

CONSTRUINDO UM DATA WAREHOUSE - PARTE 01

Objetivo do artigo:

Mostrar na prática o processo de construção do Data Warehouse. Neste primeiro artigo iremos abordar o tema: Levantamento de requisitos de negócio que lhe dará dicas que poderão ser úteis no desenvolvimento de projetos de Business Intelligence(BI)

As técnicas aqui apresentadas é uma visão pessoal com base em experiências na construção de projetos de BI.

O que preciso saber para o melhor entendimento?

É necessário conhecimento prévio sobre: Data Warehouse, Data Marts, Dimensões, Fatos, Métricas.

O que vou aprender?

Ao final do artigo é esperado que adquiriam conhecimento para iniciar o processo de construção da **matriz de necessidade** que poderá ser uma ferramenta importante no levantamento de requisitos de negócio auxilia-lo nos projetos de BI que vierem a desenvolver.

Aos inscritos no site, será disponibilizado um modelo da **Matriz de necessidade**. Inscreva-se gratuitamente [clikando aqui](#).

Muitos dos problemas começam com a pergunta: Como uma ferramenta OLAP (X,Y,Z) poderá consultar minha base de dados relacional? na cabeça de muitos é comum achar que você só precisa configurar a ferramenta para acessar sua base relacional e a partir daí é só gerar os relatórios, gráficos, *Dashboards* entre outros. Apesar de algumas ferramentas apresentarem essa característica não é uma prática comum das ferramentas de *Business Intelligence*.

No primeiro de 5 artigos, irei falar sobre o processo de levantamento de requisitos de negócio, onde através de experiência em projetos de BI mostro uma técnica que utilizo para que nos ajude a ultrapassar com sucesso essa difícil etapa.

A Figura 1 mostra um passo a passo proposto para a criação de um DW, existem outras etapas não contempladas no fluxo, porém é de grande importância para o nosso projeto como: (Qualidade dos dados, criação do modelo multidimensional, Controle e acompanhamento de carga), essas etapas também serão abordados nos próximos artigos.

Figura 1 Etapas para construção do DW



Desenvolvo projetos de *Business Intelligence* há 06 anos dentre esses anos utilizei diversos métodos de levantamento de requisitos de negócio, entre elas entrevistas e matriz de necessidade. Muitas das vezes a entrevista é o ponto inicial para que possamos entender a expectativa do gestor e identificar se o que está sendo esperado poderá ser respondido pelo Data Warehouse. Porém no decorrer do tempo percebemos que a entrevista como técnica principal na análise de requisitos de negócio por se só em muitos dos casos passavam a ser dificultoso, principalmente pela falta de agenda dos gestores que participavam do projeto.

Assim foi necessário buscar alternativas, no livro *The Data Warehouse Toolkit* o autor Ralph Kimball apresenta uma técnica chamada Matriz de barramento do *Data Warehouse*, nela é criado uma matriz onde as colunas representam as dimensões comuns do processo de negócio, e as linhas representam os *Data Marts* como mostra a figura 2, porém essa Matriz dá uma visão geral do projeto, ainda não era o que precisávamos pois não se tratava de levantamento dos requisitos de negócio.

PROCESSO DE NEGÓCIO	DIMENSÕES COMUNS				
	Tempo	Produto	Fornecedor	Cliente	Transportadora
Vendas	x	x		x	
Estoque	x	x	x		
Contas a Pagar	x		x		x
Constas a Receber	x			x	
Entregas	x	x	x	x	x

Figura 2 Matriz de barramento

Em 2008 tive o prazer de conhecer uma técnica que iria mudar minha forma de obter requisitos de negócio, através de um grande profissional que tive o prazer de trabalhar Grimaldo Lopes ([encontre-o no linkedin](#)) com sua experiência de mercado viu a necessidade de criar uma forma que fosse mais rápida e eficaz para os seus projetos, resolveu juntamente com sua equipe criar uma técnica conhecido como Matriz de necessidade.

