

Architecture Design Document

AMAZON SALES ANALYSIS

WRITTEN BY	HARSH DIXIT
DOCUMENT VERSION	1.0
LAST REVISED DATE	12/10/2024

Contents

1. Introduction

03

1. What is Architecture Design Document? 04

2. Scope

04

2. Power Bi Architecture

05

1. Tableau Architecture

06

2. Power Bi Service Architecture

06

3. Gateway/Load Balancer

07

4. Application Server

07

5. Power Bi Data Engine

08

6. Background Process

08

7. Data Server

09

8. Power Bi Community Flow

09

3. Deployment

10

1. Deployment Options in Power Bi

10

2. Single Node Architecture

10

3. Three Node Architecture

11

4. Conclusion

12

4. Five Node Architecture

11

1.Introduction

The purpose of this document is to describe the architecture and deployment of the Amazon Sales Analysis Dashboard using Power BI. This project aims to provide insights into the sales performance, customer demographics, and product trends on the Amazon marketplace.

Power BI is a powerful data visualization tool that allows business users to analyze large datasets through interactive dashboards and reports. This document outlines the architecture design, the components involved, communication flow, and deployment options for the solution.

1.1 What is an Architecture Design Document?

An Architecture Design Document (ADD) provides a detailed overview of the structure, components, and infrastructure of a software solution. It defines how various parts of the system are integrated, how data flows through the system, and what technologies are involved. The ADD is essential for understanding the technical approach of the project, how the system meets business requirements, and the key decision points for its deployment and scaling.

1.2 Scope

This document outlines the architecture for deploying the Amazon Sales Analysis Dashboard in Power BI. It covers:

- The key architectural components and services.*
- Data flow and communication between components.*
- Different deployment options (e.g., single-node or multi-node architecture).*
- Data security, performance optimization, and scalability aspects.*

2. Power BI Architecture

Power BI's architecture includes several layers and components designed to handle data ingestion, transformation, modeling, and visualization. For this Amazon Sales analysis, the key components are:

1.Data Sources:

- 1. Amazon sales data (CSV, API, database).*
- 2. Product, customer, and time-related data.*

2.Power Query:

- 1. Handles the ETL (Extract, Transform, Load) process for cleaning and transforming raw data before loading into the data model.*

3.Data Model:

- 1. Built using a star schema that includes fact tables (sales data) and dimension tables (product, customer, time).*

4.DAX (Data Analysis Expressions):

- 1. Used for calculating key metrics like total sales, profit, and time-based analysis (YoY growth, MoM growth).*

5.Visualizations:

- 1. Dashboards and reports containing interactive charts, KPIs, tables, and maps to visualize sales performance.*

6.Power BI Service:

- 1. Cloud service for hosting and sharing dashboards.*
- 2. Enables real-time collaboration, scheduled refreshes, and security controls.*

2.1 Tableau Architecture (Comparable Overview)

Before focusing on Power BI, it's important to understand that Tableau and Power BI follow similar architectural principles, though the components differ slightly.

Tableau's architecture consists of:

- **Tableau Desktop:** *The client-side tool for creating reports and dashboards.*
- **Tableau Server/Tableau Online:** *Server-based platforms for hosting, sharing, and managing Tableau dashboards.*
- **Tableau Data Sources:** *The connections Tableau makes to various data sources for analysis (databases, CSVs, APIs, etc.).*

2.2 Power BI Service Architecture

For Power BI, the architecture includes a service-based cloud infrastructure, where the key components include:

- **Power BI Desktop:** *The tool where the dashboards are created.*
- **Power BI Service:** *The cloud platform where dashboards are published and shared with end-users.*
- **Power BI Mobile:** *The mobile application for viewing reports on the go.*

2.3 Gateway/Load Balancer

In Power BI, the Data Gateway serves as the bridge between on-premise data sources and Power BI Service, allowing for scheduled refreshes. If the data resides on an on-premises server (e.g., a SQL database or a file system), the gateway ensures that data is securely transmitted to the cloud.

- Gateway:** Connects local data sources to Power BI for scheduled refresh.
- Load Balancer:** Power BI Service automatically manages traffic and report requests, balancing the load across multiple services for scalability.

