Hadoop/spark Developer Hadoop/spark Developer Hadoop/spark Developer - Verizon Wireless Fort Worth, TX ? Around 7+ years of IT experience, with over Four years of work experience in Big Data Hadoop. ? Experience in working on various components in Hadoop ecosystem like HDFS, YARN, HIVE, SPARK, MAPREDUCE, PIG, HBASE, SCOOP, FLUME, KAFKA, OOZIE, and ZOOKEEPER. ? Excellent programming skills in SCALA, JAVA and PYTHON. ? Experience in Managing scalable Hadoop clusters including Cluster designing, provisioning, custom configurations, monitoring and maintaining using Hadoop distributions: Cloudera CDH, HortonWorks HDP. ? Experience in migrating an existing on-premises application to AWS, Using AWS services like EC2 and S3 for small data sets processing and storage and Maintaining the Hadoop cluster on AWS EMR. ? Experience in using Sequence files like ORC, AVRO and Parquet file formats. ? Exposure to Data Lake Implementation using Apache Spark and developed Data pipe lines and applied business logics using Spark. ? Experience in working on Spark APIs like Spark Core, Spark SQL and Spark Streaming. ? Experience in performing Transformations and Actions on data using Spark RDDs, Datasets, DataFrames and DStreams. ? Worked on minimizing data transfers using Broadcast variables and Accumulators in Spark. ? Experience in monitoring scheduler stages and tasks, RDD sizes and memory usage, Environmental information, Information about the running executors using Spark UI. ? Experience in using dependency management tools like SBT to create a self-contained jar file. ? Well versed with developing and implementing MapReduce jobs using Hadoop to work with Big Data. ? Implemented the processing framework for converting SQL to a graph of Map/Reduce jobs and the execution time framework to run those jobs in the order of dependencies using Hive. ? Expertise in using Aggregate functions in Hive using HQL. ? Worked on partitioning Hive tables and running the scripts in parallel to reduce run-time of the scripts. ? Ability to develop Pig UDFs to pre-process the data for analysis. ? Experience in performing transformations like event joins, filter bot traffic and some pre-aggregations using PIG. ? Worked on HBase to put the data in indexed StoreFiles that exist on HDFS for high-speed lookups. ? Expert knowledge on MongoDB, NoSQL data modeling, tuning, disaster recovery backup used it for distributed storage and processing using CRUD. ? Worked on Ad hoc queries, Indexing, Replication and Load

balancing. ? Developed Sqoop Jobs to load data from Relational Database Systems like Oracle and MySQL into HDFS and Hive. ? Used Kafka to read and write streams of data like a messaging system in Hadoop. ? Experience in managing Hadoop jobs using Oozie. ? Generated various kinds of reports using Power BI and Tableau based on Client specification. ? Experience in all stages of SDLC (Agile, Waterfall), writing Technical Design document, ? Development, Testing and Implementation of Enterprise level Data mart and Data warehouses. Work Experience Hadoop/spark Developer Verizon Wireless - Fort Worth, TX March 2017 to Present The project deals with analyzing call logs and complex data from multiple sources using Apache Hadoop and spark across the data to predict the likelihood that any particular customer would leave. Hadoop helped company build more valuable customer relationships and reduce churn. Responsibilities: ? Responsible for building scalable distributed data solutions using Apache Hadoop and Spark. ? Involved in combining traditional trasactional and event data with social network data. ? Involved in loading data from relational database into HDFS using Sqoop. ? Deployed Scalable Hadoop cluster on AWS using S3 as underlying file system for Hadoop. ? Worked with spark core, Spark Streaming and spark SQL modules of Spark ? Involved in developing generic Spark-Scala functions for transformations, aggregations and designing schema for rows. ? Experienced in working with Spark APIs like RDDs, Datasets, DataFrames and DStreams to perform transformations on the data. ? Used Spark SQL to perform interactive analysis on the data using SQL and HiveQL. ? Experienced in processing live data streams using Spark Streaming with high level functions like map, reduce, join and window. ? Experienced in minimizing data transfers over Hadoop clusters using Spark optimizations like broadcasts variables and Accumulators. ? Worked on troubleshooting spark application to make them more error tolerant. ? Involved in loading the processed data into Hive warehouse. ? Stored the data in tabular formats using Hive tables and Hive Serdes. ? Implemented Static partitions, Dynamic partitions and Buckets in Hive. ? Used Oozie Operational Services for scheduling workflows dynamically. ? Used Reporting tools like Tableau to connect with Impala for generating daily reports of data. ? Designed, documented operational problems by following standards and procedures using JIRA. ? Collaborated with the infrastructure, network,

