**HANDS ON LAB**

**ETL TOOLS:**

* IBM DATASTAGE

IBM DataStage is a data integration tool built around a client-server design. From a Windows client, tasks are created and executed against a central data repository on a server. The tool is designed to support ETL and extract, load, and transform (ELT) models and supports data integrations across multiple sources and applications while maintaining high performance.

|  |  |
| --- | --- |
| Advantages | Disadvantages |
| 1) Large volume of data. Bulk transfer and complex transformation | 1) No automated mechanism for error handling and recovery. |
| 2)Refresh and synchronize data as much as needed. | 2)We don't have UNIX datastage client. |

* [Informatica PowerCenter](https://www.informatica.com/gb/products/data-integration/powercenter.html)

Informatica PowerCenter is a metadata-driven platform focused on improving collaboration between the business and IT teams and streamlining data pipelines. PowerCenter parses advanced data formats, including JSON, XML, PDF, and Internet of Things machine data, and automatically validates transformed data to enforce defined standards.

|  |  |
| --- | --- |
| Advantages | Disadvantages |
| 1) Informatica Power center is an innovative software that works with ETL-type data integration. Connectivity to almost all the database systems. | 1) It doesn't have much scheduling options and the tool is not very much capable if we schedule more jobs. |
| 2) Great documentation and customer support. | 2) Debugging the workflows and mappings are very hard with Informatica PowerCenter. |

**ELT TOOLS:**

* Airflow

An Airflow workflow is a sequence of tasks defined using the Python programming language. These tasks can be initiated on a schedule or even by an event. The pipeline can also send reports on the status of your pipelines using email and is one of the many **ELT tools** to do so.

|  |  |
| --- | --- |
| Advantages | Disadvantages |
| 1) Airflow allows users to create workflows with high granularity and track the progress as they execute. This makes it easy to scale out the management of lots of these jobs at the same time. | 1) Advanced features such as authentication, non-standard data sources, and parallelization require specialized infrastructure knowledge, and often must be set up manually as part of the design of an effective workflow. |
| 2)  It makes it easy to do potentially large data operations | 2) The core of Airflow is typically stable and reliable but there are glitches, particularly when you’re trying to scale |

* Blendo

Blendo is known as one of the best **ELT tools** enabling customers to centralize all of their different datasets and data sources into a central location. They are in the business of building new connectors and maintaining the ones already created.

|  |  |
| --- | --- |
| Advantages | Disadvantages |
| 1) They provide a fast way to replicate your application, databases, events, and files into the fully managed and elastic cloud warehouses such as BigQuery, and Redshift | 1) |
| 2)  Limited maintenance and configuration | 2) |