Essa técnica permite fácil interação com os gestores, onde sua necessidade é adicionada a uma matriz, onde as linhas são os descritores e as colunas são compostas por métricas pertencentes a um assunto.

O Processo de construção da matriz de necessidade inicia em uma reunião com os gestores onde será definido o escopo do projeto, nessa reunião estará presente gestores de diferentes áreas como (Diretor Financeiro, Gerente de Vendas, Gerente de suprimentos) e a partir daí são definidos descritores e métricas que serão cruzadas a partir de uma marcação ilustrando as relações entre os descritores e as métricas. Esse método é composto por algumas fases vista na figura 3:

Fase 1 – É realizada uma sessão com os gestores do negócio onde é apresentada a matriz de necessidade, neste momento é explicado o seu preenchimento, nas linhas estarão todos os descritores exemplo (Tempo, Centro de Custo, Cliente, Fornecedor) e nas colunas estarão as métricas como: (Valor do Item, Quantidade de item, quantidade de produtos no estoque entre outros)

Fase 2 – Preenchimento das linhas da matriz de necessidade com os descritores

Fase 3 – Preenchimento das colunas da matriz de necessidade com as métricas

Fase 4 – Marcação dos itens que relaciona-se entre métricas e descritores (x)

Fase 5 – Separação dos assuntos com cores, por exemplo, (Vendas cor azul, Estoque cor amarela)

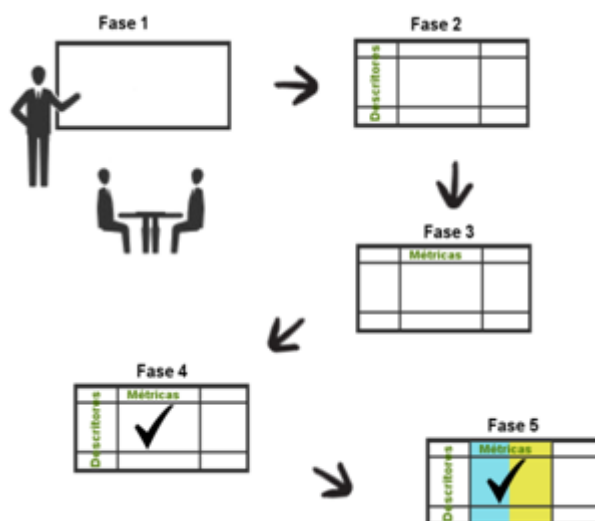


Figura 3 Criação da matriz de necessidade

Ao final é gerada como artefato a matriz de necessidade preenchida que servirá como base para a criação do nosso *Data Mart*, a Figura 4 mostra um exemplo da matriz de necessidade criada, é possível ver nas linhas os descritores que podem ser ou não uma dimensão, essa análise se dará na próxima etapa onde criaremos o modelo multidimensional, veja que separamos os assuntos por cores, a coluna de amarelo é referente ao assunto Contas a pagar e a coluna azul é referente ao assunto Vendas.

Matriz de Necessidade

DIMENSÃO		FATO			Contas a Pagar		Vendas	
		Valor Pago	QTD Item	Valor do Item	Valor Pago	QTD Item	Valor do Item	
CENTRO DE CUSTO								
Descrição do Centro de Custo		■	■	■	■	■	■	
FORNECEDOR								
Descrição do Fornecedor		■	■	■				
HISTORICO								
Descrição do Historico (Natureza Concatenado com Item)		■	■	■	■	■	■	
NATUREZA DO ITEM								
Descrição da Natureza		■	■	■	■	■	■	
ITEM								
Descrição do Item		■	■	■	■	■	■	
DATA DO PAGAMENTO								
Ano do pagamento		■	■	■	■	■	■	
Mês do pagamento		■	■	■	■	■	■	
Bimestre do pagamento		■	■	■	■	■	■	
Semestre do pagamento		■	■	■	■	■	■	
Data completa do pagamento		■	■	■	■	■	■	
AnoMes do Pagamento		■	■	■	■	■	■	
DIMENSÃO		FATO			TITULOS_PAGO		VENDAS	
		Valor Pago	QTD Item	Valor do Item	Valor Pago	QTD Item	Valor do Item	

