Rajeev Prasad

Sanctuary 36 main,
BTM 2nd Stage,
Bangalore.
rajeev.prasad.8019@gmail.com
+91 7619550012

Bigdata Hadoop/Spark & Data Science Corporate Trainer

Professional Preface

- Overall 9+ years of Experience years of Experience in Application Design, Development & Deployment of Hadoop Eco System/Java/J2EE systems with good exposure to Enterprise Architectures also.
- Currently Working as **Big Data Architect & Developer.**
- Over 5.2+ years of experience with **Hadoop Ecosystem** including **HDFS**, **Map Reduce**, **Hive**, **Pig, Kafka, YARN, Oozie**, **Zookeeper**, **Flume**, **Sqoop, R** ,**Spark & Scala, Flink**.
- I have setup **10 node cluster** for the production using **Cassandra, Spark, Kafka ,Cdh5.4.**
- **Experienced in Cassandra Admin and Developer role.**
- I am managing of 6 member team on the current project.
- Initially Started Carrier As IT-Trainer **Over 3.6 yrs** of experience in training field On **C,C++,Core Java** and **Oracle-8i/9i.**
- Have setup single node and multi node Hadoop1.0/2.0 cluster setup in LAN and AWS .
- Experienced in working with **Cassandra Query Language (CQL)** to work with file system.
- Experienced in Cassandra data modelling, cluster setup, journaling and data management.
- Experienced in migrating Map reduces to **spark Transformations** using in memory processing.
- Design and develop Spark transformations, apply spark actions to imply algorithms.
- Experienced in working with Spark QL to analyze structure data queries.
- Knowledge in implementing predictive algorithms using Spark Mlib libraries.
- Expert in implementing unified data platform to gather data from different sources using **Kafka Java Producers and consumers**.
- Experienced in design **Kafka brokers,** creating custom partitions and integrated with apache storm for transformations.
- Experienced in implementing **Kafka Simple consumers** to get data from specific partitions.
- Good knowledge of Scala programming.
- Hands on experience in **Map Reduce** programs using **Apache Hadoop** for analyzing the **Big Data.**
- Expertise in optimizing traffic across network using **Combiners**, joining multiple schema datasets using **Joins** and organizing data using **Partitioners.**
- Experience in writing **Custom Counters** for analyzing the data and testing using **MRUnit** framework
- Have Experience In Analytics.
    - Using Scala developed the spark code for capturing Twitter stream and stored in Cassandra
    - Created predictive model and recommendation engine using R/Spark.
- Good Knowledge of Data Mining and Text Mining Algorithm
- Have working exposure on security Algorithm in the Application like RSA, AES-256, Homorphic Algorithm.
- Have implemented MapReduce Algorithm using CollabFiltering Algorithm for Recommendation Engine

- Good knowledge of MapReduce Design Pattern and Pig-Design Pattern.
- Hands on knowledge with NoSQL like MongoDB, Apache Cassandra.
- Knowledge of Graphic Database Neo4j.
- Exposure to all stages of **SDLC** - Software Development Life Cycle and development projects .
- previously Comprehensive experience in building Web-based applications using **J2EE** Frame works like **Jsp, Servlet, struts, Hibernate, Webservice (Rest), JMS and SQL/PLSQL.**
- exposure to work on Jelastic cloud server and Amazon Web Server.
- Expert level skills in Designing and Implementing web servers solutions and deploying Java Application Servers like Websphere, Web Logic, configuring Apache Web Server
- Solid background in Object-Oriented analysis & design, UML and various design patterns.
- Domain Knowledge in Banking & Finance, E-commerce, Insurance & HeathCare and Retail
- good knowledge of **python & R language** .
- Knowledge of IOT project and concept

Skill Sets

Adv. Analytics Tools

Statistical Modeling Tools	R-Studio, Python,
Machine Learning	R, MLlib, Spark ML
MAchine Learning Algorithm Experience **Prediction Analytics** **Business Analysis**	**I have used the Algorithm and machine learning technique** **1. Classification Algorithm** **2. SVM** **2. KNN -Query Algorithm** **3.K-Means Clustering Algorithm** **4.Collab Filtering Algorithm** **5.Association Algorithm** **6.Regression Algorithm** **7. Decision Tree Algorithm** **8.Neural Network Algorithm**
Experience with Machine Learning tools, Data Mining Tools,	• **Spark - Mlib** • **Python-Scikit-learn** • **R-CRAN** • **Mahut**

Big Data Competencies

Big Data Technologies	Hortonworks, Cloudera, Apache Hadoop
Real Time Analytic	Apache Spark Streaming, Apache Kafka, Flink
Batch Data Processing	Spark, Hive, Pig, MapReduce
Data Collection	Sqoop, Flume, SpringXD, Impala
NoSQL Technologies	HBase, MongoDB, Cassandra, Neo4j
Messaging Framework	Kafka, Rabbit-MQ
Programming Language	Python, Scala & R,JSON

