ANAND JADHAV

✉ anandjadhavak@gmail.com　　📞 +91-8767249719　　🔗 linkedin.com/in/anand-jadhav-dataengineer/

DATA ENGINEER (3+ Years of Experience)

Engineering graduate with 3.6 years of experience in Big Data. Insightful and shrewd in understanding and applying business requirements. Experience in noting patterns, interpreting data and generating useful reports. Effectively prepare and operate databases as well as cloud-based system orchestration.

PROFILE SUMMARY

- Demonstrated trending technical skills, including comprehensive working knowledge of **Pyspark, SQL, Python and AWS Glue, Lambda, S3.**
- Closely monitored automated **ETL data pipeline** functionality and performance, proactively responding to issues.
- Designed high-performing data storage systems, optimizing **business intelligence and analytics**.
- **Validated complex data sets** within strict business deadlines, promptly identifying issues to enable swift remedy.
- Delivered **insightful reports** through successful dataset manipulation, aiding business improvement actions.

CORE COMPETENCIES

- **Hadoop Ecosystem:** HDFS, YARN, Spark, Hive, Sqoop
- **Framework**: Spark, Spark Core, Spark SQL
- **Cloud Technology:** AWS (S3, Glue, Lambda,)
- **Visualization Tool**: Python (Matplotlib)
- **Data Pipeline:** AWS Lambda
- **ETL processes using:** Pyspark, SQL, Python, AWS Glue
- **Programming Languages**: Python
- **Databases**: Oracle SQL
- **Dev. IDE:** Pycharm, Jupyter Notebook, SQL Developer
- **OS**: Windows/Unix (Centos, Redhat, Ubuntu)

PROFESSIONAL EXPERIENCE

ICORX Services Private Ltd, Aurangabad　　　　　　　　　　　　　　　　**June 2019 to Present**
Data Engineer

- Involved in gathering requirements from client and estimating time line for **developing complex queries** using Pyspark, Python, SQL logics
- Automated workflows and scheduled jobs in **AWS environment** using Glue and Lambda trigger
- Developed **data pipeline** using AWS Lambda, Glue, S3 and Apache Spark with Python
- Worked on AWS Glue Pyspark scripts to perform **ETL processes**
- Worked on importing data and customer histories **into HDFS for analysis**
- Creating necessary **policies** and assigning **different IAM roles** to provide access among different AWS services
- Developed several **User Defined Functions** in Pyspark for data transformations
- Data visualization & Data analytics with **Python (Pnadas, Numpy, Matplotlib), Power BI, Pyspark**

1. **Project I (Backend Data Pipeline):** This is e-commerce domain project and is based on the medical collectible supplier's data. The aim is to automate serverless website (backend) data pipeline and to analyze sales and hence performance of the vendor.
   Environment used: – Spark (3.1), Python (3.7), AWS: S3, Lambda, Glue

2. **Project II (Sentiment Analysis):** It was financial domain project and the aim was to classify Bank Customer reviews to understand opinion of the people about variety of services provided by the bank. Main responsibility is to automate AWS based data pipeline and transform/cleanse the input data to run machine learning sentiment analysis model.
   Apache Spark (2.4), Python (3.6), S3, Lambda, Glue, CloudWatch, SageMaker Notebook

ACADEMIC CREDENTIALS

B.E. (Electronics & Telecommunication): Dr. B.A.M. University, Aurangabad with **First class in 2008.**

Declaration – All information in this resume is true and correct to the best of my knowledge and belief.