Figura 4 Exemplo da matriz de necessidade

O layout da matriz de necessidade poderá ser modificada dependendo do seu projeto, é comum utilizar outras colunas como: observações, marcador nos descritores para que possamos identificar se um campo será histórico ou não, assim podemos definir os campos que sofrerão alteração , como também o tipo utilizado pelas dimensões *Slowly Changing Dimensions* (SCDs) que veremos no artigo sobre modelagem multidimensional até mais.

Ficou com dúvida ou gostaria de realizar um elogio, crítica ou sugestão? Nos envie sua mensagem através de nosso site: www.analiseerestado.com.br

Clique aqui e faça parte do nosso grupo no linkedin
Clique aqui e curta nossa página no facebook
Clique aqui e cadastre-se para receber novos artigos.

Artigo realizado por:
 Vinicius Oliveira – Ciências da Computação e MBA em Business Intelligence (BI)

CONSTRUINDO UM DATA WAREHOUSE - PARTE 02



Objetivo do artigo: Mostrar de uma forma prática o processo construção de um modelo multidimensional no processo de criação do Data Warehouse, este é o segundo de 5 artigos que lhe dará dicas importantes no desenvolvimento deste projeto. Será mostrado como técnica o modelo Star Schema, não contemplando o modelo Snow Flake, também iremos exemplificar os tipos de SCD
O Que preciso saber para o melhor entendimento? É preciso o conhecimento prévio sobre Data Warehouse, Dimensões, Fatos.
O Que vou aprender? Neste artigo você irá aprendo um método de criação do modelo multidimensional, e sua importância no projeto de Data Warehouse, você também irá conseguir identificar a qual modelo sua dimensão deverá ser representada. Cadastre-se no nosso site clcando aqui

No artigo anterior falamos sobre o levantamento de requisitos utilizando uma metodologia conhecida como Matriz de necessidades, a figura 1 mostra o exemplo utilizado no artigo, após o levantamento dos indicadores chegou a hora de criarmos o nosso modelo multidimensional, que será um passo importantíssimo no nosso processo de construção do Data Warehouse

Matriz de Necessidade

DIMENSÃO	FATO			Contas a Pagar			Vendas		
	Valor Pago	QTD Item	Valor do Item	Valor Pago	QTD Item	Valor do Item			
CENTRO DE CUSTO									
Descrição do Centro de Custo	☒	☒	☒	☒	☒	☒			
FORNECEDOR									
Descrição do Fornecedor	☒	☒	☒						
HISTORICO									
Descrição do Historico (Natureza Concatenado com Item	☒	☒	☒	☒	☒	☒			
NATUREZA DO ITEM									
Descrição da Natureza	☒	☒	☒	☒	☒	☒			
ITEM									
Descrição do Item	☒	☒	☒	☒	☒	☒			
DATA DO PAGAMENTO									
Ano do pagamento	☒	☒	☒	☒	☒	☒			
Mês do pagamento	☒	☒	☒	☒	☒	☒			
Bimestre do pagamento	☒	☒	☒	☒	☒	☒			
Semestre do pagamento	☒	☒	☒	☒	☒	☒			
Data completa do pagamento	☒	☒	☒	☒	☒	☒			
AnoMes do Pagamento	☒	☒	☒	☒	☒	☒			
DIMENSÃO	FATO			TITULOS_PAGO			VENDAS		
	Valor Pago	QTD Item	Valor do Item	Valor Pago	QTD Item	Valor do Item			

Figura 1 Matriz de Necessidade

Modelagem Multidimensional

É a técnica de modelagem utilizada na criação de Data Warehouse, que permite uma maior abstração do modelo de dados facilitando assim o entendimento para o usuário final.

