

AKASH KARTHIK P

Data Engineer / ETL Modernization / PySpark / Azure Databricks

Chennai, India | +91 8754064243 | connect@akashkarthik.in
linkedin.com/in/karthik-akash | akashkarthik.in

Professional Summary

Data Engineer with around 5 years of experience designing, modernizing, and supporting enterprise-scale data platforms. Experienced in building scalable ETL solutions, modernizing legacy data platforms, optimizing high-volume batch processing, and owning production-critical data systems throughout their lifecycle.

Recognized for simplifying complex engineering problems, improving platform reliability, and translating business requirements into scalable, maintainable data solutions.

Professional Experience

Neurealm | Data Engineer | 2021 – Present

Designed, developed, and supported enterprise data platforms processing workloads ranging from frequent incremental feeds to 10+ TB/day batch pipelines, integrating data across multiple enterprise systems for business-critical reporting and analytics.

Enterprise Solution Engineering

- Partnered directly with client stakeholders to gather requirements, analyze business problems, perform enterprise data mapping across multiple source systems, and deliver production-ready integration solutions.
- Designed and implemented an automated patient outreach platform with a custom Insert/Update/Delete (I/U/D) framework preserving historical records while generating accurate downstream integration extracts.
- Designed enterprise data integration workflows across clinical, operational, billing, provider, and master data systems.
- Developed secure SFTP-based extraction processes for downstream business applications.
- Redesigned workflow execution around upstream dependencies and SLA constraints.
- Delivered a solution contributing to an estimated 20–40% increase in preventive mammogram screening participation based on client feedback.

Cloud Modernization & Performance Engineering

- Joined a cloud modernization initiative after the platform had operated without a dedicated Hadoop engineer for nearly a year, rapidly reverse-engineering undocumented Hadoop MapReduce workflows.
- Redesigned legacy Hadoop processing for Azure Databricks by eliminating redundant ETL stages and adopting cloud-native processing patterns.
- Collaborated with cross-functional engineering teams to simplify migration architecture.
- Reduced batch runtime from 12–14 hours to approximately 5 hours through workflow redesign and PySpark optimization.
- Consolidated multiple MapReduce programs into a single maintainable PySpark notebook per dataset.
- Validated every migrated workload through parallel execution, production data reconciliation, defect analysis, and business-user verification.
- Supported end-to-end migration including defect resolution, deployment, post-release validation, and stabilization.

Production Engineering & Platform Ownership

- Served as ETL SME owning production support, release approvals, code reviews, incident resolution, root-cause analysis, and long-term platform stability.
- Approved pull requests, prepared release packages, resolved merge conflicts, and coordinated deployments.
- Stabilized migrated workloads across multiple enterprise data platforms.
- Authored comprehensive technical and operational documentation enabling seamless ownership transition.

Technical Skills

Programming: SQL, PySpark, Spark SQL, Shell Scripting

Cloud: Azure Databricks, Azure Data Factory, Azure Data Lake Storage

Big Data: Hadoop, Hive

Databases: Teradata, PostgreSQL, Oracle, SQL Server, Netezza

ETL: Ab Initio, IBM DataStage

Version Control: Git, Bitbucket, Azure DevOps

Methodology: Agile Scrum

Education

Bachelor of Engineering – Computer Science & Engineering

MNM Jain Engineering College

Certifications

- Azure Databricks for Data Engineering – Microsoft (Coursera)
- Big Data with Hadoop & Spark – IBM
- Introduction to Generative AI – Google