AMAZON SALES DATA ANALYSIS

ARCHITECTURE DESIGN

A. NOORJAHAN

Index

1.Introduction

1.1.What is Architecture Design Document?

Any software needs the architectural design to represent the design of the 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 help the designer to understand the overall properties of the system.

1.2. What is Scope?

Architecture Design Document (ADD) is an architectural 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

2.1 Power BI Architecture

1. Power BI is a business suite that includes several technologies that work together. To deliver outstanding business intelligence solutions.
2. Microsoft Power BI technology consists of a group of components such as:
    - Power Query (for data mash-up and transformation)
    - Power BI Desktop (a companion development tool)
    - Power BI Mobile (for Android, iOS, Windows phones)
    - Power Pivot (for in-memory tabular data modelling)
    - Power View (for viewing data visualizations)
    - Power Map (for visualizing 3D geo-spatial data)
    - Power Q&A (for natural language Q&A)

In simple terms, a Power BI user takes data from various data sources such as files, Azure source, online services, Direct Query or gateway sources. Then, they work with that data on a client development tool such as Power BI Desktop. Here, the imported data is cleaned and transformed according to the user's needs.

Once the data is transformed and formatted, it is ready to use in making visualizations in a report. A report is a collection of visualizations like graphs, charts, tables, filters, and slicers.

2.2 Components of Power BI Architecture

2.2.1 Data Sources

- An **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.

2.2.2 Power BI Desktop

Power BI Desktop is a client-side tool known as a companion development and authoring tool. This desktop-based software is loaded with tools and functionalities to connect to data sources, transform data, data modelling and create reports.

3. Deployment

3.1 Power BI Service

Power BI Service is a web-based platform from where you can share reports made on Power BI Desktop, collaborate with other users, and create dashboards. It is available in three versions:

- Free version
- Pro version
- Premium version

3.2 Power BI Report Server

The Power BI Report Server is similar to the Power BI Service. The only difference between these two is that Power BI Report Server is an on-premise platform. It is used by organizations who do not want to publish their reports on the cloud and are concerned about the security of their data..

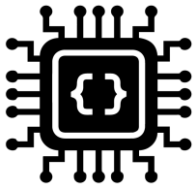
3.3 Power BI Gateway

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.

3.4 Power BI Mobile Apps

Using Power BI Mobile Apps, you can stay connected with on-premises data from anywhere. Power BI apps are available for iOS, Windows, and Android platforms.

3.5. Power BI Embedded



Power BI Embedded is an On-premises service in Azure. It offers APIs for embedding the reports and dashboards into custom applications. Till now, we have been discussing major components of the Power BI, and now, we will talk about the remaining components of Power BI as well.

Here is the list of the remaining Power BI Components.

3.6. Power BI Query

Power Query is the data connectivity that enables the business users to access the data which is stored in multiple data sources and redesign it to satisfy their business requirements. Power Query offers custom connectors SDK so that third-party users can create their data connectors.

3.7 Power Maps

Power BI Query is used to display how the values vary in proportion across the region. It also shows differences with the shadings that range from dark to light. It offers a 3D geospatial Data Visualization Tool.

