Abhineet Kumar Sinha

Senior Data Engineer | Schlumberger | Ex-GE | Ex-RIL | IIT (ISM) Dhanbad
Mumbai, Maharashtra, India | Total Exp: 8 yrs
abhineet.sinha1992@gmail.com | Mob.: 08809628590 | Linkedin: https://www.linkedin.com/in/abhineet-kumar-sinha-8122262b

SKILLS

- Data Warehousing
- Talend
- Data Lake Storage
- Azure Data Factory
- Hadoop Batch and Realtime Data Pipelines
- Spark
- Databricks
- Scala/Java
- Azure Data Paltforms
- Kafka
- Azure Stream Analytics
- ETL performance tuning
- Advanced SQL programming
- Snowflake
- Powershell
- PySpark
- Spark SQL
- Data Modelling Tools
- Azure PaaS Services
- Power BI
- Python/R
- DevOps
- CI/CD
- Jenkins
- Git/Gitlab
- Agile/Scrum

LICENSES AND CERTIFICATIONS

- Architecting Big Data Applications: Batch Mode Application
  Issued Sep 2021

- Azure Administration Essential Training
  Issued Oct 2021

- Big Data Analytics with Hadoop and Apache Spark
  Issued Oct 2020

- Dataiku Advanced Designer
  Issued Oct 2020

- Dataiku ML Practitioner
  Issued Oct 2020

- TCA-TIBCO Spotfire
  Issued Aug 2022

- AWS Essential Training for Architects
  Issued Sep 2021

- Dataiku ML practitioner
  Issued Oct 2021

- Cloud Architecture: Advanced Concepts
  Issued Apr 2019

- Mastering Microservices with Java
  Issued Sep 2021

- Neural Networks and Convolutional Neural Networks Essential Training
  Issued Sep 2021

CAREER OBJECTIVE

Data driven software engineer with around 8 years of experience in data sciences and data engineering. Highly analytical and process oriented data analyst with indepth knowledge of database types; research methodologies; and big data capture, curation, manipulation and visualization. Furnish insights, analytics and business intelligence used to advance opportunity identification, process reengineering and corporate growth

EXPERIENCE

SCHLUMBERGER | SENIOR DATA ENGINEER
Oct 2021 - Present | Pune, India

- Provided technical leadership and hands on implementationin areas of data techniques including data access, integration, modelling, visualization and implementation.
- Design and build data processing pipelines using tools and frameworks in the Hadoop ecosystem.
- Design and build ETL pipelines to automate ingestion of structured and unstructured data.
- Working on data warehousing and data lake and integration with legacy platforms.
- very strong understanding and experience with Java, Scala, GoLang and Python
- Master data management, integration , flow and mapping from as-is to to-be systems.
- Design and Build pipelines to facilitate data analysis, data modelling and building ETL pipelines.
- Proficiency in programming language - Python, Java, or Scala
- Working in Big Data Ecosystem with experience in Java, Hadoop, Spark, Snowflake, Cassandra and NoSQL.
- Working with distributed architecture and MPP engine (Spark, Impala etc.)
- Building scalable and highly available distributed systems in production.
- Implementation of CI/CD (Jenkins, Maven), Log aggregation/Monitoring/alerting for production system.
- Experience on Azure and AWS as PaaS.

GE OIL & GAS | DATA ENGINEER
Dec 2020 - Oct 2021 | Bangalore, India

- Design and build data processing pipelines using tools and frameworks in the Hadoop ecosystem.
- Responsible for building and maintaining data architecture and batch/stream data loads for wide-scale customer reporting platform.
- Deployment of ETL load stack on Linux based on-prem enterprise grade servers.
- Creating batch and realtime data pipelines on Azure Data Platform using Logic Apps, Automation Run Books, Azure Data Factory, Azure Databricks, Azure Event Hub, Kafka, Azure Stream Analytics.
- Performance tuning of complex ETL mappings of relational and non-relational workloads.
- Advanced SQL programming on SparkSQL, PySpark, PL/SQL or T-SQL.

- Unix scripting and Power shell scripting for Azure PaaS Services.

RELIANCE INDUSTRIES LIMITED | Data Engineer
Apr 2015 - Dec 2019 | Mumbai, India

- Big Data, Data Analytics, Data Lake ETL experience using Spark scripting, AWS EMR, AWS Redshift, AWS S3, AWS EC2, AWS CloudWatch, AWS IAM, Microsoft SQL, Microsoft SSIS, Java, GitLab, DevOps (CI/CD).
- Handle old and new SSIS jobs related to new Procurement Data Warehouse (PDW), code changes and issue resolution.
- Code management tools (Git/GitHub).
- Cloud deployments of BI solutions including use of AWS eco-systems.
- Experience with SQL-on-Hadoop technology (Hive, Impala, Spark SQL, and Presto).

RELIANCE INDUSTRIES LIMITED | IT-Business Analyst

- Develop business architecture using requirements such as scope, processes, alternatives, and risks.
- Derive meaningful insights after creating and validating Data Model using ML and explain the business value to senior management.
- Building end to end models related to Forecasting, Optimization Models. Etc.
- Evaluating and defining key metrics.
- Automating analyses and authoring pipelines via SQL and Python-based framework.
- Interface with multiple stakeholders and understand business requirements and develop and implement best in class analytics solutions.
- Develop analytical models for enterprise asset management, for MIS reporting and project/shutdown planning, budget planning/allocation and cost reports.
- Create various algorithms for asset performance optimization, predictive analysis and develop solutions based on insights captured from data.
- Data analysis and insights deduction based on perusing of enterprise production data.
- Structure concise reports with insights supported by solid analytics and logical thinking.
- Translate business requirements into tangible solutions specifications and high quality, on time deliverables.
- Conducted data mining, data modeling and statistical analysis to develop suitable simulation models for optimizing production facilities and process reengineering to drive production, enhance system reliability and eliminate production bottlenecks.

PROJECTS:
- Develop and implement a machine learning model to predict equipment/process failure based on historical data and propose corrective action beforehand to mitigate asset breakdown.
- Using clustering algorithm to classify multivariate data sets related to equipment to train data and develop model to predict incident of equipment failure based on the learned data.
- A 15
- Improved production forecasting that reduced deviation from actual production by 17
- Plant-wide optimization model to de-bottleneck process units to maximize profit.