NAVNATH JARARE

AZURE BIG DATA ENGINEER

PHONE | (+91) 9665246184

EMAIL | navnath.azure@gmail.com

LOCATION | Pune, INDIA

EXPERIENCE | 8 Years 0 Month

Key Skills

- Data Lake
- Hdfs
- DSA
- Ci/Cd
- Framework
- Data Pipeline
- Spark
- Pyspark
- Azure Logic Apps
- Azure Key Vault
- Azure Databricks
- Azure Data Factory
- Azure Data Lake
- Azure Blob Storage
- Hadoop
- Big Data
- Python
- Pandas
- SQL
- Big Data Engineer
- ETL
- Databricks Delta Live Table
- RAG Langchain Vector Db Fine Tuning
- RAG
- QMS System
- Rest API Integration

Profile Summary

I am an Azure Data Engineer with over 8 years of experience, specializing in Azure services, BI reports, and Gen AI PoC technology. I have extensive expertise in pyspark for banking data analysis, as well as in Azure Databricks, Azure Data Factory, Azure Data Lake, and Azure SQL Database. I am skilled in Python, SQL, Big Data, Azure Function Apps, Azure Key Vault, Azure Logic App, data warehousing, data modeling, and SCD. My background includes 5+ years of hands-on experience with Azure services and 2 years in BI reports. I am passionate about leveraging my skills to drive impactful data engineering solutions.

Work Experience

Azure Big Data Engineer

Globant

03/2024 - Present

Working as Azure Big Data Engineer in Globant Designing scalable Azure data infrastructure (ADLS, Databricks, ADF) for efficient large-scale data processing and ETL/ELT workflows.Drove experimental Azure and Gen AI initiatives (PoCs) leveraging Azure AI stack while optimizing storage/retrieval to support analytics and emerging AI solutions.

Senior Data Engineer

Happiest Minds Technologies

10/2023 - 03/2024

Designed and implemented scalable Azure Data Factory (ADF) and Databricks pipelines for enterprise clients like Pepsico and Spencer Gifts, enabling

- Agentic Ai
- Kafka
- Airflow

Certification

- Azure Data Engineer DP-203
- Salesforce Data Engineer
- GCP Data Engineer
  Valid upto December 2027

Languages

- English
- Hindi

efficient retail data migration and ETL processes.
- Spearheaded PII data security by developing encryption/decryption frameworks for Spencer Gifts, ensuring compliance with data protection regulations (e.g., GDPR).
- Optimized data workflows, reducing processing time by 30% through parallel execution and Delta Lake integrations.

Big Data Engineer
NielsenIQ
02/2022 - 10/2023

Built reusable PySpark pipelines on Azure Databricks for large-scale data ingestion (100+ TB), transformation, and Delta Lake storage, serving clients like Wells Fargo.
- Engineered data validation and risk-processing pipelines for regulatory compliance, improving data accuracy by 25%.
- Collaborated with cross-functional teams to align data architecture with business requirements for financial analytics.

Design Engineer
Valiant Tms Systems
09/2017 - 01/2022

Developed interactive Power BI dashboards and reports for manufacturing operations, enabling data-driven decision-making.
- Modeled ERP data (e.g., SAP) into star schemas to streamline reporting and ad-hoc analysis.
- Partnered with engineering teams to extract insights from IoT and production data, reducing downtime by 15%.

Education

B.Tech/B.E. - Mechanical
2015

Pune University

Projects

PepsiCo Data Cleaning & migration to Azure
11 Months

Project:Centralized Data Platform Migration & Analytics | PepsiCo US Retail

Role:Azure Big Data Engineer
Migrated multi-source retail data (1,900+ stores) using ADF for extraction/ingestion into ADLS. Engineered Azure Databricks for complex transformations to enable analytics on customer behavior, operations, and revenue streams.
Key Contributions:
* Designed ADF pipelines to extract source data (merchandise, credit ops) and land raw datasets in ADLS.
* Led heavy-lifting transformations using Azure Databricks (cleansing, enrichment, aggregation).
* Developed scalable Databricks jobs processing large-scale data, optimizing performance/cost.
* Built reusable ADF pipelines orchestrating data flow between sources, ADLS, and Databricks.
* Created analytics-ready data models supporting KPIs: customer acquisition, operational efficiency.
* Ensured data integrity through source team collaboration.
* Enabled self-service analytics for data-driven decisions across retail portfolio.

BANK project
7 Months

- Optimized and maintained Big Data pipelines for risk reporting and board preparation at HSBC Bank, utilizing PySpark, SQL, and Spark for efficient data processing and transformation of large-scale financial data.

- Managed pipeline automation and version control through CI/CD, GitHub, and Jupyter Notebooks for collaborative data analysis and code deployment, ensuring streamlined, reliable reporting and improving the performance and scalability of financial data processing workflows.Utilized PySpark, Python, Hadoop, and Delta Lake to handle and process extensive amounts of data for the project, achieving a 30% improvement in processing speed.Executed data pipelines for efficient processing and analysis of large datasets, resulting in a 25% reduction in processing time and a 15% improvement in data accuracy.

AI Multi Agent application
5 Months

Project Description:
Developed a GenAI-powered multi-agent system

designed to streamline and automate end-to-end project management workflows. The system uses a network of collaborative AI agents that communicate.

Key Features & Capabilities:

- **Multi-Agent Collaboration:** Designed and implemented multiple specialized AI agents (e.g., Requirement Agent, Document Agent, Task Planner Agent, Jira Agent) that work in tandem, sharing context and outcomes to ensure seamless task execution.
- **Automated Requirement Analysis:** Agents parse and analyze incoming business documents (PDFs, emails, requirement specs) to extract structured project requirements.
- **Proof of Concept (PoC) Generation:** The coding agent generates initial PoC code structures, boilerplates, or prototype components based on requirement types and tech stacks

Tech Stack:
Python, LangChain, OpenAI API (GPT-4), Vector DB (Pinecone/FAISS), Jira REST API, FastAPI, fine tuning, prompt engineer

Impact:
Reduced project setup time by over 60%

Salesforce, SAP Members firm extraction project
7 Months

Project: Global Member Firm — Metadata-Driven File Delivery Pipeline
Built metadata-driven ADF pipelines that dynamically controlled source-to-destination flows using SQL Server-based metadata table
Used Azure Logic Apps and Function Apps to move generated files from Azure Data Lake Storage Gen2 to SharePoint Online,
Integrated logging, error handling, and alerting mechanisms tied to metadata to support scalable operations.

Tools/Technologies Used: Azure Data Factory, Azure SQL MI, SQL Server, Azure Data Lake Gen2, Salesforce, SAP HANA, Logic Apps, Function Apps, Graph API, JSON, CSV