

Answer1)

In today's digital economy, organizations rely on **real-world data sources** to make informed decisions and gain a competitive advantage. These data sources can come from multiple domains and formats such as structured, semi-structured, or unstructured. The integration of these sources with tools like **Power BI** or modern data platforms (Azure, AWS, GCP, Snowflake, Databricks) enables effective analytics and visualization.

Examples of Real-World Data Sources

1. Retail Domain

- **Point of Sale (POS) Systems** → Capture transactional data like sales, discounts, and payment methods.
- **E-commerce Platforms (Amazon, Flipkart APIs)** → Provide customer purchase history, product trends, and reviews.
- **Customer Loyalty Programs** → Data from reward points, offers, and personalized purchases.
- **Inventory & Supply Chain Systems** → Stock levels, shipment tracking, and vendor details.
- **Social Media Data (Twitter, Instagram, Facebook APIs)** → Consumer sentiment and brand engagement.

2. Healthcare Domain

- **Electronic Health Records (EHRs)** → Patient demographics, diagnoses, prescriptions, and treatment histories.
- **Medical Devices & IoT Sensors** → Wearable devices (Fitbit, Apple Watch) providing real-time vitals like heart rate, glucose, and oxygen levels.
- **Hospital Management Systems (HMS)** → Appointment scheduling, billing, and resource allocation.
- **Government & Public Health Datasets** → WHO, CDC, and government portals with disease outbreak statistics, vaccination data, etc.
- **Research Databases** → Clinical trial data and published journals.

Integration with Tools like Power BI & Modern Data Platforms

• Data Extraction & Integration

Tools such as **ETL pipelines** (Extract, Transform, Load) built on **Azure Data Factory, Talend, or Informatica** can pull data from diverse sources. APIs and connectors simplify integration with Power BI.

- **Data Storage**

Data lakes (Azure Data Lake, AWS S3, Google BigQuery) and warehouses (Snowflake, Redshift, Synapse Analytics) are used for large-scale data storage.

- **Data Processing**

Platforms like **Databricks (Spark-based)** or AWS Glue transform raw data into structured formats.

- **Data Visualization & Reporting**

Power BI connects directly to SQL databases, cloud storage, or APIs. It provides dashboards for sales forecasting, healthcare monitoring, and supply chain optimization.

Conclusion

Real-world data integration enables organizations to transform raw information into actionable insights. In the retail domain, it helps optimize inventory and improve customer experience, while in healthcare it supports patient care, diagnosis, and preventive measures. Tools like Power BI act as the front-end enabler for storytelling and decision-making.

Answer 2

<https://github.com/A-kash1101/Akash-Autade-AIEC2-Data-Engineering-Sem5>