database, application and BA teams to ensure data quality and availability.   Environment: Hadoop 2.x, Spark Core, Spark SQL, Spark Streaming, Scala, Hive, Sqoop, Oozie, Amazon, EMR, Tableau, Impala, JIRA. Hadoop Developer Molina Healthcare - Long Beach, CA September 2015 to March 2017 The project deals with developing a centralized data repository for maintaining medical information and performing relevant transformations on the data using Hadoop and spark ecosystems to obtain optimal solutions and provide further data query capabilities to improve medical and economic value.   Responsibilities: ? Developed the code for Importing and exporting data into HDFS and Hive using Sqoop. ? Deployed multiple clusters(Cloudera's) CDH distributions on AWS ? Automated Sqoop incremental imports by using Sqoop jobs and automated the jobs using Oozie. ? Experience in creating Hive tables to apply schema on read and view the data in structured format. ? Worked on various compression and file formats like Avro, Parquet and Text formats ? Responsible for writing Hive Queries for analyzing data in Hive warehouse using HQL. ? Responsible for creating complex dynamic partition tables using Hive for best performance and faster querying. ? Experienced in developing Hive User Defined Functions in java, compiling them into jars and adding them to the HDFS and executing them with Hive Queries. ? Developed several advanced Map Reduce programs in Java as part of functional requirements for Big Data. ? Submitted MapReduce jobs to Queues, as collection of jobs, to allow the system to provide specified functionality. ? Developed multiple POCs using Spark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata. ? Developed spark scripts by using Scala shell as per requirements. ? Developed Kafka producers and consumers, Spark clients along with components on HDFS, Hive. ? Worked on NoSQL databases like HBase for read/write access on random real-time data. ? Involved in defining job flows using Oozie for scheduling jobs to manage apache Hadoop jobs.   ? Tested and reported defects in an Agile Methodology perspective. Environment: Hadoop, HDFS, Hive, MapReduce, Spark, Scala, Kafka, HBase, Oozie, Java, Linux, Cloudera. Hadoop Developer Bank of America - Charlotte, NC February 2014 to July 2015 The project is to collect separate data warehouses from multiple departments and combined them into a single global repository in Hadoop for analysis to construct a new and more accurate score of the

risk in its customer portfolios. This accurate score allowed bank to manage its exposure better and hence increased the revenue and improved customer satisfaction.   Responsibilities: ? Collected and aggregated large amounts of structured and complex data from multiple silos and combined the data and look for patterns. ? Involved in using Apache Flume and stored the data into HDFS for analysis. ? Implemented multiple MapReduce Jobs in java for data cleansing and pre-processing of data. ? Developed MapReduce programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in the EDW. ? Monitored workload, job performance and Capacity planning using Cloudera Manager. ? Involved in defining job flows, managing and reviewing log files. ? Worked on analyzing the web log data using the HiveQL to extract number of unique visitors per day, page views and returning visitors. ? Experienced in creating Hive schema, external tables and managing views. ? Responsible for data loading involved in creating Hive tables and partitions based on the requirement. ? Created Hive queries that helped market analysts to spot emerging trends by comparing fresh data with EDW reference tables and historical metrics ? Experienced in optimizing ETL workflows using PIG. ? Developed spark scripts by using python shell commands as per the requirement. ? Involved in data migration from various databases to Hadoop HDFS and Hive using Sqoop. ? Worked on Oozie workflow engine for job scheduling. ? Used zookeeper for various types of centralized configurations. ? Good experience in working with Tableau visualization tool using Tableau Desktop.   Environment: Map Reduce, Hive, PIG, Spark, Flume, Zookeeper, Oozie, Tableau, Java, Python, Unix. Java Developer Hierarchy Technologies September 2012 to November 2013 Hierarchy Technologies has been designed to facilitate the business process of tracking a Bill of Materials Order through its life cycle. This encompasses the initial forecasting of the project through to the required reports being available to the relevant Finance and Billing teams for verification. The solution is designed to automate as much of the business process and eliminate as many manual steps as possible.   Responsibilities: ? Followed Agile Rational Unified Process throughout the lifecycle of the project. ? Involved in requirements analysis and gathering and converting them into technical specifications using UML diagrams. ? Implemented core framework components for executing workflows using Core Java, JDBC, and

