

Apache Hadoop with 10 Years of Experience

Professional Overview:

- Big Data Hadoop Developer, Java, Data Analytics and Data Scientist with 10 Years' Experience.
- Knowledgeable about classical and modern tools and technologies (Spark, Scala, SQL, Python, Java, R, Hadoop).
- Excellent understanding of OOP'S (Object Oriented Programming).
- Around **10** years of experience as a Software Developer Cum Team Lead in ERP domain and have **4.7** year hands on experience in Hadoop ecosystems, Data Science, Big Data Analytics, Scala, Spark, Databricks, R, Python Machine learning with Deep learning and Tableau.
- Good hands on knowledge in Data Analytics, Data Sciences – Machine Learning, Deep Learning and Statistics.
- Proficient in working on various data sets and related models using programming languages R, Python.
- Hands on experience of working on a couple of live projects involving real life data sets such as HR related area (Employee attrition) etc.
- Adept in handling large data sets of various types for data processing, analysis, interpretation and eventually prediction using various machine learning techniques.

Technical Expertise:

- **Programming Languages /Technology** : Java, R, Python, PHP, JavaScript, Django
- **Big Data Framework**: HDFS, Spark, Zappeline Notebook, Databricks
- **Big Data Scripting**: Pig, Hive, SCALA, SPARK-SQL, Pyspark
- **Data Integration**: Sqoop, Flume, Spark Streaming, Apache NIFI
- **WorkFlow and Dataflow Scheduling**: Oozie, Apache NIFI, Azure -Logic apps, Apache NIFI, Azure Data Factory
- **Hadoop Data Platform**: HortonWorks HUE
- **Data Storage and Search Application**: HDFS, SOLR, Elastic Search
- **NoSql**: Hbase
- **Big Data Processing**: MapReduce JAVA API
- **Cloud Based Platforms**: Amazon AWS (EC2, S3, EMR), Azure, Azure Blob, **Dataricks, delta table**
- **Data Science**: Machine Learning algorithms such as Regression, Classification, Clustering, Market Basket Analysis, Deep Learning (ANN) for Supervised/Unsupervised learning, Exploratory Data analysis, Visualization and Inferential Statistics including Test of Hypothesis.
- **Languages Known**: C, C++, Data Structure, VB, Android
- **Database**: SQL-Server, MYSQL, Sqlite3, postgres
- **Operating Systems**: Windows XP/7/8/10, Linux
- **Tools**: Eclipse, Netbeans, Notepad++, Pycharm
- **BI Tool or Data Visualization & Reporting**: Tableau

Models Used:

Regression	SLR, MLR, Polynomial Regression, Interaction Term, Ridge Regression, Lasso, Time Series
Classification	Logistic Regression, KNN, Decision Tree, Bagging, Boosting, Naive Bayes Classifier, SVM
Clustering	K-means, Hierarchical, DBSCAN
Association Rule Mining	Market Basket Analysis, A-priori Algorithm
AI/Deep Learning	Support Vector Machine, Artificial Neural Network, Tensorflow, Scikitlearn, Keras

Educational Qualification :

- B.E. Information Technology
- Diploma Information Technology
- PG Diploma(Data Science Big Data Hadoop)

Key Projects:

Project 1

Product Name: Alarm Log Processing

Role/Responsibilities: Design, Coding -writing business logic as peer KPI, Data Pipeline - Workflow Designing, Database Handling Requirement analysis and clarification for the new interfaces and Enhancements. Preparation of FRD and Work BreakDown Structure, Client Meeting, Requirement Analysis and making Functional Specification understanding to team.

Technology Used/Backend/User Interface: Java, .Net, Hadoop, R, Python, HDFS, Sqoop, Flume, Spark, Scala, Pyspark, Tableau, Cloudera, AWS, EMR, AZURE, S3, Azure Blob Storage, DATABRICKS notebook, zappeline, Postgres DB, Logic Apps, Winscp., Apache Solr, Elasticsearch

Operating Systems: Linux



Description: Data Analysis, Continuous data / Normal end to HDFS, Blob OLTP, and Reporting (2G,3G,LTE)analysis OLAP to the client.

Project 2

Product Name: ROBO Trainer Simulation System for Clearing Alarm

Role/Responsibilities:

Installation of Hadoop, Creation of Node

Storing data into HFS

Making Front end and Machine Learning Model.

Technology Used/Backend/User Interface: Mysql, R, Python, HDFS, Sqoop, Flume, Java, Vmware, spark, scala, pyspark, POSTGRES, Databricks, zappeline, Data Factory.

