

Experience Summary

I have 4 years of experience in total IT and relevant into bigdata 3.2 strong technical skills in **Hadoop and Spark**, PySpark, Python, **HDFS, AWS (Glue, Athena, Airflow, S3) Hive, MapReduce, Sqoop ,etc.) include designing, developing applications using Hadoop Ecosystem components.**

- Overall, 4 years of experience with strong emphasis on Design, Development, Implementation, Deployment of applications.
- 3.2 years of experience in Hadoop ecosystem & Big-Data Analytics.
- Experienced in Spark ecosystem components like Spark batch, Spark streaming, SparkSQL, Spark-core.
- Experienced in using HDFS, MapReduce, Hive, Sqoop.
- Good knowledge of Hadoop Architecture and Yarn Architecture.
- Good knowledge of Spark Architecture and AWS.
- Good knowledge and Understanding of Spark-standalone-Architecture and Spark-Yarn Architecture.
- Good experience in handling JSON/XML files as source, process the JSON/XML data using Spark for the master and transactional data
- Experienced in working Hive data warehouse system, developing data pipeline, implementing complex business logic and optimizing business logic.
 - understanding about data sets, read data different source files like csv Parquet, Avro, Json, Sequence files and store it in HDFS/Hive tables using Spark
- Good knowledge of AWS, S3, Mongo dB .
- Experienced Hadoop Developer, have a strong background with file distribution systems in a big-data. Understands the complex processing needs of big data and have experience in developing codes and modules to address those needs
- Knowledge of Map Reduce programs to perform analysis and MR unit jobs for testing.
- Excellent communication, organizational skills, attention to detail & the ability to meet tight deadlines.

Skills Summary	
Domain	Banking
Programming Languages	Python
Operating System / ERP Version	Windows 10, Ubuntu (Linux)
Tools / DB / Packages / Framework / ERP Components	Spark, PySpark, Python, Hive, HDFS, Sqoop, MapReduce, Linux, AWS (Athena, Glue, Airflow)

Professional Certifications/ Trainings

Attended various trainings on Hadoop Bigdata, Spark and ,Python I
have Certified with Cloudera-CCA175

Work Experience

Project 1

Project Name	Llyods Bank	Team Size	12
Start Date	Jan 2019	End Date	Aug 2020
Project Description	<p>With overall million of customers like rats(ora-std) MTG, PCL, CRD, customers and a finance, data Optima. analytics is a critical component of Llyods Bank business. Bank collects huge amount of usage component of Llyods Bank business. Llyods Bank collects huge amount of usage information about all their products, services and sites such as business data, Financial Data, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various Llyods Bank Business units.</p> <p>compliance, business and marketing to improve the services, measure The collected data is aggregated and analysed for adhere legal and analyse the usage and to help to find services and advertising more relevant to the customers.</p>		
Role & Contribution	<ul style="list-style-type: none">▪ Data is pulling from Oracle using Sqoop to Hive tables.▪ Reading data from source(hive) LZ.▪ Understand business story and analyse functional specifications & design feasible solutions.▪ Using Pyspark For processing and Applying Analytics on data per client Requirement.▪ Developed code to write the logic data into Hive External Tables.▪ Involved in configuring and creating Autosys schedule.▪ Involved in various log files to fix the issues during unit testing. and performance tuning & involved in developing the PySpark		
Technology & Tools	Spark and Python, AutoSys, Arcadia Data, Hadoop Hive, Sqoop, Yarn, Bitbucket, PyCharm, Windows, Linux, JMS, Jenkins, RLM.		

Work Experience			
Project 2			
Project Name	Procter & Gamble (P&G)	Team Size	10
Start Date	Sept 2020	End Date	AUG-2022
Project Description	<p>P&G Business such as P&G Products, enterprise customers and a huge global Products, data analytics is a critical component of P&G business. P&G collects huge amount of usage information about all their products, services and sites such as call records, wireless location, network traffic data, product and device specific information and other similar data and that will consumed for Business Analytics and build Big Data Solutions</p> <p>for various P&G Businessunits. The collected data is aggregated and analyses for adhere legal compliance, businessand marketing to we improve the services, measure and analyze the usage and to help to find services and advertising more relevant to the customers.</p>		
	<ul style="list-style-type: none"> • Data Reading from Mount Path to Spark for processing data. • Understand business story and analyses functional specifications&Design&feasible solutions. • Based on business logic at applying cleaning the data, applying transformations and analytics with Spark. • Developed code to write the logic data into Athena External. Involvedin configuring and creating Airflow schedule • Involved in setting up the properties for Apache Airflow. • Involved in various log files to fix the issues during unit testing andperformance tuning&Involved in developing the spark. • Tuning &Optimizing Involved in developing the PySpark. 		
Technology & Tools RLM.	Apache Sqoop,Hadoop,Hive,Spark,PySpark, AWS, S3, Athena, Glue, Shell Scripting, Data Bricks, Shell Scripting, Data Bricks,JMS, Jenkins,		

Educational Qualification	
Education & Credentials	Electrical and Electronic Engineering (EEE) in 2018