

Professional Summary

Result Driven Data Engineer/ETL developer with 6+ years experience in building Datawarehouse/Data migration applications in finance/telecom sector. Worked with different clients to design, develop and implement different ETL pipelines using various technology stacks. Having good knowledge on hosting Analytical datawarehouse application on prem infrastructure and on cloud platforms. Designed and developed data ingestion and data extraction frameworks for code reusability. Known for providing quick, effective solutions to automate and optimize data management processes. Currently working on sourcing data from different enterprise allegation systems to support the bank’s conduct risk analytical and reporting needs and other high priority projects under allegations space and focus on learning technologies in the DevOps space. Always eager to learn and looking for scope of innovation in day to day work to keep motivated.

Skills

- PySpark
 - Ab Initio
 - SQL
 - Python 3.6 (pandas, SQL Alchemy)
 - AWS(EC2, EMR, Redshift, RDS, S3, Data pipeline)
 - Talend
 - Devops: Docker, Kubernetes, GitLab(CI/CD)
- Teradata
 - Unix
 - Jenkins
 - Autosys
 - MS SQL (SSIS)
 - File Formats(text, csv, parquet, s3objects etc.)

Work History

Data Engineer, 07/2020 to Current

Lockheed Martin Corporation – Huntsville, AL

- Working on Analyzing and converting existing Legacy Ab Initio ETL Data flows to spark(pyspark (RDD, DataFrame, SparkSQL)).
- Worked on ingesting data from different enterprise platforms in to datalake and to Analytical Data reporting databases (Teradata).
- Developed common ingestion and data validations frames works which are driven by YAML, JSON files.
- Worked on optimizing the the spark based data processing jobs by analyzing logs and applying techniques like data salting, broadcasting, repartitioning, etc.,
- Worked on converting user Huge Sandbox SQL’s to ETL to do the heavy lifting of data for reporting.
- Worked on building repots using Microstrategy b connecting to teradata.
- Implemented SCD(slowly changing dimension tables) type 2 data versioning using Spark and SQL.
- Apply data munging to check for no harm process and data salting for join performance improvements.
- Perform end to end data validations by writing test cases.
- Containerize the applications using docker and schedule to run using kubernetes.
- Developed, implemented, supported and maintained data analytics protocols, standards and documentation.
- Collaborated with different Technology teams on ETL (Extract, Transform, Load) tasks, maintaining data integrity and verifying pipeline stability.
- Worked in agile team for rapid application delivery and enhancements.

Application Systems Engineer (Data), 11/2016 to 07/2020

Leidos – Boise, ID

- Worked as a ETL consultant for Wells Fargo, SunTrust banks.
- Displayed responsibility, ownership and accountability for deliverables in development lifecycle.
- Executed various Ab Initio parallelism techniques and developed Ab Initio graphs using pipeline data and component parallelisms.
- Developed conversion and system implementation plans.
- Identify and document all the technical and system specifications by analyzing the existing ETL Code to identify the true data sources for ingestion.
- Adhere to Do-No-Harm approach and migrate to retrofit the existing ETL applications into Data lake to use HDFS storage.
- Capture and ensure End-to-End data flow lineage in Metadata Hub. Utilize features / graphs prebuilt using Ab initio components in the Data Lake Environment.
- Modifying the existing applications to ensure proper integration with Hadoop File System.
- Build ETL packages to load from data lake to IBM DB2 tables for ad hoc query analysis and reporting.
- Worked with off shore team to discuss, guide and track the progress of development. Work on preparing the cut over of the application on to Lake.

Ab Initio ETL Developer, 07/2013 to 12/2013

Sinitron Technologies – City, STATE

- Executed
- Created database objects like Procedures, Triggers and Views using T-SQL in Development and Production environment in SQL Server 2008 R2.
- Created SSIS packages for transferring data from various data sources Excel, *Txt file, *CSV files.
- Conditional Split, and SCD (Slowly Changing Dimension) and Derived column, transformation
- Designed the ETL process, including data read/write, data transformations, errorhandling, process design, logging and Development.
- Generated multiple Enterprise reports using SSRS from SQL Server Database included
- Involved in setting up Ab Initio environment
- Create Generic graphs and Ab initio ETL data flows.
- Analyze the existing SSIS packages and migrate it to Ab initio

Education

Master of Science: Computer Science, 09/2019

Northeastern University - Boston

Certifications

- Azure Administration Essential Training
- AWS Certified Solutions Architect - Associate.
- Databricks Essentials for Spark Developers (Azure and AWS).