Operating Systems: Windows 10, Linux

Description: Educational software to replace nursery teacher, software that will improve system learning experience and motivation to learning. This is an interactive system that takes response from both standard I/O devices as well as voice response to and from users to train the system. This makes it better than conventional educational software. This is build on basic languages, provides easy to use GUI for System to clear Alarm. It includes modules likeParser,Substitutor,Predictor, Knowledge Module,Orchastrator etc

Project 3

Product Name: Data DE duplication

Role/Responsibilities: Coding, Database Handling, Error handling.

Technology Used/Backend/User Interface: C, Spark, Scala, AWS, EMR, AZURE, S3, DATABRICKS, zappeline, Postgres

Operating Systems: Linux

Description: To Minimizes the space used by duplicated Characters of string on Primary Memory

Project 4



Product Name: Pachashil Meeting Tracking System & MOM Analysis

Role/Responsibilities: Design, Coding, Database Handling. Requirement analysis and clarification for the new interfaces and Enhancements. Preparation of Functional specification and Work Break Down Structure, R&D, MOM and Task Analysis using R.

Technology Used/Backend/User Interface: . Net Sharepoint, JavaScript, jQuery, MYSQL,R

Operating Systems: Linux

Description: This project is about Meeting Track and Assigning task to Attendees and their downline and Keep record of Every meeting vs task assign and its status.

Project 5

Product Name: AWS Downloading&Uploading Using Buckets

Role/Responsibilities: Design, Coding, Hdfs Handling, Requirement analysis and clarification for the new interfaces and Enhancements, Preparation of FRD and Work Break Down Structure.

Technology Used/Backend/User Interface: Java, Flume, Sqoop, HDFS, Hbase, AWS, Cloudera, Aws, EMR, AZURE, S3, DATABRICKS, zappeline, Postgres

Operating Systems: Linux,Windows

Description: User can Upload and download the files on AWS, to secure the file, this product especially design for Jwellers, DMART, Super Markets to Upload or stream their daily sales or important data on aws server for high availability and security.

Project 6

Product Name: ERP Companies and Colleges/Sale-Purchase-Expenses Analysis.

Role/Responsibilities: Design, Coding, Database Handling, Requirement analysis and clarification for the new interfaces and Enhancements, Preparation of FRD and Work Break Down Structure, Client Meetings, Requirement Analysis and making Functional Specification understanding to team.

Technology Used/Backend/User Interface: Java, .Net, Hadoop, R, Python, HDFS, Sqoop, Flume, Hotornowrks, nifi, HUE, Spark, Scala, Tableau, Cloudera, Azure, Databricks, zappeline, postgres.

Operating Systems: Linux

Description: ERP for Manufacturing Companies

Project 7

Product Name: DMART -Sale-Purchase-Expenses with Sentimental Analysis.

Role/Responsibilities: Design, Coding, Database Handling, Requirement analysis and clarification for the new interfaces and Enhancements, Preparation of FRD and Work Break Down Structure, Client Meeting, Requirement Analysis and making Functional Specification understanding to team.

Technology Used/Backend/User Interface: Java, .Net, Hadoop, R, Python, HDFS, Sqoop, Flume, Spark, Scala, Tableau, Cloudera, AWS, apache nifi, EMR, AZURE, S3, DATABRICKS, zappeline, Postgres

Operating Systems: Linux

Description: DMART Data Analysis, Continuous sales at end to HDFS, and sentimental analysis for Political Party from customer data.

Project 8

Product Name: Employee Attrition

Role/Responsibilities: Problem Formulation, Data preparation (Missing, wrong and outlier data), Exploratory data analysis and Visualization, Test of Hypothesis, Selection of best model/algorithm, Implementation in programming language like R, Python, Sqoop- for transferring data to hdfs for analytical purpose.

Technology Used/Backend/User Interface: Java, Hadoop, R,Python,HDFS,Sqoop,Flume,Spark Pyspark,Scala,Tableau, Cloudera, Aws, EMR, AZURE, S3, Azure Storage,DATABRICKS, zappeline, Postgres

Operating Systems:Linux

Description: Employee Attrition why the employee is leaving.

IEEE Projects Done

Rain Removal from Image – Framework-Java, Python(2017)

Facebook Friends Relationship analysis (2017-2018)

Certifications & PG Diploma:



- Coordinator of java competition event Xceed 2010 and 2011
- Runner up in college level programming competitions, cricket and swimming.
- Pursuing PGDiploma in Data Science and Big Data Hadoop from ETL HIVE.
- Certification of National aptitude in which secured 97.5% conducted by NIIT.