

Senior Data Engineer with over 3 years of experience designing and scaling robust data pipelines, modernizing data warehouses, and implementing secure data governance frameworks. Skilled in building distributed systems using **Python** and **Scala**, and proficient in data engineering tools and platforms including **Snowflake**, **Google Cloud Platform(GCP)**, **Cloudera Data Platform (CDP)**, and **Kafka**. Proven track record in developing multi-tenant, high-scale solutions for real-time and batch processing across diverse data sources. Strong understanding of **RESTful API** design, and hands-on experience with databases like **OracleSQL**, **PostgreSQL** and **HBase**. Experienced in collaborating with cross-functional teams to deliver scalable, customer-facing analytics solutions that support real-time and historical data analysis. Passionate about leveraging data to drive smarter strategies and business outcomes

EXPERIENCE

MBB Labs Private Ltd

Senior Data Engineer

Universal Data Lake

Bengaluru, India

June 2022 – Present

- **Optimized distributed data processing engines** using **PySpark** by enhancing **incremental load** and **Slowly Changing Dimensions (SCD)** utilities, improving historical data accuracy and reducing SLA execution time by 45%
- **Migrated legacy Oracle procedures** to a scalable, **Spark-based Data Quality (DQ)** framework, significantly accelerating real-time data validation and cutting job execution time by **over 92%**, as part of a broader **data governance initiative**
- **Built and maintained robust ETL/ELT pipelines** to support large-scale data integration and migration between **Snowflake** and the **Universal Data Lake**, enabling cross-platform data accessibility and consistency
- **Led the performance tuning and optimization** of high-volume data pipelines, achieving a 30% reduction in runtime through effective troubleshooting and scalability improvements
- **Lead the migration to Cloudera Data Platform (CDP)** from on-prem infrastructure, ensuring minimal downtime and improved operational scalability for future growth.
- Communicated effectively with product engineers and cross-functional teams to deliver a backend reconciliation framework aligned with customer requirements
- Developed **automated SSH configurations using shell scripting**, streamlining secure server connectivity processes and reducing setup overhead for engineering teams by 95%

Profitability Management

- **Designed and optimized high-throughput database solutions** in Oracle, processing over 100 million records in under 2 minutes, significantly improving system performance and reliability at scale
- **Engineered a scalable data delivery mechanism** using SPOOL, managing up to 300 million records across 34 files within 40 minutes, ensuring timely and accurate data availability
- **Implemented CI/CD pipelines with Jenkins**, automating deployment workflows and minimizing production downtime
- Developed a **secure password encryption framework** leveraging symmetric cryptography to safeguard sensitive application credentials and align with security best practices

Net Zero Carbon Cities

- **Analyzed business requirements and built scalable ODI pipelines** with configuration-driven logic, enabling reusable and future-ready data solutions for cross-country rollouts
- Built SQL queries and **integrated reporting logic into configuration modules**, ensuring automated, seamless generation of carbon compliance reports across regions like **Malaysia, Singapore** and **11 Overseas Countries**

Basel Committee on Banking Stabilization

- **Designed and deployed data integration pipelines** using Oracle Data Integrator (ODI) to consolidate multi-source data and generate 14 regulatory reports, supporting business compliance initiatives
- Used **Infogram** for building **custom data visualizations**, enhancing decision-making with interactive reporting dashboards

- **Automated critical workflows** for software installer preparation and reconciliation using Python, reducing manual effort by over **80%** and increasing operational efficiency
- Developed a **proof-of-concept ODI integration**, streamlining metadata extraction workflows and optimizing internal data handling and reporting processes

TECHNICAL SKILLS

- **Programming Languages:** Python, Scala, SQL, PL/SQL, HQL, Shell Scripting
- **Big Data & Cloud Platforms:** Apache Spark (PySpark), Hadoop, Hive, **Snowflake**, Cloudera Data Platform (CDP), Google Cloud Platform (GCP), Google BigQuery, Oracle Big Data Appliance, Azure, Databricks, Teradata, Data Lake, Data Warehousing, Data Modeling
- **Databases & Storage:** Oracle Database, MongoDB, PostgreSQL, NoSQL
- **Data Integration & Workflow:** Apache **Kafka**, Apache **Zookeeper**, Apache Oozie, Oracle Data Integrator (ODI)
- **DevOps & Automation:** Jenkins, CI/CD Pipelines, Shell Scripting, **Kubernetes**
- **APIs & Data Access:** REST API design and integration
- **Security & Governance:** Key Management Systems (KMS), Data Governance, Encryption Standards

EDUCATION

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

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, Indian Institute of Technology (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
- Top Candidate in **the LLM Agents Hackathon**, organized by **UC Berkeley RDI** in collaboration with the **LLM Agents MOOC**
- Recognized as Star Performer in multiple sprints for exceptional contributions
- Awarded “**Most Dependable Developer**” for consistently delivering reliable, high-performance data solutions across multiple sprints