DATA INTEGRATION AND DATA FEDERATION

INST 733 Research Topic Presentation
Team 9

Ainur Abilbayeva
Yogesh Boricha
Angela Tseng

INTRODUCTION

- Data integration is the process of combining data from multiple disparate sources, such as databases, applications, and systems, into a unified and meaningful view
- Data federation, on the other hand, refers to the process of accessing data from multiple databases into a single, virtual database.
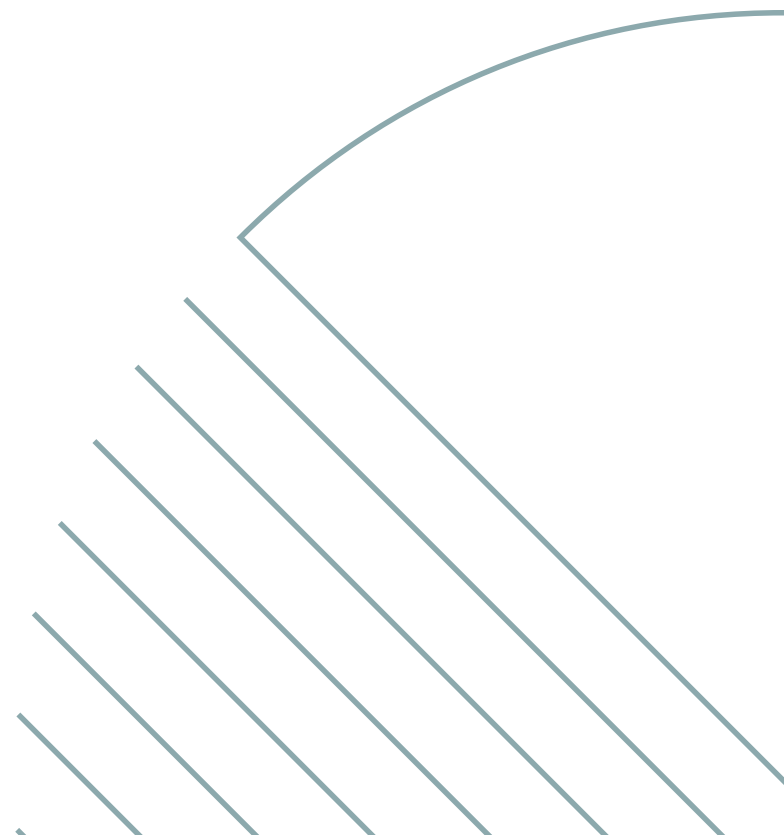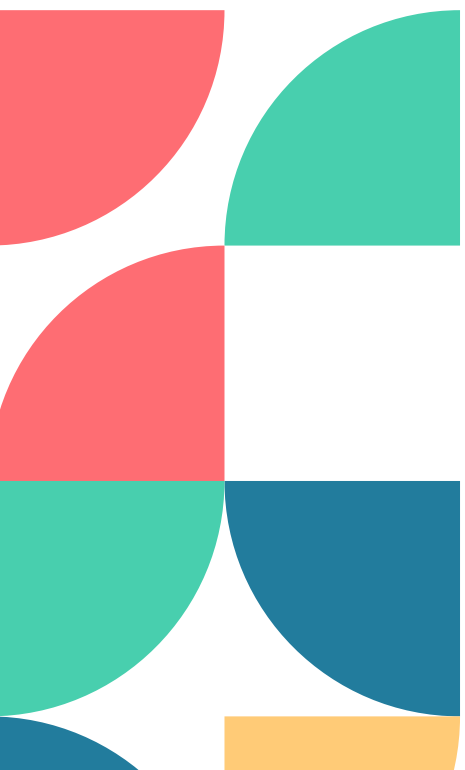
BENEFITS