Esse modelo é dividido em dimensões e fatos, onde as dimensões são os descritores e os fatos serão nossas métricas, para facilitar irei abstrair um pouco mais esse modelo como é possível ver na figura 2.



Onde: Dimensão que dá ideia de local, por exemplo: Onde nasceu um Hospede, Qual departamento vendeu determinado produto, qual classe pertence o aluno.

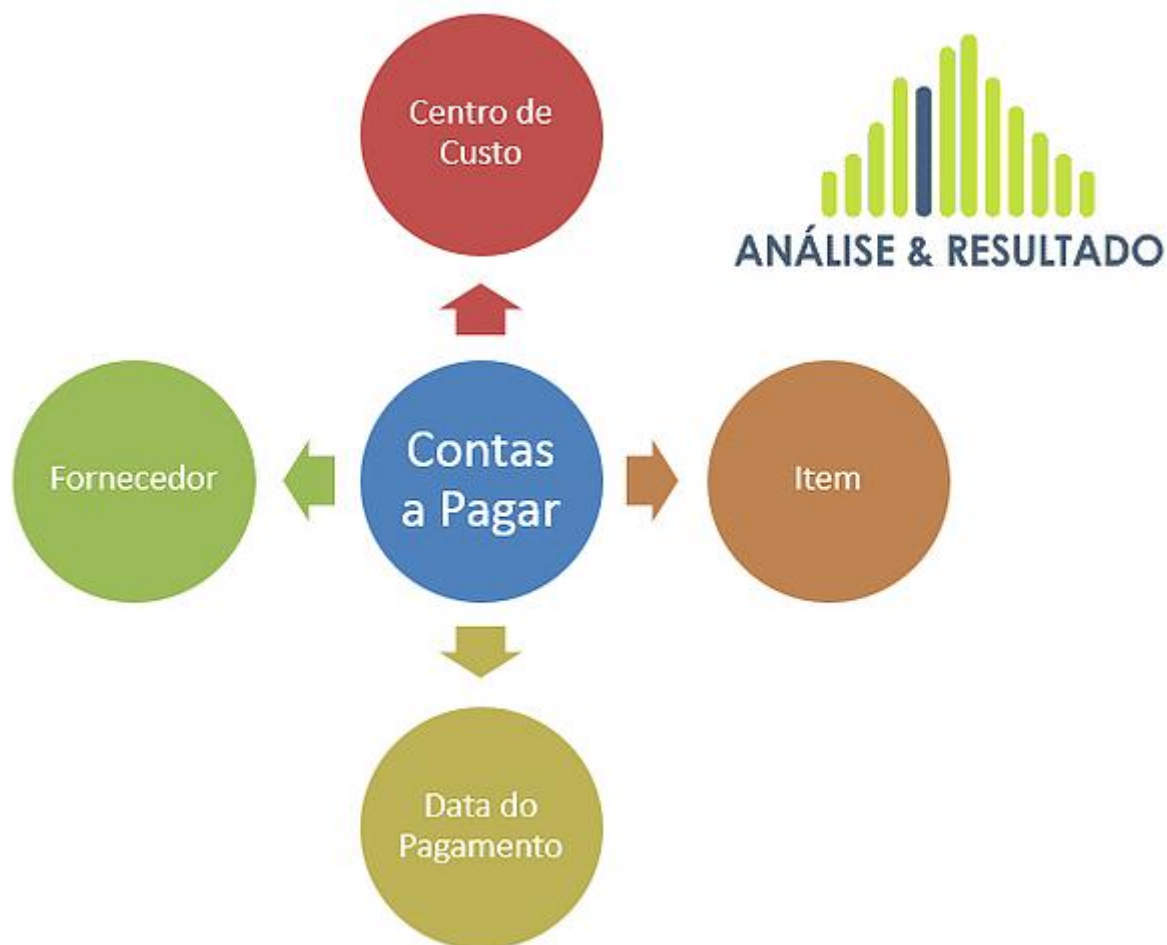
O Que: Nesse grupo encontram-se as dimensões como: O que foi vendido (Produto), O que foi alugado (Quarto, Carro).

