Sudharshan Sanjay

+49 179 1565570|sudharshansan@gmail.com|Kaggle|LinkedIn|Portfolio|Germany

SUMMARY

Results-driven Data Engineer and Software Developer with over 9 years of experience building scalable data pipelines, cloud automation, and full-stack applications across banking, insurance, and manufacturing sectors. Proficient in Python, PySpark, SQL, AWS, and Databricks, with hands-on expertise in Medallion Architecture, Delta Lake, and MLflow. Adept at driving business value through performance optimisation, automation, and Agile collaboration.

SKILLS

Programming Languages: Python(PySpark), Microsoft SQL, HTML, CSS, Java (RESTful Services using Spring Boot)

Skills & Competencies:

- **Cloud Computing Platforms:** AWS, Databricks
- **CI/CD pipelines:** Jenkins
- **Source Code Management:** GitHub, BitBucket
- **Ticketing Tools/Software:** JIRA, Youtrack, ServiceNow
- **Backend/ API Development –** REST API Services, SOAP
- **Data Engineering:** Apache Spark (PySpark), Delta Lake, MLflow, Prophet
- **Database Management:** MySQL, PostgreSQL, Microsoft SQL, Oracle
- **Agile Methodologies:** Scrum
- **Code Quality Analysis:** SonarQube
- **Workload Automation:** Control-M
- **General Tools:** Microsoft Excel, Word and Power point, Microsoft Power Apps

EXPERIENCE

Deloitte, Chennai, India

Consultant **Jan 2022 – March 2026**

- **Lead a team in performing validation and migration activities for SQL Server upgrades from 2016 to 2022.**
- Developed automation scripts to streamline and standardize the validation process.
- Designed and implemented filtered indexes to reduce stored procedure latency.
- **Created SQL archiving scripts to move data older than five years, improving database performance.**
- **Led resolution and migration of SSIS packages from SQL Server 2016 to SQL Server 2022.**
- Developed and executed SQL cleanup scripts to remove obsolete data and optimize database storage.
- **Enhanced an existing Python code to read Microsoft O365 version of excel files which were failing during macro processing and validation.**
- **Monthly and Quarterly data report generation for Business to avoid financial losses. The report generation saved about $10000 per month. Handled a Major and Minor incident which would have resulted in a loss of $3500-$3600.**
- Developed a python code for data migration from Amazon Red Shift to SQL Server Database. This code reduced manual migration and saved 10-to-15-man hours monthly.
- Wrote 10 Python scripts for respective AWS Glue Jobs.
- **Developed a glue job and wrote python code for re-run ability for any data update.**
- **Developed a control-M job that would start and stop Amazon Red Shift clusters in AWS for cost reduction. This saved the Business about 800$ weekly.**
- Database updates upon Business request to avoid incorrect data or mail alerts sent to incorrect person.
- ApplicationonPowerappsUIwasloadingslowduetoaSQLQuery.Improveditsperformance by 70% by fine tuning

the query.

- Identified an existing issue during batch runs where duplicate columns were getting created. Resolved the issue by renaming the columns. These allowed batches to get processed and helped business get the necessary report.
- **Deployed a Lambda Function which would identify long running RDS instances and stop them to save cost. This eventually saved 300-500$ per week.**
- **Set up the orchestration of admin feed pipelines throughout SDLC and Control-M jobs to automate processes to reduce manual intervention and save 18 Man hours monthly.**
- **Setup Control-M job and wrote a python code and batch script to automate data load process which reduces manual intervention up to 28 Man hours monthly.**
- **Enhanced** existing **batch processing** to accept or **process 64-bit data** in Moody's Axis Elink to improve performance and cost saving up to **200$ weekly.**

Cognizant Technologies, Chennai, India
Associate **Jun 2021 – Jan 2022**

- Worked on backend enhancements and deployments of API changes
- Fine tuning of SQL queries to enhance performance
- **SME for an Underwriting tool application from Front-end to Backend changes and deployment**

Cognizant Technologies, Chennai, India
Programmer Analyst **Dec 2017 – May 2021**

