

RAFAEL PITON

Glossário
— de —
**BUSINESS
INTELLIGENCE**

64 termos que um bom
profissional de BI deve saber





INTRODUÇÃO

O Glossário de Business Intelligence tem 64 termos com as suas definições.

Tudo de forma simples e direta, para você poder consultar e tirar suas dúvidas.

Ao longo do e-book você vai encontrar links em azul para acessar mais conteúdo.
E no topo das páginas você vai encontrar links para acessar as redes sociais, onde pode encontrar mais conteúdos como esse.

Como usar?

O glossário pode ser usado tanto como um ponto de partida para quem está começando como um material de consulta para quem já atua na área.

As definições começam em conceitos básicos e vão até os mais avançados.

Você pode usar ele da forma que melhor se adequar ao seu nível de conhecimento:

- Ler todo o glossário para começar a aprender
- Conhecer novos termos
- Descobrir o que você ainda não conhece e precisa estudar
- Consultar sempre que cruzar com alguma palavra nova



ÍNDICE

BI / Business Intelligence / Inteligência de Negócio	5	Data Integration	8
Dashboard	5	ETL	8
Métrica	5	Stage Area	8
KPI	5	Data Analysis (OLAP)	8
Dado	5	Data Visualization	8
Informação	6	Surrogate key	8
Conhecimento	6	Primary Key	9
Sabedoria	6	Natural key	9
Data Source	6	SQL	9
Data Warehouse	6	Reporting	9
Data Mart	6	Cubo	9
Big Data	7	Analysis OLAP	9
Data Mining / Mineração de Dados	7	Ad-hoc	10
Web Analytics	7	Pivot	10
Agile Data Warehouse Design	7	Slice e Dice	10
Grão	7	Drill up / Roll-up	10
Granularidade	7	Drill down	10



Drill through / perfurar os dados	10
Dados normalizados	10
Dados desnormalizados	11
Modelagem dimensional	11
Schema	11
Star Schema / Modelo Estrela	11
Snowflake Schema / Modelo Floco de Neve	11
Fact Constellation Schema / Modelo Constelação de Fatos	11
Tabela Fato	12
Tabela Dimensão	12
Atributo	12
Métrica Aditiva	12
Métrica Derivada	12
Métrica Semi-aditiva	13
Métrica Não-aditiva	13
Fato Transacional	13
Fato Agregada	13

Fato Consolidada	13
Fato Snapshot Periódico	14
Fato Snapshot Acumulado	14
Fato Sem Fato / Factless Fact Tables	14
Dimensão Hierárquica	14
Dimensão Degenerada / Degenerate Dimension	15
Static Dimension	15
Slowly Changing Dimension	15
Dimensão Não Qualificada	15
Role-Playing Dimension	16
Conformed Dimension	16
Junk Dimension	16
BUS Matrix	16



GLOSSÁRIO

BI / Business Intelligence / Inteligência de Negócio

Um conjunto de técnicas e conceitos. Esse processo passa por coleta, organização e análise dos dados, elaboração de relatórios ou dashboards e todo o acompanhamento, possibilitando a tomada de decisão inteligente utilizando dados passados.

IMPORTANTE: BI é conceito e não ferramenta.

Veja mais em: [O que é BI?](#)

Dashboard

Uma das ferramentas de visualização de dados. É um painel que apresenta visualmente as informações mais importantes e necessárias para a tomada de decisão.

Métrica

Tudo que a empresa for mensurar é uma métrica. Elas são utilizadas para metrificar algo e são sempre números, porque precisam ser contáveis. Esses números são provenientes de transações da empresa.

Veja mais em: [Tipos de métricas](#)

KPI / Key Performance Indicator / Indicador-chave de Desempenho

Índice para medir percentualmente as variações que ocorrem na empresa.

IMPORTANTE: é diferente de métrica.

Dado

Uma pequena parte sem contexto da informação, que sozinho não faz sentido.



Informação

Dados agrupados, organizados e lapidados com o objetivo de tomar decisões inteligentes.

Conhecimento

Interpretação das informações geradas que possibilita o aprendizado com elas.

Sabedoria

Toma decisões com base no leque de conhecimento.

Data Source

São as planilhas, ERPs, CRMs, etc, locais de onde os dados são retirados para serem inseridos no Data Warehouse. Os data source são geralmente compostos por dados estruturados ou semiestruturados, onde não se pode ter redundância, e são modelados para a inserção e edição dos dados, não para a consulta.

