Akshay Kadam

Data Engineer

✉ akshayanandraokadam@gmail.com
📞 +91 - 9175595335
📍 Sangli, Maharashtra.
in Linkedin/Akshay_kadam
👤 Portfolio/Akshay_kadam
🐙 Github/Akshay_kadam

EXPERIENCE

Data Engineer / Pravaig Dynamics | Bangalore | March 2023 - PRESENT

- Working on in-vehicle Data Analytics of **CAN Bus data, GPS data and Video data**.

- Designed and developed end-to-end **real time streaming data pipelines** on AWS cloud with services - EC2, S3, REDSHIFT, Spark for data processing, **Kafka** for real-time data streaming, **Airflow for workflow orchestration**, and HDFS for distributed storage.

- Utilized Airflow to schedule and monitor data pipelines, ensuring timely execution and reliability of the entire data pipeline while handling dependencies and failures gracefully.

- Worked on **Shell Scripting for Automation** of Kafka Servers and **Data Retention** for specific Interval.

- Implemented **monitoring and logging solutions** to track data pipeline health, detect issues, and provide timely alerts for proactive issue resolution

- Created **pyspark** user-defined functions (**UDFs**) to parse and decode the raw data bytes into meaningful physical values for each signal for **Vehicle Diagnostics.**

- **Data Security and Governance**: Worked on **SASL_SSL Authentication** and **Authorization** for **Kafka Servers** for secure transmission and controlled access of data within data pipelines.

- Implemented **CI/CD pipelines** using **GitLab workflows** to automate the build, test, and **deployment** processes for a cloud-based application on **AWS EC2**.

- Configured GitLab **CI/CD setting, AWS credentials** and defined **pipeline stages** to ensure continuous integration and deployment of code changes.

Data Engineer | Xoriant | Pune, Maharashtra | Dec 2019 - Feb 2023

- Experience in Analyzing & validating the completeness of business requirements. Responsible for merging data coming from different sources and loaded into HDFS. Combined data from MySQL & file source & applied various transformations. Worked on shell scripting and end-to-end data pipeline development from MySQL to Hive.

PROJECTS

1. Diamond Price Prediction with Linear Regression.

This project leverages advanced machine learning techniques to predict diamond prices based on various features. The pipeline is built using Data Version Control (DVC), MLflow, and Dagshub for efficient version control, experiment tracking, and collaboration. Deployment is automated using GitHub Actions, AWS ECR, and AWS App Runner for a seamless end-to-end solution.

GITHUB: **Link**

2. Open weather Streaming data pipeline and data visualization.

This Project is about fetching real-time weather data, stores it in the cloud, and visualizes it on a Power BI dashboard. The pipeline is built using Apache Kafka, AWS S3, Apache Airflow, and Power BI.

GITHUB: **Link**

SKILLS

Data Engineering & Big Data: Hadoop, HDFS, yarn, hive, Sqoop, Spark, Kafka, MongoDB, HBase, Airflow, AWS Databricks, Pyspark.

Data Visualization: Power-BI.

Cloud: AWS – IAM, EC2, S3, RDS, Redshift, Athena, Glue, Lambda, EMR, VPC, Aws devops.

CI/CD: GitHub, GitLab, Git, Jenkins.

Devops: Docker, Ansible.

Programming Languages: Python, Scala.

Data Science: Machine Learning, Deep Learning.

Others: Agile, Jira, Database, Data warehouse, Shell Scripting.

CERTIFICATIONS:

1. Spark - Grate Learning: **Link**

2. Databricks – Lakehouse: **Link**

3. Databricks – Platform Architect: **Link**

4. Basics of Machine Learning: **Link**

5. Machine Learning Algorithms: **Link**

EDUCATION

iNeuron AI | Bangalore, Karnataka.

Masters of Data Science (May 2023 - Present) **ZCOER**

| Pune University | Pune, Maharashtra Bachelor of

Engineering (June 2012 - June 2016)