|  |
| --- |
| **RAJESH C** **E-mail:** crajesh55@gmail.com  **Mobile:** +91- 9948485837 |

Professional Summary:

* Have over 12 **years and 2 months** of work experience, with 4+ years of good Knowledge in working with **Apache** **Spark ,Elastic search, Talend ,Hadoop Framework (HDFS & Map Reduce), Hive, IMPALA & HBASE ,Oozie,Flume,mongoDB ,SparkStreaming,kafka ,Machine Learning, DataRobot, Core Java ,Oracle and Tableau** technologies in different areas like development, analysis ,reporting, implementation and support of business requirements.
* Having good understanding about Machine Learning, used Regression Models, Classification models, and Recommendation models.
* Have understanding of Deep Learning(Neural networks)
* UNIX Scripting, scheduling jobs using Crontab
* Having understanding of Data Warehousing and ETL tool like Data Stage
* Ample knowledge on Oracle and MySQL in SQL and PL/SQL technologies which is inclusive of procedures and functions etc.
* Good awareness on organizational culture, responsibilities in individualistic and business perspective
* Decent knowledge in Telecom domain including VoIP product
* Excellent Communication , Problem Solving Skills along with good team skills

Experience Summary:

* Working in BankOfAmerica from 24th Aug-2015 to till date.
* Worked for Valuelabs from 7th July-2014 to 19th Aug-2015 date.
* Worked for **TATA Consultancy Services** (TCS) from 31st MAR 2009 to 27th June-2014 which is over 5 years and 3 months of work experience.

Certifications Summary:

* Cloudera Certified Hadoop Developer (CCDH-410)
* Sun Certified Java Programmer (SCJP) for Java SE 5
* Oracle Certified Associate (OCA) for Oracle 9i including SQL and PL/SQL.
* Oracle Database 10g Administrator Certified Associate
* Received Delivery Excellence certificate from CTL client.

Educational Qualification:

* B.TECH (Electrical and Electronics) passed in First Class from JNTU University, Hyderabad on May 2006 with aggregate of 68%.
* XII (MPC) passed in First Class with distinction on Mar 2002 with aggregate of 91%.
* SSC passed in First Class with distinction on Mar 2000 with aggregate of 88%.

Technical Skill Set:

|  |  |
| --- | --- |
| **Technologies** | Apache Spark,Haddop ,Oracle PL/SQL,UNIX ,Scala,Core Java and Python. |
| **Frameworks** | Spark Mlib, Elastic search,Hadoop eco system(Map/Reduce, Hbase, Hive Pig), Zookeeper, Sentry,Flume, sqoop, Oozie, Apache Solr, Spark streaming and Kafka. |
| **Databases** | Oracle, MySQL, HBASE, Hive, Impala |
| **Platforms** | Windows XP, Linux, Ubuntu-11.10,CHD 5.10 |
| **Development** **Tools** | Intellij,Autosys,Talend, Tableau reporting, Eclipse, Maven, JUNIT, Apache Tomcat ,Toad ,PL/SQL Developer, MySQL GUI,Tableau desktop,Jenkins,Nexus,BitBucke,Jira,Postmen,CICD |
| **Big Data** | Hadoop, Hive, HDFS, OOZIE, Spark-core, Spark-Sql ,Spark Streaming,Deep learning ,neural networks, machine Learning |
| **Virtualization Tools**  | VM Ware Workstation, Virtual Box |

PROJECTs SUMMERY:

|  |  |
| --- | --- |
| **Project Name**  |  Search Analytics |
| **Customer** |  Bank |
| **Period** |  Mar-2019 to Till now |
| **Role** |  Developer |
| **Technologies** | Apache Spark,Scala,Hive,Autosys,Unix |
| **Tools**  |  IntelliJ IDE,Autosys,Putty,CDH 5.10,CICD,Ansible,Jenkin,Git,Insomnia |
| **Description** | Search Analytics is a User Interface similar to Google,which used by bank clients to search through customers data like Accounts,Demographic,Transaction,SAR,CASE,TRMS to find out are there any fraudulent transaction happening which used to mitigate Risk. |
| **Role & Responsibilities** | * Worked on Creating Customer and Accounts json using Spark scala with Different attributes having customer details ,Account details,entity details,CASE,SAR and TRMS.
* Implemented business ligic using complex aggregations and grouping on different data sets using Spark SQl
* Did tuning to spark code to handle memory issues and to handle large datasets.
* Created Autosys jobs to schedule the Spark jobs.
* Created Ansible templates to deploy spark code and autosys jils automatically
* Developed unit test case to test the sparkcode.
* Implemented Data Controls to validtae differet types of validation using Spark .
 |

