KUNAL MISHRA (Senior Data Engineer, BigData, Cloud & Machine Learning)
Kunalmishra676@gmail.com
7376195622, 8299718405

Summary

- Over 5 years of Extensive experience in BigData domain with Cloud technologies like PySpark, Python, Spark Structured Streaming, Hive, HQL, Hadoop, NoSQL, Spark, AZURE, GCP. Currently Working as **Senior Big Data Engineer.**

- Expertise in Data pipeline and workflow creation using Airflow and oozie

- Good analytical skills in writing complex Hive queries.

- Good experience in OOPs and Functional Programming paradigm.

- Analytical Experience in Machine learning techniques Regression, Classification, Clustering

- Working with Agile methodology using Jira, rally software and daily scrum meeting.

- Experience in version control system like GIT and TFS.

EXPERIENCE

- Working with Dunnhumby, as Senior Data Engineer

- Worked with TELUS, Noida as Senior Software Engineer from NOV'2018 to MAY' 2020.

- Worked with Wipro Technologies, Bangalore as Project Engineer from Jun'2016 to Nov'2018.

TECHNICAL SKILLS

Programing Language:	**Primary**- Python, HQL (SQL) **Secondary**- Scala
Big Data :	PySpark, Python, Airflow, Hadoop(Hive, oozie, Sqoop)
Cloud Technology :	GCP, AWS, AZURE
Database/Warehouse:	HIVE, BIGQUERY, Synapse Analytics, Oracle
Machine Learning :	Sklearn, PySpark MLIB
IDE Tools :	Pycharm, Jupyter, Intelj
Version Control (Repo):	GIT

Dunnhumby – Customer Data Science Company

Dunnhumby is a product base customer Data Science Company works for top global retailers, BFSI Organization. It aims to create the AI applications for enterprise using BigData. The application aims to enrich data to make organization in decision making.

JUN 2020- PRESENT

PROJECT 1 #WALMART AND TESCO   **JUNE 2020- PRESENT**

Project	:	Development
Location	:	Gurgaon, India
Role	:	Senior Data Developer
Domain	:	**RETAIL**
Technologies	:	Pyspark, Python, GCP, AZURE, Airflow, GIT

Project Description:
The WALMART and TESCO project aim to create the AI applications for enterprise using BigData.
The application aims to enrich data to make Digital recommendation system.

Responsibilities:
- Developing complex optimized PySpark batch job using Python to extract, transform and load data into data mart tables (Hive/Cloud).
- Creating data-lake for Cloud and on-premise cluster.
- Creating complex hive queries for analytical computation as per business requirements.
- Data pipeline creating using Airflow.

TELUS - COMMUNICATIONS COMPANY   **NOV2018– MAY 2020**

PROJECT 2# TELUS CUSTOMER 360

Project	:	Development
Location	:	Noida, India
Role	:	Senior Software Engineer
Domain	:	Telecom
Technologies	:	Pyspark Python, Hadoop(Hive,Sqoop,OOzie)

Project Description:
The customer360 project aim to create the BigData applications for enterprise. The application aims to enrich data to make organization in decision making.

Responsibilities:
- Understanding of the computation equation.
- Developing optimized PySpark batch job with Python to compute data using the analyzed equations
- Spark structured streaming with PySpark
- Writing hive queries for business logic implementations and transformations.
- Writing sqoop jobs to import export data to and from Oracle database.
- Creating pipeline using oozie for scheduling spark applications and hive actions
- Unit test cases and Involved in making impact analysis for any changes or new implementation.

PROJECT# 3 LLOYDS BANKING BIGDATA FRAUD DETECTIONS

Organization	:	Wipro Technologies
Client	:	Lloyds Bank,London
Domain	:	BFSI
Role	:	Project Engineer
Duration	:	10 Months
Team Size	:	6
Technologies	:	Bigdata,Pyspark,Python, Hadoop, Spark/Scala, Hive,HQL,Maven,GIT,Jenkins

Project Description:
This is the project which aims to provide fault detections information using Bigdata technologies
Using the data of the banking files like Transaction,Branch,Currency etc.

Responsibilities:
- Understanding the design, analyzing the requirement/mapping sheet, performing complex transformation and computation using Spark/Scal,PySpark,Python,Hive HQL,Object Oriented programming and Functional Programming language
- Writing Spark File streaming code.
- Unit testing ,code commit, peer review

PROJECT #4 T-MOBILE DATA ANALYTICS

Organization	:	Wipro Technologies
Client	:	T-mobile USA
Domain	:	Telecom
Role	:	Project Engineer
Duration	:	3 Month
Team Size	:	10
Technologies	:	Bigdata Hadoop, Spark/Scala , Hive, HQL, Python

Project Description:
This project aims to provide analytics output to the client using cdrs and the customers data
Which aims to provide better services to its customer by analyzing the final data report that leds to better throughput to the organization

Responsibilities :
- Understanding the design, analyzing the requirement/mapping sheet, performing complex transformation and computation using spark/scala/Python,Pyspark, Object Oriented programming and Functional Programming language
- and Hive HQL
- performing complex transformation and computation using spark/scala and pyspark/Python
- Developing complex hive queries for transformation and Analytics
- Unit testing ,code commit, peer review

Organization	:	Wipro technologies
Client	:	Vodafone
Domain	:	Telecom
Role	:	Project Engineer
Team Size	:	12
Environment	:	Bigdata Hadoop,Spark/Scala,Pyspark,Hive,HQL,Sqoop

Project Description:

The project aims to migrate the data from legacy system/databases to Hadoop ecosystem as the that started flowing landed up into categorey of bigdata,More over,it aims to provide end to end analytics an reporting using the data from hadoop system.

Responsibilities:

- Writing Sqoop jobs to fetch data from traditional databases to Hadoop(HDFS/Hive)
- Developed complex spark jobs/application for business data computation, transformation and analytics
- Using hive query language to write jobs for analysis over customer data of different files.
- Transforming data using spark/scala and making it suitable for third party application for further analysis
- Unit testing, code commit, peer review

EDUCATIONAL QUALIFICATION

- BTECH Computer science engineering from SRM University, Chennai

(KUNAL MISHRA)