essa situação ocorra, deve ser informada ao gestor para que o entendimento seja alinhado sobre o dado.

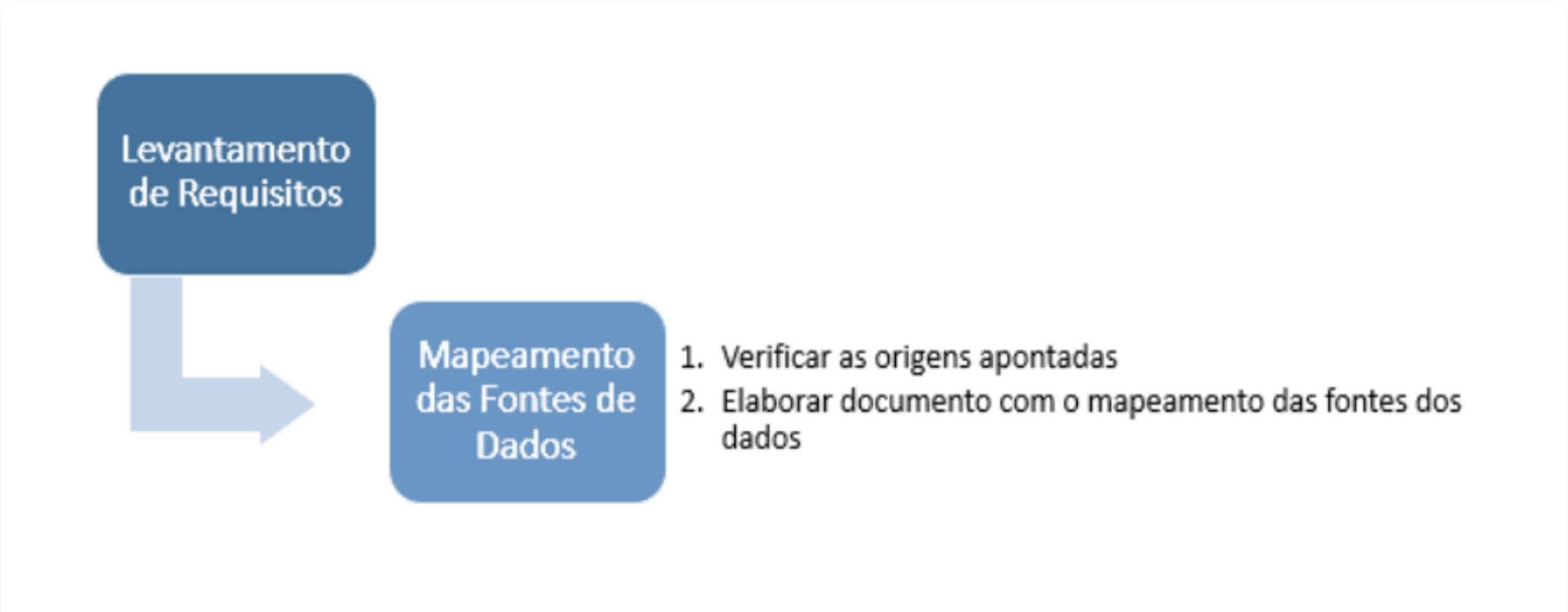
A Figura ilustra um exemplo de apontamento da origem do dado.

Dado	Sistema origem	Tabela	Tela
Produto	SisVendas	Produto TB_Produto	

 Exemplo de apontamento da origem do dado. Fonte: próprio autor

# Mapeamento das Fontes de Dados

Veja agora os passos do Mapeamento das Fontes de Dados.



Fonte: O autor.

## Passo 1 – Verificar as origens apontadas

O mapeamento das fontes dados é uma verificação mais detalhada da origem dos dados mapeados nas etapas anteriores. Com a especificação da necessidade e os conceitos definidos, o analista que realiza essa tarefa poderá localizar o dado no sistema origem, conhecer sua real localização, com o nome da tabela que será acessada, o nome, o tamanho e o tipo de dado do campo. Se o sistema transacional for muito antigo ou não houver documentação sobre ele, a investigação mais profunda da origem pode trazer surpresas que precisão ser tratadas e contornadas.

