




























# Pipeline Dados

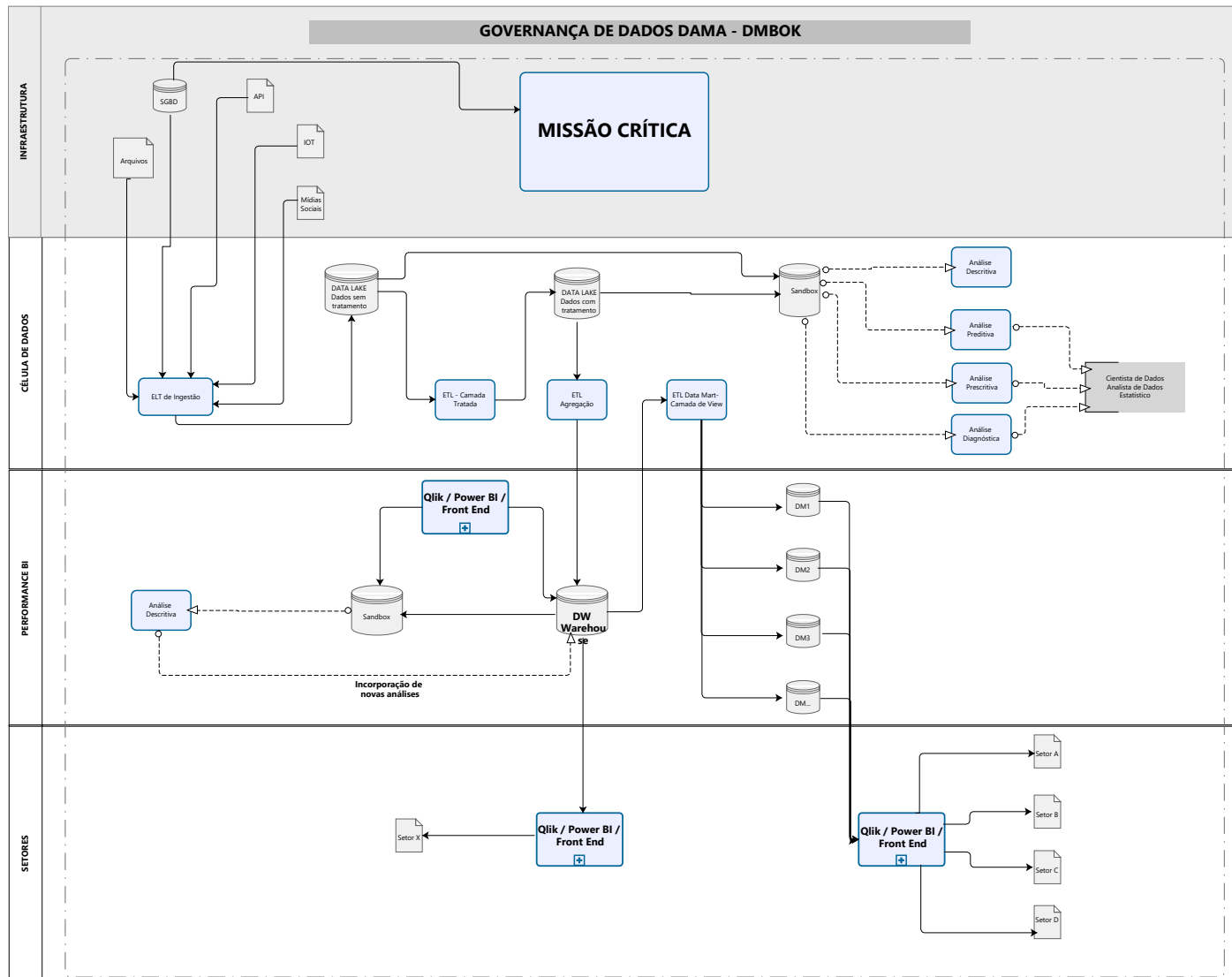
Bizagi Modeler

## Índice

PIPELINE DADOS .....	1
BIZAGI MODELER .....	1
1 ESTRUTURA DE DADOS ANALÍTICA .....	4
1.1 CÉLULA DE DADOS .....	6
1.1.1 Elementos do processo .....	6
1.1.1.1  DATA LAKE Dados sem tratamento .....	6
1.1.1.2  DATA LAKE Dados com tratamento .....	7
1.1.1.3  Sandbox .....	7
1.1.1.4  ELT de Ingestão .....	7
1.1.1.5  ETL Data Mart- Camada de View .....	7
1.1.1.6  Análise Preditiva .....	8
1.1.1.7  Análise Descritiva .....	8
1.1.1.8  ETL - Camada Tratada .....	8
1.1.1.9  ETL Agregação .....	8
1.2 INFRAESTRUTURA .....	8
1.2.1 Elementos do processo .....	9
1.2.1.1  Arquivos .....	9
1.2.1.2  IOT .....	9
1.2.1.3  SGBD .....	9
1.2.1.4  MISSÃO CRÍTICA .....	9
1.3 PERFORMANCE BI .....	9
1.3.1 Elementos do processo .....	10
1.3.1.1  Sandbox .....	10
1.3.1.2  DW Warehouse .....	10
1.3.1.3  DM... .....	10
1.3.1.4  DM3 .....	10
1.3.1.5  DM2 .....	10
1.3.1.6  DM1 .....	11
1.3.1.7  Análise Descritiva .....	11
1.4 SETORES .....	11
1.4.1 Elementos do processo .....	11
1.4.1.1  Setor X .....	11
1.4.1.2  Setor D .....	11
1.4.1.3  Setor B .....	12

1.4.1.4	 Setor C.....	12
1.4.1.5	 Setor A.....	12
1.4.1.6	 Qlik / Power BI / Front End .....	12
1.4.1.7	 Qlik / Power BI / Front End .....	12
2	RECURSOS .....	13
2.1	RONALDO COSTA (FUNÇÃO) .....	13
2.2	CHRISTIAN ROCHA (FUNÇÃO) .....	13

# 1 ESTRUTURA DE DADOS ANALÍTICA



**Versão:**

1.0

**Autor:**

ronaldo

## 1.1 CÉLULA DE DADOS

### Descrição

Novo setor que terá como finalidade estruturar o novo modelo de dados proposto, bem como efetuar análises dos dados.

A Matriz RACI explica com mais detalhes o setor e suas atribuições.

### Executantes

Christian Rocha