|  |  |
| --- | --- |
| **Project Name**  |  Equities Content Distribution Model |
| **Customer** |  Bank |
| **Period** |  Feb-2018 to Mar-2019 |
| **Role** |  Developer |
| **Technologies** | DataRobot,Impala,Talend ETL |
| **Tools**  |  Talend 6.1.1,mremoteNG,CDH 5.10 |
| **Description** | The DataRobot – Equities Content Distribution Model was developed with the intent of aiding the Equity Sales business by curating a custom tailored list of Equity research reports for their client contacts. Currently within Equities Sales there are two primary ways in which clients receive bank generated research – either directly from analysts/salespeople or through research subscriptions. Via the latter method, any report fitting some user-defined criteria is automatically emailed to the client at the time of publication, often resulting in an excessive number of research reports being distributed. Analysis of these subscriptions found that readership of Equity reports sent via this method was quite poor, with only 1% of sent reports being read. While the former method of having analysts and the sales team manually send research to their clients is a more accurate approach, it is also very time consuming, since coverage for one individual typically spans hundreds of clients, thus making it impossible to personally tailor research to low and high touch clients alike. The DataRobot – Equities Content Distribution Model aims to address these issues by leveraging predictive analytics to identify a client’s readership patterns to output recommendations specifically tailored to their interests. The model outputs a score corresponding to the “likelihood” of report readership for different combinations of clients and newly published Equity research reports. Based on these scores, report recommendations are made. |
| **Role & Responsibilities** | * Developed talend jobs using Talend Open studio (TOS) to pull data from different source systems to hdfs and scheduled these jobs to run at particular time in TAC.
* Used context params in talend while creating jobs.
* Data sets preparation : Readership Data,Research Report Metadata ,Research Subscriptions,Trade Data,Client Metadata,GICS Industry Data,Touchpoint Data,Beatfactor Data, Market Data,Securities of Interest Data,Regional Data
* Data Quality and Reconciliation and Data Exclusion.
* Dependent Variable(s) Definition and Independent/Predictive Variable(s) Definitions
* Data aggregation queries at different levels
* Missing values imputed
* Tuning of the model
 |

|  |  |
| --- | --- |
| **Project Name**  |  Research Now Recommendations |
| **Customer** |  Bank |
| **Period** |  MAY-2017 to Feb-2018 |
| **Role** |  Developer |
| **Technologies** | Hadoop ,Machine learning ,Spark Mlib (Python),Elastic search |
| **Tools**  |  Talend 6.1.1,mremoteNG,CDH 5.10 |
| **Description** |  Research Now provides Recommended Research Reports, Trending search terms, Trending reports and featured videos to user in rl pageUses Spark ml libraries (lr) to provide recommendations and impala for providing top search terms and trending reports. |
| **Role & Responsibilities** | * Developed talend jobs using Talend Open studio (TOS) to pull data from different source systems to hdfs and scheduled these jobs to run at particular time in TAC.
* Used context params in talend while creating jobs.
* Written spark Mlib programs using python for generating recommendations.
* Developed hive scripts for creating external tables.
* Developed elastic scripts to create indexes and store data into them.
* Developed hive scripts using common table expressions (CTE) for generating trending reports.
 |

|  |  |
| --- | --- |
| **Project Name**  |  Research Recommendation engine(RRE) .v2 |
| **Customer** |  Bank |
| **Period** |  June-2016 to Apr-2017 |
| **Role** |  Developer |
| **Technologies** | Hadoop Machine learning ,Spark Mlib (Python),Elastic search |
| **Tools**  |  Talend 6.1.1,mremoteNG,CDH 5.10 |
| **Description** |  RRE is a model, which is developed to provide trending companies, reports, analysts’ names, search terms and gives recommendation to users based on their readership data which is developed using apache machine learning.Here source system is RDWDB (MS Sql) where we pulls data and do some pre-processing and provides this input to machine learning to get recommendation. On the generated output we do some post processing and stores the final results in MS Sql. In RRE v2 we are replacing machine learning with spark MLib which is fast and simple in processing. And final results which are generated by model are going to store in elastic indices instead of DB which is very faster while querying the data. |
| **Role & Responsibilities** | * Developed talend jobs using Talend Open studio (TOS) to pull data from different source systems to hdfs and scheduled these jobs to run at particular time in TAC.
* Used context params in talend while creating jobs.
* Written spark Mlib programs using python for generating recommendations.
* Developed hive scripts for creating external tables.
* Developed elastic scripts to create indexes and store data into them.
* Developed hive scripts using common table expressions (CTE) for generating trending reports.
 |

