



Shitiz Gagrani

✉ shitiz.gagrani@gmail.com

☎ +919769826264

in [linkedin.com/in/shitizgagrani](https://www.linkedin.com/in/shitizgagrani)

Profile

Data Engineer with 8+ years of experience in designing large-scale data pipelines, ETL workflows, and data transformation. Specializing in Big Data, Cost Management and Project Management, proficient in utilizing Python, SQL, Spark, Databricks, and AWS to deliver scalable big data solutions.

Skills

Big Data Technologies: Spark, PySpark, Spark SQL, Hive, Hadoop, YARN.

Cloud Computing : Databricks, AWS EMR, Kinesis, AWS Glue, Amazon S3

Streaming Technology: Spark Structured Streaming, Kafka

Team Practices: Agile Framework, SCRUM.

Orchestration: Apache Airflow, Jenkins

Programming Languages: Python, SQL

Cost Optimization & Performance Tuning - Consistent watch on cost and optimization of resources such as unutilized or under- utilized jobs, services, images etc.

Excellent communication, Project Management & Analytical skills

Work Experience

NIKE

02/2022 - Present

Senior Data Engineer

Bengaluru, India

- Lead a team of 3 Data Engineers and SPOC for all the reporting and analytics. Closely working with different Stakeholder to drive execution of business plan and product development.
- Handled streaming data using autoloader in Databricks which reduced the end report refresh rate by 10%. Utilised Spark Streaming with Kinesis to transform and load data to multiple sinks (Redshift and S3).
- Collaboration with Architects on various use cases and cost optimization. Actively involved in the Solution Design and architecture of platform to build Core Layer and different components using Micro Services Architecture.
- Designed and triggered corresponding ETL jobs using Databricks using pool instances to reduce the starting time of a job.
- Responsible for Optimizing existing spark jobs which reduced the infrastructure cost by 15% by utilising the correct resources in Databricks.
- Planning & organizing scrum meetings, facilitating sprint ceremonies, sprint planning & review meetings, sprint retrospective meetings, sprint backlog, story estimation, daily stand-ups, iteration & release planning and reviews.
- Spearheaded a major data transformation initiative which included migration from Databricks to AWS EMR resulting in a 20% improvement in system performance and reduction of overall data platform cost.
- Combined Kafka and Spark Streaming to develop a real-time data processing system. This integration handled streaming data, boosting throughput by more than 10% and cutting down latency in analytics reporting by 25%
- Implemented airflow DAGs to submit PySpark jobs onto AWS EMR cluster.

PENTAIR INC.

09/2020 - 02/2022

Technical Lead (Data Engineering)

Noida, India

- Manage project planning, define expected work scopes, and develop detailed timelines to ensure the on-time execution of projects and POCs.
- Frequent client interactions on new opportunities, requirement gathering, issues or concerns. Acting as Scrum Master,

coordinating efforts among geographically dispersed scrum teams to ensure project alignment.

- Reduced SQL query time by 15% by implementing sorting, new block sizing, and strategic partitioning for Parquet data.
 - Vendor management and scaling teams to deliver complex data projects.
 - Developed a Real-Time Streaming Analytics framework that processes over 2 million IoT sensor data points per hour, enhancing operational insights by 20% through real-time analytics like filtering and transforming.
 - Followed the Medallion architecture, building robust, idempotent data pipelines across Bronze, Silver, and Gold layers to ensure data integrity and accuracy.
 - Extract Real-time feed using Kafka and Spark Streaming and process data in the form of Data Frame and save the data as Avro format.
 - Implemented the complete Pyspark framework using Spark SQL and Dataframes for Aggregation with real-time processing.
-

WNS

11/2019 - 09/2020

Senior Data Analyst

Mumbai, India

- The project deals with uploading files, extracts and historical data from on-prem to AWS S3 and performing all ETL operations through Glue, adding triggers via Lambda. Copying data(S3 copy) in Redshift, developing and migrating stored procedures(PL/SQL).
 - Working with python SDK for integrating redshift with glue which reduced cost for querying the database.
 - Developing configurable customized files(json) for SQL queries, business rule which reduced the cost to 35%.
 - Ability to work with serverless architecture, deploying using Jenkins pipeline and code quality with Sonarqube.
-

ACCENTURE

12/2016 - 11/2019

Application Analyst

Bengaluru, India

- Create or modify existing database objects like stored procedures/packages/functions.
 - To find the opportunities for query performance tuning and automation in the applications to reduce repetitive manual tasks and increase productivity of the team.
 - To generate and analyse reports using SQL or Analytics Tools like AWS Quicksight.
 - Understand and explain the relevant information observed from the data by creating dashboards.
 - Experience in doing RCA for any issue in infrastructure or application by debugging logs, source code.
 - Used JIRA for collaboration on code with inline comments and pull requests.
-

Certifications

- AWS Certified Solutions Architect – Associate
 - Databricks Lakehouse Fundamentals - Accrediation
 - Aviatix Certified Engineer
 - Agile Foundation Certified
 - Splunk Certified User
-

Education

SVKM Narsee Monjee University (NMIMS)

04/2012 - 04/2016

B.Tech (Electronics and Telecommunication)

Mumbai, India