Ajinath Gobare Data Engineer

gobareajinath@gmail.com
M.no - 7276362770

Objective

- To be associated with a progressive organization that provides an opportunity for a challenging and rewarding career by applying my knowledge, skills, and potential in the profession. I would also like tomake a positive contribution to your organization by promoting team spirit.

❖ Professional Summary

- ➢ A **Data Engineer** having **4.4** years of experience in design, implement and Support different bigdataapplications by using Apache Spark, Hadoop and AWS.
- ➢ Having experience on Spark SQL, include integration of the Spark with applications.
- ➢ Having experience in designing Data Pipelines and Data Handling.
- ➢ Having experience in all stages of the project including requirements gathering, designing & documenting, development, performance optimization, data extraction, cleaning and reporting.
- ➢ Exposure to AWS cloud technologies such as EC2, S3, EMR, RDS, IAM, Redshift, Athena,Dms etc.
- ➢ Having experience in Hadoop ecosystems and Spark include Hive, Hbase, Spark SQL,Pyspark alongwith various distributions such as Cloudera.
- ➢ Knowledge of programming language Python and its libraries related to bigdata.
- ➢ Knowledge of Spark using python, Spark core and Spark-SQL for faster processing of data.
- ➢ Worked on structured, semi-structured data frame.
- ➢ Also having knowledge of scheduler Airflow.
- ➢ Familiar with PyCharm and Jupyter notebook IDE.
- ➢ Exposure to MySQL, Terradata, Oracle, SQL Server.
- ➢ Comprehensive knowledge of Agile methodology, coupled with excellent communication skills.
- ➢ A good team player, good communicator.
- ➢ Experienced in branching, tagging and maintaining the version across the Environments using tools likeGit, GitHub.
- ➢ Experienced in Kafka data streaming technology.
- ➢ Experienced working with Airflow data as a orchestration tool.
- ➢ Experience in working with different spark API like scala, python.
- ➢ Extensive Hands -on experince in **snowflake** Data Warehouse**.**

❖ Technical Skill Set

- **Operating System** :Unix, Windows
- **Databases** :Mysql, Oracle.
- **Hadoop Ecosystem** :HDFS, Hive, Sqoop, Spark-Core, Spark-SQL, Pyspark
- **IDE** :PyCharm, Jupiter notebook, Databricks.
- **Programming Language** :Python
- **Cloud Technologies** :AWS (RDS,EMR,EC2,S3,IAM, Redshift,Glue,DMS)
- **Data warehouse** : Snowflake
- **Orchestration Tool** : Airflow, Step-funtion
- **Other tools and technologies** :Jira, Git, Github, REST, DBT

❖ Educational Qualification:

- Completed **BCS** from Pune University with 62%
- Completed HSC from Maharashtra State Board, in with 62%
- Completed SSC from Maharashtra State Board, in with 72%

❖ Professional Experience:

- **Organization** : StepStrong Software Pvt.Ltd
- **Designation** : Data Engineer
- **Location** : Pune
- **Duration** : DEC 2020 to till date

Industry Projects:1

From Web to Warehouse: Scraping, Migration & Transformation Mar 2023-Present

- Proficient in writing Python APIs and Python and Pyspark scripts for data extraction and processing.
- Implemented an end-to-end data ingestion pipeline using Python libraries such as Beautiful Soup, Requests, and Selenium to scrape and fetch data from various sources (websites, APIs).
- Designed workflows leveraging AWS EventBridge, Glue, Athena, Lambda Functions, and S3 Buckets along with Azure Blob Storage for seamless and automated data ingestion and management.
- Integrated live streaming data ingestion in batches, enabling real-time data processing alongside scheduled batch jobs.

- Responsible for applying data cleaning techniques to identify and remove inaccurate or corrupted data, significantly enhancing data quality and ensuring the integrity of processed datasets for downstream analytics.
- Involved in data processing by converting various data formats into (e.g., Parquet, XML) and leveraging multiple compression techniques tailored to business requirements.
- Experienced in data processing with Pandas, NumPy, and other Python libraries for efficient manipulation, transformation, and analysis of large datasets.
- Experienced with Apache Airflow and AWS Step Functions for orchestrating and automating complex workflows, managing dependencies, and ensuring reliable execution of data pipelines.
- Implemented data validation, quality checks, and profiling processes to ensure the accuracy and consistency of data, filtering and transforming raw data to generate meaningful insights that aid in client decision-making and business analysis.
- Awarded Best Team Lead for outstanding leadership in managing resources and guiding a high-performing, cross-functional team towards the successful delivery of complex projects

Project :2 **May-2022 to till now**

Data Migration from source (SQL Server) to target (Snowflake).

Migration of data present in various file formats. Migrated data from RDS to Snowflake.
Transformation of data according to ETL specification and business logic.

Roles and Responsibility:

- ➢ Experience in writing pyspark scripts for data extraction.
- ➢ Good experience in spark architecture including spark core and spark Sql.
- ➢ Experience in building optimized data pipeline using S3 ,Glue, PySpark, Athena and Redshift on cloudpremises.
- ➢ Developed batch processing, integrate solutions and process structured and unstructured data.
- ➢ Experience in data ingestion, transformations, and performance tuning
- ➢ Responsible for applying data cleaning techniques and remove inaccurate and corrupted data toimprove data quality which significantly improves data quality
- ➢ Worked with parquet file formats and used various compression techniques to leverage the storage
- ➢ Filtered data and performed multiple operations in order to get meaningful insides that help our client indecision making and analysis
- ➢ Understanding the upstream source nature and work with business cases.
- ➢ Experienced working with Airflow data as a orchestration tool.
- ➢ Experience working with different spark API like scala, python

Environment: Pyspark, AWS (S3, Glue, Athena, RDS), SparkSQL, DMS, Snowflake.

Project :3 **Dec 2020 to Feb 2022**

Financial Intelligence Unit

Credit Card and Debit Card Data in the recent past has seen a huge surge in volume, various productofferings and various incidences of frauds. Banks and Financial Institutions dealing in credit card require information on the credit and debit card on account of the unique nature of credit card exposures.

❖ **Roles and Responsibility:**

- Understand the upstream source nature and work with business cases.
- Involved in Collecting Business Requirements from Business Users, Translate into Technical Design (DataPipelines and ETL workflows).
- Written spark jobs (python) to extract records from downstream source.
- Work closely with the business and analytics team in gathering the system requirements.
- Define all the possible Test Cases along with the Test Data.
- Import and export batch data and delta into Phoenix (Hbase) and Hive using Pyspark.
- Involve in creating Hive tables, loading with data and writing Hive queries.
- Raise JIRA tickets for Infrastructure, Platform issues.
- Resolve JIRA tickets raised by scrum master.
- Perform UT of Big Data implementation.
- Analyzed Data in data warehouse using HUE.
- Reconcile the imported or transformed data as per business requirement.
- Responsible for branching, tagging and maintaining the version across the Environments using SCMtools like SCM as Git, GitHub.
- We have to filter data and perform multiple operations in order to get meaningful insights that will helpthe client in decision making and analysis.
- Implement Data Validation, Quality Checks, Profiling.
- Developed rest api in python.

Environment: Spark SQL, Python, Phoenix, Hbase, Hive, Hadoop, GitHub, AWS (EMR, RDS), S3,REST API