|  |  |
| --- | --- |
| **Project Name**  |  Investors interest Dashboard |
| **Customer** |  Bank |
| **Period** |  Feb-2016 to May 2016 |
| **Role** |  Developer |
| **Technologies** | Impala 2.7.0, CDH 5.10 |
| **Tools**  |  Talend 6.1.1,Tableau  |
| **Description** |  Investors Interest dashboard is tableau report where an analyst can see all the kinds of data related to an investor, which includes his trade data, touch points data, contact data.There are different source systems (Vertica, Oracle DB, and servers) from where we pulled data using talend jobs placed it in hdfs. On top of these data we have created impala external tables pointing to this location.Used impala join queries to do final aggregations.Tableau reports used this final aggregated data as source and created bar graphs other vies for final repots.We have jobs scheduled in Talend Admin console (TAC) where each job is scheduled to run at pre-set time and tableau reports will be updated based on the aggregated data.  |
| **Role & Responsibilities** | * Developed talend jobs using Talend Open studio (TOS) to pull data from different source systems to hdfs and scheduled these jobs to run at particular time in TAC.
* Developed final impala script by using Common table expression (CTE) which joins different tables and produces required data for report.
* Developed impala scripts for creating external tables.
* Did basic reporting in Tableau.
 |

|  |  |
| --- | --- |
| **Project Name**  |  ATA |
| **Customer** |  Bank |
| **Period** |  Aug-2015 to Jan 2016 |
| **Role** |  Developer |
| **Technologies** | Apache Spark Streaming, Kafka, HIVE, Java,Oozie,flume,CDH5.5 |
| **Tools**  |  Eclipse and Maven |
| **Description** |  Advisory Trading Analytics(ATA) is a real time trades analytics tool.Which streams sales data from different stock exchanges and performs aggregation on Sales Trader wise, Sector wise, Client Type.Spark streaming used to consume messages from kafka brokers and aggregation is done in Spark using Java.Messages were received in Avro format they were decoded for aggregation and after doing aggregation messages were published to kafka in avro format.Final reports were displayed in front end. |
| **Role & Responsibilities** | * Developed Spark Streaming application to consume messages from kafka topic and written decoders to convert messages to string format.
* Written aggregation logic by maintaining state for a particular key and converted messages to avro format
* Aggregated avro messages were published to kafka topic through a java client.
* Written flume agents to consume messages from kafka and store them in hdfs and configured them in Cloudera Manager.
* Developed hive scripts for creating external tables.
 |

|  |  |
| --- | --- |
| **Project Name**  |  AMS |
| **Customer** |  TATA Sky |
| **Period** |  Jan-2015 to till date |
| **Role** |  Developer |
| **Technologies** |  Hadoop, HIVE, Java,Oozie,flume,CDH5 |
| **Tools**  |  Eclipse and Maven |
| **Description** |  Tata Sky is a [direct broadcast satellite](https://en.wikipedia.org/wiki/Direct_broadcast_satellite%22%20%5Co%20%22Direct%20broadcast%20satellite) television provider in India, using [MPEG-4](https://en.wikipedia.org/wiki/MPEG-4%22%20%5Co%20%22MPEG-4) digital compression technology, transmitting using [INSAT 4A](https://en.wikipedia.org/wiki/Indian_National_Satellite_System%22%20%5Cl%20%22INSAT-4A%22%20%5Co%20%22Indian%20National%20Satellite%20System) and GSAT-10 satellite.[[2]](https://en.wikipedia.org/wiki/Tata_Sky%22%20%5Cl%20%22cite_note-2) Incorporated in 2004, Tata Sky is a Joint venture between the [Tata Group](https://en.wikipedia.org/wiki/Tata_Group%22%20%5Co%20%22Tata%20Group) and [21st Century Fox](https://en.wikipedia.org/wiki/21st_Century_Fox%22%20%5Co%20%2221st%20Century%20Fox).TataSky is the one of the largest DTH service in INDIA right now.The aggregated data will need to be available at the following levels:* Customer viewing behavior channel wise.
* Customer viewing behavior program wise.
* Channel wise minutes viewership.
* Program wise minutes viewership.
* Written JUNIT test cases.
 |
| **Role & Responsibilities** | * Wrote HIVE scripts to load data from CSV(Ex:AMS,SAF) files to hive tables.
* Wrote Java XML parser application to parse PVR xml file.
* Designed schema for HIVE tables.
* Developed custom flume interceptor to parse AMS\_log files with the frequency of more than one million per day.
* Written Oozie workflows to schedule hive scripts and java parser application with the help of co-ordinator.
* Written different hive join queries to generate final output aggregation data.
 |

