

Dhanush Shet KP

With 3+ years of experience in data engineering, specializing in building efficient pipelines, optimizing SQL queries, and ensuring data quality. Proficient in Spark technologies, with a strong focus on migrating Hive QL to Spark for improved performance and scalability.

✉ dhanush24shet@gmail.com

☎ +91-8105144193

📍 Madikeri, Karnataka, India

🌐 LinkedIn

WORK EXPERIENCE

Nineleaps Technological Solutions

Data Engineer

08/2021 - Present

Uber Cloudlake Platform

- Build and optimize data pipelines using Apache Spark, ensuring efficiency, scalability, and low resource consumption.
- Develop insightful and interactive dashboards using Tableau to provide business insights and track key metrics.
- Implement real-time alerts for data pipelines, ensuring timely detection and resolution of issues.
- Ensure data quality by validating and troubleshooting issues within the pipeline, maintaining accuracy and integrity.
- Migrated on-prem pipelines to the cloud to leverage cloud services, utilized APIs when necessary, and ensured data quality was maintained properly.
- Debug and resolve pipeline errors, providing comprehensive solutions to prevent future occurrences.

Uber Finance

- Led Spark 3.0 migration, reducing processing costs by 60% and improving data processing time by 40%.
- Optimized ETL processes, reducing resource utilization by 30% and decreasing runtime by 50%, enhancing data pipeline scalability.
- Increased data accuracy by 35% through comprehensive pipeline testing, ensuring compliance with SLA commitments.
- Developed and maintained data pipelines using big data technologies: HDFS, Spark, Hive, and Kafka.
- Tuned Spark and SQL queries to enhance data processing efficiency and performance, ensuring SLA adherence.
- Implemented Docker containerization, improving scalability, performance, and deployment processes.

Uber Customer Obsession Data Analytics

- Managed 70+ tier 1 and tier 2 datasets, handling over 10B events and serving 1M+ queries for internal customers/services.
- Developed ETLs for petabyte-scale data, optimizing pipelines to reduce runtime and resource utilization, improving efficiency and scalability.
- Orchestrated data workflows using Airflow to enhance pipeline performance and automate processes.
- Leveraged Google Cloud services (e.g., Compute Engine, Cloud Storage, BigQuery, Cloud Functions) for data infrastructure management.
- Monitored pipeline performance using scorecards and key metrics, providing data-driven optimization insights.

EDUCATION

Manipal School of Information Sciences

Master of Engineering, Embedded Systems

2021 - 2021

Canara Engineering College

B.E, Electronics and communication

2019 - 2019

PROJECT

Student Management System

- Developed a comprehensive Student Management System leveraging SQL and MySQL databases to handle and organize student records efficiently.
- Enable seamless creation, retrieval, updating, and deletion of student data to keep the system up-to-date.
- Implemented user roles (e.g., Admin, Teacher, Student) to ensure secure data management and accessibility.

Data Analysis and Dashboarding using Pandas

- Developed a data analytics pipeline and interactive dashboards for visualizing Key Performance Indicators (KPIs).
- Leveraged Pandas to preprocess and structure raw data for advanced analytics.
- Created insightful visualizations using Matplotlib and Seaborn to track student engagement, performance trends, and operational efficiency metrics.
- Implemented basic models to predict future trends such as potential dropouts or performance dips.
- Designed the solution to handle increasing datasets, ensuring robust and future-ready dashboards.

SKILLS

Python, SQL, Apache Spark, Hive, Hadoop, Apache Airflow, Power BI, MySQL Workbench, PyCharm, Visual Studio Code, Jupyter Notebook, Data Modeling, System Design, Problem Solving, Collaboration, Big Data, Data Engineering, ETL building

ACHIEVEMENTS

- Represented my college in university-level cricket tournaments and emerged as champions.
- Qualified for state-level badminton tournament.
- Participated in 100m and 400m at district level.