Unit 1: Introduction to Business Intelligence

Mrs.Manasi Pandharkar

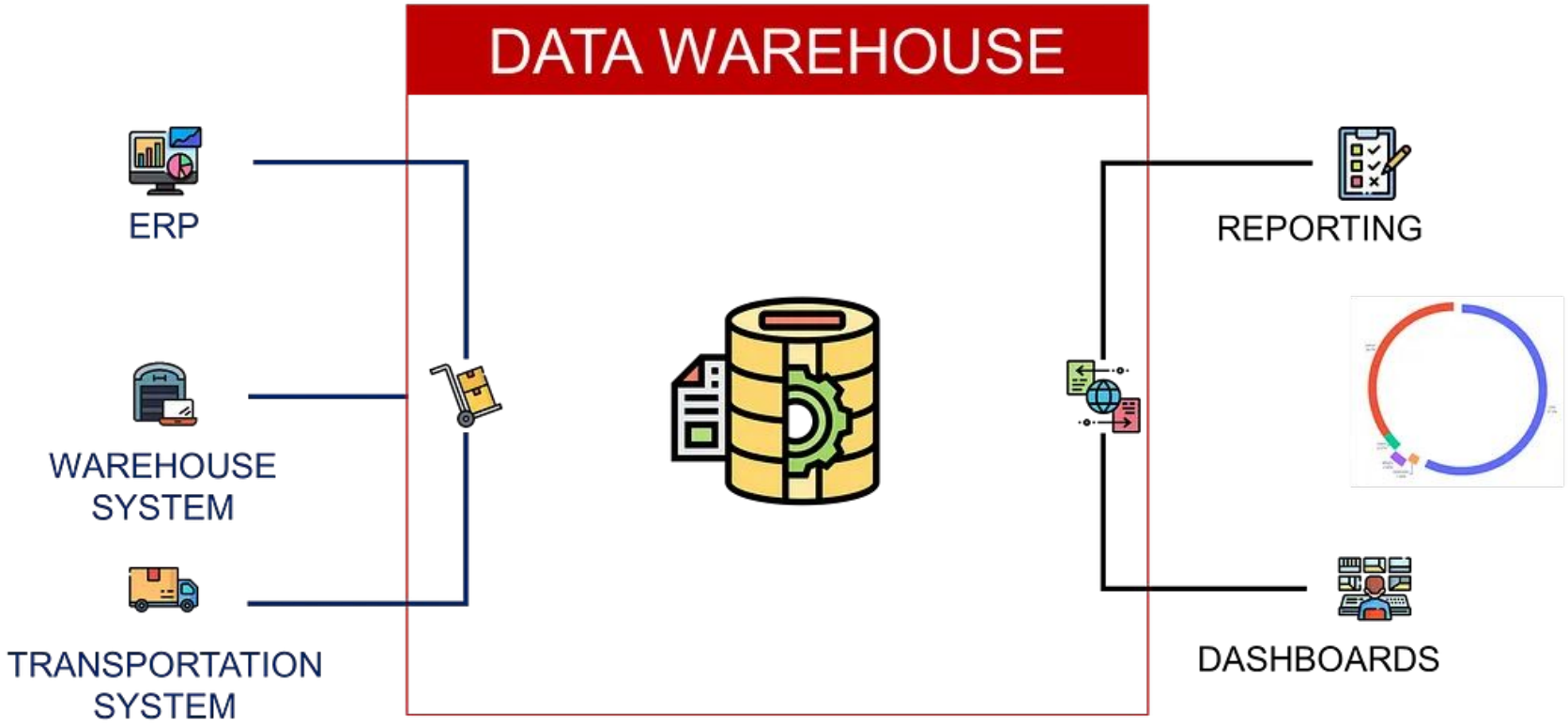
Outline

- Overview of Business Intelligence & Data-Driven Decision Making
- BI vs. Data Analytics vs. Data Science
- Key Components of BI: **Data Warehousing, ETL, Data Mining, Reporting & Dashboards**
- Real-World BI Use Cases in Various Industries
- Open-Source BI Tools: **Metabase, Redash, Apache Superset**

Introduction

- Business Intelligence (BI) is the process of turning data into actionable insights to help businesses make better decisions.
- The key components are data gathering, data standardization, data analysis, and reporting, and the process flow involves collecting, integrating, storing, and analyzing data to derive insights.
- Business Intelligence is a process that leverages software and services to transform data into actionable intelligence supporting decision-making.
- This intelligence helps businesses gain insights into their operations, improve process efficiency, and create a competitive market advantage.

