Sankalp Singh

Data Engineer

sankalp.s@worksemails.com | +1 (857) 384 8799 | Boston, MA

SUMMARY

- 3+ years of experience in information technology as a Data Engineer with an expert hand in the areas of Database Development, ETL Development, Data modeling, Report Development, and Big Data Technologies.
- Experienced in working in SDLC, Agile, and Waterfall Methodologies.
- Proficiency in Designing Business Intelligence Solutions with Microsoft SQL Server and using MS SQL Server Integration Services (SSIS), MS SQL Server Reporting Services (SSRS), and SQL Server Analysis Services (SSAS).
- Experience working with Amazon Web Services (AWS) cloud and its services like Snowflake, EC2, S3, RDS, VPC, IAM, Elastic Load Balancing, Lambda, RedShift, Auto Scaling, Cloud Front, Cloud Watch, Data Pipeline, ETL and other AWS Services.
- Well-versed in Data Mining solutions to various business problems and generating data visualizations using Tableau, and Power BI.
- Hands-on experience in developing and deploying enterprise-based applications using major Hadoop ecosystem components like MapReduce, YARN, Hive, HBase, Flume, Sqoop, Spark SQL, and Kafka.
- Hands-on experience in SQL and NoSQL databases such as Snowflake, HBase, Cassandra, and MongoDB.

SKILLS

Programming Languages: Python, SQL, Scala, Java
Database Technologies: MySQL, PostgreSQL, SQL Server, Oracle, MongoDB, Cassandra
Big Data Technologies: Hadoop, Spark, Kafka, Flink, Hive, HBase
Cloud Platforms: AWS (Amazon Web Services), Azure, Google Cloud Platform (GCP)
Data Processing Tools: Apache Airflow, Apache NiFi, Apache Beam, Talend, Informatica
ETL/ELT Tools: Apache Spark, Talend, Informatica, AWS Glue, Google Cloud Dataflow
Data Warehousing: Snowflake, Amazon Redshift, Google BigQuery, Azure Synapse Analytics, Apache Hive
Data Modeling: ER Diagrams, Dimensional Modeling, Star Schema, Data Vault
Version Control: Git, SVN
Containerization: Docker, Kubernetes
Scripting: Bash Scripting, PowerShell
Operating Systems: Linux, Windows

EDUCATION

Master in Analytics | Northeastern University, Boston, MA
Bachelor in Electronics and Telecommunication Engineering | University of Mumbai, India

EXPERIENCE

State Street, USA | Data Engineer — Aug 2023 - Current

- Responsible for building scalable distributed data solutions using Big Data Technologies.
- Designing and developing end-to-end ETL Solutions and processing applications using Spark, and Hive to perform Streaming ETL.
- Developing Spark using Scala for data extraction and transforming the data as per the business needs.
- Extract transactional data from MySQL Databases and load the data into Hive tables using Sqoop.
- Created Tableau reports with complex calculations and worked on Ad-hoc reporting using Power BI.
- Used Apache airflow in GCP composer environment to build data pipelines and used various airflow operators like bash operator, Hadoop operators, and python callable and branching operators b) Deployed application to GCP using Spinnaker(rpm-based).
- Optimize SQL queries, resulting in a 40% reduction in query execution time and enabling faster data retrieval for analytics purposes.
- Contributed to an Agile team of developers focused on data ingestion across multiple sources.

Nevina Infotech Pvt. Ltd, India | Data Engineer — June 2019 - Nov 2021

- Implemented data quality checks in Apache Airflow, reducing data errors by 25% and ensuring data consistency across all pipelines.
- Performed data extraction, transformation, loading, and integration in a data warehouse, and operational data stores.
- Aggregated daily sales team updates to send reports to executives and organize jobs running on Spark clusters
- Loaded application analytics data into a data warehouse at regular intervals of time
- Involved in Data Migration using SQL, SQL Azure, Azure Storage, and Azure Data Factory, SSIS, PowerShell.
- Created action filters, parameters, and calculated sets for preparing dashboards and worksheets using Power BI.
- Measured Efficiency of the Hadoop/Hive environment ensuring SLA was met.
- Tuned database configurations in Amazon Redshift, improving query performance by 50% and reducing overall processing time for analytics workflows.
- Implemented a Continuous Delivery pipeline with Docker, Git Hub, and AWS.
- Built performant, scalable ETL processes to load, cleanse, and validate data
- Refactored existing ETL pipelines using Apache NiFi, resulting in a 20% improvement in pipeline performance and reliability.