Other Technical Skills

Cloud Platform & Tech.	Amazon AWS, jelastic cloud server
Programming Languages	C,C++,Java, AndroidSDK3.2
RDBMS	Oracle8i/9i,MySQL5.6
Mobile Technologies	Android
Web Technologies	Servlet, JSP, XML, XSL, XSLT, XPath, WebServices, JSON
Design Tools	StarUML, EDraw
Web server	Apache Tomcat, JBoss, WebSphere
Testing Tools	JUnit
Version Control	GIT, Tortoise
Project Management Tools	Maven
Build Tools	ANT, log4J
Other Tools	Eclipse, Android Studio

Project Undertaken

Quantum-Sales-IB	
Roles	**Big Data Architect & Developer**
Project Description	This project deals with the sales and support team for the customer and product sales. In this project every day 50GB data is being generated for the sales. This project has been developed using ruby-rails and MySQL.
Technologies Used	**CDH5, cassandra3.0(DataStax), Zookeeper3.4.6, ApacheKAfka, Spark1.3.0, Scala2.10, Kafka-2.10.0.9.0.1,ruby-rails, RDH6.0, mysql5.6, CloudServer-10Node**
Responsibilities	• Design the Cluster Architecture using 10 Node cluster • Setup 10 Node Cluster Setup on Cloud Server. • Multi-node cluster setup for spark and Cassandra, Kafka • Design Cassandra data model for importing the data • Setting Scala, SBT and Spark.

	• Integrate Apache Kafka with Multinode using cloud server. • Implementation of Spark Streaming using Apache Kafka to Cassandra • Involved In Testing for cluster data • Client interaction on the project • Manage 6 member team • use Spark/MLib for Machine Learning Algorithms • use K-means ,Expectation -maximization ,power Iteration Clustering Algorithm . • Write higher order user defined function in Scala. • Perform Analytics using Spark • Use Kafka for the live data streaming on hourly basis in the project • Adding Recommendation listing • understanding customer Recommendation behavior • Perform Analytics
Predict shipment delivery failures	
Company/Project Duration	**Client : DHL** **Roles : Big Data Consultant** **Duration : May 2014 to October 2015**
Client Location	**Enrich-IT,DLF Building, Gachi Bowli, Hyderabad-500019**
Project Description	Problems concerning increased shipment re-deliveries and customers refusing the shipments due to additional charges had become concern for DHL. The redelivery of shipments not only added costs but also consumed extra storage space in facilities for the un-delivered goods. Work was conducted with objective to, Predict chances of customers refusing 'to be' delivered shipments Reduce the number of multiple last mile journeys Provide insights into different paradigms of failed shipment delivery
Technologies Used	**MapR2.0, SQOOP1.4.6,HIVE0.13.0,RHadoop, RMR2,RHDFS,ApacheSolr3.6,Flume1.5.0,Jdk1.6 .**
Responsibilities	*Data ingestion* • With SQOOP I transferred bulk data related to shipments, billing, customers, global references, tracing, claims, complaints, & checkpoints [sensor data] to HDFS & HIVE • Designed & developed HIVE tables using JOINS, Partitioning & Bucketing. Made the data ready for performing future analysis. Execution of HIVE was done within SPARK engine *Analysis& reports* • Using R programming language I performed predictive analysis on the shipments data. Applied Decision Trees and Random Forest supervised learning technique to predict failure of shipment delivery • Using Qlickview I facilitated comprehensive customer relationship management dashboards which provided the insights into different paradigms of failed shipment delivery & customer problems **Email analysis** • High volumes of customer queries in the form of emails are received by DHL. Managing emails, understanding customer queries, providing timely responses, tracing, searching & storing emails became a challenge & concern for DHL. Work was undertaken with objective to, • Perform sentiment & text analysis on the emails, using

