

# Sai Prasad Padhy

+91-9550003197 | saiprasadpadhy@gmail.com  
LinkedIn | GitHub

## PROFILE

---

Data Engineer with 5+ years of experience building large-scale data pipelines, ETL processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark, Databricks, and ADF** to develop multi-terabyte scalable big data solutions for Fortune 100 manufacturing and pharmaceutical companies.

## TECHNICAL SKILLS

---

**Programming Languages:** Python, SQL  
**Big Data Technologies:** Spark, PySpark, Spark SQL, YARN, Kubernetes, Hadoop, Hive, Impala  
**Streaming Technology:** Kafka, Spark Structured Streaming  
**Cloud computing:** ADF, Databricks, ADLS, Amazon S3, Event Hub, Azure VM  
**Data Engineering Tools:** Azure DevOps, Data Modelling, ETL/ELT data Pipeline, Delta Lake  
**Orchestration:** ADF, Airflow, Redwood  
**Familiar:** FastAPI, Synapse, Microsoft Fabric

## EXPERIENCE

---

- Tech Mahindra** January 2022 - Present  
*Senior Data Engineer* Hyderabad, India
- Led a team of 2 Data Engineers in designing, developing, and optimizing data pipelines and ETL processes.
  - Developed an automated system by aggregating data from multiple sources where business can easily find reliable customers for procuring raw materials.
  - Integrated and processed Procurement and shipment data from multiple sources including **AWS, Azure, SQL Server, and SAP**, handling various file formats like JSON, CSV, and Parquet.
  - Implemented Spark optimization techniques such as **caching, multithreading, broadcast joins and Delta lake optimizations**, resulting in a 20% decrease in processing time for handling a daily load of around **3 Million records**.
- Tech Mahindra** January 2022 - Present  
*Data Engineer* Hyderabad, India
- Migrated the entire data platform from on-prem to Azure Databricks, including code migration, code optimization, and data validation between on-premises and Databricks, achieving a 30% cost reduction.
  - Designed and developed advanced scheduling capabilities using **Azure Data Factory (ADF)**, reducing manual intervention and streamlining workflows.
  - Developed a robust framework to handle data from various file formats, including **CSV, Excel, JSON, and Parquet**, with support for multiple load types such as **full, append, and merge**.
- Cognizant** June 2019 - January 2022  
*Data Engineer* Chennai, India
- Delivered a project to migrate legacy on-premise processes to the cloud using Big Data technologies (Spark), reducing processing time by 20%.
  - Developed Azure Data Factory pipelines to perform daily data migration from Microsoft Dynamics to SQL Server, handling approximately 40 million records across 25 tables.
  - Optimized overall process performance through **Spark performance tuning**, improving job run times by 20% and efficiently managing a **15Terabyte(TB)** dataset containing approximately **12 billion** records.
  - Created Power Automate, Azure functions to ingest data from SFTP, SharePoint to ADLS

## EDUCATION

---

**Lovely Professional University** Jalandhar, Punjab  
*Bachelor of Technology in Mechanical Engineering* July 2015 - June 2019

**Sri Chaitanya Junior College** Hyderabad, Telangana  
*Senior Secondary School* April 2013 - Mar 2015