Quando: Essa é a dimensão mais importante do nosso DW, conhecida como dimensão tempo, porém gosto sempre de lembrar que o tempo não está restrito ao nosso calendário, além das datas comuns como: Data de Nascimento, Data de criação, Data da compra, Hora da compra, também temos o tempo que não é contemplado no nosso calendário por exemplo: Semestre acadêmica ou período letivo, esses poderão variar em relação ao ano por exemplo 4 unidade de 2014 poderá está ocorrendo no ano de 2015.

Quem/Qual: Nesse grupo iremos encontrar as dimensões como exemplo: Quem realizou a venda, ou qual aluno tirou nota acima de 7.0.

Vamos utilizar a matriz de necessidade para criarmos o nosso modelo multidimensional:

Durante o levantamento da Matriz percebemos que um Item possui uma Natureza, logo criamos somente uma Dimensão ITEM onde terá um descritor de natureza, assim eliminaremos os descritores Natureza do Item e Histórico do Item formando apenas uma dimensão.



Onde: Centro de Custo

Quem: Fornecedor

Quando: Data do Pagamento

O que: Item

Fato (Métricas): Qtd Item, Valor do Item e Valor Pago



No processo de criação das dimensões deveremos identificar qual tipo SCD (Slowly Changing Dimension) será utilizado. Nos próximos artigos, iremos utilizar como exemplos os SCD tipo 2 e tipo 6, abaixo segue uma breve explicação:

SCD 1 – Substitui os valores velhos pelos novos e não mantém histórico

SK_VENDEDOR	CODIGO	NOME	TELEFONE
1	1	VINICIUS OLIVEIRA	(99)9999-9999



SK_VENDEDOR	CODIGO	NOME	TELEFONE
1	1	VINICIUS OLIVEIRA	8888-8888

SCD 2 – Adiciona Data de início e final de vigência do dado, mantendo assim o histórico da informação

SK_VENDEDO R	CODIGO	NOME	TELEFONE	DTC INICIO	DTC FIM
1	1	VINICIUS OLIVEIRA	9999-9999	01/01/1900	



SK_VENDEDO R	CODIGO	NOME	TELEFONE	DTC INICIO	DTC FIM
1	1	VINICIUS OLIVEIRA	9999-9999	01/01/1900	Hoje
201	1	VINICIUS OLIVEIRA	8888-8888	Hoje	

OU

SK_VENDEDO R	CODIGO	NOME	TELEFONE	VERSÃO
1	1	VINICIUS OLIVEIRA	9999-9999	0
201	1	VINICIUS OLIVEIRA	8888-8888	1

SCD 3 – Dimensões de alteração lenta, onde é adicionado duas colunas o estado original e a data de vigência

SK_VENDEDOR	CODIGO	NOME	TELEFONE
1	1	VINICIUS OLIVEIRA	8888-8888



SK_VENDEDOR	CODIGO	NOME	TELEFONE ORIGINAL	TELEFONE ATUAL	DATA DE VIRGÊNCIA
1	1	VINICIUS OLIVEIRA	8888-8888	9999-9999	10/06/2014

SCD 6 – Híbrido, Combina os métodos 1, 2 e 3 (1+2+3 = 6), esse método que será utilizado no artigo e também é implementado no PDI.