3.8 Power Pivot

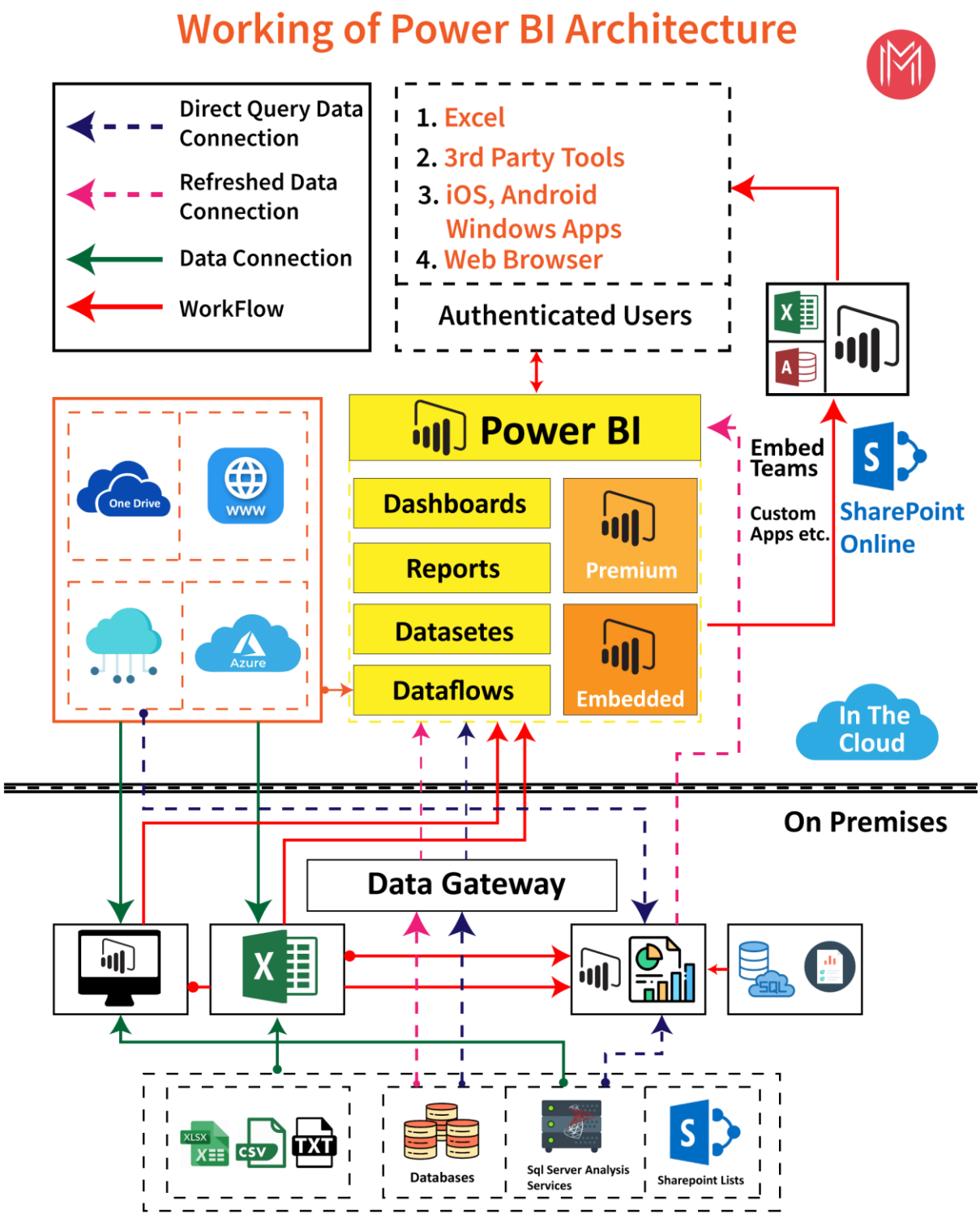
Power Pivot is an element that stores the information in memory and allows highly compressed data storage and incredibly quick aggregation and calculation. It is also accessible as part of Excel and can be used within an Excel workbook to build a data model. Power Pivot can load information on its own, or Power Query can load information into it. It is highly comparable to the tabular model of SSAS (SQL Server Analysis Services), which is like a Power Pivot server-based variant.

3.9. Power View

Power View offers interactive visualization that enables a drag-and-drop interface for users to create visualizations quickly and effectively in their Excel workbooks (using the Power Pivot data model).

3.10 Power Q&A

Power Q & A is the feature of Power BI, and it enables you to explore your data in your own words. In other words, you can use natural language and ask a question to get an answer from your data.



4 Power BI Service Architecture

In the previous section, you have learned how to publish the created reports in the Power BI Service.