- Fast Data Access
- Improved Data Quality
- Cost Savings
- Increased Efficiency
- Better data flows
- Better Decision Making
- Gain a holistic business view
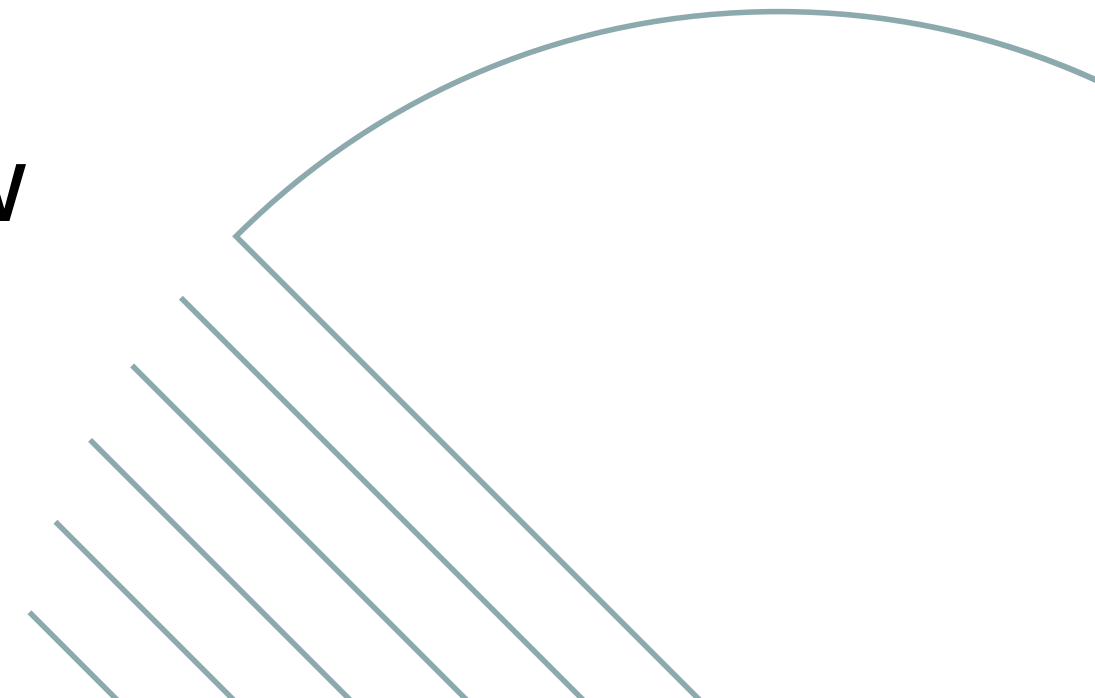
The different Data Integration approaches are:

- Data Consolidation
- Data Propagation
- Data Virtualization

Federated database system

- A federated database system (FDBS) is a type of meta-database management system (DBMS), which transparently maps multiple autonomous database systems into a single federated database.
- Federated database systems provide a uniform user interface that enables users and clients to store and retrieve data from multiple noncontiguous databases with a single query.

