**Armazém de Dados**

**O ciclo da exploração:**

1. Extração de dados
2. Visualização de resultados
3. Análise de resultados

As estruturas dimensionais de dados (hipercubos) são os responsáveis pelo suporte de qualquer processo de exploração de dados e, como tal, vitais para o desempenho adequado de qualquer sistema orientado para a tomada de decisão.

**Desenvolvimento de sistemas de data warehousing**

**Modelação dimensional**

A modelação dimensional de dados é uma das atividades mais relevantes que usualmente se desenvolve no âmbito de um projeto de conceção e implementação de um sistema de data warehousing.

É a atividade relacionada com o desenvolvimento de esquemas para sistemas de dados, especialmente orientados para o suporte a processos de tomada de decisão.

Os modelos dimensionais são os alicerces de todos os processos de tomada de decisão suportados por um data warehouse. Esquemas dimensionais complexos exigem sistemas de povoamento complexos.

**Desenvolvimento de SDW**

Os SDW requerem metodologias de conceção especificas, adequadas à sua própria natureza e objetivos. Mesmo quando a nível operacional, os SDW devem acolher e refletir as necessidades de suporte à decisão requeridas pelos agentes empresariais.

**Processo de desenvolvimento**

Os SDW devem também ser rápidos, apresentando excelentes desempenhos na satisfação desses requisitos, o que implica que as suas estruturas de dados devem estar orientadas especificamente para o fornecimento dos dados pedidos.

**Ciclo de desenvolvimento**

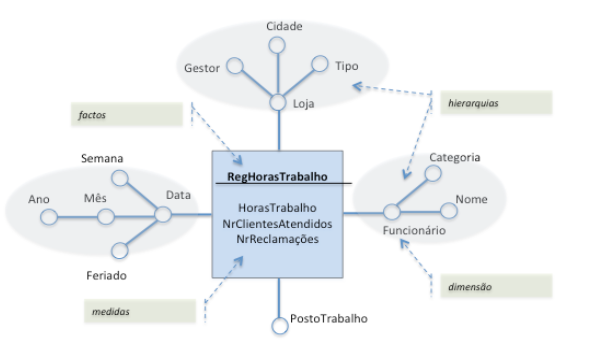
1. Planeamento do projeto
2. Definição dos requisitos do negócio
3. Modelação dimensional
4. Desenho físico
5. Projeto e desenvolvimento da área de retenção

**Etapa da modelação dimensional**

1. Construção da matriz de decisão
2. Seleção da data Mart a desenvolver
3. Escolha do grão das tabelas de factos
4. Escolha das dimensões de analise
5. Desenvolver o diagrama das tabelas de factos
6. Documentar as tabelas de factos
7. Projetar o detalhe das dimensões

**Etapa da análise das fontes:**

1. Identificar as fontes de dados candidatas
2. Analisar o conteúdo das fontes de dados – dados e metadados
3. Desenvolver uma tabela com o mapeamento dos dados entre as diversas fontes de dados operacionais e os dados da data warehouse – source-to-target map.
4. Estimar o número de registos envolvidos futuramente no processo de povoamento
5. Rever o projeto e tratar da aceitação por parte dos seus futuros utilizadores.

**Construção de data Warehouses**

**Adoção de um modelo**

**Processo de modelação**

Uma das formas mais usadas de fazer um desenvolvimento de um esquema dimensional é através da utilização do método dos “4 passos”. Pressupõe o desenvolvimento do SDW de baixo para cima(bootom-up).

**Método dos “4 passos”**

1. Seleção da área de suporte à decisão a implementar.
2. Definição do detalhe dos factos (o grão) do processo selecionado.
3. Seleção das dimensões de analise sobre as quais se pretende analisar os factos.
4. Definição das medidas a integrar na estrutura de cada facto.

**Seleção da área de suporte à decisão**

O processo de modelação dimensional inicia-se com a identificação e caracterização da área de negócio em que desejamos desenvolver as nossas atividades de tomada de decisão.

**Data mart** – são subconjuntos departamentais focados em determinados assuntos e podem ser compostos por um ou mais cubos de dados. Um data mart de marketing, por exemplo, pode armazenar dados referentes a clientes, produtos, vendas, etc.

Identifica-se o data mart para a matriz de decisão.

**Grão, factos e a sua representação**

**O grão de uma tabela de factos** – define a estrutura base dos registos. Está intimamente ligado com a estrutura da tabela de factos.

