

Hadoop Developer Hadoop Developer Hadoop Developer - Verizon Dallas, TX Having more than 7+ years of experience in IT industry, involved in Developing, Implementing, testing and maintenance of various web based applications using J2EE technologies and Big Data ecosystems experience on Linux environment. Including comprehensive experience as a Hadoop, Big Data & Analytics Developer. Expertise on Hadoop architecture and ecosystems such as HDFS, Map Reduce programming paradigm. Experience in installation, configuration, supporting and monitoring Hadoop clusters using Apache, Cloudera distributions and AWS. Knowledge in installing, configuring, and using Hadoop ecosystem components like Hadoop Map Reduce, HDFS, HBase, YARN, Oozie, Hive, Sqoop, Pig, Zookeeper and Flume. Good understanding of NoSQL databases like HBase, Cassandra and MongoDB. Experience in Spark and scala. Experience in importing and exporting data using Apache Sqoop from HDFS to Relational Database Systems / Non-Relational Database Systems and vice-versa. Extending Hive and Pig core functionality by writing custom UDFs. Experienced in analyzing data using HiveQL, Pig Latin, and custom Map Reduce programs in Java. Experience in building, maintaining multiple Hadoop clusters (prod, dev etc.,) of different sizes and configuration and setting up the rack topology for large clusters. Worked on NoSQL databases including HBase, Cassandra and MongoDB. Experienced in SYBASE, Teradata. Experienced in Datawarehousing systems. Experienced in data integration and Datamodeling. Experienced in job workflow scheduling and monitoring tools like Oozie and Zookeeper Strong experience with XML, JSON, SQL, DB2, Oracle. Experience on the Linux operating system. Experienced in designing, developing and implementing connectivity products that allow efficient exchange of data between the core database engine and the Hadoop ecosystem.

Experienced in Data warehousing and using ETL tools like Informatica and Pentaho. Expert level skills in developing intranet/internet application using JAVA/J2EE technologies which includes Struts framework, MVC design Patterns, Chrodiant, Servlets, JSP, JSLT, XML/XLST, Java Script, AJAX, EJB, JDBC, JMS, JNDI, RDMS, SOAP, Hibernate and custom tag Libraries. Experience using XML, XSD and XSLT. Experience with web-based UI development using jQuery UI, jQuery, ExtJS, CSS, HTML, HTML5, XHTML and JavaScript. Extensive experience in middle-tier

development using J2EE technologies like JNDI, JSP, Servlets, JSF, Struts, Spring, Hibernate, JDBC, EJB. Possess excellent technical skills, consistently outperformed schedules and acquired interpersonal and communication skills. Authorized to work in the US for any employer Work Experience Hadoop Developer Verizon - Dallas, TX January 2017 to Present Verizon's employee portal provides ability to search employee browsing habits at work. To increase employee productivity and ensure security of company information, Verizon's employee portal tracks employee email and internet browsing collects various data like urls, transmitting data size into Hbase/HDFS. Hadoop ecosystem was used collect the Big Data and analyze the data. Flume servers were installed on the network proxy servers to collect the data and store into HDFS. Responsibilities Involved in full life-cycle of the project from Design, Analysis, logical and physical architecture modeling, development, Implementation, testing. Responsible to manage data coming from different sources and involved in HDFS maintenance and loading of structured and unstructured data. Collected data was analyzed using Map Reduce Jobs and presented to employee portal built using Spring MVC. Processed Big Data using a Hadoop cluster consisting of 40 nodes. Designed and configured Flume servers to collect data from the network proxy servers and store to HDFS. Developed data pipeline using Flume and Java map reduce to ingest employee browsing data into Hbase/HDFS for analysis. Used agentE2EChain for reliability and failover in flume. Designed and implemented Map Reduce jobs for analyzing the data collected by the flume server. Actively involved in working with Hadoop Administration team to debugging various slow running MR Jobs and doing the necessary optimizations. Designed and implemented RESTFul APIs to retrieve the data from Hadoop Platform to Employee Portal Web Application. Optimized the full text search function by connecting MongoDB and Elastic Search. Utilized AWS framework for content storage and Elastic Search for document search. Creating Hive tables and working on them using Hive QL Created concurrent access for hive tables with shared and exclusive locking that can be enabled in hive with the help of Zookeeper implementation in the cluster. Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions. Wrote MRUnit tests for unit testing the Map Reduce jobs. Performed functional

requirement review. Worked closely with Risk & Compliance Team and BA. Facilitated Knowledge transfer sessions. Worked in an agile environment. Hadoop Developer Kroger Co - Cincinnati, OH April 2014 to December 2016 Retail-Logistics is in-store analytics program which provides comprehensive information on how to resolve the shopper's traffic. It provides the detailed analysis to understand shoppers' behavior, purchase pattern. By using these insights it is used to increase in-store traffic, optimize staffing, improve merchandising and display performance. Responsibilities