A conceituação obtida com os gestores auxiliará na identificação do dado no sistema origem e é utilizada na integração de dados, caso venham de sistemas diferentes. Durante o mapeamento das origens podem ser definidas regras a serem aplicadas na etapa de construção do ETL.

## Passo 2 - Elaborar documento o mapeamento das fontes dos dados.

O documento de Mapeamento das Fontes de Dados pode ser uma versão estendida do documento de apontamento da origem de dados, acrescentado as informações levantadas pelo analista técnico.

Dado	Sistema origem	Tabela	Tela	Campo	Tipo de Dado	Tamanho	Observação
Produto	SisVendas	Produto TB_Produto	<div>Cadastro do Produto</div> <div><div>Código do Produto</div><div>Qtd</div><div>Cadastrar produtos</div><div><div>Código</div><div>Nome do produto</div><div>valor Unitário</div><div>Preço de Compra</div></div></div>	ID_Produto	Inteiro		
				DS_Produto	String	200	
				ID_Fabricante	Inteiro		
				ID_Categoria	Inteiro		

Fonte: O autor.



# Metadados

O banco de Metadados reúne informações importantes que apoiam o desenvolvimento do DW. Conhecido como Dados sobre Dados, ele permite o entendimento sobre o que é o dado, quais são suas características (tipo, tamanho), qual a sua origem, qual o seu destino, quais são suas relações, entre outros.

Conforme definido por Barbieri (2020), Metadados podem ser classificados em três tipos:

1

**Metadados de Negócios**  
Contextualiza o dado na visão do negócio.

2

**Metadados Técnicos/Físicos**  
Diz respeito a estrutura, tamanho, domínio, entre outros.

3

**Metadados Operacionais**  
Relacionado a processos, rotinas, tratamentos associado aos dados, entre outros

Metadados contribuem com o processo de ETL, com os analistas, com os usuários e se comportam como um dicionário que norteia o entendimento sobre todos os dados contidos no DW e sobre os processos que disponibilizam no ambiente analítico.

Atenção

Desde o início do levantamento de requisitos, informações estão sendo coletadas e farão parte do banco de Metadados do projeto. Sua manutenção deve ocorrer durante todo o ciclo de desenvolvimento do projeto, quando ocorrer manutenção ou crescimento do DW.

Estudamos, nesta aula, a importância do levantamento de requisitos para um projeto de DW/DM, como realizar o entendimento sobre a necessidade dos usuários e a importância de documentar o conhecimento adquirido. Vimos também que os documentos elaborados são utilizados pelos analistas que participam da construção do DW/DM, pelos usuários que farão suas análises no ambiente e pelas pessoas que futuramente possam interagir com o ambiente analítico.

## Atividades

---

1. O desenvolvimento de um projeto é dividido em fases e possui início e fim. Para iniciar qualquer atividade que envolva várias fases, você precisa planejar a execução delas. O Planejamento é a primeira fase do ciclo de vida de um projeto de Data Warehouse. Nessa fase são definidos:

- a) O escopo do projeto, o processo ETL, as tarefas a serem desenvolvidas no projeto e o mapeamento das fontes de dados.
  - b) A viabilidade de recursos, as tarefas a serem desenvolvidas no projeto e o encadeamento delas e as consultas predefinidas.
  - c) O escopo do projeto, a viabilidade de recursos, a matriz de granularidade e o encadeamento das atividades do projeto.
  - d) O escopo do projeto, a viabilidade de recursos, as tarefas a serem desenvolvidas no projeto e o encadeamento delas.
  - e) O escopo do projeto, a viabilidade de recursos, as tarefas a serem desenvolvidas no projeto e a documentação de visões e indicadores.
- 

