

SCHOOL OF COMPUTER ENGINEERING AND TECHNOLOGY

BTech Capstone Project details Year 22-23

- ❖ **Title of the Project:** Standardization of ETL Process
- ❖ **Name of the Students**

1032190030	Utkrist Agrawal
1032190041	Mehul Pansari
1032190048	Prabhat Panwar
1032190109	Nikunj Padia
- ❖ **Group ID:** B56
- ❖ **Name of the Internal Guide:** Dr. Rashmi Phalnikar
- ❖ **Company name:** Symphony Technologies
- ❖ **External Guides:** Mr. Ajay Ghatpande and Mr. Ravi Khare

Summary of Work:

Abstract: The ETL (Extract, Transform, Load) standardization project aims to develop an integrated solution that enables efficient management, processing, and analysis of large volumes of data specific to company's data requirements. The project includes researching and developing an ETL application along with standardization module for interconversions of various data formats, leverages ETL tools and technologies to extract data from various sources, transform it into a format suitable for analysis, and load it into a data warehouse. The project also includes the development of a user-friendly data visualization tool that provides intuitive and interactive access to the data warehouse's contents. The project's objective is to enable the company to make informed decisions based on accurate and up-to-date data, leading to improved business performance and maintain data coming from various sources. The project utilizes an Agile development methodology to ensure flexibility and adaptability to changing requirements and stakeholders' feedback. Overall, the ETL, visualization, and data warehousing project addresses the critical need for effective data management, processing, and visualization in today's data-driven business landscape.

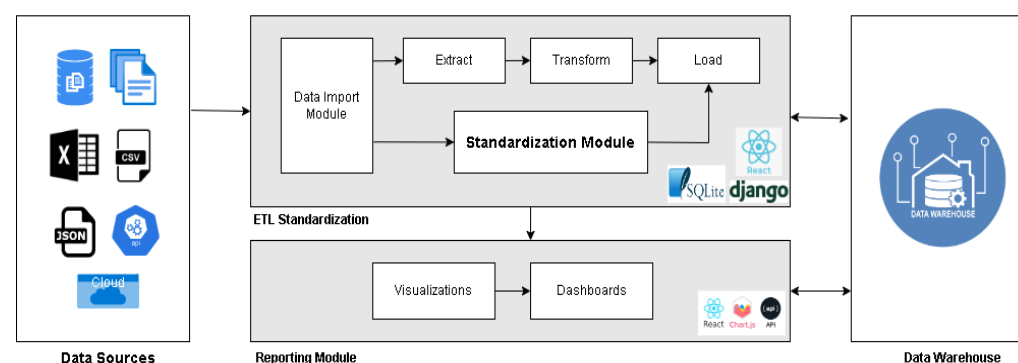
Objectives:

1. To research and arrive at Proof of Concept for ETL standardization application in accordance with company's data requirements.
2. To perform technical assessment of various web technologies along with feasibility study of data warehouse for the given application.
3. To develop an application which will act as intermediary between different data formats and store standardized data in data warehouse.
4. To experience complete Software development Life cycle for the given project domain.
5. To enable the user to import data from any data source such as online or offline data files, process the data and convert it into standard format as per user requirement

Methodology: ETL Application: The ETL application will be responsible for extracting data from various sources, transforming it, and loading it into a target data store. The scope of the ETL application includes identification and integration of data sources.

Standardization: The ETL standardization application would allow the user to convert the data file into specified data format such as JSON, CSV.

Visualization Module: The scope of the visualization application includes data exploration using various interactive visualizations such as charts.



Results: Following are some of the benefits the ETL application would provide to the end-users once deployed:

Accurate Data: The ETL application extracts data from various sources, transforms it into a standard format, and loads it into a database, ensuring that the data is accurate, consistent, and reliable.

Improved Data Analysis: The data standardization component ensures that the data is comparable and can be easily analysed. The visualization tool enables stakeholders to view the data in a way that is easy to understand, identify trends and patterns.

Enhanced Efficiency: The ETL application automates the data processing, reducing manual effort and errors. The data standardization component eliminates the need for manual data cleaning, saving time and resources.

Better Decision Making: With accurate, standardized data and powerful visualization tools, stakeholders can make more informed decisions. The data can be used to identify trends, predict outcomes, and optimize operations.

Conclusion: An ETL application project that includes data standardization and visualization components can be a valuable tool for organizations to gain insights from their data. The project involves extracting data from various sources, transforming it into a useful format, standardizing the data, and loading it into a database. With this foundation, powerful data visualizations and reports can be created to help users better understand and analyze the data. Successfully implementing this project requires a deep understanding of data processing, database management, data standardization techniques, and visualization tools.