Workings on bigdata infrastructure build out for batch processing as well as real-time processing. Worked on Hadoop, Hive, Oozie, and MySQL customization for batch data platform setup. Worked on implementation of a log producer in SCALA that watches for application logs, transforms incremental logs and sends them to a Kafka and Zookeeper based log collection platform. Implemented a data export application to fetch processed data from these platforms to consuming application databases in a scalable manner. Design, deploy, Manage cluster nodes for our data platform operations (racking/stacking) Install and configure cluster. Setting up puppet for centralized configuration management. Monitoring Cluster using various tools to see how the nodes are performing. Developed Spark scripts by using Scala Shell commands as per the requirement. Expertise in cluster task like adding Nodes, Removing Nodes without any effect to running jobs and data. Write scripts to automate application deployments and configurations. Monitoring YARN applications. Troubleshoot and resolve cluster related system problems. Wrote map reduce programs to clean and pre-process the data coming from different sources. Implemented various output formats like Sequence file and parquet format in Map reduce programs.

Also, implemented multiple output formats in the same program to match the use cases. Developed Hadoop streaming Map/Reduce works using Python. Performed benchmarking of the No-SQL databases, Cassandra and HBase. Created data model for structuring and storing the data efficiently. Implemented partitioning and bucketing of tables in Cassandra. Implemented test scripts to support test driven development and continuous integration. Converted text files into Avro then to parquet format for the file to be used with other Hadoop eco system tools. Experienced on loading and transforming of large sets of structured, semi structured and

unstructured data. Developed and implemented core API services using Scala and Spark. Exported the analyzed data to HBase using Sqoop and to generate reports for the BI team. Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. Participate in requirement gathering and analysis phase of the project in documenting the business requirements by conducting workshops/meetings with various business users. POC work is going on using Spark and Kafka for real time processing. Developed a data pipeline using Kafka and Storm to store data into HDFS. Populated HDFS and Cassandra with huge amounts of data using Apache Kafka. POC work is going on comparing the Cassandra and HBase NoSQL databases. Worked with NoSQL databases like Cassandra and Mongo DB for POC purpose. Implement POC with Hadoop. Extract data with Spark into HDFS. Environment MapReduce, HDFS, Hive, Pig, Hue, Oozie, Core Java, Eclipse, Hbase, Flume, Spark, Scala, Kafka, Cloudera Manager, Cassandra, Python, Greenplum DB, IDMS, VSAM, SQL*PLUS, Toad, Putty, Windows NT, UNIX Shell Scripting, Pentaho, Talend, Bigdata, YARN, Mongo DB Sr. Java Developer Huntington Bank - Columbus, OH January 2012 to March 2014 Risk Rating is developed for calculating the risks on the loans given by the Huntington Bank to their customers. This is the rating system for the loans, which takes the borrower and customer type information and performs a series of ratings and calculations. This system gives the rating of the risk associated with that loan. Frontend is built using a Spring MVC framework in the presentation layer. The back end functionality was built using Oracle as data persistence layer and Hibernate as data access layer. Responsibilities Reviewed functional and non-functional requirements. Implemented a highly visible POC using Hadoop ecosystem. Responsible for Identifying risks from the discussions happens during onsite hours and high light the risk by articulating risk and mitigation option to project managers. Responsible for leading performance test preparations and execution. And leading the performance issues from identifying the root cause, have discussions with offshore tech leads and architects to put a fix in place. Developed application service components, created Hibernate mapping files and generated database schema Designed and implemented UI module using JSPs and Spring MVC framework. Implemented design patterns like MVC, Factory, Composite and Singleton Implemented