**Tipo de tabelas de factos:**

1. Transacional
2. Instantâneo
3. Acumulativo

O tipo de uma tabela de factos revela-nos algumas características muito pertinentes, nomeadamente:

* a forma como é povoada
* o tipo de elementos de dados irá acolher
* a forma como os seus dados serão extraídos das fontes de informação
* a cadência do seu povoamento
* a periodicidade de refrescamento

**Dimensões e perspetivas de analise**

**Dimensões**

**Dimensão** – suporta uma dada perspetiva de análise, categorizando um dado objeto de dados que pode ter uma ou mais referencias em tabelas de factos. Uma dimensão disponibiliza um meio privilegiado de exploração das várias medidas de uma tabela de factos. Os atributos de uma dimensão são definidos de acordo com as necessidades apresentadas pelos agentes de decisão.

**Hierarquias**

A definição e a caracterização das hierarquias numa dimensão permitem indicar os caminhos de agregação – desagregação que se podem seguir a partir de um dado nível de um dado atributo de uma dimensão. Uma hierarquia define uma sequência lógica de atributos.

Uma hierarquia é uma arvore direcionada na qual os nodos são constituídos por atributos de dimensões e os ramos representam relacionamentos do tipo N:1 entre pares de atributos dimensionais. São agregações, por exemplo agrupar por titulo, por editora,..

Por exemplo, os ramos da hierarquia:

• H1: LivroId-­‐>Título-­‐>Editoar-­‐>ALL

• H2: LivroId-­‐>Título-­‐>Língua-­‐>ALL

• H3: LivroId-­‐>Título-­‐>Subgénero-­‐>Género-­‐>ALL

**Tipos de dimensões e suas variantes**

**Tipos de dimensões:**

* com variação
* conforme ou partilhada
* degenerada
* gigantesca
* mini dimensão
* sub dimensão
* com diferentes papéis
* de controlo ou auxiliar
* de origem externa

A categorização de uma dimensão está muito dependente do tipo de serviço que esta presta ao SDW durante as suas fases de povoamento, refletindo:

* a forma como os dados vão evoluindo ao longo da vida da data warehouse;
* a forma como é realizada a exploração de dados
* **Dimensões com variação**

As tabelas de dimensão apresentam uma característica que as diferencia de todas as outras tabelas existentes num data warehouse:

* Os valores dos seus atributos podem variar ao longo do tempo.

Isto contrasta claramente com uma das características base de um data warehouse: a sua **não volatilidade – os dados são factuais e como tal não devem ser atualizados, apenas podem ser lidos.**

**Volátil** – sofre constantes mudanças

Numa dimensão, os atributos foram escolhidos para fazer a sua caracterização, suportar as diferentes perspetivas de exploração de dados sobre a dimensão.

As dimensões não constituem qualquer elemento factual relacionado com qualquer uma das áreas de decisão em que estão envolvidas, fornecendo elementos de dados que comprovem esta ou aquela situação – simplesmente atuam como elementos de descrição, de filtragem ou de agregação.

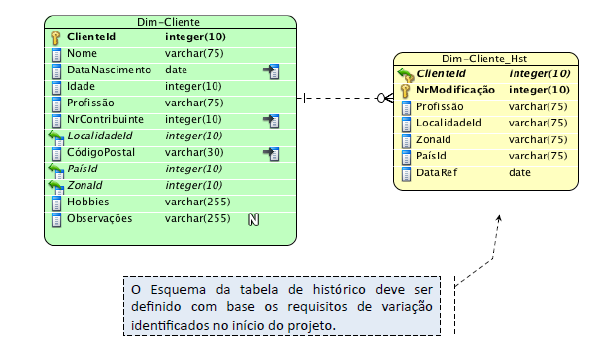
Mas, as dimensões com variação categorizam-se em 4 tipos:

* Tipo 1 – reescrita simples dos dados afetados
* Tipo 2 – criação de novo registo na tabela base
* Tipo 3 – criação de novos atributos
* Tipo 4 – criação de tabelas de histórico

Caso um cliente mude de casa, teremos de atualizar o registo. Cria-se uma dimensão histórica que contem atributos temporais, como por exemplo, a profissão, a localidade, a cidade, etc.

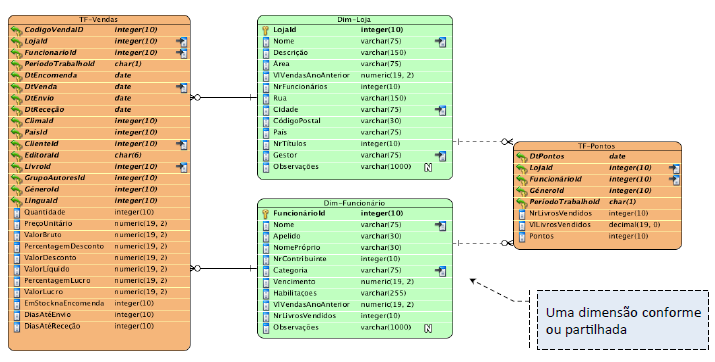
Ao caracterizarmos uma dimensão com variação devemos indicar sempre, para cada um dos seus atributos, a forma como queremos que o futuro sistema de povoamento atue sobre cada um deles. Para isso devemos indicar, atributo a atributo, os seguintes aspetos:

* Variação (S/N)
* História (S/N)
* Periodicidade (R/D/W/M/T…)



* **Dimensões conforme**

Quando se lida com mais do que uma área de suporte à decisão é possível que aconteça que mais do que uma tabela de factos incorpore uma mesma dimensão na sua estrutura.