What is BI


DATA WAREHOUSE
ERP
WAREHOUSE SYSTEM
TRANSPORTATION SYSTEM
REPORTING
DASHBOARDS

Business Intelligence is not Advanced Analytics

	BUSINESS INTELLIGENCE	ADVANCED ANALYTICS
	▪ What happened? ▪ When? ▪ Why? ▪ Who?	▪ Why did it happen? ▪ Will it happen again? ▪ What should we do?
	▪ Dashboards (KPIS, metrics) ▪ Automated Monitoring and Alerting ▪ OLAP (cubes, slice ▪ Ad hoc query ▪ Retail Time Visualization	▪ Statistical Analysis ▪ Process Mining ▪ Predictive Modelling ▪ Prescriptive Modelling ▪ NLP ▪ Digital Twins

Continue

- Business Intelligence provides Descriptive and Diagnostic Analytics solutions focusing on “understanding past events”.

- What happened? How many orders have been delivered with delay?
- When? Has the order 1878497 been loaded at the warehouse?
- Who? Which carrier delivered store 12 last week?
- Why? Why did order 1878497 arrive at the airport 1 hour late?

BUSINESS INTELLIGENCE

DESCRIPTIVE

Which events happened?

Sales jumped by +20%

DIAGNOSTIC

Why did these events happen?

National Holidays Pushed the Sales

ADVANCED ANALYTICS

PREDICTIVE

What can happen?

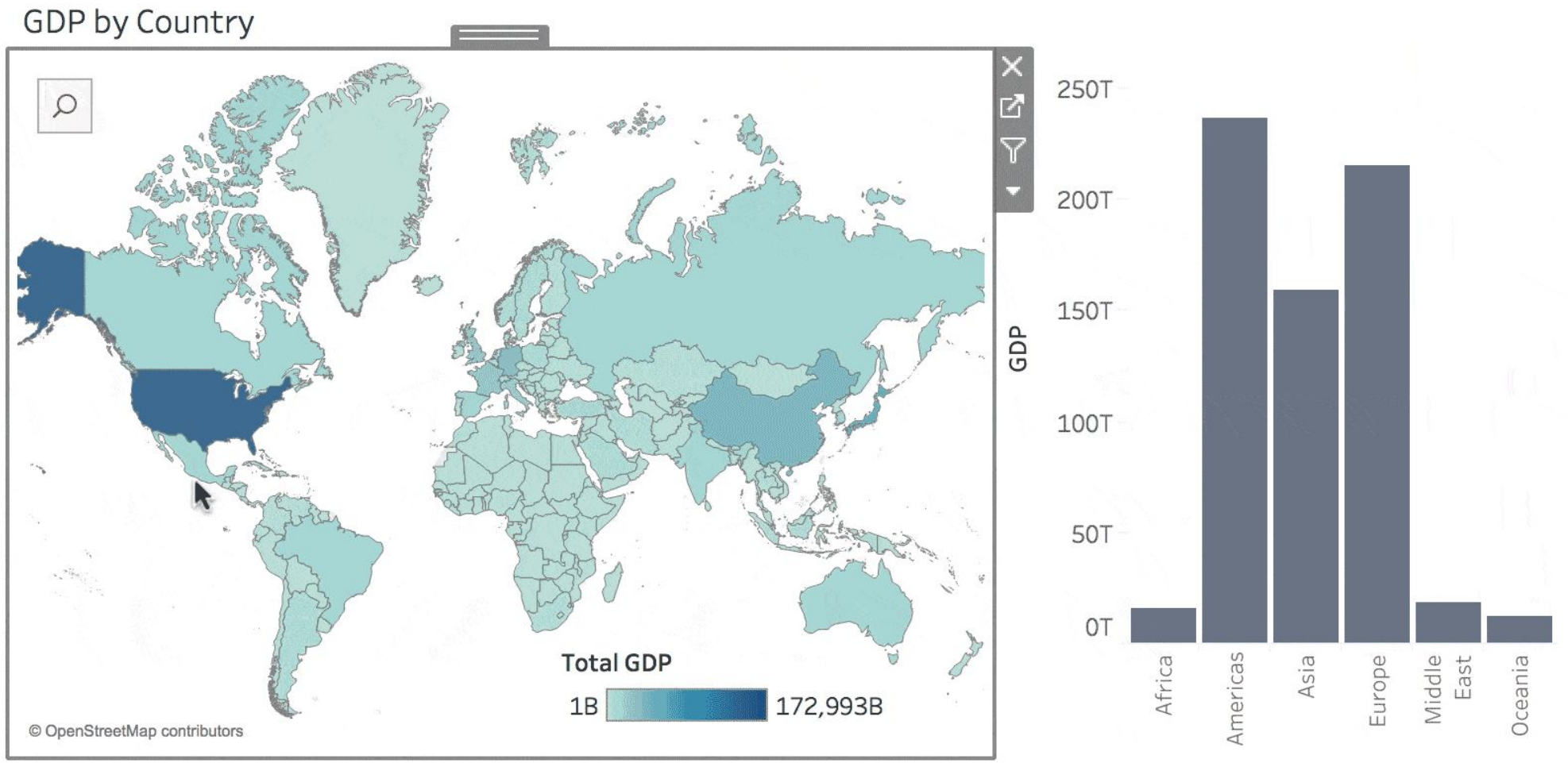
Sales will increase by +10% next week

PRESCRIPTIVE

What should we do?

