

B.Com.(Business Analytics) V Semester
Advanced Data Visualization
Unit-III

1. What is Mashup? Explain types of Mashups.

Mashup: A mashup is a combination of separate, typically heterogeneous elements into a single, compound, and useful system. The data sources for mashups can be internal, external, or a combination. The end goal is to produce a resource that is more efficient and effective than what is currently available, in order to enhance or automate a routine business process. In **BI systems, mashups** are defined as a process of **combining heterogeneous data from multiple sources**. Data Mashing is the process of **integrating business-related heterogeneous application data from numerous sources** to give a more unified view. Mashing involves aggregating similar elements from different data sources. Good example of a mashup is an **API(Application Programming Interface)**. Developers have the ability to plug in APIs to build a new application. The concept of mashups has reduced the development overhead of rewriting already existing code, as developers can reuse the code in the form of APIs or web services. As the use of technology and social media have increased, mashups have gained in popularity.

Types of Mashups:

❖ **Business or Enterprise Mashup:**

Business (or) enterprise mashups define applications that combine their own resources, application and data, with other external Web services. It focuses data into a single presentation and allow for collaborative action among businesses and developers. An enterprise mashup is the integration of heterogeneous digital data and applications from multiple sources for business purposes. In an **enterprise mashup**, the product is typically a combination of internal corporate data and applications with externally sourced data, **SaaS (software as a service)**, and Web content

❖ **Consumer Mashups:**

Consumer mashups combine data from multiple public sources in the browser and organize it through a simple browser user interface (e.g., Wikipedia vision combines **Google Map** and a **Wikipedia API**). A consumer mashup is aimed at the general public. When offered with a web-based mashup facility, consumer mashups prove to be an effective means for customer personalization of data/viewing. Users can use their web browser to combine and reformat data according to their needs.

❖ **Data Mashups**

Data mashups combine similar types of media and information from multiple sources into a single representation. The combination of all these resources creates a new and distinct web service that was not originally provided by either source. Within the BI industry, a small number of vendors create applications that allow the combination or mashup of disparate data sources to be improvised without necessarily relying on a middle step of **ETL (Extract, Transform, Load)** and Data warehousing.

2. What are the benefits of Mashups? Give three beneficial examples

The following are the benefits of Mashups:

Interactive and Dynamic Features:

- ❖ Interactive and dynamic mashups are very useful to business end-users.
- ❖ Interactivity comes in the form of drill-down or drill-through capabilities, as well as the ability to use web-based features or solutions as a dedicated desktop application.
- ❖ Mashups can be made dynamic through the use of Flash technology (e.g., enabling key performance indicators, graphs, or charts to update dynamically when the page is opened or refreshed).

Extending BI Capabilities:

- ❖ Mashups become particularly beneficial in business intelligence, extending reporting and analysis capabilities.
- ❖ Reporting applications can draw directly from any web-based information source, providing richer and deeper insight by combining intelligence gathered from across the web with data that resides in enterprise systems.
- ❖ A BI dashboard can be enhanced by incorporating elements of Google Maps or other web applications into the environment, allowing end-users to dynamically present report output through those vehicles.

Embedding BI into Other Applications:

- ❖ Mashups enable embedding BI into other applications.
- ❖ Through a mashup of various APIs, end-users can conduct more detailed and comprehensive reporting directly from within a CRM or ERP application.
- ❖ Report output can become fully embedded within automated workflow or business process management systems, where the generation of a report can dynamically trigger a subsequent event based on its content.

Decision-Making & Process Execution:

- ❖ Mashups greatly extend and enhance BI applications.
- ❖ With the correct policies and procedures in place, companies can fully leverage mashups to ensure intelligence drives the most effective decision-making and process execution possible.

Examples:

1) Purchase Funnel:

The purchase funnel is an example of a data mashup used by sales and marketing teams to monitor their customer acquisition process by measuring the number of people at each stage and examining what percentage of people move from one stage to the next. Filling in numbers for each stage gives a clear understanding of customer flow and helps in understanding organizational strengths and weaknesses.

2) Social Interactions and Web Visits:

Measures the success of social campaigns by correlating social interaction information with web analytics. Social interactions may boost traffic indirectly and increase search and direct traffic to a website. Combines the data sources into a single chart (Google Analytics, Facebook, and Twitter) to demonstrate the correlation between viral posts and increased website traffic.

3) Server Response Time and Help Desk Tickets

Used by SaaS (Software as a Service) organizations to monitor the strong correlation between application performance and the number of help desk tickets submitted. Displays average response time for each hour across the servers

3. Explain about POWER BI and its tools

Power BI is a **cloud-based Business Intelligence** platform offered by **Microsoft** that provides business users with **tools for visualizing, sharing, aggregating, and analysing data** from various sources through a user-friendly interface. Power BI is a **Business Intelligence and Data Visualization tool** for converting data from various data sources into **interactive dashboards and analysis reports**. Power BI offers cloud-based services for interactive visualizations with a simple interface for end-users to create their own reports and dashboards.

