

Name: Dilip Kumar

Email: kumadil26789@gmail.com

Mobile: 9356008175 / 7264814837

Education Qualifications

Degree	University
Bachelor of Engineering (B.E)	LNCT&S(Bhopal)

Professional experience

Profile

- ❖ Having Total 8 years of Experience in IT Industry including (3 years of experience in **AWS**, 2.5 years of Experience in **Azure Databricks**, 2 years of experience in **Spark** with **Scala** and about 2.5 years worked on microservices, Docker, data science.
- ❖ Hands on Experience on AWS services used (**Glue, Lambda, CloudFormation Template, S3, Kinesis, DynamoDB, Step Function**)
- ❖ Hands on experience in Azure Databricks Services Used (**Blob Storage, Delta Storage, Azure Databricks Notebook with azure data factory, key vault, Azure Kubernetes service**)
- ❖ Good understanding of Spark Architecture with Databricks, Structured Streaming. Setting up AWS and Microsoft Azure with Databricks, Databricks Workspace for business analytics.
- ❖ Hands on experience data extraction (extract, Schemas, corrupt record handling and parallelized code)
- ❖ Good understanding on **Data Wrangling and Data analyst using Python, NumPy, Pandas & Matplotlib libraries.**
- ❖ Defined and Designed **Kafka clustering for real-time streaming** data process.
- ❖ Strong perseverance and diligence towards attaining challenging goals and possess good communication skills and quick learning ability.
- ❖ Good communication skills, interpersonal skills, self-motivated, quick learner, team player.

Key Skills

Bigdata Cloud/Technologies/	AWS, Microsoft Azure, Spark, Kafka, Kinesis, Hive, Oozie, Impala
Programming & Scripting	Python, Unix-Shell Scripting, Scala, Java
Scientific Packages	Numpy, Pandas
Database	Oracle, Mysql, MongoDB, DynamoDB
Code Repository	GitHub
Visualization tools/Libraries	Matplotlib libraries

Career History

Organization	Duration
IBM	Nov 2021 to till date
Deutsche Bank	Jul/2021 to Nov/2021
Cognizant	Oct/2019 to Jul/2021
Tech Mahindra	Sep/2016 to Sept/2019
Magna Infotech Limited	Jan/2016 to May/2016
HPCL	Aug/2011 to May/2013

Projects undertaken

1. NGCB(Next Generation Corporate Banking)

Technologies: PySpark, Python, DynamoDB, PITR, AWS Kinesis, AWS DynamoDB, Hudi, Cloudformation, Lambda Function, Jenkins, Git,

Description:

Data from different sources pushed to DynamoDB on daily basis, Pyspark application written in python read data from kinesis stream , and store data to hudi table , PITR is generated for each business day and after that other glue script calling pyspark application read data from hudi and PITR and create outbound for downstream with required business transformation.

Roles and Responsibilities:

- Cost optimization for AWS services used.
- Write Pyspark Streaming jobs.
- Write Glue jobs, lambda function, Deployment of jobs through Jenkins.
- Understanding requirement assigning tasks assisting team and completing task assigned to me on time.

2. Total Energy

Technologies: PySpark, Data Frame, Hue, Python, Databricks, Data Lake, Azure data factory

Description:

The Big Data platform will be used for storing all the relevant data for Business Intelligence reporting. It is using Spark for batch processing and store data into data lake. The data has been read from multiple source systems like Akpo, Amenam, Girassol, Daliya, etc and need to process these data coming from these sources. Stored in data lake.

Roles and Responsibilities:

- Responsible for creating DataFrame and performing operations with python script as per business requirements.
- Understanding of concepts in Hive and designed both managed and external tables in Hive to optimize performance.
- Maintaining the Realtime and historical data jobs and uploading data in data lake.
- Maintaining the Azure pipelines

3. Prediction Of Equipment Failure in Chemical Manufacturing Plant

Technologies: Python, Tableau, ML Algorithm, Deep Learning, Numpy, Pandas, Docker

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, H2s, CO2 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.

4. IWM (International Wealth management)

Technologies: PySpark, Python, Aws Glue, AWS lambda, AWS Step function.

Description:

Data received from different sources ingested into hive external staging table using AWS Glue, data ingested to staging table processed and enriched by spark application and data populated to fact table after enrichment finally fact table data populated to reporting table after several data transformation logic as per business requirement.

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 as- signed.

- Write glue jobs, create step functions and deploy product using cloud formation.
- Perform unit testing and Junit testing 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.

5. F1 LTD Enrichments

Technologies: Spark, Python, Hive, Impala, Control M, Shell Script, 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.

6. 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.