Replenish the store tomorrow

The difference between traditional BI and modern BI



Business Analytics

- ***Definition:*** Business analytics (BA) is the process of collecting, analyzing, and interpreting data to gain insights that can inform better business decisions. In the context of car manufacturing, BA focuses on leveraging data from various sources to improve efficiency, quality, customer satisfaction, and profitability.

Business analytics is the process of transforming data into insights to improve business decisions. Data management, data visualization, predictive modeling, data mining, forecasting simulation, and optimization are some of the tools used to create insights from data.

Business Analytics involves

- identifying new patterns and relationships with data mining;
- using quantitative and statistical analysis to design business models;
- forecasting future business needs, performance, and industry trends with predictive modeling;
- communicating your findings in easy-to-digest reports to colleagues, management, and customers.

Business Analytics for Car Manufacturing Industry

Unveiling Hidden Insights: Through data mining, we uncover new patterns and relationships within our vast data sets. This allows us to identify areas for improvement, predict customer preferences, and gain a deeper understanding of market trends.

Data-Driven Decision Making: We leverage quantitative and statistical analysis to inform strategic business model design. This ensures our decisions are grounded in solid evidence, leading to increased efficiency and profitability.

Future-Proofing Our Business: Predictive modeling allows us to forecast market trends, anticipate industry shifts, and plan for future business needs. This proactive approach ensures we stay ahead of the curve and deliver the right vehicles at the right time.

Clear Communication is Key: Our BA team doesn't just generate data; they translate insights into clear and actionable reports.

Key Differences Between Business Intelligence (BI) and Business Analytics (BA)

Feature	Business Intelligence (BI)	Business Analytics (BA)
Focus	Understanding **past and present** data	Predicting and optimizing for the **future**
Purpose	Reporting, dashboards, monitoring business performance	Data-driven decision-making, predictive modeling
Data Type	Structured, historical, real-time operational data	Structured & unstructured data, statistical models
Techniques Used	Data visualization, OLAP, ETL, KPI tracking	Machine Learning, Statistical Analysis, Forecasting
Outcome	**Descriptive analytics** – "What happened?"	**Predictive & Prescriptive analytics** – "What will happen & what to do?"
Example	Tracking sales performance with **Power BI/Tableau** dashboards	Forecasting **customer demand** using **Python ML models**
Tools Used	Power BI, Tableau, Apache Superset, Google Data Studio	Python (Scikit-learn, TensorFlow), R, SAS, Alteryx

BI Use Case: Automobile Industry

BI Question:

- *"Which car models have the highest sales, and what factors influence customer preferences?"*

How BI Helps?

- ✔ **Sales trend analysis** across different models and regions
  ✔ **Customer behavior analytics** based on demographics and purchase history
  ✔ **Supply chain efficiency analysis** to ensure timely delivery

Continue

Process:

1. **Data Collection** – Gather data from sales records, service logs, customer feedback
2. **ETL (Extract, Transform, Load)** – Clean and integrate data from multiple sources
3. **Analysis & Reporting** – Create **interactive dashboards** for insights on best-selling models, customer reviews, and supply chain efficiency

Tools Used:

- **Power BI / Tableau** – For sales trend and customer analytics dashboards
- **Apache Superset** – Real-time performance tracking for vehicle sales
- **Google BigQuery** – Large-scale data storage for automobile industry data

BI Use Case: Healthcare Industry

BI Question:

- *"How can we reduce patient wait times and improve hospital resource utilization?"*

How BI Helps?

- ✔ Identifies **peak hours & patient inflow patterns** to optimize staff allocation
  ✔ Tracks **hospital resources (beds, doctors, equipment)** for better resource utilization
  ✔ Analyzes **electronic health records (EHR)** for disease pattern identification

Continue

Process:

1. **Data Extraction** – Collect patient appointment records, hospital occupancy data
2. **Data Processing** – Apply **predictive analytics** to forecast peak demand
3. **Dashboard Reporting** – Show real-time bed availability, doctor schedules

Tools Used:

- **Power BI / Tableau** – Hospital resource and occupancy dashboards
- **Apache Kafka / Spark** – Real-time data streaming for patient monitoring
- **SQL Server / PostgreSQL** – Database for hospital management system

BI Use Case: Sales & Retail Industry

BI Question:

- *"Which products are selling the most, and how can we optimize pricing and promotions?"*

How BI Helps?

