Suprith M

8197723170   suprithvasista829@gmail.com   LinkedIn   Portfolio

Data Engineer with over two years of experience in designing scalable data processing frameworks, modernizing data warehouses, and implementing data governance and security controls. Proficient in Python, Scala, Dart, and cloud platforms like **Google Cloud Platform** (GCP) And **Cloudera Data Platform** (CDP). Looking to leverage expertise in data migration, real-time data processing and scalability to deliver secure, scalable, and cost-effective solutions

EXPERIENCE

MBB Labs Private Ltd   **Bengaluru, India**
Data Engineer   **June 2022 – Present**
Universal Data Lake

- Enhanced data processing utilities and algorithms in Pyspark, including **Incremental Load** and **Slowly Changing Dimensions** (SCD), to optimize data updates and ensure historical accuracy in data warehouse environments, achieving a **45%** reduction in **SLA** execution time compared to the initial baseline
- Designed, developed, and migrated from Oracle Procedures to a **Spark-based Data Quality** (DQ) framework as part of a robust **Data Governance** initiative, enhancing real-time data processing, improving scalability, and reducing job execution time by **92.22%**
- Designed and implemented scalable **ETL/ELT** pipelines to support large-scale data migrations and integrations, bringing data from the **Regional Enterprise Data Warehouse** to the **Universal Data Lake** (UDL)
- Performed **performance tuning**, troubleshooting, and scalability enhancements on complex data pipelines, resulting in a **30%** reduction in job execution time
- Led the migration from on-premises systems to **Cloudera Data Platform**, ensuring seamless data transfer and enhanced scalability
- Developed the backend **reconciliation framework** in collaboration with product engineers, who focused on the frontend, to deliver customer requirements effectively
- Developed an automated and **Secure Shell** (SSH) configuration using shell scripting for seamless server connectivity, streamlining the setup process and reducing manual effort

Profitability Management

- Engineered high-throughput database solutions in Oracle, processing **100 Million** records in under **2 minutes**, enhancing system reliability and performance, and reducing operational costs
- Implemented **SPOOL** to deliver data based on configuration, efficiently handling up to 34 files and scaling to **300 million** records in under **40 minutes**
- Deployed multiple **CI/CD** pipelines using tools like **Jenkins**, automating deployments and reducing downtime
- Developed a password encryption framework for application usage, utilizing **symmetric cryptography** to securely store and manage passwords

Net Zero Carbon Cities

- Analyzed business requirements to develop an **ODI procedure** along with a configuration mechanism, ensuring the implementation of a robust solution that accommodates future requirements and scalability
- Created queries to extract relevant data for NZCC report generation and integrated them into the configuration mechanism, ensuring seamless data generation for reporting. Successfully implemented the solution across Malaysia and Singapore

Basel Committee on Banking Stabilization

- Facilitated the generation of 14 reports to meet business goals by designing a robust ODI data pipeline that consolidated and transformed data from multiple sources into a unified final table
- Utilized **Infogram (Data Visualization Tool)** to create custom report templates, enabling effective data visualization based on the generated data

MBB Labs Private Ltd   **Bengaluru, India**
Data Engineer Intern   **June 2022 - February 2023**

- Automated key workflows for **installer** preparation, and **reconciliation** processes using Python, reducing manual effort by **80%**
- Developed a proof of concept for **Oracle Data Integration** (ODI), optimizing the metadata extraction workflow to enhance data handling processes and improve overall efficiency

TECHNICAL SKIILLS

- **Programming Languages**: **Python**, Scala, SQL, Plsql, HQL, Shell Scripting, Dart
- **Technologies & Tools:** Spark, **Pyspark**, **Hadoop**, Hive, MongoDB, Oracle Database, ODI, NoSQL, Unix, Apache Kyline, Apache Kafka
- **Big Data & Cloud:** Data Lake, **Data Warehouse**, **Google Cloud Platform** (GCP), **Google Big Query,** Oracle Big Data Appliance, **Azure Cloud**, **Data Bricks,** Teradata, **Cloudera Data Platform** (CDP), Data modeling
- **DevOps**: Jenkins, **CI/CD pipelines**
- **Data Governance & Security**: **Key Management Systems** (KMS)
- **Software & Web Technologies:** Flutter, Dart
- **Other Skills: Troubleshooting**, **performance tuning**, stakeholder collaboration, recommendations for **data-driven decisions**

EDUCATION

- BE in **Computer Science and Engineering** AUG-2022 from **VTU** with **6.79 CGPA**

PROJECTS

- **Portfolio Generator and Enhancer**
  Developed and deployed a web application on Streamlit Cloud that generates user portfolios based on the **Flutter** framework with a single button click. Integrated a custom **CI/CD** pipeline for deployment on **Vercel** and added **Gemini** for dynamic project section rewriting. Utilized **Google Big Query API** for seamless data updates.
- **Recursive Job Matching**
  Developed and deployed a web application on Streamlit Cloud that delivers customized resume feedback by leveraging data from job portals and the **Google Gemini Pro API** for enhanced job alignment
- **Data Consolidation Utility**
  Developed and deployed a data consolidation utility in Python to the **PyPI community**
- **Custom File Loader Widget**
  Created a custom Streamlit file uploader widget that delivers upload paths for metadata-focused use cases, contributing it to the **Streamlit community** for enhanced functionality

TRAINING AND CERTIFICATION

- Machine Learning with Python from E & ICT Academy, IIT Kanpur

ACHIVEMENTS

- Earned membership in the prestigious **Google Cloud Innovator Club** by excelling in a competitive Google **hackathon** focused on **Generative AI**, demonstrating innovative problem-solving and technical expertise
- Recognized as Star Performer in multiple sprints for exceptional contributions
- Awarded Best Employee for outstanding teamwork and project delivery
- Most **Dependable Developer** award with a Silver GT pin
- Successfully participated in and completed **the LLM Agents Hackathon**, organized by **UC Berkeley RDI** in collaboration with the **LLM Agents MOOC**