SUDHIR SINGH

New Delhi, India • +91 8587001379 • work.crepantherx@gmail.com • de.crepantherx.in • LinkedIn • GitHub

---

- **Specialized** in optimizing **Apache Spark** Jobs, Azure Databricks, SQL & developing CICD pipelines, with a seasoned perspective honed around **4 years** of collective industry wisdom in crafting **Data Ingestion Frameworks and building DevOps CI/CD solutions** on cutting-edge cloud infrastructure, to the **clients across the globe**

WORK EXPERIENCE

Tiger Analytics / Data Engineering Lead - Data Ingestion Team   *Jan 2023 - Present*
Chennai, INDIA
- Leads team of 7 data engineers developing common data foundation layer across entire firm
- Processed & Migrated 10 PetaBytes of data into ADLS as a part of Common Data Foundation
- Architected & developed metadata driven object oriented PySpark Data Ingestion framework, allowing direct client configuration changes and cutting development needs by 60%.
- Implemented Azure DevOps CI/CD pipelines, automating 95% of deployments.
- Designed automated data ingestion framework for diverse file types in Raw Layer.
- Created PII obfuscation and field encryption framework for datasets.
- Designed PyDeequ framework wrapper for custom DQC integration.
- Implemented logging framework for data ingestion steps to log analytics.
- Designed Python data ingestion framework for firm-wide adoption for all big data ingestion

Tech Mahindra / Data Engineer - Data Quality Team   *Jan 2021 - Jan 2023*
Sydney, AUSTRALIA & Noida, INDIA
- Translated 2+ on-prem architecture to the cloud, incorporating Big Data capabilities.
- Architected data ingestion from 15+ sources into ADLS, processing with Spark
- Implemented Python project exceeding 15,000 lines, after in-depth data analysis, enhancing data-driven decisions for 75% increase in technician allocation efficiency in telecom.
- Optimized complex SQL queries, reducing operational costs by 10%.
- Analyzed partitioned and bucketed data from 15+ sources ingested into Hive using Hive queries.
- Created 15+ notebooks for intricate security anomaly analysis using Regular Expressions.
- Proficient in containerization and orchestration services, deploying 3+ CI/CD pipelines.
- Established 3+ ML CI/CD pipelines on AWS using Kubernetes, Jenkins, and GitHub.
- Led a 3-month mentorship program for 50+ employees, focusing on Databricks, SQL

CERTIFICATIONS

- **Microsoft** / Microsoft Certified: Azure Data Scientist Associate, **DP-100 Exam**   *Jun 2022*
- **Microsoft** / Microsoft Certified: Azure Data Engineer Associate, **DP-203 Exam**   *Jun 2022*
- **Google** / Google Certified: Data Analytics Professional   *Dec 2021*
- **Amazon** / AWS Certified: Cloud Practitioner Trainee   *Jul 2021*

ACHIEVEMENTS

- **Tech Mahindra** / Bravo Award - Exceeding client expectation   *Mar 2022*
- **Tech Mahindra** / Pat On The Back Award - Trained 50+ new employees   *Dec 2021*
- **Tech Mahindra** / Bravo Award - For showing outstanding development results   *Jun 2021*

EDUCATION

Gautam Buddha University / Bachelor's Degree, Information Technology   Aug 2017 - Jul 2021

SKILLS

- **Data Science**: Scikit-learn, TensorFlow, Keras, PyTorch, Matplotlib, NumPy, SciPy
- **Big Data Processing Framework**: Apache Spark, Apache Hive, Apache Hadoop, Apache Kafka
- **Distributed Data Storage System**: Apache Cassandra, Amazon S3, ADLS, Apache HDFS
- **Data Warehousing**: Snowflake, BigQuery, and Redshift
- **Databases**: Microsoft SQL Server, MySQL, PostgreSQL
- **Language & Libraries**: Python, PySpark, SparkSQL, Pandas, Seaborn, Spacy, NLP
- **DevOps**: Azure DevOps, Splunk, Kubernetes, Kubeflow, Jenkins, JFrog, Docker, GIT
- **Linux Skills:** File System, System Permissions, VIM Editor, Data Processing, Data Excursion
- **Tools**: SSMS, DBeaver, Ambari, Apache Zeppelin, Jira, Confluence, Agile Project Management