persistence logic in ORM technology using Hibernate OR mappings. Developed hibernate mapping file and Criteria queries for retrieving data from the database Created various Data Access Objects for Addition, modification and deletion of records using various specification files and also using CRUD manager concept in Hibernate for the above operations. Designed and implemented risk rating engine. Used Log4J logging framework for logging messages. Performed Unit Testing of Various Modules by generating the Test Cases. Environment Java 1.6, Eclipse Indigo, JBoss 5.0, JSP, JQuery, Maven, JUnit, Log4J, Visio, Oracle, SQL Developer, HDFS 0.20, SVN, Unix, Hibernate 3.3.1 Java Developer InfoTech Software solutions May 2009 to December 2011 The FNS is low cost internet and mobile banking based delivery channel solution which allows intuitions to greatly enhanced customer service levels and to streamline banking services offered to customers and allows customers the flexibility to take control of banking activities via the internet via leveraging the rich functionality and uninterrupted 24/7 availability of the core banking. The rich functionality of Financial Connect include business banking, payroll scheduling and international funds transfer, together with standing payments-pay, domestic funds transfers, account enquiries, online loan simulations and loan applications. Responsibilities Developed Controllers for request handling using Spring framework Involved in Command controllers, handler mappings and View Resolvers. Designed and developed application components and architectural proof of concepts using Java, EJB, JSP, JSF, Struts, and AJAX. Participated in Enterprise Integration experience web services Configured JMS, MQ, EJB and Hibernate on Web sphere and JBoss Focused on Declarative transaction management Developed XML files for mapping requests to controllers Coded Spring Portlets to build portal pages for application using JSR 286 API Hibernate templates were used to access database Coded JDBC calls in the servlets to access the Oracle database tables. Responsible for Integration, unit testing, system testing and stress testing for all the phases of project. Prepared final guideline document that would serve as a tutorial for the users of this application. Use the DAO in developing application code Developed stored procedures. Extensively used Java Collection framework and Exception handling. Environment Java, J2EE5, Spring, JSP, XML, Spring TLD, JSP, Servlets, Hibernate

Criteria API, XSLT, CSS, JSF, JSF RichFaces, WASD, Java Swing, Web service, AXIS Server, WSDL, XML, Glassfish, JSR 286 API, UML, EJB, Java script, JQuery, Hibernate, SQL, CVS, Agile, JUnit, Oracle, SQL Developer. Hadoop Developer RCFile March 2003 to May 2003 Environment Hadoop 0.20.2 MR1, CDH3U6, HDFS, Hbase 0.90.x, Flume 0.9.3, Sqoop 1.x, AWS (EC2, S3, Elastic Search), Hive 0.7.1, Java 1.6, Linux, Spring 3.x, Eclipse Juno, XML, REST, JSON, Maven, Avro, RCFile, SVN Education Bachelor's Skills JAVA (8 years), SQL (7 years), APACHE HADOOP HDFS (6 years), Hadoop (6 years), HADOOP (6 years), Hadoop 0.20.2 MR1, CDH3U6, HDFS, Hbase 0.90.x, Flume 0.9.3, Sqoop 1.x, AWS (EC2, S3, Elastic Search), Hive 0.7.1, Java 1.6, Linux, Spring 3.x, Eclipse Juno, XML, REST, JSON, Maven, Avro, RCFile, SVN (8 years), Drupal, Hybris, Cobol, Servicenow Additional Information TECHNICAL SKILLS Hadoop/Big Data HDFS, Mapreduce, HBase, Pig, Hive, Sqoop, Oozie, YARN, ZooKeeper, NoSQL DB, CDH3, CDH4, Apache Hadoop. Java & J2EE Technologies Core Java, Servlets, JSP, JDBC, JNDI, Java Collections IDE's Eclipse, Net beans Frameworks MVC, Struts, Hibernate, Spring Programming languages Java, Python, Ant scripts, Linux shell scripts Build Management Tools Maven, Apache Ant Version control SVN, github Databases Oracle 11g/10g/9i, MySQL, DB2, MS-SQL Server Web Servers Web Logic, Web Sphere, Apache Tomcat Web Technologies HTML, XML, JavaScript, AJAX, SOAP, WSDL Network Protocols TCP/IP, UDP, HTTP, DNS, DHCP ETL Tools Informatica, Pentaho Testing WinRunner, QTP, Selenium

Name: Becky Hill

Email: juanli@example.com

Phone: 9623290826