Data Integration and Federation in Healthcare
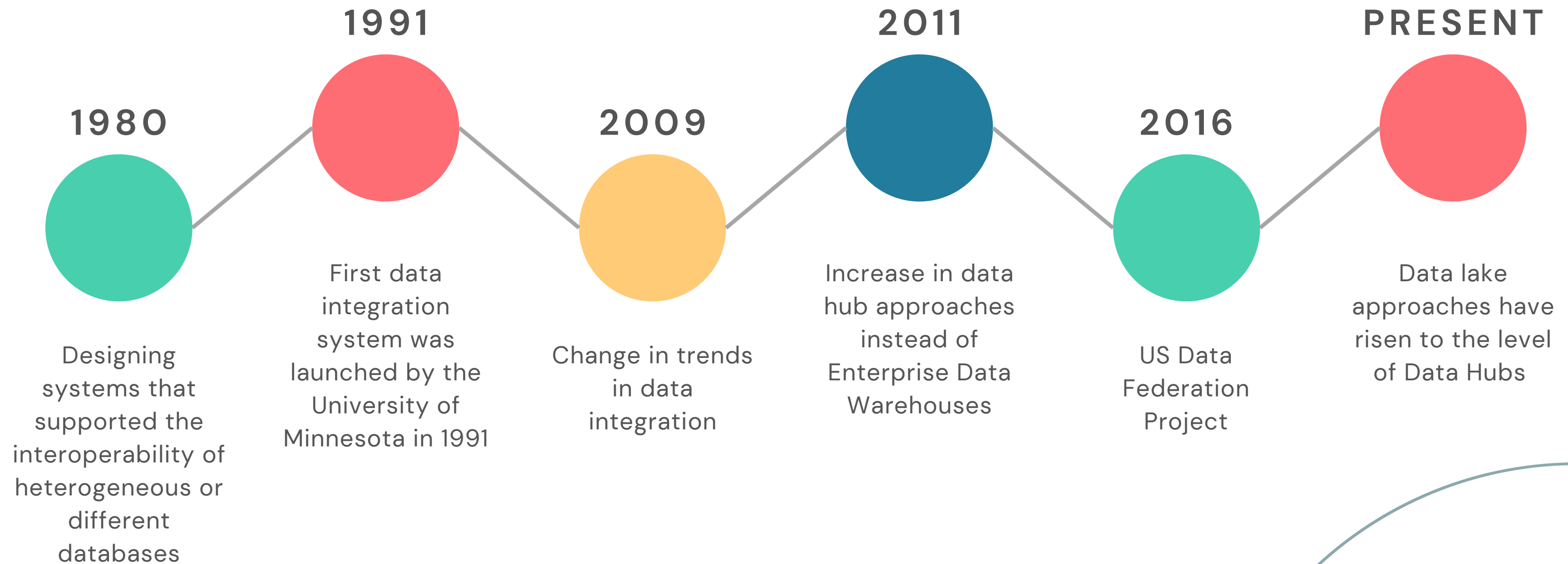
Hospitals have data in the forms of :

- Electronic health records (EHRs)
- Medical imaging systems
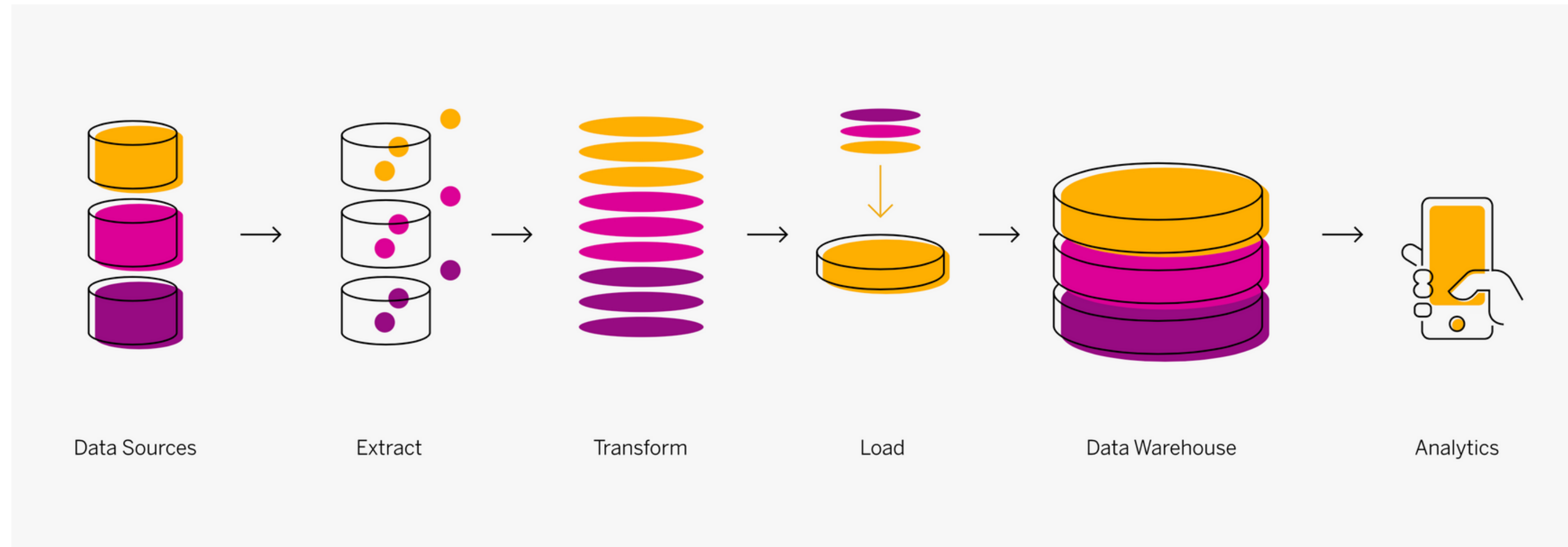- Laboratory information management systems (LIMS)