- **Enhanced SQL query performance** through **fine-tuning techniques**, resulting in **improved efficiency**.
- Served as the **Subject Matter Expert (SME)** for a **financial insurance underwriting application**, overseeing **full-stack development**, including **front-end**, **back-end**, and **deployment** processes.
- Utilized **Java**, **Java Server Pages (JSP)**, and **Spring2.xx** to build the application, ensuring **robust functionality**.
- Managed **PostgreSQL database optimization**, implementing **table indexing**, **data archiving**, and **Clean-up** of **unnecessary data** using **DML drop** and **delete commands**.
- Achieved **significant performance improvements**, reducing **web page loading time by 35%** and database query retrieval time by 66.67% (from 2 minutes to 40seconds).
-

Cognizant Technologies, Chennai, India
Programmer Analyst Trainee **Nov 2016 – Nov 2017**

- Contributed to **backend enhancements** and **API deployments**, focusing on **improving performance** and **functionality**.
- **Designed and developed REST services**, implementing **GET** and **POSTAPI requests** in **Java** for an **insurance-based project** to retrieve **agent details** and **insurance claims** related to **natural disasters**, displaying the data using **Google Maps API**.
- Worked on **UI enhancements** to improve the **existing interface** based on **business requirements**, utilizing **JavaScript**, **HTML**, and **CSS**.
- Built and stored the **Maven artifacts** on the **cloud**, conducted **code quality analysis** using **SonarQube**, and resolved **code duplication errors** to ensure **clean, maintainable code**.

PROJECTS

- **AIML Capstone Project–Industrial Safety and Health Analytics Database**:
  https://drive.google.com/drive/folders/17Mc6ZnlTNhiTzaImHE7rkyOXePpDtRo9?usp=sharing
  - Utilized Python with Numpy, Seaborn, pandas to complete EDA on the Dataset.
  - Upsampled the DataSet post the EDA.
  - Designed, trained and tested RNN and LSTMs on the preprocessed dataset.
- **Lang Chain Data Science Assistant**:
  https://drive.google.com/file/d/1dCtv-SZJEEGh70Mnw_M6OcQTWFqpCBxB/view?usp=sharing
  - Set up environment for the Open AI module to be used.
  - Reading data set and completing EDA.
  - Designing, Training and Testing a Logistic Regression Model through LLM.
  - Designing, Training and Testing a Binary Classification model through LLM.
  - Designing, Training and Testing a Binary Classification model using Neural Networks through LLM.
  - Implemented Retrieval Augmented Generation to get data from Web if the LLM is unable to provide information.
- **Medallion Architecture – Databricks & Delta Lake Data Engineering Pipeline**:
  https://github.com/SudharshanSanjay/medallion-architecture-databricks.git
  - Implemented Bronze → Silver → Gold Medallion Architecture on Databricks using Delta Lake 3.2 and Unity Catalog.
  - Ingested 3,500 rows from CSV via Databricks DBFS Volume and PostgreSQL via JDBC over ngrok TCP tunnel.
  - Built idempotent Silver pipeline with window function deduplication and MERGE INTO upserts.
  - Developed Gold aggregation views (revenue by region, product, order status, monthly trend) with Delta table data quality constraints.
  - Implemented Prophet time-series forecasting model with MLflow experiment tracking, pytest suite, and GitHub Actions CI/CD.

EDUCATION

Great Lakes Executive Learning
Generative AI for Natural Language Processing — **Jan 2024 – May 2024**

Deakin University
Australia
Master of Data Science — **Aug 2022 – Nov 2023**

Great Lakes Executive Learning
Post Graduate Program in AIML — **Aug 2020 – Nov 2021**

SRM Institute of Science and Technology,
Kattankulathur, India — **GPA (8.10/10)**
Bachelor of Technology in Computer Science Engineering — **July 2012 – July 2016**

CERTIFICATIONS

- AWS Certified AI Practitioner — **Amazon Web Services**
- Generative AI for NLPs — **Great Lakes Executive Learning**
- Post Graduate Program in AIML — **Great Lakes Executive Learning**
- Databricks Fundamentals Accreditation — **Databricks Academy**