### Responsável

Ronaldo Costa

---

### 1.1.1 ELEMENTOS DO PROCESSO

#### 1.1.1.1 DATA LAKE Dados sem tratamento

### Descrição

Camada de armazenamento dos dados brutos. Aqui é onde todos os dados que estão chegando via ingestão (streaming, Batch) vão ser armazenados no primeiro momento. A ideia aqui é trazer o dado no estado original dele, exemplo, se você está capturando dado do MongoDB, traga o json, se está capturando de um outro HDFS, traga o parquet, orc, avro, etc. Sem tratamento, o objetivo é capturar o mais rápido possível.

#### PRINCIPAIS CARACTERÍSTICAS DO DATA LAKE:

1. Ingestão em alta velocidade com grandes volumes de dados;
2. Possibilidade do armazenamento de diversas estruturas de dados (Estruturados, semi-estruturados e não-estruturados);
3. Centralizar todos os dados da organização em um único local;
4. Democratização dos dados para a empresa;
5. Baixo custo de armazenamento. (Isso não pode ser considerado uma regra);
6. Desacoplamento entre o armazenamento e o processamento da produção. Isso faz muito sentido quando falamos em escalabilidade e performance;
7. Dependendo da arquitetura e tecnologia adotada, possibilita a centralidade do consumo em batch e near-real time, onde todos os consumers podem beber da mesma água;

8. Criação de plataforma completa de produção para aplicações que exigem processamento em pipelines de dados, como Machine Learning (ML), Inteligência Artificial (IA), BI e analytics em geral;

#### 1.1.1.2 DATA LAKE Dados com tratamento

##### Descrição

Zone (Stage): A camada de stage ou trusted é utilizada como tratamento e enriquecimento dos dados, aplicando governança, padronizações de data types, mascaras e limpeza de dados.