|  |  |
| --- | --- |
| **Project Name**  |  Snapfish Picture Management. |
| **Customer** |  SNAPFISH |
| **Period** |  SEP-2014 to DEC-2014 |
| **Role** |  Developer |
| **Technologies** |  Hadoop, HIVE,PIG, SQOOP, Java, MySQL, HBASE |
| **Tools**  |  Eclipse |
| **Description** |  Snapfish is a web-based photo sharing and photo printing service that is owned by Hewlett-Packard. Members can upload files for free, and are given unlimited photo storage.Most Snapfish features are free. However, if members want to download high resolution or original copies of their own uploaded images, Snapfish charges a per-image fee for each download. From 26-Jan-2012 Snapfish stopped charging the members for downloading their picturesSnapfish need below reports from existing large volumes of data in different oracle servers.* Count of number of photos uploaded day wise, website wise and upload type.
* Sum of storage taken by H.R and L.R type photos
* Count of pictures by product wise by joining with main table.
 |
| **Role & Responsibilities** | * Wrote HIVE scripts to load data (pipe delimited) from text file to hive tables.
* Wrote scripts to create HIVE-HBASE mapping tables.
* Designed schema for HBASE and HIVE tables.
* Wrote HBASE Map reduce programme to aggregate data in HBASE tables.
* Implemented HBASE joins using map reduce.
 |

PROJECT SUMMERY:

|  |  |
| --- | --- |
| **Project Name**  |  Allianz Insurance quotation processing |
| **Customer** |  ALLIANZ |
| **Period** |  JULY-2014 to AUG-2014 |
| **Role** |  Developer |
| **Technologies** |  Hadoop, HIVE,PIG, SQOOP, Java, MySQL, HBASE |
| **Tools**  |  Eclipse |
| **Description** |  Allianz is a German multinational financial services company headquartered in Munich. Its core business and focus is insurance. As of 2013, it was the world's largest insurance company, the 11th-largest financial services group and 25th-largest company, as well as the largest financial services company when measured by 2012 revenue.Its Allianz Global Investors division ranks as a top-five global active investment manager, with specialized asset managers including PIMCO (bonds), RCM (equities) and Degi (real estate).Allianz has below query from the existing data* Aggregated count of Day wise, Week wise, Month wise and Year wise quotes received by customers and sum of premium amount for a product.
 |
| **Role & Responsibilities** | * Setup and benchmarked Hadoop/HBase clusters for internal use
* Wrote JOIN queries in IMPALA/HIVE to generate aggregated reports Day wise, Week wise Month wise and Year wise.
* Implemented map reduce for loading the JSON format data to HIVE tables.
* Wrote Map-reduce jobs to interact with HIVE
 |

PROJECT SUMMERY:

|  |  |
| --- | --- |
| **Project Name**  |  Telecom data Predictive Analysis |
| **Customer** |  Century Link |
| **Period** |  SEP-2013 to JUNE-2014 |
| **Role** |  Developer |
| **Technologies** |  Hadoop, HIVE,PIG, SQOOP, Java, JUnit, MySQL, HBASE |
| **Tools**  |  Eclipse |
| **Description** |  CTL is one of the largest telecom service providers in USA. It provides Business VoIP and Consumer VoIP products to its customers. CTL is one of the clients to TCS. To improve its business, CTL is looking for market analysis/predictive analytics based on existing users and their orders. CTL would like to develop a POC on small scale data. CTL has below queries to be analyzed from existing BIG Data (for past 10 years).1. What are the peak& off-peak hours (i.e. at what hour of the day most orders are getting placed)?2. Who are the top 10 customers placed the most number of the orders?3. What types of orders have eroded out most?4. Predict the number of "Disconnect" orders will be placed in the year 2013? |
| **Role & Responsibilities** | * Setup and benchmarked Hadoop/HBase clusters for internal use
* Loading the data to HDFS and HIVE through Sqoop from various Mysql databases of various policy sources.
* Writing Map-reduce jobs to interact with Hbase
* Writing Hive UDF (User Defined Functions).
* Used Hbase and Hive to store/retrieve data using HQL.
 |