Use of Data Integration in Healthcare

- Data integration and federation enables healthcare organizations to integrate patient data from various sources into a single central repository.
- Sources of patient data can include EHRs, medical imaging systems, LIMS, and other systems.
- The integrated patient data provides a unified view of patient health, enabling the organization to identify trends and patterns and monitor patient health more effectively.
- This unified view of patient data also enables the organization to provide more personalized and effective care to patients.

HISTORY

1980

Designing systems that supported the interoperability of heterogeneous or different databases

1991

First data integration system was launched by the University of Minnesota in 1991

2009

Change in trends in data integration

2011

Increase in data hub approaches instead of Enterprise Data Warehouses

2016

US Data Federation Project

PRESENT

Data lake approaches have risen to the level of Data Hubs

STEPS: DATA INTEGRATION

Data Sources   Extract   Transform   Load   Data Warehouse   Analytics

STEPS: DATA FEDERATION

SOURCE DATA & DATABASES

TRANSACTIONAL DATA
CURRENT DATA

DATA FEDERATION

VIRTUAL
DATABASES

BUSINESS INTELLIGENCE

HISTORICAL DATA

ETL

DATA WAREHOUSE
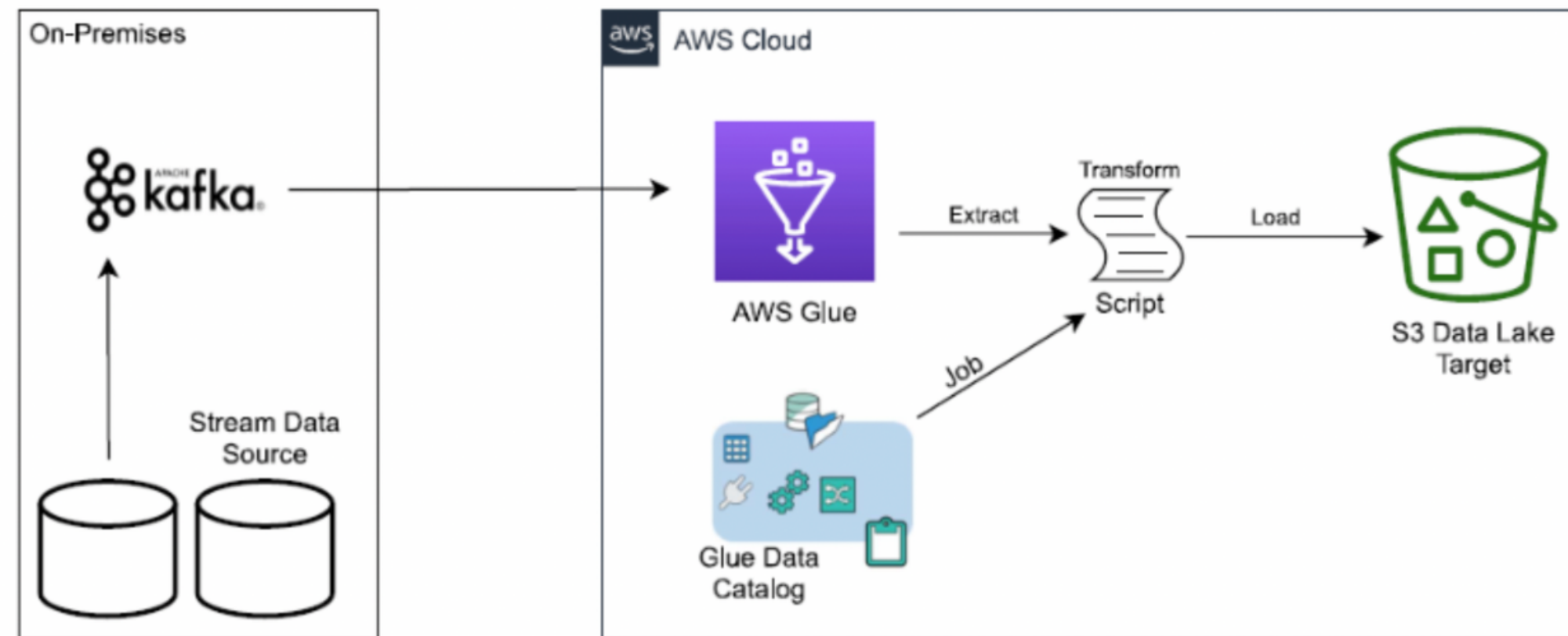AND DATA MARTS

1. Identify Data Sources
2. Create a Virtual Layer
3. Define Metadata
4. Query Data
5. Optimize Performance

ETL