2. O Levantamento de requisitos é uma importante etapa no desenvolvimento de projetos:

- a) Que desenvolve sistemas de apoio operacional.
  - b) Para todos os projetos, menos projetos de Data Marts, pois possuem um escopo menor.
  - c) Para todos os projetos, inclusive projetos de Data Warehouse.
  - d) Caso o objetivo e as fontes de dados sejam conhecidos pelos analistas de BI, não é necessário realizar o levantamento de requisitos.
  - e) É importante em outros projetos, mas não se aplica em projetos de Data Warehouse.
- 

3. O levantamento de requisitos é apoiado por técnicas que auxiliam a condução das entrevistas. Sobre o BABOK, é possível afirmar que:

- a) São técnicas utilizadas para orientar a condução de projetos de Data Warehouse.
  - b) É um guia que reúne os principais conceitos e técnicas que apoiam a Análise de Negócios.
  - c) É um guia que reúne documentos para a elaboração das consultas e da matriz de granularidade.
  - d) É o conjunto completo de documentos elaborados durante o levantamento de requisitos.
  - e) É um dicionário de dados utilizado para a construção do Metadados.
- 

4. A Matriz de Granularidade é um documento que:

- a) Relaciona Visões e Indicadores, apresentando a relação entre eles e explicita o grão dos dados nas análises do DW/DM.
- b) Relaciona Visões e Indicadores, apresentando a relação entre eles e define as consultas que deverão ser construídas.

- c) Relaciona os Indicadores com as consultas em que eles serão apresentados.
  - d) Apresenta o grão contido no sistema de apoio operacional, fonte do DW/DM desenvolvido.
  - e) Relaciona Visões que serão desenvolvidas com seus conceitos e explicita o grão dos dados contidos no DW/DM.
- 

5. Metadados são muito importantes para sistemas de BI e mantêm informações relevantes sobre os dados. O banco de Metadados de um projeto de BI:

- a) Documenta os processos de extração, conceitos e histórias dos usuários da organização.
  - b) Documenta os dados contidos no DW/DM, os tratamentos sobre os dados, o relacionamento entre eles, o entendimento de conceitos e definições e a verificação das regras de negócios aplicadas sobre os tratamentos realizados.
  - c) Documenta os processos de extração, conceitos e definições de negócio e os erros que ocorrem nos sistemas transacionais que são fontes para os sistemas de BI.
  - d) Documenta o mapeamento dos processos de extração, os resultados obtidos pelas consultas, mas não registram regras de negócio e conceitos.
  - e) Não apresenta conhecimento sobre o ambiente e sim as estatísticas das execuções de consultas realizadas pelos usuários.
- 

## Notas

Título modal <sup>1</sup>

---

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos.

Título modal <sup>1</sup>

---

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos.

## Referências

---

BARBIERI, C. 2020. **GD 2.0-Governança de Dados-Parte 8- Metadados**. <https://www.linkedin.com/pulse/gd-20-governan%C3%A7a-de-dados-parte-8-metadados-carlos-barbieri>. Acesso em: 20 set. 2020.

KIMBALL, M. R. R. **The Data Warehouse Toolkit - The Definitive Guide to Dimensional Modeling**. 3. ed. Indianapolis, Indiana: John Wiley Sons, 2013.



- Modelagem de Dados Multidimensional;
- Diferenças entre o modelo relacional e dimensional;
- Esquema estrela e floco de neve.

## Explore mais

---

- Conheça mais sobre guia BABOK e aprofunde seus conhecimentos sobre a análise de requisitos, por meio do Portal de Análise de Negócios para o público brasileiro - IIBA (International Institute os Business Analysis):  
<http://www.iiba.org.br/index.php/sobre-o-iibabr>.
- Conheça mais sobre Metadados no artigo escrito por Carlos Barbieri: GD 2.0-Parte 8-GD2.0-Metadados.