|  |  |
| --- | --- |
| **Project Name**  |  Web log Analysis |
| **Customer** |  Century Link |
| **Period** |  MAR-2013 to SEP 2013 |
| **Role** |  Developer |
| **Technologies** |  Hadoop, HIVE, PIG, HBASE, SQOOP, Java 1.6, Ubuntu-11.10. |
| **Tools**  |  Eclipse |
| **Description** |  The POC is based on web log analysis to analyze the data in the web server logs. Analysis of the data collected from the website visitors will help increase the return on investment of the website. The main purpose for web log analysis has traditionally been to gain a general understanding of what is happening on the site. The current report is based on the number of hits. There are also reports on number of hits received from various countries. The domain names like .com, .net, .in etc. are used for identifying the country. The ultimate goal of the web housing exercise is to make the anonymous web traffic information into meaningful analytical information. |
| **Role & Responsibilities** | * Analyzed the functional specification.
* Developed Map/Reduce code using Hadoop Framework.
* Used Hbase and Hive to store/retrieve data using HQL.
 |

|  |  |
| --- | --- |
| **Project Name** |  VoIP Reports Application  |
| **Customer** |  Century Link |
| **Period** |  MAY-2012 to MAR 2013 |
| **Role** |  Developer |
| **Technologies** |  Java , JSP ,JDBC and Oracle |
| **Tools**  |  Eclipse, Apache Tomcat , Toad |
| **Description** |  VoIP Report Application is web application to report Business and Consumer VoIP Order ,Customer and Telephone number status in OMWF It takes order number or Telephone number or any other key attributes as Input and ports the details and state of the order  Produces List of failed, due date crossed and porting orders for Consumer VoIP Customer Provides DML tool to perform any DML operations on the OMWF Data Base  |
| **Role & Responsibilities** |  Held the role of Developer for this project with below responsibilities:  Writing code to develop GUI and business logic for most screens of the application  Writing the SQL queries and PLSQL procedure and functions required to produce different types of orders like filed ,due date crossed and porting orders  Writing unit test cases and performing unit testing  Participated in daily status calls |

|  |  |
| --- | --- |
| **Project Name** |  CTL AMT BMG MNT HYD |
| **Customer** |  Century Link |
| **Period** |  SEP-2009 to MAY 2012 |
| **Role** |  Maintenance and Developer |
| **Technologies** |  Oracle SQL and PL/SQL |
| **Tools**  |  Toad , PL/SQL Developer , Altova XML Spy, Remedy |
| **Description** |  OMWF is one of the key provisioning application for Century Link which takes cares of activating and deactivation of VoIP Numbers for different kinds of products Integrated Access , Hosted VoIP and networkx supporting different protocols like SIP , MGCP ,HVDS and SIP Trunking and providing emergency 911 service for the customers  |
| **Role & Responsibilities** |  Held the role of Maintenance and Developer for this project with below responsibilities:* Understanding the requirements of every release writing PL/SQL procedures inside packages to which will gets called by OMWF java application
* Providing support to clear the data related issues regarding OMWF application
* Creating the procedures to perform the group of related activities and modifying the same to cope up with the new release if required
* Understanding the data design for the new products and platforms that Century link offering and resolve any issues during the production
* Interacting with application development tem when found any production issues and trying to resolve the issue
* Interacting with external third party teams when there are issue with provisioning or 911 emergency service
* Supporting VoIP portal in case of any data issues from OMWF perspective
* Performing data pulling, grouping and analysing the quarterly 911 audit data to guarantee the emergency service for all the users

  |

Personal Details

|  |  |
| --- | --- |
| Mobile Number | +91701383551 |
| Date of Birth | 16-05-1985 |
| Email Id | crajesh55@gmail.com |
| Nationality | Indian |
| Sex | Male |
| Marital Status | Married |
| Designation | Analyst-I |