



Synopsis

- Result-driven software developer **with 5+ years of experience** in the domains of **Data Engineering & Data Science in both on-prem and cloud environments.**
- Expertise in **SPARK, Python, SQL, Databricks, Azure and AWS.**
- Demonstrated proficiency in **data engineering, analysis, and reporting.**
- Skilled in managing complex datasets and delivering **data pipelines using AWS glue and orchestration.**
- Skilled in high volume **ETL, PySpark, data extraction and data manipulation using AWS Glue.**
- Experienced in working with cross functional, multi-geography teams and managing stakeholders' expectations.
- Team player and an avid problem solver.
- Enthusiastic about new opportunities to use generative AI, building AI enabled solutions.

Experience

Senior Data Engineer– ACL Digital

[Mar 2022 – Present]

Project 1: To build a new reporting tool from scratch (ETL to Reporting) that will enable users to get tailored reports from a legacy Order Management System. Lead a 2-member team to build multiple reports.

Technology Used: Pyspark, S3, AWS Glue, Postgres, AWS Lambda, SQL, AWS gateway, Python, AWS SQS, AWS SES, LDAP Authentication

- Developed a delta Ingestion pipeline on AWS glue that pulls data from SFTP and ingests into Postgres.
- The target Datawarehouse DB to the ETL serves as the base platform for all future reports.
- Developed APIs enabled via AWS Lambdas that supports a UI to generate reports from data in Postgres.
- Pioneered the first report built on this DB. Solving all teething hurdles for future reports.
- Developed APIs that enable users to download and schedule a report based on selected parameters in xlsx format.

Project 2: To Migrate an Application from Java to PySpark. The application aimed to factor in data quality during ingestion through uncompromising validation that ensures clean data for tailing use cases and quality management.

Technology Used: PySpark, EC2, Docker, S3, Oracle, SQL

- Developed a data Ingestion pipeline on PySpark that replicates the validations done by an outdated system in terms of tech stack.
- Invalidate records based on pre-defined conditions and populate reason for invalidation on the fly.
- Deliver reports with the invalidated records with appropriate reasons.
- Ensure the valid records are ingested to the database.

Project 3: PoC on building an AI driven voice interactive digital assistant that can answer any questions based on documents (csv, word, pdf)

- Developed and delivered an AI system using **streamlit, openAI and langchain** that achieved all the above requirements.

Member Technical Staff (Python) – OpsVeda Asia

1 year 6 months [Oct 2020 – Mar 2022]

Product Description: The company focuses on Operations Management System powered by an AI infused and automation capable platform. The platform uses algorithms and rules to monitor business events in real-time and alerts users to risks and opportunities in the context of customer relationships and financial outcomes.

Technology Used: Built on PostgreSQL. Data solutions developed in Python via plpython3u.

- Developed solutions to streamline supply chain data and generate periodic reports via email in python.
- Developed solutions to pull data to perform binning in python.
- Developed deployment pipelines (QA & Prod) in shell for PostgreSQL deployment using Liquibase and Jenkins.

Data Engineer (Intern) – Reliance Industries

11 months [June 2019 – May 2020]

Technology Used: Databricks notebooks, Data Factory pipelines, Azure bricks Delta Lake, Hadoop, Hive, Sqoop, Scala, Phoenix, Hbase.

Project 1: Data Ingestion Framework (ETL Tool)

- Increased the full load data Ingestion's performance by seventy-five by optimizing Sqoop & Spark code.
- Increased server query response time by 40% by optimizing Hive & Phoenix on Hbase for CDC Tables.

Project 2: Analytics using Databricks.

- Developed reports on data from multiple sources using Databricks notebooks, Data Factory pipelines.
- Handled Ad-hoc production support requests on the Azure Data bricks Delta Lake.

Software Developer – Value Global

1 year [April 2017 – May 2018]

Certifications

S.No.	Certification	Issued on
1	Data Virtualization Champion – Denodo	Mar 2024
2	Databricks Lakehouse Fundamentals	Feb 2024
3	Langchain & Vector Databases in production	Oct 2023
4	Dataiku ETL Advanced Designer	Jan 2023
5	Dataiku ETL Core Designer	Dec 2022
6	Oracle Certified Professional (PL/SQL Developer)	Jan 2017
7	Oracle Certified Associate (SQL)	Dec 2016

Education

M.E. Big Data & Data Analytics (CGPA: 9.05/10)

June 2018 – July 2020

Manipal School of Information Sciences, Manipal University

B.E. Electronics and Communication Engineering

Aug 2012 – May 2016

Sri Venkateswara College of Engineering (SVCE), Anna University