

Sanjana Sanjay Bobade

Data Engineer

(816) 979-2509 | sanjanabobade1298@gmail.com | Portfolio | [Linkedin](#) | [Github](#)

TECHNICAL SKILLS

Programming Languages: PySpark, SQL (MySQL, SQL Server), Python, Scala, Linux, HiveQL, SSIS, NoSQL

Big Data Technologies: Snowflake, Databricks, Hadoop (HDFS, MapReduce, YARN, Airflow), Apache Spark, Hive, Sqoop, Pig, Kafka, Cloudera, Data Warehouse, Data marts, Data Mesh

Cloud Technologies: Microsoft Azure (ADLS, Synapse Analytics, Data Lake, Data Factory), AWS

Developer Tools: Airflow, Power BI, Tableau, VS Code, JIRA, Azure DevOps, CI/CD, Jenkins, Autosys

soft skills: Leadership, Cross Functional Collaboration, Problem Solving, Stakeholder Communication, Documentation

EDUCATION

University of Central Missouri, Warrensburg, USA

May 2024 – Jul 2025

Master's of Science in Computer Science

CGPA: 3.7/4.0

EXPERIENCE

Data Analytics Engineer

Oct 2023 – Apr 2024

Deloitte, IDFC First Bank

Remote

- Built and maintained end to end **ETL** and **ELT** pipelines using **Azure Data Factory, Databricks, SQL, Python** and **PySpark** to process **50M+** records for seamless **data ingestion** and transformation
- Developed and optimized **SQL** and **T-SQL** (procedures, functions, views, CTEs), **Spark** and **Scala** (joins, caching, repartitioning) queries and partitioned **Hive** tables to boost performance by **20%** and built reusable data models, fact and dimension structures in **Snowflake**
- Built **RESTful APIs** using Azure Data Factory, integrating Databricks notebooks, Synapse pipelines for governed analytics and **Unity Catalog** for data discovery, governance, metadata, access control, and lineage.
- Designed **Power BI** dashboards tracking **8+ KPIs** to enhance operational visibility for database analytics.
- Resolved **10+ data quality** issues early in the **SDLC**, performed peer reviews, root cause analysis, **data validation**, data quality checks, implemented trials in UAT and collaborated with cross-functional teams.

Associate Data Engineer

Jun 2021 – Oct 2023

Cognizant, Molina Healthcare Inc. (USA)

Remote

- Developed, managed, and optimized data pipelines for **500M+ healthcare** records using **Azure Databricks** and **Spark** for **5M+ members** and **claims** (Medicare, Medicaid, Marketplace) with **HIPAA** compliance.
- Optimized ETL jobs via **SQL, Python, SparkSQL, PySpark** and **Databricks**, cutting runtime by **45%**
- Migrated **ELT** workflows from **Cloudera Hadoop (HDFS)** to **Azure cloud** via **Databricks, Azure Data Lake Storage, Delta Lake, Delta Live Table** and **Azure Data Factory** with **CI/CD** via **Airflow**
- Developed and managed historical datasets MCC, Cigna and Affinity performing **data transformation** using **SQL, Spark** and **Hive** for member, membermonth and provider pipelines with full load and integrated with ongoing dataset for incremental loads.
- Led requirements **gap analysis** for membermonth and claims dataflows built BRDs and logic maps in Azure.
- Automated Data Validation** framework for jobs with email notification triggers in case of data discrepancies
- Engineered the Agewell acquisition dataset for the member-month pipeline and integrated it with incremental data to ensure accurate and timely updates in the enterprise layer.
- Automated L3 file extracts** for FRG and **vendor extract** pipelines via SFTP using shell scripts, Scala, Excel and reduced manual efforts by **80%**
- Reduced backlog **JIRA** tickets across Member(**40%**), Provider(**10%**), and Claims(**15%**) modules in **Agile methodologies**, handling releases and documentation with **Azure DevOps** and **GIT** for **workflow orchestration**
- Maintained data lineage documentation, source to destination mapping documentation and QA frameworks using **Git**, ensuring 100% reproducibility and compliance.

Data Engineering Intern	Jan 2021 – Jun 2021
<i>Cognizant</i>	<i>Remote</i>

- Developed scalable ELT models in **Snowflake** using SQL, stored procedures, and task orchestration to support high-volume enterprise reporting.
- Built **SSIS** ETL pipelines and **SSRS** reports on **SQL Server** using statistics to improve business decision making
- **Led the team** for a final internship project: Timesheet Management interface to record employee status daily
- Built a Python Flask based application with Registration, Login and Timesheet Entry UI using HTML5, CSS3 for frontend and **MySQL** for database handling.
- Leverage knowledge of **Hadoop** architecture, **HDFS** commands and Pig to build data marts and data models

ACADEMIC PROJECTS

MailMatrix AWS S3, Lambda, EventBridge, SES, IAM roles, RDS, CloudWatch	Sep 2024 – Dec 2024
----------------------------------------------------------------------------------	---------------------

- Built a serverless bulk email automation system using **AWS (Lambda, S3, SES, EventBridge, RDS, IAM)** with CSV based personalization, monitoring via **CloudWatch**, and error alerts via **SNS**

Employee TimeTracker AWS Cognito, RDS, CloudTrail, CloudWatch, Python, HTML	May 2025 – Jul 2025
--------------------------------------------------------------------------------------	---------------------

- Developed a Python Flask-based timesheet application with **AWS (Cognito, Aurora RDS, CloudFront, CloudWatch, CloudTrail, KMS)** for secure authentication, scalable storage, and monitoring.

CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate DP-203

Microsoft Certified: Azure Fundamentals AZ-900

Microsoft Certified: Azure Data Fundamentals DP-900