|  |  |
| --- | --- |
| **Name** | Shubham Golwal |
| **UID no.** | 2020300015 |
| **Experiment No.** | 08 |

|  |  |
| --- | --- |
| **AIM:** | To study various ETL tools used in data warehouses. |
| **Program 1** | |
| **PROBLEM STATEMENT :** | To study various ETL tools used in data warehouses between Talend Open Studio and Oracle Data Integrator. |
| **Theory :** | **Talend Open Studio:**  Talend Open Studio is an open-source data integration software that enables users to integrate, transform, and manage large volumes of data from various sources. It provides a graphical interface and drag-and-drop tools for building complex data integration workflows without requiring users to write code.  Some of the key features of Talend Open Studio include:   1. Data Integration: Talend Open Studio supports various data integration techniques, such as Extract-Transform-Load (ETL), Extract-Load-Transform (ELT), and Extract-Load-Update (ELU). 2. Data Mapping: Talend Open Studio provides a graphical interface and visual mapping tools for designing and managing data integration workflows. 3. Data Quality: Talend Open Studio offers data profiling, cleansing, and matching capabilities, which helps users ensure data accuracy and consistency. 4. Cloud Integration: Talend Open Studio supports cloud-based data integration, enabling users to manage and process data from various cloud-based services. 5. Real-time Data Streaming: Talend Open Studio supports real-time data streaming and processing, which helps users to manage and analyze large volumes of data in real-time. 6. Development Environment: Talend Open Studio provides an intuitive development environment that enables users to design, develop, and deploy data integration workflows.   Talend Open Studio supports a wide range of data sources, including databases, flat files, cloud services, and enterprise applications. It also offers a vast library of connectors and pre-built components, enabling users to integrate data quickly and easily from various sources.  Overall, Talend Open Studio is a powerful and flexible data integration tool that can help organizations to streamline their data integration processes and improve data quality. It is suitable for organizations of all sizes and can be used for various use cases, including data warehousing, business intelligence, and data migration.  **Oracle Data Integrator:**  Oracle Data Integrator (ODI) is a comprehensive data integration platform that provides a powerful set of tools for managing and transforming data from various sources. It enables users to extract, load, and transform data from various sources, including databases, flat files, and enterprise applications.  Some of the key features of Oracle Data Integrator include:   1. Data Integration: Oracle Data Integrator offers a comprehensive set of data integration tools that enable users to extract, load, and transform data from various sources. 2. Data Mapping: Oracle Data Integrator provides a graphical interface and visual mapping tools that allow users to design and manage complex data integration workflows. 3. Data Quality: Oracle Data Integrator offers a comprehensive set of data quality tools that enable users to profile, cleanse, and match data from various sources. 4. Cloud Integration: Oracle Data Integrator provides cloud-based data integration capabilities, which enable users to manage and process data from various cloud-based services. 5. Real-time Data Streaming: Oracle Data Integrator supports real-time data streaming and processing, which allows users to manage and analyze large volumes of data in real-time. 6. ETL Capabilities: Oracle Data Integrator provides Extract-Transform-Load (ETL) capabilities, which enable users to extract data from various sources, transform it, and load it into target systems. 7. Development Environment: Oracle Data Integrator provides a comprehensive development environment that enables users to design, develop, and deploy data integration workflows.   Overall, Oracle Data Integrator is a powerful and comprehensive data integration platform that can help organizations to streamline their data integration processes and improve data quality. It is suitable for organizations of all sizes and can be used for various use cases, including data warehousing, business intelligence, and data migration. |
| **Output** | **Talend Open Studio working:**  **Reading movies information from a CSV file**    **Adding and linking components**    **Preparing the movies metadata**            **Configuring and executing your Job**        **Filtering the movies information**  **Preparing directors file metadata**            **Duplicating the existing Job**      **Gathering rejected movies information and saving processing results to a database** |
|  |  |
| | **Feature** | **Talend Open Studio** | **Oracle Data Integrator (ODI)** | | --- | --- | --- | | Cost | Free and open source | Proprietary software with paid licenses | | Data integration | Supports various data integration techniques and tools | Provides comprehensive data integration and transformation | | Data mapping | Provides visual mapping tools and graphical user interface | Offers comprehensive mapping capabilities and graphical tools | | Data quality | Supports data profiling, cleansing, and matching capabilities | Provides comprehensive data quality features and functions | | Real-time data streaming | Supports real-time data streaming and processing | Provides real-time data integration and transformation | | Cloud integration | Offers cloud-based data integration capabilities | Provides cloud-based data integration and migration | | Development | Provides drag-and-drop interface and intuitive development | Offers a comprehensive development environment | | Integration with ETL | Provides Extract-Transform-Load (ETL) capabilities | Offers ETL capabilities | | Scalability | Highly scalable and supports large-scale data integration | Offers highly scalable and flexible integration capabilities | | |
| **Conclusion**  To create a data warehouse, a lot of data from various sources is required. This data available at different sources may not fit our design. Hence, ETL tools come to rescue from this problem. ETL tools extract data from various sources and transform them into required format making our  work easier and faster. They also load the formatted data into data warehouse.  **References:**  [**https://www.talend.com/products/talend-open-studio/**](https://www.talend.com/products/talend-open-studio/)  [**https://www.dremio.com/resources/guides/adv-types-etl-tools/**](https://www.dremio.com/resources/guides/adv-types-etl-tools/) | |