SK_VENDEDOR	CODIGO	NOME	TELEFONE	DTC INICIO	DTC FIM	VERSAO
1	1	VINICIUS OLIVEIRA	9999-9999	01/01/1900		1



SK_VENDEDOR	CODIGO	NOME	TELEFONE	DTC INICIO	DTC FIM	VERSAO
1	1	VINICIUS OLIVEIRA	9999-9999	01/01/1900	Hoje	1
201	1	VINICIUS OLIVEIRA	8888-8888	Hoje		2

Conclusão

A Modelagem Multidimensional diferentemente dos modelos relacionais, permite uma maior facilidade na abstração das informações onde as tabelas são divididas entre dimensões e fatos, Nos Fatos serão as métricas já as dimensões nossos descritores. Não necessariamente os descritores criados na matriz de necessidade serão nossas dimensões, é possível identificar o que temos em comum junta-las ou até mesmo separá-las. É muito importante entendermos o modelo de versionamento das dimensões, assim iremos compreender se será necessário manter ou não o histórico das informações, você já imaginou o problema que poderia ocorrer caso um departamento pertencente a uma gerência passe a ser subordinada a outra? Se você não manter o histórico, todos os custos, receitas e despesas passará para a nova gerência podendo ser um problema na hora de tomar decisão.

Aguardo vocês no próximo artigo sobre fonte de dados.

Ficou com dúvida ou gostaria de realizar um elogio, crítica ou sugestão? Nos envie sua mensagem através de nosso site: www.analiseeresultado.com.br

Clique aqui e faça parte do nosso grupo no linkedin
Clique aqui e curta nossa página no facebook
Clique aqui e cadastre-se para receber novos artigos.

Artigo realizado por:
Vinicius Oliveira – Ciências da Computação e MBA em Business Intelligence (BI).

CONSTRUINDO UM DATA WAREHOUSE - PARTE 03



Objetivo do artigo:

Mostrar na prática o desenvolvimento da fonte de dados para desenvolvimento de um Data Warehouse.

Para os leitores cadastrado no site, foi disponibilizado um modelo da fonte de dados criada neste artigo, se não é cadastrado **clique aqui** cadastre-se

O que preciso saber para melhor entendimento?

Primeiro o leitor deverá ter lido os dois artigos anteriores **Levantamento de requisitos e Construção do modelo multidimensional**. Também é necessário conhecimento em DER (Diagrama de Entidade e Relacionamento)

O que vou aprender?

Neste artigo você irá aprender na prática o desenvolvimento do documento de fonte de dados, que servirá como base para o seu processo de carga ;-)

Nos artigos anteriores, aprendemos na prática como funciona o levantamento de requisitos, e posteriormente criamos o modelo multidimensional. Todos nossos artigos são práticos, esse é o terceiro de cinco etapas onde mostro na prática a construção do projeto de Data Warehouse.

Os projetos de Data Warehouse tem no seu processo de Extração Transformação e Carga (ETL), uma das etapas mais complexas do projeto. Já participei de diversos projetos com diferentes formas de construção do nosso processo de ETL, porém encontrávamos sempre diferentes dificuldades, assim foi necessário criar uma metodologia que facilitasse esse processo tão importante.

Em muitos projetos, a construção da sua fonte de dados, é a partir do analista de sistema, que tem o conhecimento do banco de dados, muitas das vezes contamos apenas com um caderno e uma caneta anotando as possíveis relações entre tabelas, campos, chaves...

E como funciona esse método de fonte de dados? Bom nesse método que você acabou de receber no seu e-mail, mostro todo o caminho que precisamos para realizar a carga dos dados.

E onde a Fonte de Dados me ajuda? Bom lembre-se que a Fonte de Dados será um documento, nela iremos realizar juntamente com o conhecedor da base de dados, as relações entre as tabelas dos sistemas relacionais com o modelo do nosso Data Warehouse, identificando suas chaves, campos e relacionamentos, ao final você terá um documento mostrando todos os passos que precisa para execução da carga, assim evitando os esquecimentos na hora de escrever no papel, ou até

mesmo ouvir aquela velha frase “Eu ensinei certo mostrei todos relacionamento inclusive as regras, mas ele que entendeu errado :P”

Na figura 1, mostro um diagrama de entidade e relacionamento fictício para exemplificar a criação do documento de fonte de dados.