Veja mais em: [Business Intelligence Framework](#)

Data Warehouse

Um repositório de dados. É quem centraliza os dados da empresa e elimina os ruídos de comunicação entre os departamentos, deixando tudo unificado. O objetivo principal do Data Warehouse é garantir a entrega de dados confiáveis para dar suporte à tomada de decisão.

Veja mais em: [O que é um Data Warehouse?](#)

Data Mart

Uma parte do Data Warehouse que abrange determinado assunto ou departamento.



Big Data

Uma grande quantidade de dados não estruturados, como dados de redes sociais, web logs e dados de texto. No BI ele entra como mais uma fonte de dados, que precisa passar pelo processo de transformação e ser armazenados no Data Warehouse para ser analisado.

Veja mais em: [BI vs. Big Data](#)

Data Mining / Mineração de Dados

Enquanto o BI supre as necessidades já conhecidas do negócio, o Data Mining busca por informações que não estão sendo monitoradas ainda, percorrendo os dados em busca de padrões e anomalias.

Web Analytics

Coleta de dados de navegação e interação e análise sobre a experiência de usuários em um site com o objetivo de otimizar seu uso.

Agile Data Warehouse Design

Uma metodologia de desenvolvimento ágil para criação de Data Warehouse. A ideia é diminuir as horas de reuniões improdutivas e focar nos requisitos que realmente precisam ser levantados para desenhar o Data Warehouse.

Veja mais em:
[Passo a passo para modelar um Data Warehouse](#)

Grão

O menor nível da hierarquia da dimensão. É a informação base, o menor detalhe da informação.

Granularidade

O nível de detalhamento do dado. Alta granularidade é o grão, o menor nível do dado.



Data Integration

Estágio onde acontece o ETL / EL-T. É a etapa onde os dados são retirados das fontes de origem, transformados de forma que façam sentido juntos e inseridos no Data Warehouse.

ETL / Extract, Transform, Load / Extrair, Transformar, Carregar

Processo de extrair os dados do data source, transformar e carregar no Data Warehouse ou Stage Area.

Stage Area

Uma área temporária que geralmente está em um banco de dados relacional e fica desacoplado da origem. Ela tem tabelas soltas e sem relacionamento, onde os dados são transformados para serem enviados ao Data Warehouse.

Veja mais em: [Business Intelligence Framework](#)

Data Analysis (OLAP)

Onde entram os cubos e as análises começam a ser feitas, identificando padrões e fazendo projeções. Tudo em tempo de execução. Serve para analisar e identificar padrões passados e problemas futuros.

Veja mais em: [Business Intelligence Framework](#)

Data Visualization

Etapa onde a informação é apresentada, com dashboards, gráficos e reports.

Veja mais em: [Business Intelligence Framework](#)

Surrogate key

A Surrogate Key nada mais é que o campo de Primary Key da dimensão, uma chave artificial e auto incremental. É a chave utilizada para conectar a dimensão na fato.

Veja mais em: [Para que serve a Surrogate Key?](#)



Primary Key

Coluna utilizada para identificar cada linha da tabela de forma única. Garante a integridade dos dados.

Natural key

A Natural Key é uma chave utilizada na dimensão que serve para identificar o código daquele dado na origem, de onde ele foi retirado.

SQL / Structured Query Language / Linguagem de Consulta Estruturada

Linguagem utilizada para realizar consultas em banco de dados relacional.

Reporting

Transforma uma variedade de fontes de dados em informações compreensíveis que possam ser apresentadas em diversos formatos.

Cubo

Serve para manipular e analisar um grande volume de dados sob múltiplas perspectivas e hipóteses. Os Cubos permitem filtrar, cortar e pivotar os dados em tempo real, como em uma tabela dinâmica.

IMPORTANTE: cubo é conceito.

Analysis OLAP

São formatos de cubos.

OLAP = Online analytical processing / Processamento analítico em tempo real (online)

IMPORTANTE: OLAP é a tecnologia utilizada no cubo.



Ad-hoc

Quando o usuário final tem acesso ao cubo e pode montar suas próprias análises. Normalmente o usuário de negócio recebe uma tela em branco com o cubo e monta suas análises selecionando as tabelas que precisa, também conhecido como Self-service BI e Data Discovery.

Pivot

Cruzar os dados no cubo, explorando combinações e possibilitando novas perspectivas para análise.

Slice e Dice

Operações que criam um novo sub-cubo a partir de uma seleção de dados do cubo original.