* **Dimensões degeneradas**

É uma dimensão que não possui qualquer atributo que a caracterize para além daquele que integra a chave da tabela de factos.

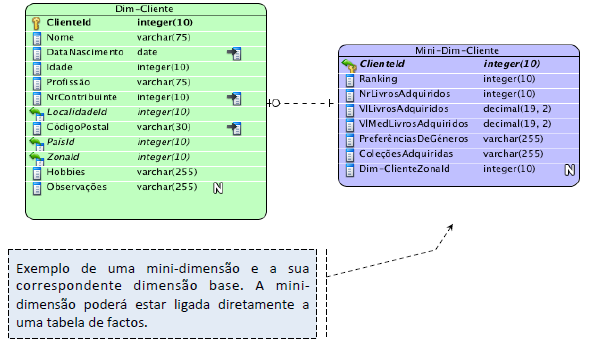
A sua existência num esquema dimensional é revelada apenas por um ou mais atributos integrados na estrutura de uma tabela de factos.

Uma imagem com mesa

Descrição gerada automaticamente

* **Mini dimensões**

Por vezes, na caracterização de uma dimensão integramos um número de atributos tal, que a sua estrutura atinge uma dimensão bastante significativa.

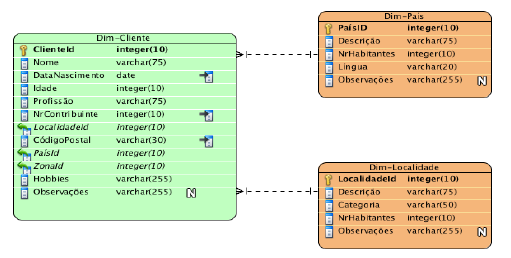


* **Dimensões gigantescas**
* **Sub dimensões (outriggers)**

As sub-dimensões, também designadas por outriggers, são tabelas definidas especificamente para apoiarem a caracterização de uma dimensão designada como principal que num esquema dimensional está ligada diretamente a pelo menos uma tabela de factos.

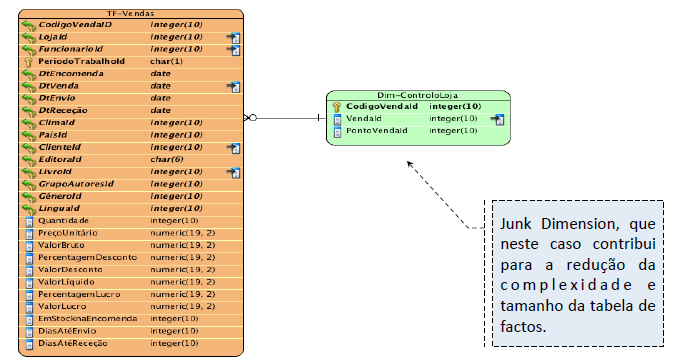
Uma sub-dimensao não desempenha um papel de primeiro plano nos processos de exploração de dados realizados sobre o esquema dimensional em que está localizada.

Em geral, são tabelas criadas **para reduzirem o nível de redundância existente na dimensão.**



* **Dimensões com diferentes papeis**

Uma dimensão com diferentes papeis é uma tabela de dimensão regular, mas que está referenciada mais do que uma vez numa mesma tabela de factos. Cada uma dessas referências dá sentido a um contexto de aplicação diferente.



* **Dimensões de controlo**

Por vezes, é necessário atuar no sentido de reduzir o tamanho de uma tabela de factos, porque:

* Pode estar a provocar algum tipo de estrangulamento no desempenho do sistema na satisfação das queries.
* Queremos simplificar os processos de monitorização e controlo desenvolvidos sobre as estruturas de dados implementadas
* Precisamos de assegurar contextos transacionais associados com uma dada de tabela de factos;
* Se quer obter um pouco mais de comodidade nos processos de manipulação de dados sobre o esquema dimensional em causa.

Uma imagem com mesa

Descrição gerada automaticamenteUma dimensão de controlo é, pois, uma dimensão abstrata especialmente concebida para acolher dados que não têm diretamente a ver com as atividades de análise.

* **Dimensões de origem externa**

As dimensões de origem externa são assim designadas por materializar uma vertente de análise de negócio específica, que é suportada exclusivamente por um conjunto de atributos cujos valores são povoados a partir de fontes de informações externas à empresa.

