

# Sahith Jangiti *Azure Data Engineer*

✉ sahith889@gmail.com

☎ +919533887177

📍 Bangalore

## PROFILE

---

Azure Data Engineer with **3.6 years of experience** in designing and delivering end-to-end data analytics solutions. Proven expertise in **data ingestion, ETL pipeline development, and large-scale data processing** using **PySpark and Spark DataFrames**. Skilled in building **scalable, reliable, and automated data workflows** using the Microsoft Azure ecosystem, including **Azure Data Factory, Azure Data Lake, Databricks, and Power BI**. Adept at optimizing data pipelines, implementing CI/CD practices, and collaborating with cross-functional teams to deliver business-critical insights.

## SKILLS

---

### Big Data Technologies

Spark, Data Modelling, Data Warehousing, Delta Lakh, Data Migration.

### Programming

Python, SQL

### Cloud Services

Azure Cloud, Azure Data Factory, Azure Databricks, Azure Storage, Azure ADLS, Azure Synapse Analytics, Azure SQL Databases, Azure Logic Apps.

### Operating Systems:

Windows, Linux, Ubuntu

### Spark API's

PySpark, Spark SQL

### Data Manipulation

SQL, PL/SQL

### Data Visualisation

Power BI

## KEY ACHIEVEMENTS

---

### Data Migration Success

Achieved 100% success rate in data migration with zero downtime from on-prem to Azure

### SQL Optimization

Reduced SQL procedure runtime from 2 hours to 20 minutes by optimizing queries and indexing

### Stakeholder Engagement

Increased stakeholder satisfaction by 15% through real-time Power BI dashboards

## EXPERIENCE

---

### Azure Data Engineer

Flipkart Ltd

04/2022 – present

### Project#2: Analytics Platform for Flipkart Wholesale (Best Price)

Aug 2023 - present.

**Tech Stack:** Azure Data Lake Storage Gen2, Azure Data Factory, Azure SQL DB, T-SQL, Azure Databricks, PySpark, Power BI, Azure Monitor, Git, Azure DevOps.

**Description:** Best Price is Flipkart's B2B platform catering to wholesale buyers across India. The objective of this project was to build a robust and scalable analytics platform that could handle large volumes of transactional and operational data from wholesale stores. The solution aimed to provide actionable insights to business stakeholders by leveraging the Microsoft Azure ecosystem for data ingestion, transformation, and reporting.

**Responsibilities:**

- Ingested raw data securely from on-premise/private networks into **Azure Data Lake Storage Gen2 (ADLS Gen2)** using **Azure Data Factory**.
- Transformed raw data using **SQL** and designed **analytical views, fact, and dimension tables** for reporting purposes.
- Utilized **Azure Databricks** (PySpark) for data validation, ad-hoc analysis, and performance tuning during early stages of development.
- Optimized long-running **SQL stored procedures**, improving performance from 2 hours down to 20 minutes.
- Built interactive **Power BI dashboards** to visualize key business KPIs across wholesale operations.
- Automated daily data refreshes and extract deliveries using **ADF pipelines** with **email alerts**.
- Implemented **CI/CD pipelines** with **Azure DevOps**, managing all pipeline and code deployments through Git integration.

### **Project#1: Unified Data Platform (UDP) for Customer Service and Support (CSS)**

*Developed the UDP system to centralize and streamline Flipkart Customer Support analytics.*

**April 2022 - July 2023**

**Tech Stack:** Azure Data Lake Storage Gen2, Azure Data Factory, Azure Databricks, PySpark, Delta Lake, Azure SQL DB, OLAP Cubes, Power BI, Git, Azure Monitor, Azure DevOps, Key Vault

**Description:** The Unified Data Platform (UDP) is a centralized big data platform designed to consolidate Flipkart Customer Service and Support (CSS) data from multiple sources into a unified Azure-based system. The platform supports advanced analytics, root cause analysis, and performance monitoring for internal support operations.

#### **Responsibilities:**

- Extracted large volumes of raw data from multiple enterprise applications and ingested into **Azure Data Lake Storage Gen2** using ADF.
- Transformed datasets using **PySpark** on **Azure Databricks** to produce clean, structured **Dimension and Fact Tables**.
- Tuned and optimized PySpark notebooks, reducing pipeline execution time and improving cluster utilization.
- Loaded processed data into **Data Marts** and **OLAP Cubes** for downstream reporting.
- Interacted with end-users and business analysts to gather detailed requirements and translate them into reusable **metrics and dimensions**.
- Managed notebooks and pipeline versions through **Git integration** with Databricks and monitored job health using **Azure Monitor** and **Log Analytics**.

### **Data Analytics Intern**

10/2021 – 03/2022 | Bangalore

#### **MK Corporate Sol**

- Created scalable data models using SQL to analyze and monitor business performance.
- Developed interactive dashboards in Power BI and Tableau to visualize trends and KPIs.
- Automated data cleaning and transformation workflows using Pandas, reducing manual reporting time by 50% and improving accuracy.
- Performed exploratory data analysis (EDA) to uncover patterns and support data-driven decisions.
- Contributed to lightweight ETL pipeline development, improving data ingestion speed and reliability.

## **EDUCATION**

---

### **B-Tech (E.C.E)**

03/2022 | Warangal,Telangana

#### **Vaagdevi Engineering College**