Data Sources

PRESENTED BY DR.MYO

DAY 3

What is Data Source?

- **A data source is the location where data that is being used originates from.**

- A data source may be a database, a flat file, live measurements from physical devices, scraped web data.

- Example of a data source in action.

- *Imagine a fashion brand selling products online. To display whether an item is out of stock, the website gets information from an inventory database. In this case, the inventory tables are a data source, accessed by the web application which serves the website to customers.*

Data Source Nomenclature

- Databases remain the most common data sources, as the primary stores for data in ubiquitous relational database management systems (RDBMS).

- An important concept is the Data Source Name (DSN).

- The DSN is defined within destination databases or applications as a pointer to the actual data, whether it exists locally or is found on a remote server.

- For example, within the Java software platform, a 'Datasource' refers specifically to an object representing a connection to a database (like an extensible, programmatically packaged DSN)

Data Source Types

- The diversity of content, format, and location for data is only increasing with contributions from technologies such as IoT and the adoption of big data.

- **Machine Data Sources**
  - Machine data sources have names defined by users, must reside on the machine that is ingesting data, and cannot be easily shared
  - Machine data sources provide all the information necessary to connect to data, such as relevant software drivers and a driver manager
  - Users need only ever refer to the DSN as shorthand to invoke the connection or query the data.

Data Source Types (Cont'd)

- **File data sources**

- File data sources contain all of the connection information inside a single, shareable, computer file (typically with a .dsn extension).

- Users do not decide which name is assigned to file data sources, as these sources are not registered to individual applications, systems, or users, and in fact do not have a DSN like that of machine data sources.

- Each file stores a connection string for a single data source.

- File data sources, unlike machine sources, are editable and copyable like any other computer file.

- This allows users and systems to share a common connection (by moving the data source between individual machines or servers), and for the streamlining of data connection processes (for example by keeping a source file on a shared resource so it may be used simultaneously by multiple applications and users).

How data sources work

- Data sources are used in a variety of ways. Data can be transported thanks to diverse network protocols, such as the well-known File Transfer Protocol (FTP) and HyperText Transfer Protocol (HTTP), or any of the myriad Application Programming Interfaces (APIs) provided by websites, networked applications, and other services.

- Many platforms use data sources with FTP addresses to specify the location of data needed to be imported.

- For example, in the Adobe Analytics platform, a file data source is uploaded to a server using an FTP client, then a service utilizes this source to move and process the relevant data automatically.

- SFTP (The S stands for Secure or SSH) is used when usernames and passwords need to be obfuscated and content encrypted, or FTPS may alternatively be used by adding Transport Layer Security (TLS) to FTP, achieving the same goal.

The purpose of a data source

- data sources are intended to help users and applications connect to and move data to where it needs to be.

- They gather relevant technical information in one place and hide it so data consumers can focus on processing and identify how to best utilize their data.

- The purpose here is to package connection information in a more easily understood and user-friendly format.

- This makes data sources critical for more easily integrating disparate systems, as they save shareholders from the need to deal with and troubleshoot complex but low-level connection information.

Thank you

QUESTIONS?