2.4 Application Server

In the Power BI Service architecture, the Application Server handles all the backend operations such as user authentication, report rendering, and handling data requests from users.

- Authentication:** Supports integration with Azure Active Directory (AAD) for single sign-on and secure access.
- User Management:** Power BI Service provides user-level security and permissions to access datasets and reports.

2.5 Power BI Data Engine

*Power BI uses a highly optimized data engine known as **Vertipaq**, which compresses data and allows for in-memory processing. The **Vertipaq Engine** enables fast query responses and supports large datasets, making it ideal for analytical reports.*

- In-Memory Data Model:** Vertipaq stores data in a highly compressed format in memory, optimizing speed and performance.*
- Columnar Storage:** Data is stored in a columnar format, improving aggregation speed and enabling complex calculations.*

2.6 Background Processes

*In Power BI, **Background Services** are responsible for tasks like data refresh, email notifications, and dashboard updates.*

- Scheduled Data Refresh:** Power BI Service uses background processes to periodically refresh the data from connected sources (e.g., daily or weekly).*
- Alerting and Notifications:** Users can set up alerts based on specific KPIs or thresholds, and Power BI will send notifications when these conditions are met.*

2.7 Data Server

The Data Server within Power BI Service handles data storage and access for reports. Power BI offers storage options including:

- Direct Query: Live connection to the data source (e.g., SQL Server).*
- Import Mode: Data is imported and stored in Power BI's cloud for faster query performance.*

2.8 Power BI Communication Flow

The communication flow for Power BI involves several key components, especially when working with on-premise data:

- 1.Power BI Desktop: The user creates reports and dashboards using desktop software.*
- 2.Data Sources: Data is fetched from Amazon sales data (CSV, database, API) via Power Query.*
- 3.Data Gateway (if applicable): For on-premise data, the gateway securely connects to the data source.*
- 4.Power BI Service: Reports are published to the Power BI cloud, enabling users to view dashboards.*
- 5.End User: The end-users interact with the reports through a web browser or mobile app.*

3. Deployment

Deployment in Power BI is cloud-based and can be done through:

- **Power BI Service:** Hosted in the Microsoft cloud, it allows for sharing, collaboration, and real-time analytics.
- **On-Premise Power BI Report Server:** For organizations that need to keep all data and analytics on-premises.

Deployment Options in Power BI

3.1 Single Node Architecture

In a single-node architecture, all Power BI components (gateway, report server, etc.) are installed on a single machine. This is typically used for small-scale deployments or proof of concept (POC) setups.

- **Components on One Machine:** Power BI Desktop, Report Server, and Gateway are on a single server.
- **Best For:** Small organizations or initial testing environments.

3.2 Multi-Node Architecture (Three Node)

*A multi-node architecture improves performance and scalability by distributing components across multiple machines. In a **three-node architecture**, we have:*

- 1. **Node 1:** Handles Power BI Gateway and ETL processes.*
- 2. **Node 2:** Hosts Power BI Report Server.*
- 3. **Node 3:** Handles scheduled data refreshes and background jobs.*
- **Best For:** Medium to large organizations with increasing data loads.*

3.3 Five Node Architecture

For large-scale deployments, a five-node architecture distributes components even further to optimize load balancing and high availability.

- 1. **Node 1 & 2:** Two nodes handle Power BI Gateways for redundancy.*
- 2. **Node 3:** Power BI Report Server.*
- 3. **Node 4:** Data Engine (for DAX calculations and storage).*
- 4. **Node 5:** Backgrounder for scheduled refresh and alerts.*
- **Best For:** Enterprise-level deployments with high user concurrency and data volume.*

15. Conclusion

*The architecture design document for the **Amazon Sales Analysis Dashboard** built with Power BI demonstrates a scalable and secure data visualization solution. Depending on organizational requirements, the architecture can be deployed using various configurations, from single-node setups to multi-node enterprise-level environments. Through Power BI's rich set of features like data modeling, DAX calculations, and cloud service capabilities, this project will enable data-driven decision-making for sales performance and customer insights.*