- ✔ Analyzes **sales performance across regions** to identify best-performing products
  ✔ Uses **customer purchase behavior** to recommend personalized discounts
  ✔ Predicts **demand trends** to optimize stock levels

Continue

Process:

1. **Sales Data Collection** – Gather data from **POS (Point of Sale) systems**
2. **Customer Segmentation** – Use **BI dashboards** to track purchase behavior
3. **Sales Forecasting** – Apply **predictive models** for stock and pricing optimization

Tools Used:

- **Tableau / Power BI** – Sales performance analysis dashboards
- **Google Analytics** – Customer behavior tracking
- **Python (Pandas, Scikit-learn)** – Sales forecasting models

BI Use Case: Finance & Banking Industry

BI Question:

- *"How can we detect fraudulent transactions and assess credit risk?"*

How BI Helps?

- ✔ Monitors **real-time banking transactions** to flag anomalies
  ✔ Uses **machine learning models** to detect fraud patterns
  ✔ Helps in **credit risk assessment** by analyzing financial history

Continue

Process:

1. **Transaction Data Collection** – Gather transaction logs from bank systems
2. **Fraud Detection Model** – Use **AI-powered BI tools** to find anomalies
3. **Risk Assessment Dashboard** – Show flagged transactions for further review

Tools Used:

- **Power BI / Tableau** – Fraud detection dashboards
- **Python (Scikit-learn, TensorFlow)** – AI-based fraud detection models
- **Apache Spark / Kafka** – Real-time banking transaction processing

BI Use Case: Education Industry

BI Question:

- *"Which students are at risk of dropping out, and how can we improve academic performance?"*

How BI Helps?

- ✔ Tracks **student attendance and grades** to identify drop-out risks
- ✔ Monitors **teaching effectiveness** through student feedback analysis
- ✔ Provides **real-time performance insights** to academic institutions

Continue

Process:

1. **Data Collection** – Extract student records, attendance logs, exam results
2. **Data Processing** – Use **BI tools** to identify struggling students
3. **Predictive Analytics** – Apply **AI models** to predict drop-out risks

Tools Used:

- **Google Data Studio / Tableau** – Student performance dashboards
- **Python (Pandas, Scikit-learn)** – Predictive analytics for drop-out risk
- **SQL / Snowflake** – Large-scale academic data management

BI Use Case: Supply Chain & Logistics

BI Question:

- *"How can we optimize delivery routes and reduce transportation costs?"*

How BI Helps?

- ✔ Tracks **real-time shipment status** to identify delays
  ✔ Analyzes **warehouse efficiency** and stock availability
  ✔ Uses **route optimization models** to reduce fuel and logistics costs

Continue

Process:

1. **GPS Data Collection** – Gather vehicle tracking data from IoT sensors
2. **Route Optimization** – Use **BI algorithms** to find shortest delivery paths
3. **Logistics Dashboard** – Provide real-time tracking for supply chain managers

Tools Used:

- **Power BI / Tableau** – Route optimization and delivery tracking dashboards
- **Google Maps API + Python** – Predictive analytics for traffic and fuel efficiency
- **AWS Redshift / Snowflake** – Logistics database for supply chain analytics

BI Use Case: Media & Entertainment Industry

BI Question:

- *"What type of content is performing the best, and how can we improve engagement?"*

How BI Helps?

- ✔ Tracks **viewership patterns** to identify popular content
  ✔ Monitors **social media engagement** to optimize marketing strategies
  ✔ Uses **predictive analytics** to recommend personalized content