#### 1.1.1.3 Sandbox

##### Descrição

Um Analytics Sandbox é um ambiente separado que faz parte da arquitetura geral do data lake, o que significa que é um ambiente centralizado destinado a ser usado por vários usuários e mantido com o suporte de TI.

O ambiente é controlado pela Célula de Dados.

Permite que eles instalem e usem as ferramentas de dados de sua escolha.

Permite que eles gerenciem a programação e processamento dos ativos de dados.

Permite que os analistas explorem e experimentem dados internos e externos.

Pode conter e processar grandes quantidades de dados com eficiência de muitas fontes de dados diferentes; big data (não estruturado), dados transacionais (estruturados), dados da web, dados de mídia social, documentos, etc.

#### 1.1.1.4 ELT de Ingestão

##### Descrição

ELT que faz a leitura dos dados dos sistema de produção e ingestão no Data Lake.

Os dados não sofrem nenhum tratamento nesta fase.

##### Responsável

Ronaldo Costa

#### 1.1.1.5 ETL Data Mart- Camada de View

##### Descrição

Faz parte do Pipeline de dados.

Principal função é a segmentação de dados por setor, dividindo o grande volume de dados corporativo em dados menores orientados ao atendimento setorial.

#### 1.1.1.6 Análise Preditiva

##### Descrição

Análise preditiva é a utilização de modelos estatísticos para apontar como serão os dados no futuro, ou como se comportam em condições diversas. Utiliza-se para tal modelos de machine learning e inteligência artificial, que vão desde modelos clássicos de regressão até modelos e algoritmos complexos que envolvem deep learning, métodos de redução de dimensionalidade, ensemble learning e swarm intelligence

#### 1.1.1.7 Análise Descritiva

##### Descrição

A análise descritiva é um dos instrumentos imprescindíveis quando se busca retratar comportamentos e tendências a partir de fenômenos reais.

Ela vem a ser a base sobre a qual os métodos estatísticos se aplicam, o que faz dela uma prática fundamental no universo dos negócios.

Sua utilização em favor de seus objetivos exige dominar não só o conceito como as técnicas, ferramentas e os processos relacionados a esse tipo de avaliação.

#### 1.1.1.8 ETL - Camada Tratada

##### Descrição

Faz parte do Pipeline de dados.

Processamento que trata as informações, efetuando enriquecimento, agregação, slice e outros processamentos de ETL.

#### 1.1.1.9 ETL Agregação

##### Descrição

Faz parte do Pipeline de dados.

Processamento que trata as informações, efetuando enriquecimento, agregação, slice e outros processamentos de ETL.

## 1.2 INFRAESTRUTURA

##### Descrição

Área de Infraestrutura



---

## 1.2.1 ELEMENTOS DO PROCESSO

### 1.2.1.1 Arquivos

#### Descrição

Arquivos utilizados para complemento ou enriquecimento de dados

### 1.2.1.2 IOT

#### Descrição

Internet das Coisas (IOT)

### 1.2.1.3 SGBD

#### Descrição

Um Sistema de Gerenciamento de Banco de Dados (SGBD) — do inglês Data Base Management System (DBMS) — é o sistema de software responsável pelo gerenciamento de um ou mais bancos de dados. Seu principal objetivo é retirar da aplicação cliente a responsabilidade de gerenciar o acesso, a persistência, a manipulação e a organização dos dados.

### 1.2.1.4 MISSÃO CRÍTICA

#### Descrição

Sistemas de missão crítica da empresa, a saber:

CM

OPERA

Multiclubes

Multivendas

#### Executantes

Christian Rocha

#### Responsável

Ronaldo Costa

## 1.3 PERFORMANCE BI

#### Descrição

Setor que efetua o desenvolvimento de Dashboard a partir do Data Warehouse.