Drill up / Roll-up

Quando você sobe no nível de hierarquia do dado, diminuindo a granularidade e o nível de detalhamento.

Drill down

Quando você desce no nível de hierarquia do dado, aumentando a granularidade e o nível de detalhamento.

Drill through / perfurar os dados

Ao invés de mover verticalmente, como drill down e drill up, o drill through se move horizontalmente, passando de um relatório para outro enquanto analisa a mesma amostra de dados.

Dados normalizados

Dados que estão na terceira forma normal, utilizada em banco de dados relacional. Eles permitem um armazenamento consistente e acesso eficiente aos dados. Esse estilo de banco tem como foco inserir, alterar e deletar os dados.



Dados desnormalizados

Dados organizados de forma a melhorar o desempenho das consultas. Não garante a consistências dos dados e deixa o banco maior que com os dados normalizados.

Modelagem dimensional

Uma forma de modelagem de dados que busca simplificar o banco de dados e tornar as consultas mais rápidas para sistemas de apoio à decisão.

Veja mais em: [O que é modelagem dimensional?](#)

Schema

É uma descrição lógica de todo o banco de dados.

Star Schema / Modelo Estrela

O modelo estrela é composto no centro por uma tabela fato que é rodeada por dimensões. E por isso tem o nome de Star Schema, porque parece uma estrela.

Veja mais em: [O que é Star Schema?](#)

Snowflake Schema / Modelo Floco de Neve

O modelo Snowflake também tem uma tabela fato rodeada por dimensões, mas segue o princípio de normalizar as dimensões, removendo atributos de baixa cardinalidade e criando tabelas separadas.

Fact Constellation Schema / Modelo Constelação de Fatos

Modelo com múltiplas tabelas fato que compartilham dimensões, também conhecido como Galaxy Schema.



Tabela Fato

A principal tabela do Data Warehouse, ela fica no centro do Star Schema e é rodeada por dimensões. A tabela fato armazena o que ocorreu, é o fato propriamente dito.

A fato armazena 2 coisas:

- As métricas
- As chaves das dimensões

Veja mais em: [Tipos de fatos](#)

Tabela Dimensão

Descreve o fato ocorrido, ela contém as características do evento. Ela vai qualificar, classificar ou descrever as métricas que estão na fato.

A dimensão armazena 3 coisas:

- A Surrogate Key
- A Natural Key
- Os atributos

Veja mais em: [Tipos de dimensões](#)

Atributo

São os campos das dimensões e servem para qualificar as métricas, porque elas sozinhas na fato não fazem sentido.

Métrica Aditiva

São as métricas que permitem operações matemáticas como soma e subtração por todas as dimensões. Dentro da fato há diversas linhas, e as métricas aditivas devem poder somar todas elas. Ela tem que fazer um cruzamento completo e perfeito na linha da fato, então a métrica precisa fazer sentido com cada uma das dimensões sozinha.

Veja mais em: [Tipos de métricas](#)

Métrica Derivada

É uma métrica calculada. Esse cálculo é sempre em cima de métricas que já estão na fato, não no que está no legado

Veja mais em: [Tipos de métricas](#)



Métrica Semi-aditiva

Métricas que podem ser somadas por todas as dimensões exceto a tempo. Você só vai conseguir somar pela tempo se colocar um filtro que diga que só seja usado o último registro.

Veja mais em: [Tipos de métricas](#)

Métrica Não-aditiva

São métricas tipo percentual, algum cálculo feito em tempo de execução, e não podem ser somadas por nenhuma dimensão.

Veja mais em: [Tipos de métricas](#)

Fato Transacional

Fatos transacionais são as mais comuns. A maioria dos bilhões de linhas que temos no Data Warehouse são de tabelas fato transacionais. Elas geralmente utilizam métricas aditivas.

Veja mais em: [Tipos de fatos](#)

Fato Agregada

É uma fato que consolida informações de uma outra fato em um tempo diferente. Ela serve para quando você não quer analisar o tempo no nível do grão.

IMPORTANTE: essa fato tem a função de acelerar o desempenho das consultas.

Veja mais em: [Tipos de fatos](#)

Fato Consolidada

Serve para consolidar duas fatos, mas não é feito nenhum join com elas. No processamento do ETL, na hora de carregar a fato, você vai carregar uma, carregar a outra, e misturar as duas.

IMPORTANTE: o grão das fatos precisa ser o mesmo.

