Architecture Design

AMAZON SALES ANALYSIS

Written By	Harshit Soni , Aniee Sapra , Sujal Pandey , Harshit Nagila
Document Version	1.0
Last Revised Date	24/09/22

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	23- Sept - 2022	Harshit Nagila	Introduction and architecture defined
0.2	24 - Sept - 2022	Harshit Soni	Architecture & Architecture description appended and updated.
0.3	24- Sept - 2022	Aniee Sapra	Deployment of PowerBi

Reviews:

VERSION	DATE	REVIEWER	COMMENTS
0.1	24- sept 2022	Sujal Pandey	Unit test cases to be added

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY		APPROVED BY	COMMENTS

Contents

1. Introduction

1.1 What is Architecture design document?

Any software needs the architectural design to represents the design of software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures.

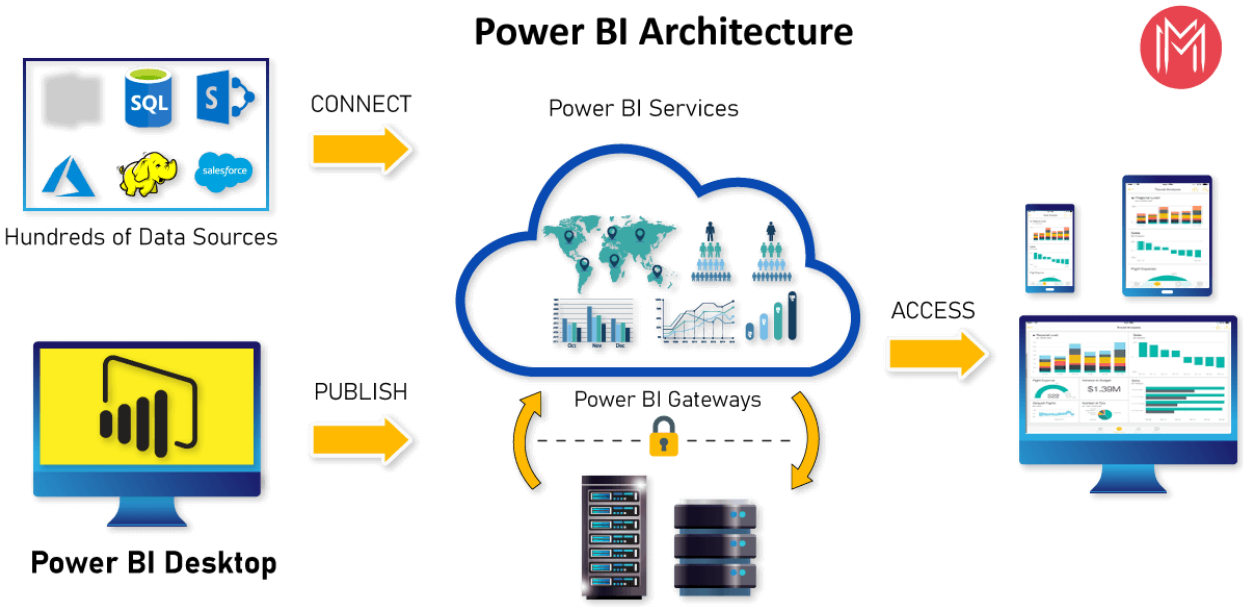
Each style will describe a system category that consists of :

- A set of components (eg: a database, computational modules) that will perform a function required by the system.
- The set of connectors will help in coordination, communication, and cooperation between the components.
- Conditions that how components can be integrated to form the system.
- Semantic models that help the designer to understand the overall properties of the system.

1.2 Scope

Architecture Design Document (ADD) is an architecture design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.