Continue

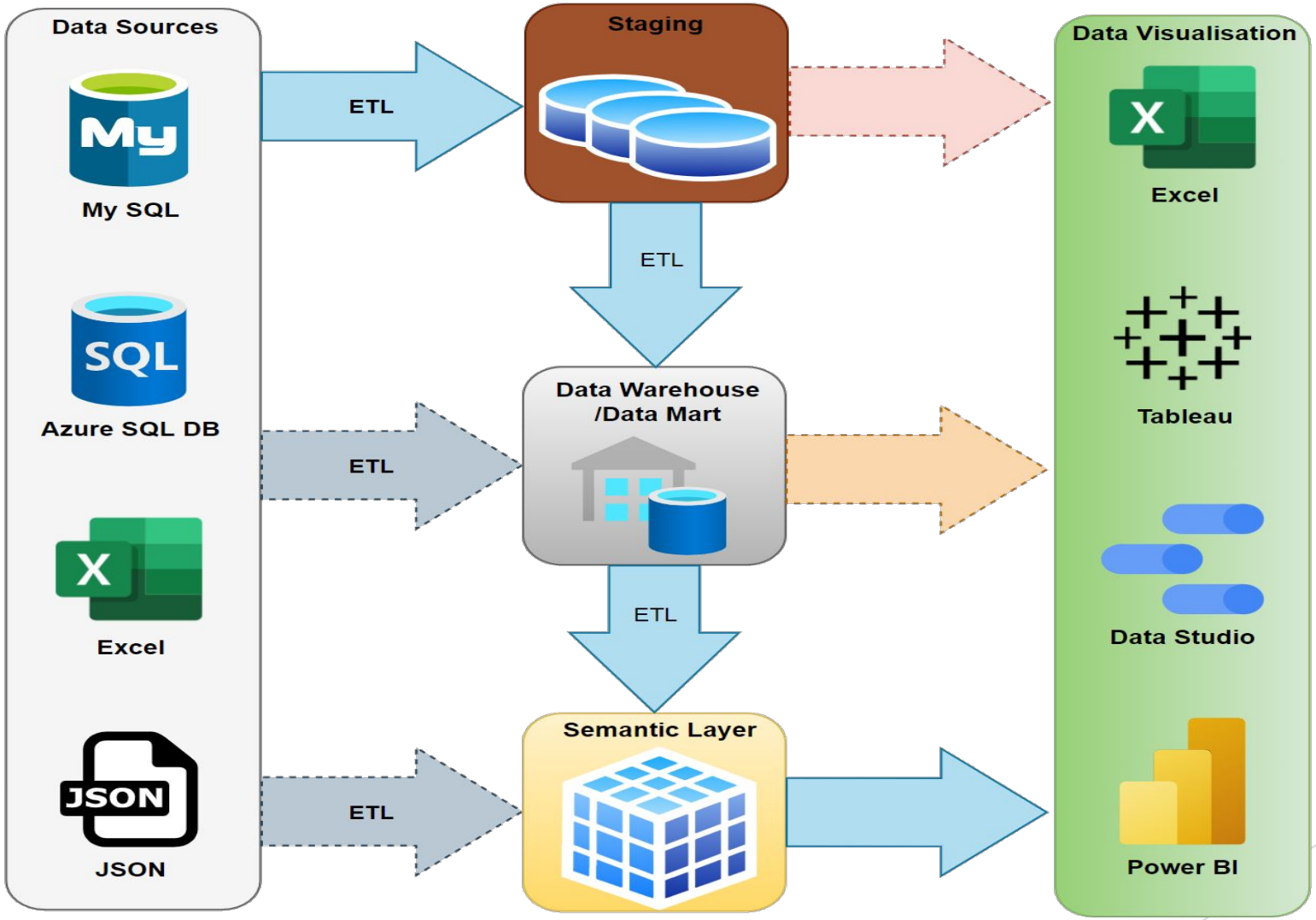
Process:

1. **User Data Collection** – Gather streaming habits, social media interactions
2. **Engagement Analysis** – Track content performance across platforms
3. **Recommendation Engine** – Apply AI-based **content suggestion models**

Tools Used:

- **Google Analytics / Tableau** – Content engagement tracking dashboards
- **Python (Pandas, NLP models)** – Sentiment analysis for audience insights
- **AWS S3 / Google BigQuery** – Large-scale media data processing

BI Components



BI Components

1. Data Sources (Raw Data Collection)

What It Does?

- BI begins with collecting **raw data** from multiple sources, including:
  ✔ Databases (SQL, NoSQL)
  ✔ Cloud storage (AWS, Google BigQuery)
  ✔ APIs (Web services, third-party apps)
  ✔ CRM/ERP systems (SAP, Salesforce, Oracle)
  ✔ Social media and IoT devices

Example:

- A **retail company** collects **sales transactions** from **POS (Point of Sale) systems**, **website logs**, and **customer feedback**.

BI Tools Used:

- ✔ **Database Management Systems**: MySQL, PostgreSQL, MongoDB
  ✔ **Data Lakes**: AWS S3, Google BigQuery, Azure Data Lake

Continue

2. ETL (Extract, Transform, Load)

What It Does?

- **ETL (Extract, Transform, Load)** processes **raw data** and prepares it for analysis.
  ✔ **Extract** – Pulls data from different sources
  ✔ **Transform** – Cleans, filters, formats, and merges data
  ✔ **Load** – Stores processed data into a **data warehouse** for analysis

Example:

- A **bank** extracts **customer transaction logs**, transforms them to remove errors and duplicates, and loads them into a **central database** for fraud detection.

BI Tools Used:

- ✔ **ETL Tools**: Talend, Apache Nifi, Alteryx, Informatica
  ✔ **Cloud-Based ETL**: AWS Glue, Google Dataflow

Continue

3. Data Warehousing (Centralized Data Storage)

What It Does?

- A **Data Warehouse** is a **centralized storage system** that holds **structured data** for historical analysis and reporting.
- It enables **fast querying**, **data integration**, and **business intelligence applications**.

Example:

- A **healthcare provider** stores **patient records**, **treatment history**, and **insurance claims** in a **data warehouse** for long-term analysis.

