Dilip kumar

Email: *kumadil26789@gmail.com*
Ph.: **7264814837/9356008175**

Professional Summary

- Total 8.5 years of Experience in AWS, Aws lambda, Glue,Aws stepfunction, Dynamodb, Cloudformation, Spark, Scala, Hive, Python big data technologies, docker, flask.

- Involved in ETL design , convert business requirement into technical design document.

- Involved in preparing data model.

- Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.

- Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's.

- Consumed the data from Kafka using Apache spark.

- Performed text analytics and processing, using the in-memory computing capabilities of Spark.

- Expertise in managing entire data science project life cycle and actively involved in data acquisition, data cleaning, data engineering, features scaling, features engineering, statistical modelling.

- Hands on experience in implementing Naive Bayes, Random Forests, Decision Trees, Linear and Logistic Regression, Clustering .

- Experience in Docker.

- Excellent interpersonal and communication skills, creative, research-minded, technically competent and result-oriented with problem solving and leadership skills.

Technical skill

Hadoop / Big Data Technologies/Cloud services	Pyspark, Python ,AWS Glue, DynamoDb, AWS Lambda , Step function, Cloud formation Templates,
Programming languages	Python,
Databases	MongoDB, MYSQL, DynamoDB, Oracle, Postgres
Tools	IntelliJ Idea, Eclipse, Docker, Git, Atlassian Jira, Jenkins, Pycharm

Work Experience

Organisation	Duration	Designation
IBM	Nov/2021	Architect
Deutsche Bank	Jul/2021 to Nov/2021	Team lead
Cognizant	Oct/2019 to Jul/2021	Associate
Tech Mahindra	Sep/2016 to Sept/2019	Senior software Engineer
Magna Infotech Limited	Jan/2016 to May/2016	Senior Engineer

Education			
	Year	Institute	Result
Secondary	2003 to 2004	CBSE	71%
Higher Secondary	2004 to 2006	State Board	61.5%
Graduation	2007 to 2011	RGPV(Bhopal)	75%

Projects undertaken

1. **fortitude life insurance & annuity**

Technologies: Oracle, Postgres, AWS Glue, S3, Python, Athena

Description:

Data migration of millions of policies with different products from aws oracle to postgresql including policy details, client details, policy status, terminated policy bank details, contact details , address etc

Roles and responsibility

- ➢ **Design ETL document.**
- ➢ **Understand business requirement and create technical design document.**
- ➢ **Attending meeting with client and business team to create data models.**
- ➢ **Creatting mapping documents from source to target and reference data mapping.**
- ➢ **Designing and writing AWS Glue jobs , job optimization**
- ➢ **Writing Reconciliation job for data validation**
- ➢ **Mentoring Team.**
- ➢ **Code Review.**
- ➢ **Creating deployment document.**
- ➢ **Performance improvement of ETL Jobs.**

2. **IWM (International Wealth management)**

Technologies: PySpark, Python, AWS, DynamoDB, Glue, Cloud formation template, Step Function, lambda function

Description:

Data received from different sources ingested into DynamoDb table extract data from dynamodb table using spark application , glue jobs used to call spark application and these jobs are scheduled using step function.

Roles and Responsibilities:
- ➢ Attending BA calls to understand the business requirement.
- ➢ Convert Business Requirement into stories and assign tasks and subtasks.
- ➢ Write Spark application and perform unit testing and attach test evidences to corresponding tasks assigned.
- ➢ Perform unit testing and with code coverage.
- ➢ Close all tasks before end of sprint.
- ➢ Attending scrum call on daily basis and provide daily updates.
- ➢ Work on performance enhancement of application.

3. **F1 LTD Enrichments**

Technologies: Spark, Python, Glue, Aws Kinesis, Lambda functions, Hive, Git, Atlassian Jira
Description:

Collect Data from different sources populate data into staging table by writing shell script for data ingestion into staging area after that data enrichment of staging area done and populated to fact tables from multiple fact tables Finally data is populated to reporting table reporting table data is used for data analytics.

Roles and Responsibilities:
➢ Writing data ingestion script.
➢ Writing Spark job.
➢ Write Scripts to configure spark job.
➢ Performance tuning.
➢ Creating Hive Table.
➢ Working with Business Analyst team to understand the requirement.
➢ Scheduling Spark Jobs through ControlM.

4. Prediction Of Equipment Failure in Chemical Manufacturing Plant
Technologies: Python, Tableau, ML Algorithm, Deep Learning
Description:
The Equipment were failed every 2 to 3 years and to maintain production shut down for 2 to 3 days failures comes mainly due to corrosion in pipelines. Corrosion depends on two things pipes material composition and fluid passing pipe . Pipe features are diameter, carbon content and length of pipe fluid features are pressure, flow rate, fluid composition, H_2s, CO_2 content ratio, PH value of fluid.

Roles and Responsibilities:
➢ Collecting historical Data and data cleaning.
➢ Feature scaling of data.
➢ Data QA and QC checks.
➢ Selecting Model and training of model validating and testing.

5. Gaia
https://www.techmahindra.com/gaia.html

Technologies**: Python, docker, Data Science/ ML, Flask**

Description:

Gaia is an application developed by tech Mahindra, to process huge volume data, train model, test model and deploy model on cloud environment by using spark, Kafka and machine learning.
This solves the requirement of fast data processing and generates alert message to service engineers before any probability of server failures based on ml model trained with historical data.

Roles and Responsibilities

➢ Writing scalable code for spark application.
➢ Analyzing the business requirement of client.
➢ Data processing using ml tools.
➢ Model Selection and training.