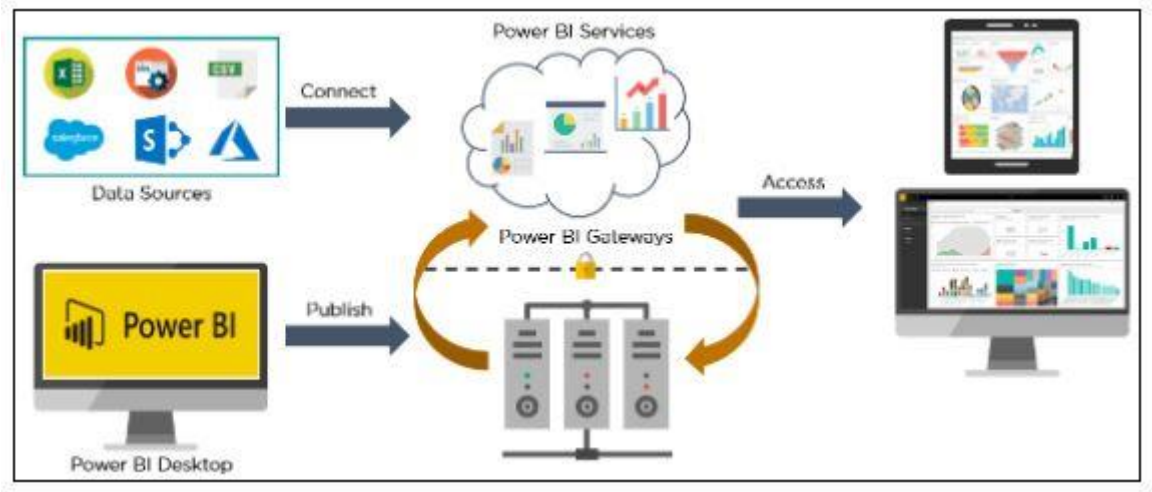
2. Architecture



PowerBi Server Architecture

Power BI is a business platform that includes several technologies to work together. It delivers outstanding business intelligence solutions. Power BI Architecture contains four steps.

The following diagram shows Powerbi Server's architecture:



Powerbi Server is internally managed by the multiple server processes.

Let us discuss these four steps giving insightful information about each one of them.

1. Data Integration
2. Data Transforming
3. Report & Publish
4. Creating and Dashboard

1. Data Integration:

Data is extracted from different sources which can be different servers or databases. The data from various sources can be in different types and formats. If you import the file into the Power BI, it compresses the data sets up to 1GB, and it uses a direct query if the compressed data sets exceed more than 1GB. Then the data is integrated into a standard format and stored at a place called a staging area. There are two choices for big data sets. They are as follows.

- Azure Analytics Services
- Power BI premium

2. Data Transforming:

Integrated data is not ready to visualize data because the data should be transformed. To transform the data, it should be cleaned or pre-processed. For example, redundant or missing values are removed

from the data sets. After data is pre-processed or cleaned, business rules are applied to transform the data. After processing the data, it is loaded into the data warehouse.

3. Report & Publish:

After sourcing and cleaning the data, you can create the reports. Reports are the visualization of the data in the form of slicers, graphs, and charts. Power BI offers a lot of custom visualization to create the reports. After creating reports, you can publish them to power bi services and also publish them to an on-premise power bi server.

4. Creating Dashboards:

You can create dashboards after publishing reports to Power BI services, by holding the individual elements. The visual retains the filter when the report is holding the individual elements to save the report. Pinning the live report page allows the dashboard users to interact with the visual by selecting slicers and filters.

Components of Power BI Architecture:

1. Data Sources

Power BI can supply information from different online sources and file types. Import the information into the Power BI or establish a live service to receive the information. If you import the file into the Power BI, it compresses the data sets up to 1GB and, uses a direct query if the compressed data sets exceed more than 1GB. Here is the list of Data Sources supported in Power BI.

Here is the list of Data Sources supported in Power BI.

- **File Types:** Power BI supports XML, txt/CSV, Excel, JSON, and Share point folder type files.
- **Database:** It supports SQL Server Analysis Services Database, SAP HANA Database, SQL Server Database, SAP Business Warehouse server, Access Database, Google BigQuery (Beta), Amazon Redshift, Snowflake, Impala, Oracle Database, IBM Informix database (Beta), Teradata Database, MySQL Database, IBM Netezza (Beta), Sybase Database, PostgreSQL Database.
- **Azure:** Azure SQL Data Warehouse, Azure Blob Storage, Azure Analysis Services database (Beta), Azure SQL Database, Azure Data Lake Store, Azure Table Storage, Azure HDInsight (HDFS), Azure Cosmos DB (Beta), Azure HDInsight Spark (Beta).