BI Tools Used:

- ✔ **Cloud Data Warehouses**: Amazon Redshift, Google BigQuery, Snowflake
- ✔ **On-Premise Data Warehouses**: Microsoft SQL Server, Oracle Data Warehouse

Continue

4. Data Mining (Pattern Recognition & AI/ML Models)

What It Does?

- **Data Mining** is the process of discovering **hidden patterns** in large datasets using **machine learning** and **statistical models**.
- Helps in **forecasting trends, anomaly detection, and customer segmentation**.

Example:

- **E-commerce platforms** use **data mining** to **analyze customer behavior** and recommend products.

BI Tools Used:

- ✔ **Machine Learning Libraries**: Python (Scikit-learn, TensorFlow), R
- ✔ **Data Mining Platforms**: Orange, RapidMiner, KNIME

Continue

5. OLAP (Online Analytical Processing)

What It Does?

- **OLAP (Online Analytical Processing)** allows users to perform **multi-dimensional analysis** of data.
- Enables users to **drill down, roll up, slice, and dice** data for deeper insights.

Example:

- A **retail business** uses OLAP to analyze **sales data by region, product category, and customer demographics**.

BI Tools Used:

- ✔ **OLAP Tools**: Microsoft SQL Server Analysis Services (SSAS), Oracle OLAP
- ✔ **Cloud-Based OLAP**: Google BigQuery, AWS Athena

Continue

6. Data Visualization & Reporting (Dashboards & KPI Tracking)

What It Does?

- **Data visualization tools** turn complex data into **interactive dashboards, charts, and reports**.
- Helps businesses **track Key Performance Indicators (KPIs)** in real time.

Example:

- A **finance company** uses **Power BI dashboards** to monitor **monthly revenue, fraud alerts, and customer complaints**.

BI Tools Used:

✔ **BI Dashboards**: Power BI, Tableau, Google Data Studio, Apache Superset
✔ **Reporting Tools**: Microsoft Excel, Crystal Reports

Continue

7. Business Performance Management (BPM)

What It Does?

- BPM focuses on **measuring and improving business performance** using data-driven insights.
- Uses **forecasting models** and **scenario analysis** to optimize business strategies.

Example:

- A **manufacturing company** uses BPM to optimize **supply chain logistics, reduce costs, and improve production efficiency**.

BI Tools Used:

✔ **BPM Platforms**: SAP BusinessObjects, IBM Cognos Analytics
✔ **Cloud-Based Solutions**: Google Looker, Microsoft Power BI

Continue

8. Advanced Analytics (AI & Predictive Analytics)

What It Does?

- Uses **machine learning, artificial intelligence (AI), and predictive modeling** to analyze future trends.
- Helps businesses with **fraud detection, customer recommendations, and demand forecasting**.

Example:

- A **telecom company** uses **AI-powered analytics** to predict **customer churn and optimize marketing campaigns**.

BI Tools Used:

- ✔ **AI/ML Frameworks**: Python (TensorFlow, Scikit-learn), R
- ✔ **Big Data Processing**: Apache Spark, Google Vertex AI

Continue

9. Cloud BI & Real-Time Data Processing

What It Does?

- **Cloud BI tools** store and process data in **real-time**, eliminating hardware limitations.
- Enables **real-time dashboards, streaming analytics, and automated data processing**.

Example:

- **Ride-hailing apps (Uber, Lyft)** use **real-time analytics** to match **drivers with passengers**, **calculate dynamic pricing**, and **optimize routes**.

