



# BlueShift

**Business Blueprint**

**Caieiras, 14 de julho de 2022**

CONTROLE DE VERSÃO			
Autor	Versão	Data	Descrição
Felipe Gustavo de Abreu	1.0	13/07/2022	Criação do Documento.

## **1. Introdução**

Este documento visa detalhar as necessidades do cliente Pocco Pamonhas, do ponto de vista técnico referente ao procedimento de estruturação do seu Data Warehouse, bem como a migração dos fluxos de trabalho e o armazenamento dos dados e a visualização deles através de um Dashboard.

## **2. Solicitação**

O Cliente Pocco Pamonhas deseja uma solução para extração e processamento de dados para geração de relatório com indicadores de desempenho, e para isso eles necessitam de uma nova estruturação em sua pipeline de dados. A entrega do projeto deverá ser dividida em três partes. A primeira consistirá na extração dos dados do Blob Storage via Databricks para o Azure SQL database. A segunda parte consistirá na construção de um Data Warehouse, via Stored Procedure dentro do Azure Data Factory. A terceira e última parte consistirá na cópia dos dados do Azure SQL database para um Dashboard construído no Power BI.

## **3. Premissas da Solução**

A seção abaixo apresenta as premissas da solução

Origem e especificação dos dados

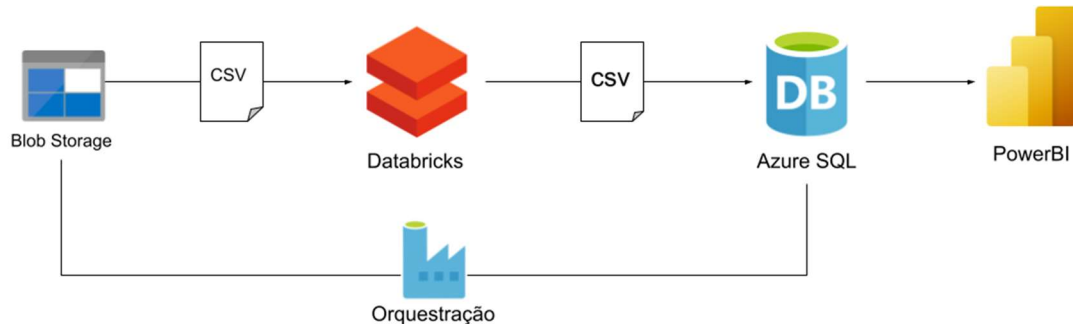
- Os dados estão sendo disponibilizados por um arquivo CSV armazenado em um Blob Storage no Azure.

Ambiente de desenvolvimento

- O cliente deverá disponibilizar acessos aos ambientes de desenvolvimento em todas as ferramentas específicas na arquitetura proposta neste documento para a Blueshift Brasil.

#### 4. Modelo da arquitetura sugerida

A imagem abaixo apresenta a arquitetura da solução proposta com base no levantamento de requisitos e entendimento do negócio.



#### 5. Extração dos dados do Blob Storage

Será desenvolvido uma Pipeline no Azure Data Factory, e nele será inserido primeiramente um Notebook do Databricks, na qual será coletado os dados das ordens de vendas do arquivo CSV armazenado no Blob Storage, e a realização do ETL contendo a conversão e tratamento dos tipos dos dados, e após o consumo dos dados deste arquivo, essas informações serão inseridas no banco de dados criado no Azure SQL.

#### 6. Modelagem dos Dados

Para este projeto estamos considerando uma tabela de stage para armazenar os dados extraídos do arquivo, e a criação de um Data Warehouse composto por 4 tabelas, sendo 3 tabelas Dimensão e 1 tabela Fato, que irão receber os dados convertidos da tabela Stage através da Storage Procedure. A estrutura das tabelas listadas abaixo.

## - Tabela Stage:

CAMPO	TIPO
Region	VARCHAR(50)
Country	VARCHAR(50)
Item_Type	VARCHAR(50)
Sales_Channel	VARCHAR(50)
Order_Priority	VARCHAR(50)
Order_Date	Date
Order_ID	INT PRIMARY KEY
Ship_Date	DATE
Units_Sold	INT
Unit_Price	FLOAT
Unit_Cost	FLOAT
Total_Revenue	FLOAT
Total_Cost	FLOAT
Total_Profit	FLOAT

## - DW Tabela DIM Regiões:

CAMPO	TIPO
Id_Region	INT PRIMARY KEY IDENTITY
Region	VARCHAR(50)

## - DW Tabela DIM Países:

CAMPO	TIPO
Id_Country	INT PRIMARY KEY IDENTITY
Country	VARCHAR(50)

## - DW Tabela DIM Canais de venda:

CAMPO	TIPO
Id_Sales_Channel	INT PRIMARY KEY IDENTITY
Sales_Channel	VARCHAR(50)

- DW Tabela FATO Vendas:

CAMPO	TIPO
Region	VARCHAR(50)
Country	VARCHAR(50)
Item_Type	VARCHAR(50)
Sales_Channel	VARCHAR(50)
Order_Priority	VARCHAR(50)
Order_Date	DATE
Order_ID	INT PRIMARY KEY
Ship_Date	DATE
Units_Sold	INT
Unit_Price	FLOAT
Unit_Cost	FLOAT
Total_Revenue	FLOAT
Total_Cost	FLOAT
Total_Profit	FLOAT

## 7. Storage Procedure

Os dados extraídos do arquivo no Blob Storage foram inseridos no banco de dados na tabela Stage já com a conversão dos dados feita no Databricks. Para realizar inserção dos dados nas tabelas do DW, será criado uma “STORED PROCEDURE” no Azure SQL Database, e introduzindo-a no Data Factory após o primeiro Copy Data. A Storage Procedure irá popular as tabelas DW conforme mostrado acima.

## 8. Power BI

A última etapa do projeto, será criado um Dashboard em Power BI, e será adicionado após o Stored Procedure na pipeline, sendo este Dashboard construído com as bases de dados do DW, através de uma conexão com o Azure SQL Server. O respectivo Dashboard deverá conter informações e indicadores para a tomada de decisões, sendo alguns deles:

- O acumulado de vendas do último ano por Região e País. Ele gostaria de ter essa visão através de um Mapa Mundial diretamente no Relatório.
- Quantidade de vendas dos últimos 10 dias através de um gráfico de colunas.
- Quantidade de vendas e a Quantidade acumulada de vendas dos últimos 30 dias.
- Uma visão acumulada das vendas do último ano por Canal e País. De forma que seja possível ver a distribuição das vendas um determinado país por canal.

O relatório a ser desenvolvido no Power BI será composto pelas seguintes tabelas em um modelo dimensional “esquema em estrela”, apresentados abaixo.

- Dimensão com as regiões.
- Dimensão com país.
- Dimensão com Canais de venda.
- Dimensão Calendário.
- Fato com as vendas

Com isto, o dashboard a ser desenvolvido irá apresentar todas as visões que atendam as necessidades apresentadas pela empresa.