**Project Scenario: Cloud-based Data Integration with IBM DataStage**

You have been tasked with building a data integration solution using IBM DataStage, a popular ETL (Extract, Transform, Load) tool, in a cloud environment. The goal of the project is to extract data from multiple sources, transform it, and load it into a cloud-based data warehouse for analysis and reporting.

**Requirements:**

1. **Data Sources:** You will need to integrate data from the following sources into the cloud-based data warehouse:

• **CSV Files:** Data in CSV format from various sources that contain information about customers, products, orders, and sales.

• **Relational Database:** Data from a relational database (e.g., DB2, Oracle, etc.) that contains customer information.

• **REST API:** Data from a REST API that provides real-time data about stock prices.

2. Data Integration Jobs: Design and implement data integration jobs using IBM DataStage to extract data from the data sources, transform it, and load it into the cloud-based data warehouse.

• **Extract:** Use IBM DataStage to connect to the data sources and extract data using relevant connectors such as Sequential File stage, Database stages, and HTTP stages.

• **Transform:** Apply data transformations using IBM DataStage stages such as DataStage Transformer stage or DataStage Aggregator stage to clean, validate, enrich, or aggregate the data.

• **Load:** Load the transformed data into the cloud-based data warehouse using appropriate DataStage stages such as Database stages, Data Warehouse stages, or Cloud-based data warehouse stages.

3. **Cloud-based Data Warehouse:** Use a cloud-based data warehouse platform such as IBM Db2 on Cloud, Amazon Redshift, or Google BigQuery to store and manage the data.

• **Create Data Warehouse:** Create a cloud-based data warehouse instance with appropriate configurations, such as storage capacity, performance, and security settings.

• **Data Warehouse Configuration**: Configure data warehouse settings such as tables, schema, indexes, and partitions to optimize data storage and query performance.

• **Data Warehouse Security:** Implement security measures such as encryption, access control, and authentication to ensure data security and compliance.

4. **Error Handling and Monitoring:** Implement error handling mechanisms and monitoring solutions to ensure the reliability and performance of the data integration jobs.

• **Error Handling**: Implement error handling mechanisms in IBM DataStage, such as error logging, notifications, and alerts, to handle errors and exceptions during data integration.

• **Monitoring**: Set up monitoring solutions using IBM DataStage Director, cloud-based data warehouse monitoring tools, or other relevant tools to monitor the health, performance, and status of the data integration jobs and the data warehouse.

5. **Data Processing and Transformation**: Apply data processing and transformation techniques using IBM DataStage to clean, validate, enrich, and aggregate the data as per business requirements.

• **Data Cleaning**: Use IBM DataStage stages to clean and validate the data by applying data quality rules, filtering, and data validation techniques.

• **Data Enrichment**: Use IBM DataStage stages or custom code to enrich the data with additional information or perform data enrichment tasks such as data enrichment with external APIs, data lookup, or data enrichment with reference data.

• **Data Aggregation**: Use IBM DataStage stages or custom code to aggregate the data by applying data aggregation functions, grouping, or pivoting operations.

6. **Documentation**: Create documentation for the data integration jobs and the cloud-based data warehouse, including design specifications, mapping documents, and user guides, to facilitate future maintenance and support.

**Deliverables:**

• IBM DataStage jobs that extract, transform, and load data from the data sources into the cloud-based data warehouse.

• Cloud-based data warehouse configured with appropriate settings for data storage and management.

• Error handling mechanisms and monitoring solutions implemented