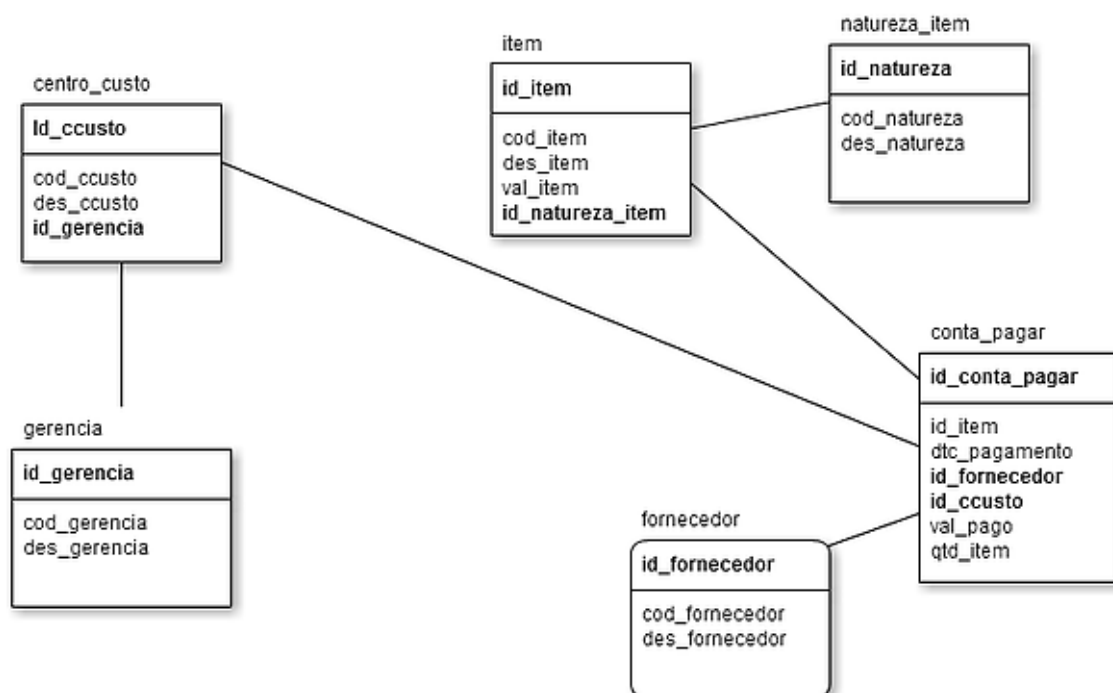


Figura 1 - DER Contas a Pagar

Na nossa matriz de necessidade, é possível ver que as informações de Item e natureza do item estavam separadas, assim adicionamos essas informações diretamente em uma única dimensão chamada Item, como foi visto no artigo de modelagem multidimensional. No nosso exemplo não contemplamos a tabela gerência, logo se seguirmos o que foi levantado na Matriz de necessidades essa informação não iria conter no nosso DW, mas como somos analistas precavidos, vamos adicioná-la na nossa dimensão centro de custo ;-)

FONTE DE DADOS - ANÁLISE E RESULTADO						
Dimensão	Campo	Chave	tabela_origem	chave tabela	coluna tabela	Observação
DIM_FORNECEDOR	cod_fornecedor	X	fornecedor		cod_fornecedor	
	des_fornecedor		fornecedor		des_fornecedor	
DIM_CENTRO_CUSTO	cod_ccusto	X	centro_custo		cod_ccusto	
	des_ccusto		centro_custo		des_ccusto	
			centro_custo	id_gerencia		
			gerencia	id_gerencia		
	cod_gerencia		gerencia		cod_gerencia	
	des_gerencia		gerencia		des_gerencia	
DIM_ITEM	cod_item	X	centro_custo		cod_ccusto	O histórico é a concatenação dos campos des_natureza e des_item
	des_item		centro_custo		des_ccusto	
			centro_custo	id_gerencia		
			gerencia	id_gerencia		
	cod_natureza		gerencia		cod_gerencia	
	des_natureza		gerencia		des_gerencia	
des_historico						

Figura 2 - Construção da Fonte de dados das dimensões

Na figura 2 mostro a criação da nossa fonte de dados que é dividida em duas partes (Dimensão e Fato). Nesta primeira parte mostro a criação da fonte de dados das dimensões, esse é um exemplo prático que poderá ser adaptado de acordo com a sua necessidade.

