

# SUDHIR SINGH

New Delhi, India · work.crepantherx@gmail.com · +91 8587001379

[LinkedIn](#) · [GitHub](#) · [Website](#)

## SUMMARY

---

**Specialised in Apache Spark** with 3.6 years of expertise in designing and **delivering Big Data ingestion workflow and DevOps CI/CD pipelines** on Azure Cloud to numerous clients across the globe. Additionally, **Mentored 60+ working professionals in SQL** over a period of 4 months.

## SKILLS

---

- **Transactional DBMS:** Microsoft SQL Server, MySQL, PostgreSQL
- **Distributed Storage System:** ADLS, Amazon S3
- **Big Data Processing & Streaming:** Apache Spark, Apache Kafka
- **Data Orchestration:** Apache Airflow, Azure Data Factory
- **Data Processing:** Python, PySpark API, SparkSQL API, Pandas, SpaCy, NLP
- **Data Warehousing:** Snowflake, BigQuery, Redshift
- **Machine Learning:** Scikit-learn, TensorFlow, Keras, Seaborn, Matplotlib, NumPy, SciPy
- **DevOps:** Azure DevOps, Splunk, Kubernetes, Kubeflow, Jenkins, JFrog, Docker, GIT

## WORK EXPERIENCE

---

### Tiger Analytics

*Data Engineer*

Remote

Jan 2023 - Present

- **Migrated and processed 10 PetaBytes of data** into Azure Data Lake Storage as a part of the Common Data Foundation, streamlining the data foundation layer for improved accessibility and efficiency, resulting in a more robust data ecosystem.
- **Engineered & Delivered a metadata-driven, object-oriented PySpark Data Ingestion framework paired with Azure DevOps CI/CD pipelines** that empowered automated deployment of client configuration changes, reducing re-development needs by 90% and minimizing further operational cost.

### Tech Mahindra

*Data Engineer*

Sydney, Australia

Jan 2021 - Jan 2023

- **Delivered** a decision support system **to help technician to relocate based on forecasted ticket numbers** before, backed up by Machine Learning model deployment and prediction, leading to 95% improvement in ETA estimation.
- **Ingested Billions of customer data** into cloud Hortonworks based Data Ingestion framework.
- **Led the establishment of 3+ ML CI/CD pipelines on AWS using Kubernetes**, Jenkins, and GitHub, resulting in improved deployment processes and increased productivity for the team.
- **Conducted** comprehensive training sessions for 60 employees on SQL Analytics, data modeling techniques, and building scalable data pipelines using PySpark and Apache Airflow.

## EDUCATION

---

- **Liverpool John Moores University**  
Master of Science in Machine Learning & AI

Liverpool, UK  
Graduation Date: Sep 2025

- **Gautam Buddha University**  
Bachelor of Technology in Information Technology

Greater Noida, UP  
Graduation Date: Jul 2021

## ACHIEVEMENTS

---

- **Tech Mahindra** / Bravo Award - Project Delivery Jun 2021 & Mar 2022
- **Tech Mahindra** / Pat On The Back Award - Trained 50+ new employees Dec 2021

## CERTIFICATIONS

---

- **Microsoft** / Azure Data Scientist Associate, DP-100 Exam Jun 2022
- **Microsoft** / Azure Data Engineer Associate, DP-203 Exam Jun 2022

**Tiger Analytics, Remote**

- **Common Data Foundation** Jan 2024 - Present
  - **Developed & migrated unified foundation layer for Retail data** coming from various source systems.
  - **Lead for delivery** of data sources migration from Databricks to on-prem frameworks.
  - **Developed** Python frameworks for **data quality checks, PII encryption, and data transformation** from raw to trusted and feature layers.
  - *Tech Stack: Azure Databricks, Python IDE based Development, API, Large Data Volume*
- **Launchpad** Feb 2023 - Dec 2023
  - **Lead the end-to-end design and development** of an object-oriented PySpark-based Data Ingestion framework, increasing data processing speed by 30%.
  - Built Azure DevOps CI/CD pipeline, automating 95% of deployments and accelerating timelines.
  - *Tech Stack: Azure DataFactory, Azure DevOps CI/CD, Databricks, Logic Apps, Log Analytics*

**Tech Mahindra, Sydney, Australia**

- **DQR-DAC Ingestion System** June 2022 - Jan 2023
  - **Implemented** a workflow to determine customers' eligibility for migration to new strategic plans.
  - **Developed & maintained** data ingestion layer, audit & configuration layer, Hadoop layer for seamless operational needs as per GDPR compliance.
  - Transformed 25+ data sources using Scala as per business & operational requirements.
  - Provided operational support in detecting any discrepancy in 25+ data sources.
  - *Tech Stack: Hadoop, Scala, Hive, Delta Tables, Linux, Zeppelin, Ambari, Jira, Confluence*
- **Pharma Insights Data Extraction & Analytics** Apr 2022 - May 2022
  - **Created ADF pipeline to fetch data from S3 to ADLS**, using ELTL approach.
  - Written Pyspark script in Azure Databricks to extract text from images.
  - Categorized data retrieved into JSON objects, loaded into ADLS.
  - Provided analytics insight report using Azure Databricks & presented it to clients.
  - *Tech Stack: SQL, Python, PySpark, Azure Data Factory, Data Bricks, Data Lake, Key Vault*
- **Decision Support For Technician Allocation** Aug 2021 - Mar 2021
  - **Improved end-to-end service processes & cut down manual deployment** by 75% by predicting proper utilization of technicians and minimizing customer wait time for best customer experience by generating 3000 recommendations per day, in near real-time, to mitigate the risk of the day.
  - **Deployed** 5+ container-based applications orchestrated with Kubernetes.
  - Established 2+ CI/CD pipelines of ML models on AWS.
  - **Written 40+ highly optimized SQL stored procedures** to transform/aggregate raw data, API integrations to fetch live data.
  - *Tech Stack: Python, SQL, ML, CI/CD, AWS, Kubernetes, JFrog, Jenkins, Kubeflow*
- **Control Tower Intrusion Detection Analytics** Jan 2021 - July 2021
  - **Worked closely with stakeholders, BA, & scrum masters** to gather the business requirements.
  - Extracted complex Fields from different types of Log files using **Regular Expressions** in Databricks.
  - **Exploratory data analysis to analyze security anomalies** from server's generated data.
  - Wrote and optimized simple and complex SQL queries to load information extracted into Splunk datasets as per business needs.
  - *Tech Stack: Azure Databricks, Azure Datafactory, CI/CD, Spark SQL, PySpark, Data Lake, Delta Tables*