Different Power BI versions like **Desktop**, **Service-based (SaaS)**, and **mobile Power BI apps** are used for different platforms. It provides multiple software connectors and services for business intelligence

The following are some of the benefits of POWER BI to business

- ❖ **Total integration:** Power BI offers seamless integration with Microsoft's and non-Microsoft's tools for better data connectivity.
- ❖ **Trends and predictions:** With AI and ML capabilities, users can analyse their data that can help businesses monitor important trends and even predict future trends.
- ❖ **Unlimited capacity:** Power BI can handle vast data sets compared to other platforms that struggle to process large amounts of data

POWER BI Tools:

- ❖ **Power BI Desktop:** Power BI desktop is the primary authoring and publishing tool for Power BI. Developers and power users use it to create brand new models and reports from scratch. Power BI Desktop is free and intended for small- to mid-size businesses. You can create, share, and publish interactive reports with visual analytics anytime, anywhere, and on any device.
- ❖ **Power BI Service:** Power BI service is a cloud-based business analytics and data visualization service that enables anyone to visualize and analyse data with greater speed, efficiency, and understanding. It connects users to a broad range of data through easy-to-use dashboards, interactive reports, and compelling visualizations that bring data to life.
- ❖ **Power BI Service Pro:** It is paid version that charges on a per-user basis with all the features that power BI can offer, suitable for a mid-size business. In contrast, the Power BI Service Premium is licensed as per the organization's capacity and intended for large businesses and enterprises.
- ❖ **Power BI Report Server:** It can host paginated reports, KPIs, mobile reports and Power BI Desktop reports. It is updated every 4 months and installed/managed by the IT team. The users can modify Power BI reports and other reports created by the development team.
- ❖ **Power BI Mobile Apps:** Power BI Mobile App is available for iOS, Android, Windows. It can be managed using Microsoft Intune. Allows the user to create reports, receive push notifications for personal data alerts, make queries, and more from any digital device, with native mobile BI apps for Windows, iOS, and Android.
- ❖ **Power BI Embedded:** Microsoft offers a version of Power BI for independent software vendors to embed Power BI in their own apps; rather than having to build an analytical tool right from scratch, you can simply embed Power BI into their apps.

Data Sources for POWER BI:

Data Source	Description
Excel (.xlsx)	A workbook can have data entered manually or data, which is queried and loaded from external data sources.
Comma Separated Values (.csv)	Every row can contain one or more values, which is separated by a comma.
Power BI Desktop (.pbi)	User can use Power BI Desktop to query and load data from external data sources.
Databases in the Cloud	It allows to connect to SQL Database & SQL DataWarehouse,
Databases On-premises	User can connect directly to SQL Server Analysis Services Relational model databases.

4. Write about Event Streaming Processing. Compare Stream Processing vs Batch Processing.

Event Stream Processing: Stream processing refers to the methodology of acting upon continuous data streams immediately as the data is produced. Incoming data streams undergo various tasks sequentially, concurrently, or both. This end-to-end workflow comprises stream data generation, real-time data processing, and final delivery to target destinations. The following are the common actions on Stream Data:

- ❖ **Aggregations:** Performing statistical computations like sums, averages, or standard deviations.
- ❖ **Analytics:** Identifying data patterns to forecast future occurrences.
- ❖ **Transformations:** Reformatting data fields (e.g., converting numerical values into standard date formats).
- ❖ **Enrichment:** Merging incoming stream data with external data sources to add context and value.
- ❖ **Ingestion:** Loading and persisting processed data directly into database systems.

- ❖ **Real-time Application Response:** Applications process input pipelines via a dedicated stream processing engine in real time, routing outputs to live streaming analytics applications or target output streams.

Stream Processing vs. Batch Processing:

- ❖ **Batch Processing (Legacy Approach):** Historically, processing occurred in discrete batches according to predetermined schedules or thresholds (e.g., scheduled nightly executions, every 100 rows, or upon reaching a 2 MB size limit).
- ❖ **Shift to Real-Time:** Batch processing is less suitable for modern big data demands, where massive volumes of continuously generated data require instant processing to remain actionable.
- ❖ **Modern Necessity:** Contemporary enterprise applications process events immediately as they happen instead of accumulating data at fixed intervals.

5. Explain the process of Analysing Trusted Data.

Analytics can transform raw data into an extensive collection of information that categorizes data to identify and analyse behavioural data and patterns. Organizations can use this information to analyse the current state of their operations and to predict future behaviour and trends by asking "what-if" questions. Additionally, analytics can help with fraud detection, image recognition, sentiment analysis, overall general employee productivity, and it also often replaces cumbersome manual processes. Traditionally, data analysis was a complex task that was carried out by engineers. Today, data analysis is more accessible to, and understood by, many people within organizations, across all teams. Power BI is an exceptional tool for quickly pulling actionable insights from data. It allows you to build visuals and metrics for your data in reports and dashboards so that you and your users can analyse data insights at a high level and drill down into those insights for more detailed information.

