* Monitored and troubleshoot various data pipeline issues to ensure accurate reporting for key stakeholders.
* Managed ETL workflows by gathering requirements, transforming data, and writing optimized code.
* Collaborated effectively with cross-functional teams to understand business requirements and delivered data solutions within data warehouses, generating insightful reports that drive business value.
* Developed solutions to resolve data discrepancies, ensuring seamless data sync between systems that communicated using the pub-sub model.
* Implemented change requests to optimize workflows and address bottlenecks within SQL stored procedures in data warehouses.
* Designed and developed scalable batch and real-time data pipelines, capable of handling large volumes of data.
* Built and deployed scalable ETL Pipelines using Python and Spark.
* Maintained technical documentation for ETL processes, cataloging implemented changes and issue resolution procedures.

**PROJECT OVERVIEW**

**NEAR REAL-TIME DATA PIPELINE**

* Implemented process improvements based on the root cause analysis findings on streaming pipeline issues, to eliminate recurring incidents.
* Managed incident escalation by coordinating with cross-functional teams to resolve critical issues in the pipeline, for smooth data migration.
* Maintained technical documentation to catalog issue resolution procedures.

**BATCH INGESTION PIPELINE**

* Engineered and executed business logic on transformed DWH models in Data Warehouse according to business requirements.
* Crafted solutions to resolve data discrepancies among integrated API sources and Power Bi reports, ensuring accurate reporting for key stakeholders.

**AGN-SAP IMPLEMENTATION (PHARMACEUTICAL)**

* Migrated over 25 TB of acquired SAP data to the client’s scalable system post-acquisition, ensuring seamless alignment with client formatting requirements.
* Validated Informatica Cloud Service mappings (ETL) in UAT by executing unit test plans for exceptional handling of data errors.

**SPARK PERFORMANCE TUNING**

* Crafted a Big Data based solution using the Hadoop ecosystem, Spark, and Python for structured & unstructured data.
* Reduced data shuffling across executors by using the best practices of Spark and optimizing resources by properly configuring the number of executors and memory.