	sophisticated text-analysis to discover trends in customer problems. • To categorize the emails and route them to right agent queue Map correlations between the unstructured data in email and transaction data. Integrate with backend systems, and customer data in single place Increase contact centre productivity and improve overall customer service, reduce handle and wrap time for replying to the emails, enforce service levels Provision for the CSA's to search emails on different email fields, TO, SUBJECT, FROM, DATE, SUBJECT, Email contents, keywords Provide management dashboards to the supervisors *Data ingestion* • Configured FLUME for fetching the emails, MORPHLINES for transformation and storing the emails into SOLR, HDFS • Designed & developed HIVE queries • Exported data from HDFS to transaction database using SQOOP *Analysis* • Using R programming language I performed text analysis & sentiment analysis on the emails. Created word cloud, discovered topics using LDA algorithm of topic modelling • Hierarchical and K-Means Clustering were performed to understand various segments of the customer queries • Exported HDFS data to the transactional database using SQOOP
Clinical Integration Acceleration	
Company/Project Duration	**Enrich-IT / August 13- April 2014**
Client/Project Duration	**OPTUM Health / May 13- April 2014**
Client Location	**Enrich-IT,DLF Building, Gachi Bowli, Hyderabad-500019**
Project Description	CDID is identified as one of the source for this Clinical Integration Acceleration Project. Data will be extracted from XML Files for this source and will be maintained in Non-relational File Hub Layer after enhancements and enrichments. CDID is UHC's capability for in-taking HL7 transactions. CDID consumes HL7 from OCIE and transforms them in the HL7 CCD v3 format Implementation details: ➢ Creating two workflows to handle Historical rejects. ➢ CDID_HistoricalLoad 2. CDID_ProcessHistoricalRejects. ➢ **CDID_HistoricalLoad:** This workflow will have a single step that will invoke a java class. HistoricalLoad**:** ▪ It will get all the rejected csv files(message id, reason..) from reject location and merge all the rejected message ids into one file. ▪ Creates a HBase table on top of the consolidated rejected message ids. ➢ **CDID_HistoricalLoad workflow:** this workflow has 3 steps. 1. GetFilesFrom Archive: 2. FilteringRejects: 3. MoveFilesTo LandingZoneDir: ➢ Once xml,xsd and control file are moved into the Landing zone directory new incremental intake workflow will be triggered.

Technologies Used	**MapR2.0,jdk1.6,,ApacheCassandra2.0,PIG0.12.0,Mapreduce, AVRO, HDFS, Zookeeper, Sqoop, spring, hibernate, MySQL, Impala**
Responsibilities	• Design the Solution Architecture for CDID Service using Edraw. • Involve in writing java code to read the input xml file from /datafabric/d_hdfs/optum/optumhealth/commondocumentintakedistribution/dq_working/20140409T055107_4075_CDD_CSP_HL7_SUBQ_FROM_CDID_IFW_DATA_484_FULL.xml. • Write Mapreduce applications which process each line as a xml file from input file. • Design and development of Data Ingestion Process services • Making changes only in the schema validation step for processing of the rejects. • write MapReduce for New Process for Schema Validation • Write a MapReduce job to read data from Cassandra and write it to Cassandra from HDFS. • Write PIG User Defined Functions (UDF) to insert the data into HDFS and read the data from HDFS . • write PIG Script to Filter the data and insert to HDFS . • Build a Script to join driverId, Name, Hours and Miles Logged • Involved into **CDID_HistoricalLoad** process and Update Ingestion History • involved Schema Change Detection using MD5 algorithm • Design and develop application for the analysis of the health data

ClubRecommendation And Analytics System(CRAAS) For Business Growth In E-Commerce	
Client/Project Duration	**MIAC Analytics, Bangalore / November 2015 to Nov 2016**
Project Description	This project deals with clubbing many applications for their recommendation and performs analytics as per market needs and current demand for the E-commerce domain. This project deals • Recommendation of the Item using Collab Filtering Algorithm and perform Analytics: these modules recommend the item while shopping based on the user rating in the e-commerce portal. This will perform analytics for rated item only and show the saled of the item in the reporting and analytics tools. We are using bestbuy.com API for testing platform. • Recommendation of Item to improve the business sales For E-Commerce Portal Using SBTA (Structural Balance Theory Algorithm). These modules also perform the analytics on the offline and online data to see the increase in the sales demand for unrating product also which is listed in the portal. • Recommendation of the product based on microbloging site like