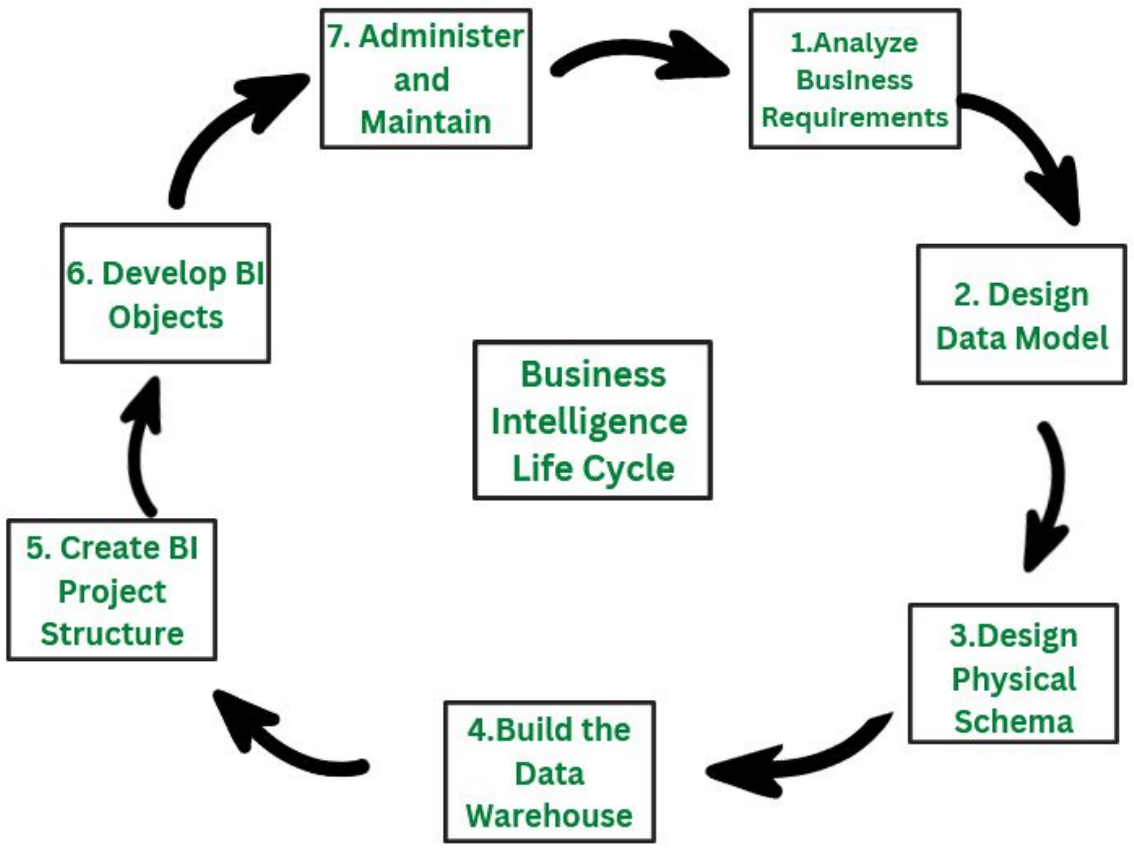
BI Tools Used:

- ✔ **Cloud-Based BI**: AWS QuickSight, Google Looker, Microsoft Power BI
- ✔ **Streaming Analytics**: Apache Kafka, Apache Flink, Google Dataflow

Summary

Component	Function	Examples
Data Source	Collects raw data	Databases, APIs, IoT devices
ETL	Cleans and integrates data	Talend, Apache Nifi, Alteryx
Data Warehouse	Stores structured business data	Google BigQuery, Snowflake, Amazon Redshift
Data Mining	Identifies patterns in data	Python (Scikit-learn), RapidMiner
OLAP (Online Analytical Processing)	Multi-dimensional data analysis	Microsoft SSAS, Oracle OLAP
Data Visualization & Reporting	Creates dashboards and reports	Power BI, Tableau, Apache Superset
Business Performance Management (BPM)	Monitors business KPIs	IBM Cognos, SAP BusinessObjects
Advanced Analytics (AI & Predictive)	Forecasts trends, detects fraud	TensorFlow, Apache Spark
Cloud BI & Real-Time Analytics	Processes live data streams	AWS QuickSight, Apache Kafka

BI Life Cycle



Business Intelligence Life Cycle

Continue

- **Phase 1: Analyze Business Requirements**
- The first step in the Business Intelligence life cycle is to analyze the business requirements. The user identifies the business requirements in order to determine the type of analysis that the user then needs to perform. Identifying the requirements, let the user decides the further action to be performed.

- For example, any retail company can analyze the sales data to figure out the products that are top-selling and the products that least sell.

Continue

- **Phase 2: Design Data Model**
- Once the requirements are identified the user needs to design the logical model according to the requirements. This logical model helps the user to analyze the relationships that exist within the data entities.

- For example, For any retail company, the data model consists of products, their customers, and the sales data

Continue

Phase 3: Design the Physical Schema

Once the logical model is prepared the next step is to design the physical schema using the data model. The physical schema describes the structure and the content of the data warehouse.

For example, in any retail company, physical schema consists of sales-related facts, product-customer relationships, and the sales transactions

Continue

- **Phase 4: Build the Data Warehouse**
- Once the logical and physical schema is designed, the next step is to build the data warehouse. The design of a data warehouse depends on the physical and logical schema. After the design of the data warehouse, the data and the content from the source system are loaded into the data warehouse for further steps.

- For example, for the retail system, designing the data warehouse consists of developing a database that would store the details of customers, products, and other requirements for the business.

Continue

- **Phase 5: Create the Project Structure (Metadata)**
- The next step after designing the data warehouse is to create a project structure also known as metadata. With the help of this created project structure, the mapping of the tables and data in the data warehouse is easier. Creating the project structure describes the further steps and types that need to be implemented.

- For example, The project structure of the retail company consists of the attributes of the data, the design, and the working flow of the system. This project structure or metadata gives a brief idea about the working of the system.

Continue

- **Phase 6: Develop The BI Objects**
- The next step is to develop the BI objects such as metrics, attributes, dashboards, reports, and facts. This step consists of developing the reports and dashboards that can be used to analyze the data in the data warehouse.
- For example, the retail company can develop reports and statistics charts that can describe the profit and loss margins.