Power BI helps to accomplish this task. You can develop quick insights and share them in reports and dashboards with different teams within the organization. The advanced analytics capabilities of Power BI will enable you to identify categories and trends, see how data changes over time, and much more. From this information, you can make predictive data models, and therefore, help your organization make more robust business decisions, plans, and forecasts.

The various kinds of analysis that can be performed on data to derive useful information are as follows:

- ❖ **Statistical Summaries:** Data is often intertwined with statistics because statistics help in exploring the data. Statistics show you the distribution of your data and help you to identify key takeaways and trends, and determine whether outliers exist. The statistical summary is the information that provides a quick and simple description of your data. Power BI has many functions that help you to conduct a statistical analysis, such as Data Analysis Expressions (DAX) functions, visuals such as histograms and bell curves, advanced analytics visuals, and statistical programming languages such as Python and R.
- ❖ **Statistical Functions:** Power BI Desktop has a number of DAX functions that you can use to get quick statistics based on your data.
 - **Histogram:** Histograms and bell curves are the most common way to display statistics about your datasets. In Power BI terms, you can represent a histogram with one of the bar or column chart visuals and represent a bell curve with an area chart visual.
 - **Top N analysis:** The TOPN DAX function returns the top N rows of a specified table. The Top N analysis is a great way to present data that might be important, such as the top 10 selling products, top 10 performers in an organization, or top 10 customers.
 - **Use the Q&A visual to find the top N:** The Q&A visual is an effective tool because it allows users to quickly get answers about the data independently, which saves time for everyone involved. The Q&A visual is unique in that it does not require knowledge of Power BI to use the visual; users can ask their question and they, too, can create insightful visuals.
 - **Use a TOPN DAX function:** You can also calculate your top 10 products in DAX by using the TOPN function. This option could be useful if you want to present the top 10 in a different context, such as how much of the top 10 best-selling products contributed toward the overall total sales.

- **Use a visual to identify outliers:** The best visual to use for identifying outliers is the scatter chart, which shows the relationship between two numerical values. Scatter charts display patterns in large sets of data and are, therefore, ideal for displaying outliers.
- **Group and bin data for analysis:** When you create visuals, Power BI Desktop aggregates your data into groups, based on the values that it finds in the underlying data. You can refine how those default groups are presented. You can also create new groups by grouping two or more data points in a visual or putting values into equal-sized groups (binning). Grouping is used for categories of data. Binning is similar to grouping, but it is used for grouping continuous fields, such as numbers and dates. You can use the grouping and binning features to ensure that the visuals in your reports display your data according to your preference.
- **Apply clustering techniques:** Clustering allows you to identify a segment (cluster) of data that is similar to each other but dissimilar to the rest of the data. The process of clustering is different to that of grouping, which you accomplished previously. The Power BI clustering feature allows you to quickly find groups of similar data points in a subset of your data. It analyses your data set to identify similarities and dissimilarities in the attribute values, and then it separates the data that has similarities into a subset of the data. These subsets of data are referred to as clusters.
- **Conduct time series analysis:** Time series analysis involves analysing a series of data in time order to identify meaningful information and trends, and make predictions. The result of time series analysis is the best data that you can use for forecasting activities. Time series analysis often involves the use of visuals such as Gantt charts, project planning, and stock movement datasets. In Power BI, you can use visuals to view how your data is progressing over time, which in turn allows you to make observations, such as if any major events disrupted your data.
- **Use the Analyse feature:** The Analyse feature provides you with additional analysis that is generated by Power BI for a selected data point. You might want to use this feature to determine if Power BI has found something that you have not seen before, or if you want Power BI to give you a different insight into your data. This feature is particularly useful for analysing why your data distribution looks the way that it does. Make use of the Analyse feature to get fast, automated, insightful analysis of your data.

SUMMARY

Mashup:

Mashups are defined as a process of combining heterogeneous data from multiple sources. The mashup data is from different applications and services. The Best example of a mashup is an API (Application Programming Interface). Types of Mashups are Business or Enterprise Mashup, Consumer Mashups and Data Mashups

POWER BI:

Power BI is a cloud-based Business Intelligence platform offered by Microsoft that provides business users with tools for visualizing, sharing, aggregating, and analysing data from various sources through a user-friendly interface. Power BI is a Business Intelligence and Data Visualization tool for converting data from various data sources into interactive dashboards and analysis reports.

Event Stream processing:

Event Stream processing refers to the methodology of acting upon continuous data streams immediately as the data is produced. Incoming data streams undergo various tasks sequentially, concurrently, or both. This end-to-end workflow comprises stream data generation, real-time data processing, and final delivery to target destinations

Trusted Data:

Trusted Data is the source of data that is accurate, reliable, and suitable for analysis. Every organization must verify the accuracy of the information obtained from data. Trusted data can be derived through Statistical Summaries and Statistical Functions.

Data Analysis:

Data analysis is the process of examining data to discover useful patterns and trends. Data analysis is more accessible to, and understood by, many people within organizations, across all teams. Data Analysis Expressions (DAX) functions, visuals such as histograms and bell curves, advanced analytics visuals, and statistical programming languages such as Python and R.