Servlets and JSPs. ? Created Responsive Designs (Mobile/Tablet/Desktop) using HTML5& CSS3. ? Applied Object Oriented concepts (inheritance, composition, interface), design patterns (singleton, strategy) ? Applied Spring IOC Container to facilitate Dependency Injection. ? Used Spring AOP to implement security, where cross cutting concerns were identified. ? Responsible for developing SOAP based Web Services consuming and packaging using Axis. ? Involved in design and decision making for Hibernate ORM mapping. ? Responsible for designing front end system using JSP, HTML, jQuery and JavaScript. ? Involved in Managing Web Services and operations. ? Used Maven as the build tool to manage the dependencies of the project. ? Implemented Stored Procedures for the tables in the database. ? Develop and perform Mock Testing and Unit Testing using JUNIT and Easy Mock. ? Involved in implementing the continuous integration using Jenkins ? Built project using Apache Maven build scripts. Environment: Java1.6/J2EE, Microsoft Visio, Web Sphere Application Server, Spring MVC, IOC, Spring AOP, Apache Axis, SOAP, Hibernate, Web services, Maven, jQuery, JUnit, Easy Mockito, Git, Jenkins. Jr. JAVA Developer Serveen Software Systems (P) Ltd June 2011 to September 2012 This is a web based project which is used to create surveys for the participants with a user friendly GUI. This tool will not record any kind of information aboutusers giving priority to the confidentiality of the user information which includes modules like Admin, Registered User and Manager. Responsibilities ? Involved in various phases of Software Development Life Cycle (SDLC). ? Used Rational Rose for the Use Case Diagrams, Object Diagrams, class Diagrams and Sequence diagrams to represent the detailed design phase. ? Front-end is designed by using HTML, CSS, JSP, Servlets, Ajax and Struts. ? Involved in writing the exception and validation classes using Struts validation rules. ? Used JavaScript for the web page validation. ? Used SOAP for Web Services by exchanging XML data between applications over HTTP. ? Created Servlets which route submittals to appropriate Enterprise Java Bean (EJB) components and render retrieved information. ? Written ANT scripts for building application artifacts. Environment: Java, J2EE, Servlets, Struts, JSP, XML, DOM, HTML, JavaScript, JDBC, Web Services, Eclipse Plug-ins. Education Bachelor's Skills DYNAMODB, HDFS, IMPALA, OOZIE, SQOOP Additional Information TECHNICAL SKILLS: Data Access Tools HDFS, YARN, Hive, Pig,

HBase, Solr, Impala, Spark Core, Spark SQL, Spark Streaming Data Management HDFS, YARN

Data Workflow Sqoop, Flume, Kafka Data Operation Zookeper, Oozie Data Security Ranger, Knox

BigData Distributions Hortonworks, Cloudera Cloud Technologies AWS (Amazon Web Services)

EC2, S3, IAM, CLOUD WATCH, DynamoDB, SNS, SQS, EMR, KINESIS Programming and

Scripting Languages Java, Scala, Pig Latin, HQL, SQL, Shell Scripting, HTML, CSS, JavaScript

IDE/Build Tools Eclipse, Intellij Business Intelligence Tools Tableau, Power Bi, MS Excel, Microsoft

Visio, SSIS, SSAS, SSRS Java/J2EE Technologies XML, Junit, JDBC, AJAX, JSON, JSP

Operating Systems Linux, Windows, Kali Linux SDLC Agile/SCRUM, Waterfall

Name: Jessica Vance

Email: nrandall@example.com

Phone: 512.746.5432x17747