

# Sai Prasad Padhy

+91-9550003197 | saiprasadpadhy@gmail.com

[LinkedIn](#) | [GitHub](#)

## PROFILE

---

Data Engineer with 6+ 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, Auto Loader  
**Cloud computing:** ADF, Databricks, ADLS, Amazon S3, Event Hub, Azure VM, AWS Lambda, AWS DMS  
**Data Engineering Tools:** Azure DevOps, Data Modelling, ETL/ELT data Pipeline, Delta Lake  
**Orchestration:** ADF, Airflow, Redwood  
**Familiar:** FastAPI, Synapse, Microsoft Fabric

## EXPERIENCE

---

- Nagarro** April 2025 - Present  
*Senior Data Engineer* Hyderabad, India
- Designed and implemented comprehensive data quality metrics for 100+ source systems, encompassing **database ingestion, file-based sources, and Kafka streams** to ensure robust data integrity.
  - Explored and built proof of concepts (POCs) on **Lakebase** and **Lakeflow**, evaluating their capabilities for enhancing modern data platform architecture.
  - Analyzed and optimized existing data pipelines, resulting in a **50% reduction in total operational costs** while maintaining performance and scalability.
- Tech Mahindra** January 2022 - March 2025  
*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**.
- 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
- ZP High School** R.Belagam, Andhra Pradesh  
*Secondary Education* April 2012 - Mar 2013