ETL, which stands for **extract, transform and load**, is a data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system.

AWS Glue is a serverless data integration service that makes it easier for analytics users to discover, prepare, move, and integrate data from multiple sources for analytics, machine learning, and application development.

Data Integration Methods and Strategies

Manual data integration

Middleware data integration

Application-based integration

Uniform access integration

Common storage integration

Which of these data integration approaches is best for business?

Data integration approaches	When to use them
Manual data integration	Merge data for basic analysis between a small amount of data sources
Middleware data integration	Automate and translate communication between legacy and modernized systems
Application-based integration	Automate and translate communication between systems and allow for more complicated data analysis
Uniform access integration	Automate and translate communication between systems and present the data uniformly to allow for complicated data analysis
Common storage integration	Present the data uniformly, create and store a copy, and perform the most sophisticated data analysis tasks

Top 10 ETL Framework Alternatives
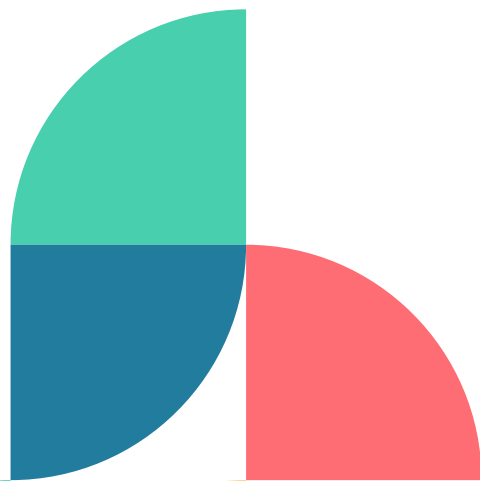
DATA FUSION IN SMART CITIES

Data fusion (DF) is collection of techniques by which information from multiple sources are combined in order to reach a better inference.