Veja mais em: [Tipos de fatos](#)



Fato Snapshot Periódico

Uma fato com fotografias de uma outra fato em determinado momento no tempo.

Ex: se a fato original for trabalhada no minuto, a fato snapshot periódico pode ser no dia.

Veja mais em: [Tipos de fatos](#)

Fato Snapshot Acumulado

Essa fato vai ser dividida em etapas de atualização (que não é update na fato). A fato snapshot acumulado também faz uma fotografia de outra fato, mas em mais de um momento.

Veja mais em: [Tipos de fatos](#)

Fato Sem Fato / Factless Fact Tables

É uma fato sem métricas. Serve para fazer uma intersecção de dimensões. Às vezes você precisa comparar ou cruzar algo somente entre duas dimensões e não tem uma métrica para fazer essas comparações.

IMPORTANTE: essa fato é a exceção, só é usada quando se precisa fazer uma intersecção entre as dimensões.

Veja mais em: [Tipos de fatos](#)

Dimensão Hierárquica

Uma forma hierárquica de organizar os dados nas dimensões. É comum existir apenas uma hierarquia por dimensão, mas não é erro ter duas na mesma. A contagem de níveis começa no grão, que é o nível zero e então vai subindo.

Veja mais em: [Tipos de dimensões](#)



Dimensão Degenerada / Degenerate Dimension

É a dimensão que não mereceu ser uma dimensão e foi inserida como coluna na fato.

Veja mais em: [Tipos de dimensões](#)

Static Dimension

Dimensões que não são extraídas do legado, mas criadas para o próprio Data Warehouse. Normalmente são carregadas manualmente ou geradas através de um processo. São dimensões como as de tempo, data ou código de status.

Veja mais em: [Tipos de dimensões](#)

Slowly Changing Dimension

É uma técnica para atualizar a dimensão. Todas as dimensões são SCD, porque elas vão precisar ser atualizadas para se manterem sincronizadas com o transacional. A única exceção são as Static Dimensions, que são chamadas de SCD tipo 0, porque depois que os dados forem inseridos, não é mais necessário atualizar.

Veja mais em: [Tipos de dimensões](#)

Dimensão Não Qualificada

Existem dois tipos de dimensões não qualificadas, a que não foi identificada e a que não se aplica.

Se for nulo: -1 Não identificado

Essa acontece quando você não encontra na dimensão a informação que vem lá do legado.

Se não deve cruzar: -2 Não se aplica

Essa acontece quando você não quer monitorar algum dado.

Veja mais em: [Tipos de dimensões](#)



Role-Playing Dimension

No Data Warehouse, as dimensões muitas vezes são utilizadas para múltiplos objetivos. Quando você quer fazer uma análise de vendas e nessa análise precisa mensurar quantidade vendida, você quer ver essa quantidade pela data do pedido, data do envio do pedido e data do recebimento do pedido.

Quando acontece essa situação, nós precisamos ter as Surrogate Keys na fato para que você possa analisar de forma separada.

Veja mais em: [Tipos de dimensões](#)

Conformed Dimension

É uma dimensão que tem o mesmo significado para todas as fatos com que se relaciona. Uma boa prática é deixar todas as dimensões conformadas, porque assim todo mundo fala a mesma língua.

Veja mais em: [Tipos de dimensões](#)

Junk Dimension

Essas dimensões têm basicamente 3 colunas, a Surrogate Key, a Natural Key e uma descrição. Ela é usada para quando você tem diversas tabelas semelhantes e cada uma com apenas uma coluna. Então você utiliza uma junk dimension para agrupar elas em uma só.

IMPORTANTE: essa solução é apenas para quando não tem onde colocar as coisas.

Veja mais em: [Tipos de dimensões](#)

BUS Matrix

Uma tabela com a relação completa de todas as fatos e dimensões e como elas se relacionam entre si, garantindo uma visão única de todas as informações da empresa. Proporciona uma visão geral do Data Warehouse e ajuda a manter a conformidade entre as dimensões.



Rafael Piton

Meu objetivo é eliminar as barreiras de acesso ao conhecimento no mundo de BI

CEO & Fundador da Raizzer, criador do BI Academy e idealizador do BI & Analytics Summit.

Arquiteto de BI & Big Data, especialista em tomada de decisões orientadas a dados e um dos principais expoentes da revolução do Agile BI & Analytics no país.

Utiliza suas mais de 35.000 horas de experiência em projetos de BI para dar suporte a quem entra nessa área hoje e educar o mercado sobre a importância da análise de dados.