	twitter and Facebook etc. Here we are using microbloging site like face book and twitter. this module makes set of the on given keywords using K-Means Clustering and Hierarchical Clustering Algorithm . It Stream all the online data as per Cluster Setup and remove the stop word. • Image Recommendation and Analytics of the Online Image For branding and model demand about the camera or mobile based on online e-commerce portal. **<u>Implementation details For Image Recommendation And Analytics :</u>** - Image dataset will be sourced from the Flickr, sample datasets crawled from web - Hadoop Image Processing API will be used to process the image - Image dataset are downloaded and imported on Hadoop . - MapReduce jobs will be configured and run against the Image Data set. All the image would be store in HadoopImageBundle(*.hib) format in HDFS. - Run the MaReduce using Mesos to read metadata of the HIB and store into HDFS using PIG . - Standard benchmark mapreduce jobs distributed with Apache Hadoop will be used to measure the performance. - Apache Hadoop installation will be applied Mesos – CPU resource aware Scheduler and will be configured accordingly. - HiBecnh – Performance measuring tool for Mapreduce will be rerun with same set of benchmark mapreduce jobs and datasets. - Recommendation and analytics would perform based on metadata of the images
Technologies Used	**CDH5, Java, HDFS, MongoDB3.2, PIG0.13.0, Mapreduce, Mesos1.0.1, ApacheNutch1.10, Sqoop, Zookeeper3.4.6, Spark1.3.0, Scala2.10, Kafka-2.10.0.9.0.1, spring, hibernate, mysql5.6**
Responsibilities	• Design the Solution Architecture for Club Recommendation • Client interaction and understand the requirement of the Recommendation engine . • setting Scala,SBT and Spark . • setting MongDB and Apache Kafka • write function for frequent patterns matching ,people to people correlation, customer review and ratings • develop recommendation from other's views • use Spark/MLib for Machine Learning Algorithms • use K-means ,Expectation -maximization ,power Iteration Clustering Algorithm • Perform Association Analysis using Frequent Pattern Mining (FPGrowth). • Population the e-commerce Data set from bestbuy.com and amazon • use content based recommendation algorithm in Scala . • Write higher order user defined function in Scala . • Perform Analytics using Spark • Exported the result to MySQL using Sqoop

	• Configured Apache Nutch and write MapReduce Program in Spark for the WebCrwalling of the image as data • Use Kafka for the Microbloging data in the project • Configure mesos with Hadoop and write MapReduce program to crawl the image and store as Hadoop Image Bundle file into HDFS. • Write MapReduce Program and PIG using UDF to read image metadata and store into HDFS. • Adding Recommendation listing • understanding Recommendation behavior • Perform Analytics
Enrich -Gloabl Payment GateWay	
Company/Project Duration	**EziDebit-AMexCard, Australia / December'April 13- July'13**
Client Location	**Enrich-IT,DLF Building, Gachi Bowli, Hyderabad-500019**
Technologies Used	Java, Hive, Sqoop
Responsibilities	• High Level Design • Project deals with Payment settlement of the card holder for the bank like Amex card Holder. • Hadoop / Hive configuration and setup. • Hive Schema design for Oracle OLAP tables. • Design / Develop Server side component.

Bank-Fusion Universal Banking (UB) – Product support	
Project Duration	July 2011 to December 2011
Company	**Agewell Technologies Pvt Ltd**
Client Location	**IBM, Manayata**
Project Description	This project involved support to onsite implementation team for various modules.
Technologies Used	Java, JBoss 4.0, Bank Fusion Platform
Responsibilities	• Discussion with client and understanding problem • Fixing Bugs. • Performance fixing for batch processes.

Bank-Fusion Universal Banking (UB) – Integration of Universal Banking	
Project Duration	December 2011 to November 2012
Company	**Agewell Technologies Pvt Ltd**
Client Location	**IBM, Manyata**
Project Description	This project involved interface integration of Universal Banking with other Misys products such as OPICS, TI etc and Intellect Flow (Polaris product)
Technologies Used	Java, .JMS, Hibernate, JBoss 4.0, IBM Web sphere MQ ,Bank Fusion Platform
Responsibilities	• Requirement understanding for integration of Universal Banking with

	Trade Innovation. • High level design for financial posting from TI to Universal Banking. • IBM Web sphere MQ configuration as part of messaging layer. • Development of accounting Entries from TI to Universal Banking. • Development of Universal Banking-TI Meridian Repository using Misys Message Manager Middle ware for message transformation and routing. • Design and development of UB with Intellect Flow(Polaris product).

Levvo Software Technologies Pvt Ltd	
Project Duration	February 2008 to June 2011
Company	**Levvo Software Technologies Pvt Ltd**
Roles	**IT - Trainer**
Technologies	**C, C++, Core Java, j2ee, struts, Oracle 9i/10g, Android, Python**
Responsibilities	❖ I used to train their fresher employee at the company premises ❖ I used to support the development team for the any programming issue in java ❖ i used to write PLSQL code in the development team ❖ I used to visit client location from company side as trainer for providing corporate training ❖ I used to take some set of workshop on technologies like java collection ,plsql to the engineering college in Bangalore ❖ I was guest trainer for SSI, NIIT, Manual Academy, Engineering colleges from my company.

Academic Credentials

Highest Degree	Year of Passing	Board/University
Master of Science (MSc.IT) –	**2004**	**Sikkim Manipal University**
B.Sc. Computer Science	**1999**	**Magadh University, BodhGaya**

Awards & Achievements

- **Project of the Quarter Award for first quarter of year 2016.**
- **SPOT Award for Excellent Performance for the Bank fusion module.**
- **Received several client appreciations for well management and delivering the project module well before the deadline.**