DATA INTEGRATION

TECHNIQUES AND SOFTWARE

WHAT IS DATA INTEGRATION?

- Process of consolidating data from multiple applications and creating a unified view of data assets

- Vital strategy to adopt as companies store information in different databases

- Offers significant assistance for use in reporting and analysis

- Core component of several mission-critical data management projects such as:

  - Building an enterprise data warehouse

  - Migrating data from one or multiple databases to another

  - Synchronising data between applications

TYPES OF DATA INTEGRATION TECHNIQUES

- Manual data integration

- Middleware data integration

- Application-based integration

- Uniform access integration

- Common storage integration/data warehousing

MANUAL DATA INTEGRATION

- occurs when a data manager oversees all aspects of the integration- usually by writing custom code

- Refers to connecting the different data sources, collecting the data and cleaning it, without automation

Advantages	Disadvantages
Reduced cost: little maintenance required and typically only integrates a small number of data sources Greater freedom: user has total control over integration	Less access: A developer or manager must manually orchestrate each integration Difficulty scaling: Scaling for larger projects requires manually changing the code for each integration, draining time Greater room for error: A manager and/or analyst must handle the data at each stage

Best suited to-
- Merge data for basic analysis between a small amount of data sources

MIDDLEWARE DATA INTEGRATION

- Software connecting applications and transfers data between them and databases

- Especially useful when integrating stubborn legacy systems with newer ones as middleware can act as an interpreter

Advantages	Disadvantages
Better data streaming: software conducts integration automatically and in the same way each time	Less access: middleware needs to be deployed and maintained by a developer with technical knowledge
Easier access between systems: software coded to facilitate communication between systems in a network	Limited functionality: Middleware can only work with certain systems

Best suited to-
- Automate and translate communication between legacy and modernized systems

APPLICATION BASED INTEGRATION

- software applications do all the work i.e. locating, retrieving, cleaning and integrating all data from disparate sources

- This compatibility makes it easy for data to move from one source to the other

Advantages	Disadvantages
Simplified processes: One application does all the work automatically Easier information exchange: Application allows systems and departments to transfer information seamlessly Fewer resources are used: Automation allows managers and/or analysts can pursue other projects	Limited access: Requires special, technical knowledge and a data manager and/or analyst to oversee application deployment and maintenance Inconsistent results: Approach is unstandardized and varies from businesses offering this as a service Complicated setup: Designing the application(s) to work seamlessly across departments requires developers, managers, and/or analysts with technical knowledge

Best suited to:
- Automate and translate communication between systems and allow for more complicated data analysis

UNIFORM ACCESS INTEGRATION

- Accesses data from even more disparate sets and presents it uniformly

- Does so while allowing the data to stay in its original location

Advantages	Disadvantages
Lower storage requirements: No need to create a separate place to store data	Data integrity challenges: Accessing so many sources can lead to compromising data integrity
Easier data access: works well with multiple systems and data sources	Strained systems: Data host systems are not usually designed to handle the amount and frequency of data requests in this process
Simplified view of data: Creates a uniform appearance of data for the end user	

Best suited to:

- Automate and translate communication between systems and present the data uniformly to allow for complicated data analysis

COMMON STORAGE INTEGRATION

- involves creating and storing a copy of the data in a data warehouse

- leads to more versatility in the ways businesses can manipulate data, making it one of the most popular forms of data integration

Advantages	Disadvantages
Reduced burden: host system isn't constantly handling data queries	Increased storage costs: Creating a copy of the data means finding and paying for a place to store it
Increased data version management control: Accessing data from one source, versus multiple disparate sources, leads to better data integrity	Higher maintenance costs: requires technical experts to set up the integration, oversee, and maintain it
Enhanced data analytics: Maintaining a stored copy allows manager and/or analysts to run more sophisticated queries without worrying about compromised data integrity	

Best suited to:
- Present the data uniformly, create and store a copy, and perform the most sophisticated data analysis tasks

TOP DATA INTEGRATION TOOLS AVAILABLE IN THE MARKET

- Hevo Data

- IRI Voracity

- Xplenty

- Informatica

- Microsoft

HEVO DATA

Capabilities

- Real-time Data Replication

- Automatic Schema Detection

- Change Data Capture

- Enterprise-Grade Security

- Detailed Alerts and Logging

- Zero Data Loss Guarantee

IRI VORACITY

Capabilities

- Multi-source, multi-action
- multi- target in same I/O
- Multiple job design options in Eclipse
- seamless execution in IRI CoSort or Hadoop engines
- Compatible with COBOL, erwin, Git, MIMB, KNIME, Splunk, etc.

XPLENTY

Capabilities

- complete toolkit for data pipelines

- no-code & low-code options

- intuitive graphic interface, etc.

INFORMATICA

Capabilities

- Integrated codeless environment

MICROSOFT

Capabilities

- Can run SQL server integration services packages directly in Azure

- Fully managed ETL service in the cloud