

# AS

## ABHINEET KUMAR SINHA





Big Data Developer |  GE O&G |  Ex-RIL | IIT(ISM) Dhanbad  
Mumbai, Maharashtra, India | Total Exp: 7 yrs  
abhineet.sinha1992@gmail.com | Mob.:08809628590 | LinkedIn:  
<https://www.linkedin.com/in/abhineet-kumar-sinha-8122262b>

























## SKILLS

### OBJECTIVE

Data driven software engineer with around 7 years of experience in data sciences and machine learning. Highly analytical and process-oriented data analyst with in-depth 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.

### LICENSES AND CERTIFICATION

-  Apache Spark Essential Training  
Issued Dec 2020
-  Big Data Analytics with Hadoop and Apache Spark  
Issued Dec 2020
-  Mastering Microservices with Java  
Issued Sep 2020
-  Cloud Architecture: Advanced Concepts  
Apr 2019

- |  |   |   |
|--|---|---|
|  <b>Data Warehousing</b>                  |  <b>Data Lake Storage</b>      |  <b>Azure data Factory</b>       |
|  <b>Hadoop</b>                            |  <b>Spark</b>                  |  <b>Scala/Java</b>               |
|  <b>Batch and Realtime Data Pipelines</b> |  <b>Azure Data Platforms</b>   |  <b>Kafka</b>                    |
|  <b>Azure Stream Analytics</b>            |  <b>ETL performance tuning</b> |  <b>Advanced SQL programming</b> |
|  <b>Snowflake</b>                        |  <b>PowerShell</b>            |  <b>Spark SQL/PySpark</b>       |
|  <b>Data modelling tools</b>            |  <b>Azure PaaS services</b>  |  <b>Power BI</b>               |
|  <b>Python/R</b>                        |  <b>DevOps</b>               |  <b>CI/CD</b>                  |
|  <b>Jenkins</b>                         |  <b>Git/GitLab</b>           |  <b>Agile/Scrum</b>            |

## EXPERIENCE

### BIG DATA DEVELOPER • GE OIL \$ GAS • Dec'19 – Present

- 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.
- Design and Build pipelines to facilitate data analysis.
- Proficiency in a programming language, ideally 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.).
- Experience with SOA & Microservice architecture.
- 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.

#### **DATA SCIENTIST • BAKER HUGHES, A GE COMPANY**

---

- Steered rapid model creation in Python using Pandas, NumPy, Scikit-Learn & plot.ly for data visualization.
- Model development, Validation, Testing and implementation of analytics products and applications.
- Implemented various machine learning techniques to build dynamic chemicals laboratory predictive models and improve engineering efficiency and streamline and expedite customer query requests.
- Worked on project to develop corrosion and scale inhibitor dosage recommendation and predict corrosion rate and scale formation prediction based on laboratory and field data via initial EDA (exploratory data analysis), feature selection and importance.
- Created multivariate regression-based attribution models & segmentation models using K-means Clustering.
- Utilized high dimensional data sets from users, field data, lab data and third-party apps via PCA, LDA & kernel approximations.
- Tested and implemented decision trees, random forests and ensemble methods using bagging and boosting.

#### **DATA ENGINEER • RELIANCE INDUSTRIES LIMITED • Apr'15-Dec'19**

---

- 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, C#, 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).

#### **IT-BUSINESS ANALYST • RELIANCE INDUSTRIES LIMITED**

---

- 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 & cost reports.
- Create various algorithms for asset performance optimization, predictive analysis and develop solutions based on insights captured from data.
- Data analysis & 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% reduction in facility failure cases, resulting in achieving 1 MMSCMD gas production.
- Improved production forecasting that reduced deviation from actual production by 17%
- Plant-wide optimization model to de-bottleneck process units to maximize profit.

## **PRODUCTION ENGINEER • RIL E&P CBM PROJECT • Jul'14 – Apr'15**

- Process engineer responsible for real time monitoring of essential utilities and critical equipment like gas dehydration unit, gas engine reciprocating compressors and gas engine generators.
- Working knowledge of control room equipment as well as on field experience. Experience in troubleshooting of different field elements.
- Used quantitative data gathered to develop an understanding of fluid behavior, pressure variance (both bottom-hole and surface) and PC pumps performance and gas flow.
- Real-time monitoring of PC pumps, performance optimization, cause & effect analysis & troubleshooting through VFD module.
- Dynamic and steady state multiphase simulation through OLGA & PIPESIM.
- Implementation of permit to work (PTW) system in the field.

## **TECHNICAL SKILLS**

- Hands-on: Azure, Python, SQL server, SSIS, Scikit-Learn, SciPy, NumPy, OpenCV, Pandas, TensorFlow.
- Database Platforms: Oracle, PostgreSQL, Teradata, AWS Redshift SQL, Microsoft Azure
- Visualization Tools: Tableau, Power BI, QlikView, SAP HANA.
- Proficient in BI tools (SAP BI, SAP Hana, MicroStrategy, Tableau & QlikView).
- Expertise in RDBMS, SQL, PostgreSQL, Queries and exploratory analysis.
- Functional programming experience with Python, R & SAS.
- Experience with Python packages like pandas, scikit-learn, TensorFlow, NumPy & NLTK.
- Experience with HADOOP stack (Map Reduce, Pig, Hive, Nifi, Spark)
- Power user of Excel, Minitab & other Microsoft office (Word, Outlook, PowerPoint, Visio).
- Experience with statistical modeling languages (e.g. R, MATLAB, Python)
- Conversant with reporting and data visualization tools like Tableau.

## **EDUCATION**

**B.Tech**, INDIAN INSTITUTE OF TECHNOLOGY (INDIAN SCHOOL OF MINES) DHANBAD (Jun'14)

OGPA: 8.41/10,

Summer Internship: IOGPT (Institute of oil and gas production technology), ONGC, Navi Mumbai

**12<sup>th</sup> (HSC)**, CENTRAL BOARD OF SECONDARY EDUCATION (CBSE) BOARD (2010)

1st CLASS: 88.2%

**10<sup>th</sup> (HSC)**, CENTRAL BOARD OF SECONDARY EDUCATION (CBSE) BOARD (2008)

1st CLASS: 94.3%

## **ACADEMIC ACHIEVEMENTS**

- Recipient of the prestigious “Shell Outstanding student award” for the year 2013 given by Shell Technology center, Bangalore on Shell day, held at ISM Dhanbad during March'13.
- Figured among the top 1% candidates successful in IIT-JEE 2010, securing an All India Rank of 4473 in my first attempt.

## **STRENGTHS**

Sincere, diligent, pragmatic, motivated, innovative, function adeptly in team, quick learner