- **Online Services:** Power BI service, Dynamics 365 (online), Microsoft Exchange Online, Common Data Service (Beta), SharePoint Online List, Visual Studio Team Services (Beta), Dynamics 365 for Financials (Beta), Microsoft Azure Consumption Insights (Beta), Salesforce Objects, Salesforce Reports, Google Analytics, Dynamics 365 for Customer Insights (Beta), GitHub (Beta), appFigures (Beta), comScore Digital Analytix (Beta), Facebook, Kusto (Beta), Planview Enterprise (Beta), MailChimp (Beta), Mixpanel (Beta), QuickBooks Online, Projectplace (Beta).
- **Other Services:** Hadoop File (HDFS), Vertica (Beta), Web, OData Feed, SharePoint List, Microsoft Exchange, Active Directory, R Script, ODBC, Spark (Beta), Blank Query, OLE DB.

2. Power BI Desktop

It is free software that enables you to connect, transform and visualize the data on your desktop. You can connect to various data sources with the help of Power BI Desktop and combine the data into a data model. This data model allows you to create a collection of images and graphics that make you share the information within the organization as records. The majority of the users who work on Business Intelligence projects use Power BI Desktop to create and share their reports with others.

3. Power BI Service

Power BI Service is an On-Cloud service with a web-based platform and used to share and publish the reports made on Power BI Desktop. It collaborates the data with other users and creates dashboards. Power BI Service is also called “Power BI Workspace”, “Power BI Web Portal”, and “Power BI Site”. Power BI Service offers wonderful features like alerts and natural language Q&A.

It is available in three versions. They are as follows:

- Premium version
- Pro version

- Free version

4. Power BI Report Server

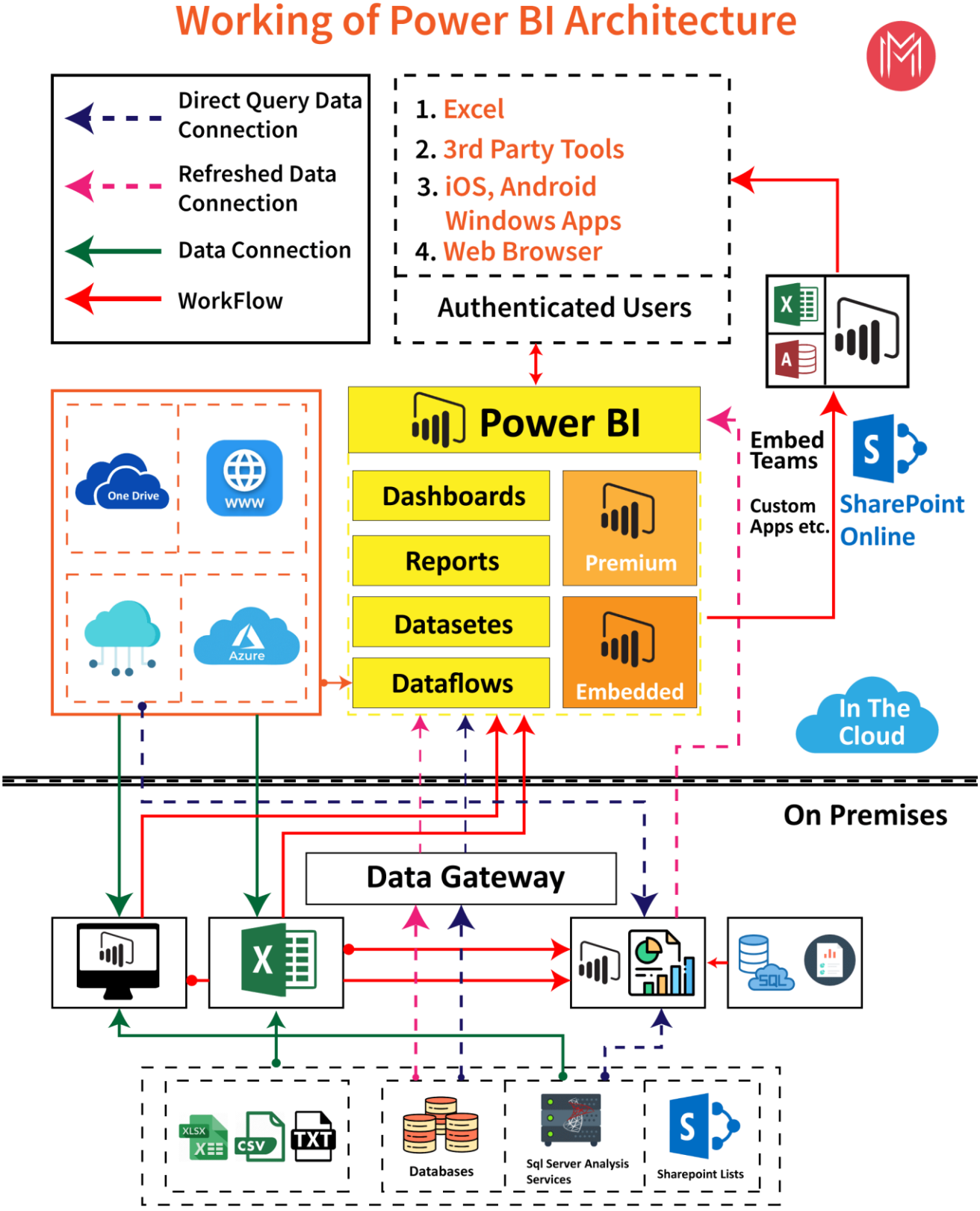
Power BI Report Server is similar to the Power BI Service. It is an On-Premises server platform. Using Power BI Report Server, organizations can secure their data. It enables the users to create reports and dashboards and allows you to share the reports with other users or organizations with proper security protocols. To use this service, you need to have a Power BI premium license.