Continue

- **Phase 7: Administer and Maintain the Project**
- The last step is to administer and maintain the project continuously as it undergoes changes. The project needs to be monitored to maintain the changes, security, and performance of the system.

- For example, the retail company needs to monitor the reports and statistics accordingly to increase the profit of the sales.

BI Tools

1. Power BI

✔ **What It Is?**
Power BI is a **business analytics tool** by Microsoft that allows users to **connect to data sources, create interactive reports, and share insights across an organization**.

✔ **Source & Owner**:

- Developed by **Microsoft**
- First released in **2015**

✔ **Versions & Open-Source Status**:

- ❌ **Not Open-Source** (Proprietary)
- **Versions**:
  - Power BI **Desktop** (Free)
  - Power BI **Pro** (Paid)
  - Power BI **Premium** (Enterprise)

✔ **When to Use?**
✔ If you need **seamless integration with Microsoft products** (Excel, Azure, SQL Server).
✔ For **drag-and-drop visualization & easy reporting** without coding.
✔ If your organization already uses **Microsoft 365 ecosystem**.

Continue

2. Tableau

✔ **What It Is?**
Tableau is a powerful **visual analytics tool** that enables users to create **interactive dashboards and reports** using real-time data.

✔ **Source & Owner**:

- Originally developed by **Tableau Software**, acquired by **Salesforce** in **2019**.
- First released in **2003**.

✔ **Versions & Open-Source Status**:

- ❌ **Not Open-Source** (Proprietary)
- **Versions**:
  - Tableau **Public** (Free, cloud-based)
  - Tableau **Desktop** (Paid)
  - Tableau **Server/Online** (Enterprise)

✔ **When to Use?**
✔ If you need **advanced, customizable visualizations** for large datasets.
✔ If your organization uses **Salesforce, cloud-based BI, or embedded analytics**.
✔ When you need **AI-driven BI with natural language queries**.

Continue

3. Google Looker (formerly Looker)

✔ **What It Is?**
Looker is a **modern, cloud-based BI tool** that offers **embedded analytics** and **SQL-based data modeling**.

✔ **Source & Owner**:

- Developed by **Looker**, acquired by **Google Cloud** in **2020**.
- First released in **2012**.

✔ **Versions & Open-Source Status**:

- ❌ **Not Open-Source** (Proprietary)
- **Versions**:
  - Looker **Cloud Enterprise** (Paid)
  - Looker **Embedded BI** (Custom pricing)

✔ **When to Use?**
✔ If your company uses **Google Cloud (BigQuery, Cloud SQL)**.
✔ If you need **real-time analytics and embedded dashboards**.
✔ For **SQL-based reporting & modeling** (great for data analysts).

Continue

5. Redash

✔ **What It Is?**
Redash is an **open-source data visualization and dashboarding tool** designed for **SQL-based reporting and collaboration**.

✔ **Source & Owner**:

- Originally an open-source project, acquired by **Databricks** in **2020**.
- First released in **2015**.

✔ **Versions & Open-Source Status**:

- ✅ **Open-Source**
- **Versions**:
  - Redash **Self-Hosted** (Open-source, free)
  - Redash **Cloud Edition** (Paid, hosted by Databricks)

✔ **When to Use?**
✔ If you need **SQL-based data querying with dashboards**.
✔ If your team **collaborates on complex data analysis**.
✔ When you need **self-hosted, secure BI for cloud-based environments**.

Continue

6. Apache Superset

✔ **What It Is?**
Apache Superset is a **highly scalable, open-source BI tool** designed for **big data visualization and exploration**.

✔ **Source & Owner**:

- Developed by **Airbnb**, now part of the **Apache Software Foundation**.
- First released in **2016**.

✔ **Versions & Open-Source Status**:

- ✅ **Open-Source** (Apache License)
- **Versions**:
  - Superset **Open-Source Edition** (Free)

✔ **When to Use?**
✔ If you need **enterprise-level BI with big data processing (Hadoop, Spark)**.
✔ If your company requires **advanced SQL-based analytics**.
✔ When you want **a free alternative to Power BI/Tableau**.

Continue

7. Google Data Studio

✔ **What It Is?**
Google Data Studio is a **free cloud-based BI tool** that helps users create **interactive reports and dashboards** using Google's ecosystem.

✔ **Source & Owner**:

- Developed by **Google**.
- First released in **2016**.

✔ **Versions & Open-Source Status**:

- ❌ **Not Open-Source** (Free but proprietary)
- **Versions**:
  - Free (Cloud-based)

✔ **When to Use?**
✔ If you need **Google Analytics, Google Ads, and Google Sheets integration**.
✔ If you want **a free, simple BI tool for reporting**.
✔ If you prefer **cloud-based, real-time dashboard sharing**.