---

## 1.3.1 ELEMENTOS DO PROCESSO

### 1.3.1.1 Sandbox

#### Descrição

Um Analytics Sandbox é um ambiente separado que faz parte da arquitetura geral do data lake, o que significa que é um ambiente centralizado destinado a ser usado por vários usuários e mantido com o suporte de TI.

O ambiente é controlado pela Célula de Dados.

Permite que eles instalem e usem as ferramentas de dados de sua escolha.

Permite que eles gerenciem a programação e processamento dos ativos de dados.

Permite que os analistas explorem e experimentem dados internos e externos.

Pode conter e processar grandes quantidades de dados com eficiência de muitas fontes de dados diferentes; big data (não estruturado), dados transacionais (estruturados), dados da web, dados de mídia social, documentos, etc.

### 1.3.1.2 DW Warehouse

#### Descrição

Zone (Refined): A cada Refined ou especializada é onde encontramos o dado já tratado, enriquecido e possivelmente onde as aplicações irão consumir. Essa camada "normalmente" é construída sob infraestrutura com banco de dados relacionais (SQL Server, Oracle, etc), devido a sua característica de consumo, que normalmente é feita via API's e sistemas transacionais.

### 1.3.1.3 DM...

#### Descrição

Idem ao DM1

### 1.3.1.4 DM3

#### Descrição

Idem ao DM1

### 1.3.1.5 DM2

## Descrição

Idem ao DM1

### 1.3.1.6 DM1

## Descrição

Data mart (repositório de dados) é sub-conjunto de dados de um Data warehouse (ou DW, armazém de dados). Geralmente são dados referentes a um assunto em especial (ex: Vendas, Estoque, Controladoria) ou diferentes níveis de sumarização (ex: Vendas Anual, Vendas Mensal, Vendas), que focalizam uma ou mais áreas específicas.[1] Seus dados são obtidos do DW, desnormalizados e indexados para suportar intensa pesquisa. Data marts extraem e ajustam porções de DWs aos requisitos específicos de grupos/departamentos.

### 1.3.1.7 Análise Descritiva

## Descrição

A análise descritiva é um dos instrumentos imprescindíveis quando se busca retratar comportamentos e tendências a partir de fenômenos reais.

Ela vem a ser a base sobre a qual os métodos estatísticos se aplicam, o que faz dela uma prática fundamental no universo dos negócios.

Sua utilização em favor de seus objetivos exige dominar não só o conceito como as técnicas, ferramentas e os processos relacionados a esse tipo de avaliação.

## 1.4 SETORES

---

### 1.4.1 ELEMENTOS DO PROCESSO

#### 1.4.1.1 Setor X

## Descrição

Setor da empresa

#### 1.4.1.2 Setor D

## Descrição

Setor da empresa

#### 1.4.1.3 Setor B

##### Descrição

Setor da empresa

#### 1.4.1.4 Setor C

##### Descrição

Setor da empresa

#### 1.4.1.5 Setor A

##### Descrição

Setor da empresa

#### 1.4.1.6 Qlik / Power BI / Front End

##### Descrição

O objetivo dos Sistemas de Análise e desenvolvimento de Dashboard é fornecer visualizações interativas e recursos de business intelligence com uma interface simples para que os analista de negócio e/ou usuário final criem os seus próprios relatórios e dashboards. Algumas das áreas onde esses sistemas operam são:

Finanças, Engenharia, Tecnologias de Informação, Marketing e Saúde.

Algumas vantagens desses sistemas relativamente às outras ferramentas de visualização são:

capacidade de carregar visualizações personalizadas através do AppSource,

acesso à informação em tempo real,

acesso de informação em qualquer aplicativo móvel,

uso empresarial, atualização mensal com novas funções e facilidade de uso.

#### 1.4.1.7 Qlik / Power BI / Front End

##### Descrição

Idem ao item Qlik / Power BI

## 2 RECURSOS

### 2.1 RONALDO COSTA (FUNÇÃO)

Descrição

Célula de Dados

### 2.2 CHRISTIAN ROCHA (FUNÇÃO)