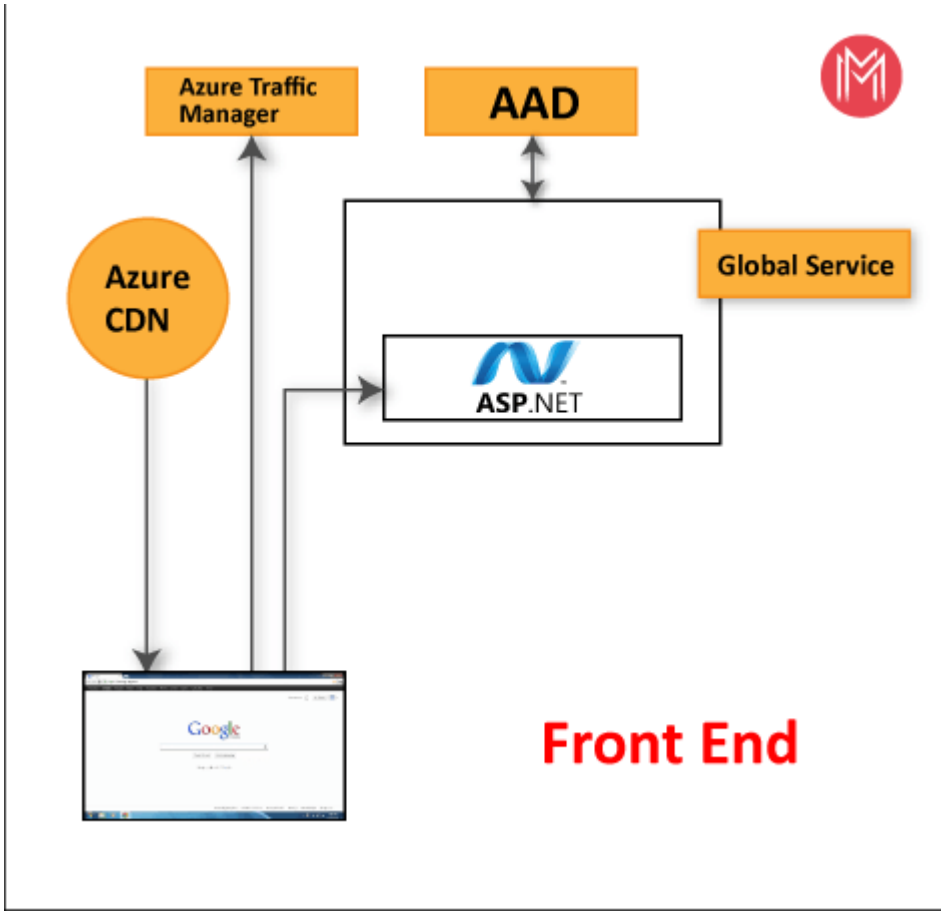
Power BI Service enables the users to create and access the reports, dashboards from the client platforms like mobile devices, websites, etc. User needs to interact with the Power BI Service whenever they want to access the data that is created on the Power BI. So, now, we will learn how the Power BI Service works.

Power BI Service Architecture consists of two clusters. The following are the two clusters.

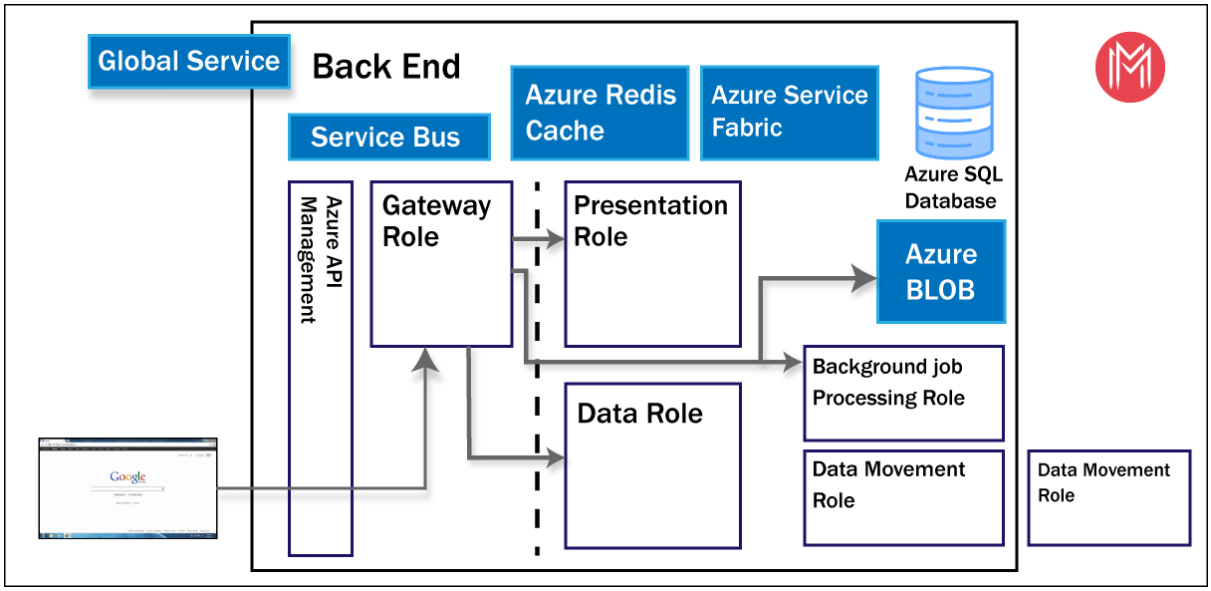
- Front End Cluster
- Back End Cluster

Now, we will discuss the two clusters in detail.

4.1. 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 centreio3s and Azure Content Delivery Network (CDN) allocates the statice files/content to the users or clients based on the geographical locations.



4.2. 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.



4.3 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.
- When you publish a Power BI Desktop file to the Power BI service, you publish the data in the model to your Power BI workspace. The same is true for any reports you created in Report view. You'll see a new dataset with the same name, and any reports in your Workspace navigator. Publishing from Power BI Desktop has the same effect as using Get Data in Power BI to connect to and upload a Power BI Desktop file.