5. Power BI Gateway

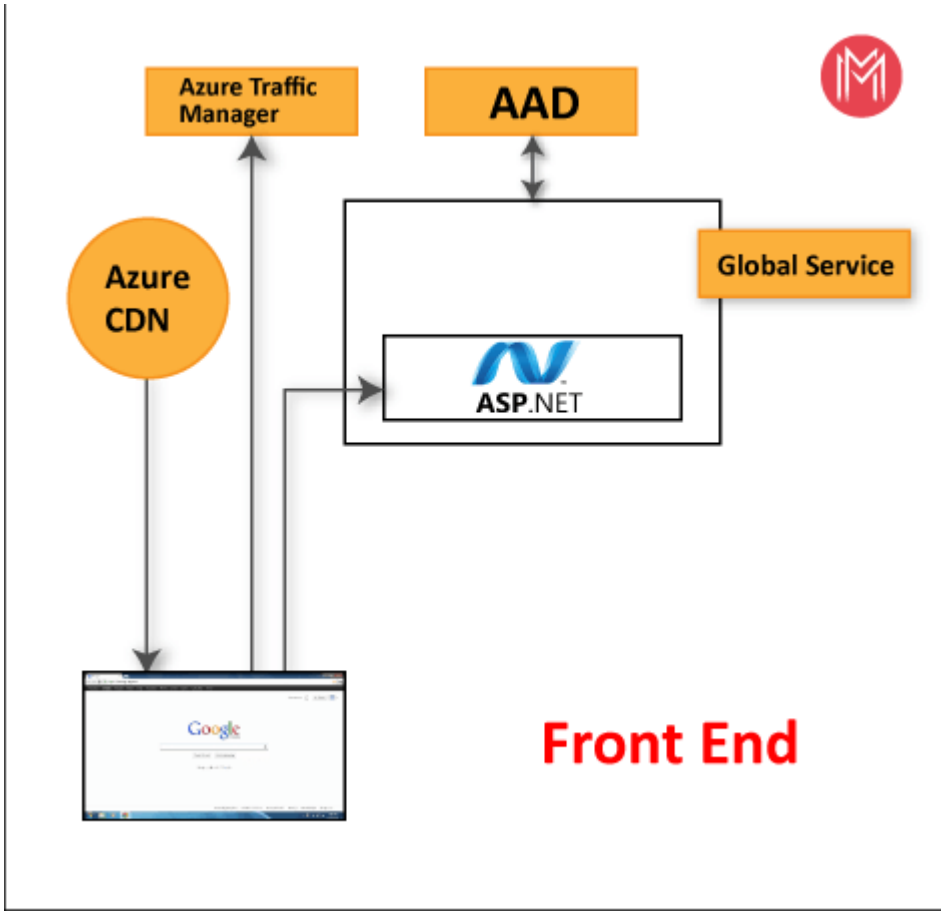
Power BI Gateway is used to maintain fresh information by connecting to your on-site data sources without transferring the data. It provides secure data and allows you to transfer the data between Microsoft cloud services and on-premise services. Microsoft cloud services include PowerApps, Power BI, Azure Analysis Services, Microsoft Flow, and Azure logic apps. By using a gateway, organizations can maintain the databases and other data sources securely in cloud services.