A tabela abaixo explica cada coluna da nossa fonte de dados.

Campo	Descrição
Dimensão	Nome da dimensão mapeada
Campo	Campo da dimensão que deverá ser mapeada
Chave	Chave da dimensão, chave oriunda do sistema transacional, essa não será a chave física da nossa dimensão, no artigo anterior falei que usaríamos as dimensões do tipo 6, logo a chave deverá ser composta a partir da chave do sistema transacional juntamente com a data de início ou efetivação, garantindo a unicidade já que o mesmo código poderá aparecer mais de uma vez devido ao versionamento da dimensão.
Tabela origem	Nome da tabela no nosso sistema transacional
Chave da tabela	Chave da nossa tabela no sistema transacional, essa informação é preenchida para identificar a junção entre as tabelas como por exemplo: Item e Natureza do Item
Coluna tabela	Coluna da tabela do sistema relacional que irá ser mapeado na coluna da dimensão "Coluna Campo da Fonte de Dados"
Observação	Qualquer observação relevante na construção da fonte de dados

Na Figura 3 apresento o mapeamento de fonte de dados da nossa fato contas a pagar, veja que nesse mapeamento adicionamos mais uma coluna chamada dimensão, lembre-se que na nossa fato criamos chaves artificiais conhecidas como surrogate key que são as ligações entre nossa fato e nossas dimensões, isso acontece por que nossas dimensões sofrem atualizações assim as chaves do sistema transacional podem se repetir na nossa dimensão assim somos obrigados a criar nossas chaves artificiais.

FONTE DE DADOS - ANÁLISE E RESULTADO								
Fato	Campo	Chave	dimensao	tabela_origem	chave tabela	coluna tabela	Observação	
contas_pagar	val_pago			conta_pagar		val_pago		
	qtd_item			conta_pagar		qtd_item		
				conta_pagar	id_item			
				item	id_item			
	val_item			item		val_item		
				item	cod_item			
			DIM_ITEM		cod_item			
	sk_item		DIM_ITEM				sk_item	
					conta_pagar	id_fornecedor		
					fornecedor	id_fornecedor		
					fornecedor	cod_fornecedor		
			DIM_FORNECEDOR			cod_fornecedor		
	sk_fornecedor		DIM_FORNECEDOR				sk_fornecedor	
					conta_pagar	id_ccusto		
					centro_custo	id_ccusto		
					centro_custo	cod_ccusto		
			DIM_CENTRO_CUSTO			cod_ccusto		
	sk_ccusto		DIM_CENTRO_CUSTO				sk_ccusto	
					centro_custo	dte_pagamento		
			DIM_TEMPO			dte_completa		
sk_tempo		DIM_TEMPO				sk_tempo		

Figura 3 - Fonte de dados da fato contas a pagar

A criação da fonte de dados é uma etapa muito importante na construção do Data Warehouse, a partir dela iremos desenvolver o nosso processo de carga, além de ser uma ótima documentação do nosso DW.

No próximo artigo iremos falar sobre o processo de carga do nosso DW, aguardo vocês!

[Clique aqui](#) e faça parte do nosso grupo no linkedin

[Clique aqui](#) e curta nossa página no facebook

[Clique aqui](#) e cadastre-se para receber novos artigos.

Artigo realizado por:

Vinicius Oliveira – Ciências da Computação e MBA em Business Intelligence (BI)