Ananya Rao

Data Engineer · 5 years · Azure · Microsoft Fabric · PySpark

ananya.rao@example.com +91 98765 43210 Bengaluru, Karnataka

SUMMARY

Data engineer with five years of experience designing batch and incremental pipelines in Python, SQL and PySpark. Built and run lakehouse platforms on Azure Data Lake Storage Gen2 and Microsoft Fabric for retail and fintech analytics. Comfortable owning a pipeline from ingestion to the Power BI model, with a focus on reliable, re-runnable loads and fast queries.

EXPERIENCE

Senior Data Engineer · Northwind Retail Analytics **Apr 2024 – Present**
Bengaluru · Retail analytics platform for 300+ stores

- Designed a Microsoft Fabric lakehouse (bronze, silver, gold) bringing sales, inventory and returns from five source systems into one reporting layer.
- Rebuilt daily incremental loads to be idempotent (watermarks moved only after success, MERGE upserts), ending duplicate and missed records on reruns.
- Served gold tables through the Fabric Warehouse to 40+ Power BI reports; cut the main sales report refresh from 25 to 6 minutes.
- Used Dataflows Gen2 for low-code clean-up and PySpark notebooks for high-volume transformations, orchestrated with Fabric pipelines.

Data Engineer · Bluefin Payments **Jun 2021 – Mar 2024**
Pune · Transaction analytics

- Built Azure Data Factory pipelines (copy, data flow and stored-procedure activities) loading 20M+ transactions a day into ADLS Gen2.
- Tuned PySpark jobs with broadcast joins, partition sizing and skew handling; brought a nightly job from 2 hours to 25 minutes.
- Wrote SQL models for fraud and settlement dashboards using window functions and incremental aggregates.

Associate Engineer · Data Services · Kestrel Infotech **Jul 2020 – May 2021**

- Maintained SQL Server ETL jobs and Python scripts; added data-quality checks and alerting.

SKILLS

Platforms	Microsoft Fabric (Lakehouse, Warehouse, Dataflows, Pipelines), Azure Data Factory, ADLS Gen2, Azure Synapse, Databricks (proof of concept)
Processing	PySpark, Spark SQL, Delta Lake, Parquet
Languages	Python, SQL (T-SQL, Spark SQL)
Modelling	Medallion architecture, dimensional models, incremental and CDC loads
Tools	Power BI, Git, Azure DevOps

EDUCATION & CERTIFICATIONS

- B.Tech, Computer Science · 2020
- Microsoft Certified: Fabric Analytics Engineer Associate (DP-600)
- Microsoft Certified: Azure Data Engineer Associate (DP-203)