3. Deployment Description:

Working of Power BI Architecture

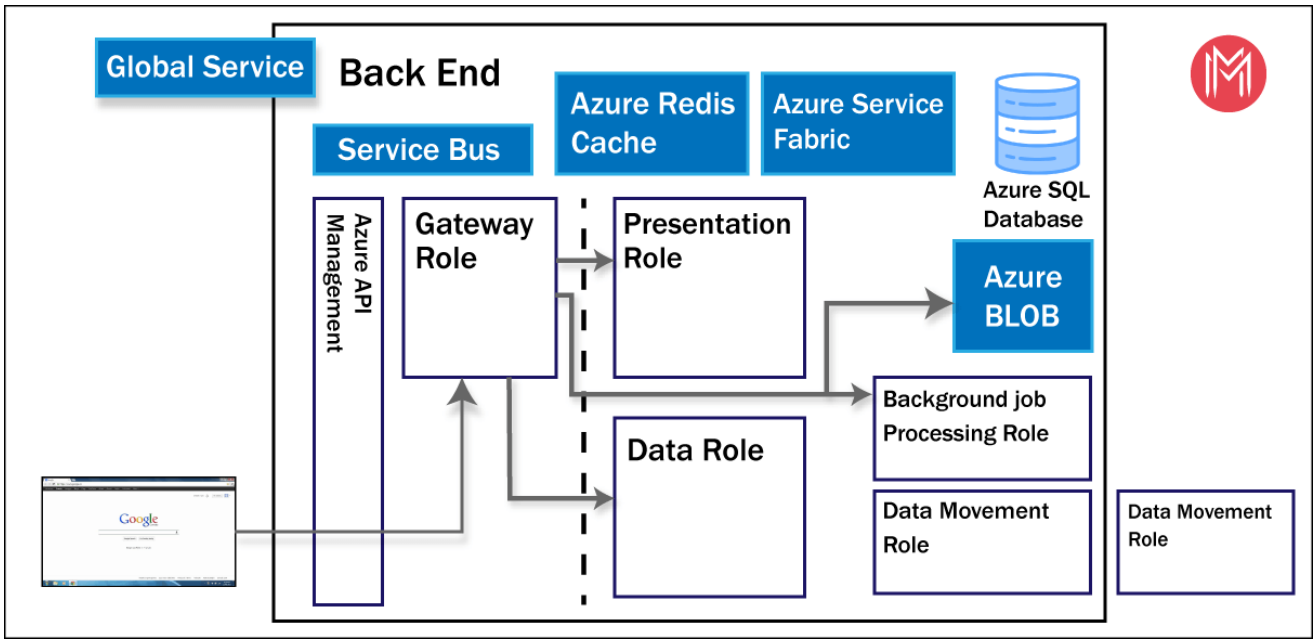


3.2 Front End Cluster



Front end cluster acts as an intermediate between the back end cluster and the clients. It is also called a Web Front End Cluster. It establishes the initial connection and authenticates the users or clients using the Azure Active Directory. After user authentication, Azure Traffic Manager directs the user requests to the nearest data centers and Azure Content Delivery Network (CDN) allocates the statice files/content to the users or clients based on the geographical locations.

3.3) Back End Cluster



It manages the datasets, reports, storage, visualizations, data refreshing, data connections, and other services in the Power BI. At the back end cluster, the web client has only two direct points to interact with the data, i.e., Gateway Role and Azure API Management. These two components are responsible for authorizing, load balancing, routing, authentication, etc.

Working Of Power BI Service:

- Power BI stores the data in two leading repositories, i.e., Azure SQL Database and Azure Block Storage. Azure Block Storage enables the users to store the datasets, and all system-related data and metadata are stored in the Azure SQL database.
- It authenticates the user requests and sends them to the Gateway Role. It processes the requests and assigns them to the appropriate components like Background Job Processing Role, Data Movement Role, Presentation Role, and Data Role.

- The presentation role manages all the associated visualization queries like reports and dashboards.
- Presentation Role sends requests to the Gateway Role to the Data Movement Role or Data Role for all relevant datasets.
- Azure Service Bus is used to connect and fetch the data from the On-Premises data sources with the cloud. It sends a request to execute the queries On-Premises data source and retrieve the data from its cloud service.
- The Azure Service Fabric allows all components and microservices which are related to the Power BI Service.
- Azure Cache helps in reporting the data that is stored in the in-memory of the Power BI system.