INTELLIGENT TRANSPORTATION SYSTEMS
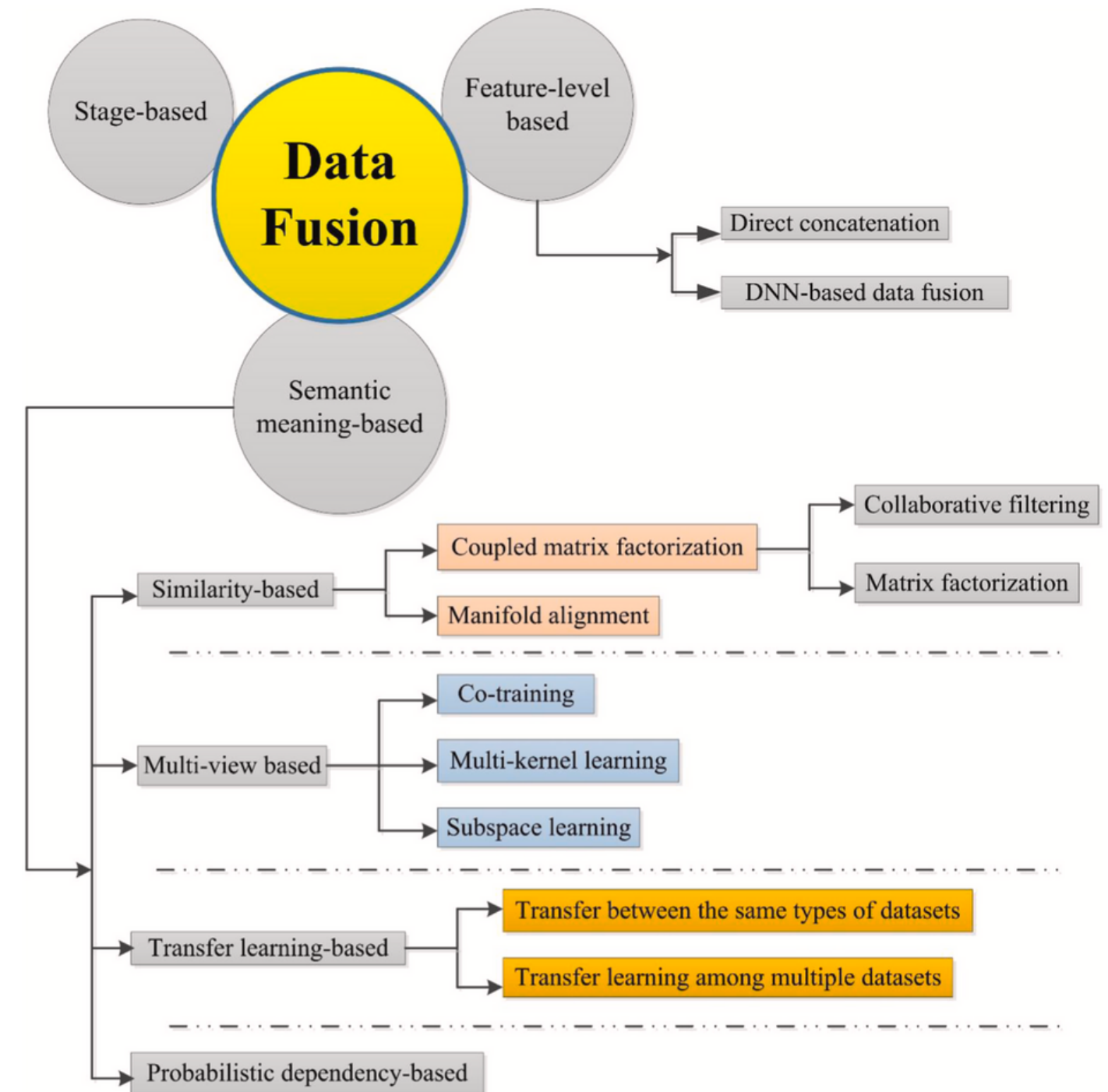
SMART CITY ONTOLOGIES

Fig. 3. Conventional data fusion methodologies diagram.

CHALLENGES

1. Different alternatives for data integration tools and many data integration techniques
2. Different explanations for how data integration developed over time
3. Heterogeneous data sources (different formats, structures, and semantics)
4. Scalability and performance concerns (large volumes of data, real-time processing)

LESSONS LEARNED

- Data Integration
  - History
  - Approaches
  - Process
  - Tools and platforms
- Data Federation
- ETL

QUESTIONS FOR CLASS

- What is the process for Data Integration?
- Mention the main Frameworks used for Data Integration and Data Federation?
- When was the first Data Integration system developed?
- What are the benefits of Data Integration?
- How does ELT differ from ETL in the context of Data Integration?
- How can data integration and federation help organizations achieve a more complete and accurate view of their data?

THANK YOU