Sr. Big Data Developer Sr. Big Data Developer Sr. Big Data Developer Charlotte, NC ? Above 9+ years of working experience as a Sr. Big Data/Hadoop Developer in designed and developed various applications like big data, Hadoop, Java/J2EE open-source technologies. ? Work Extensively in Core Java, Struts2, JSF, spring, Hibernate, Servlets, JSP and Hands-on experience with PL/SQL, XML and SOAP. ? Well verse working with Relational Database Management Systems as Oracle , MS SQL, MySQL Server ? Hands on experience on Hadoop /Big Data related technology experience in Storage, Querying, Processing and analysis of data. ? Hands on experience in working on XML suite of technologies like XML, XSL, XSLT, DTD, XML Schema, SAX, DOM, JAXB. ? Experience in working with Web Servers like Apache Tomcat and Application Servers like IBM Web Sphere and JBOSS. ? Hands on experience in advanced Big-Data technologies like Spark Ecosystem (Spark SQL, MLlib, Spark, R and Spark Streaming), Kafka and Predictive analytics ? Knowledge of the software Development Life Cycle (SDLC), Agile and Waterfall Methodologies. ? Experience on applications using Java, python and UNIX shell scripting ? Experience in consuming Web services with Apache Axis using JAX-RS(REST) API's. ? Experienced in building tool Maven, ANT and logging tool Log4J. ? Experience in Programming and Development of java modules for an existing web portal based in Java using technologies like JSP, Servlets, JavaScript and HTML, SOA with MVC architecture. ? Expertise in ingesting real time/near real time data using Flume, Kafka, Storm ? Good knowledge of NO SQL databases like Mongo DB, Cassandra and HBase. ? Strong development skills in Hadoop, HDFS, Map Reduce, Hive, Sqoop, HBase with solid understanding of Hadoop internals. ? Excellent knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MRA and MRv2 (YARN). ? Good knowledge of NoSQL databases such as HBase, MongoDB and Cassandra. ? Experience in working with Eclipse IDE, Net Beans, and Rational Application Developer. ? Experience in using PL/SQL to write Stored Procedures, Functions and Triggers. ? Expertise in developing a simple web based application using J2EE technologies like JSP, Servlets, and JDBC. ? Experience working on EC2 (Elastic Compute Cloud) cluster instances, setup data buckets on S3 (Simple Storage Service), set EMR (Elastic MapReduce). ? Hands on experience in

installing, configuring and using Apache Hadoop ecosystem components like Hadoop Distributed File System (HDFS), MapReduce, Pig, Hive, HBase, Apache Crunch, Zookeeper, Scoop, Hue, Scala, AVRO.  ? Strong Programming Skills in designing and implementing of multi-tier applications using Java, J2EE, JDBC, JSP, JSTL, HTML, CSS, JSF, Struts, JavaScript, JAXB.  ? Extensive experience in SOA-based solutions - Web Services, Web API, WCF, SOAP including Restful APIs services  ? Good Knowledge in Amazon Web Service (AWS) concepts like EMR and EC2 web services which provides fast and efficient processing of Teradata Big Data Analytics.  ? Experience in collection of Log Data and JSON data into HDFS using Flume and processed the data using Hive/Pig. Work Experience Sr. Big Data Developer Wells Fargo - Charlotte, NC January 2018 to Present Responsibilities    As a Big Data Developer, I worked on Hadoop eco-systems including Hive, HBase, Oozie, Pig, Zookeeper, Spark Streaming MCS (MapR Control System) and so on with MapR distribution.    Installed and configured Hadoop MapReduce, HDFS, Developed multiple MapReduce jobs in Java for data cleaning and Pre-processing.    Primarily involved in Data Migration process using Azure by integrating with Github repository and Jenkins.    Built code for real time data ingestion using Java, Map R-Streams (Kafka) and STORM.    Involved in various phases of development analysed and developed the system going through Agile Scrum methodology.   Worked on Apache Solr which is used as indexing and search engine.   Involved in development of Hadoop System and improving multi-node Hadoop Cluster performance.    Worked on analysing Hadoop stack and different Big data tools including Pig and Hive, HBase database and Sqoop.   Developed data pipeline using flume, Sqoop and pig to extract the data from weblogs and store in HDFS    Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.    Worked with different data sources like Avro data files, XML files, JSON files, SQL server and Oracle to load data into Hive tables.    Used J2EE design patterns like Factory pattern & Singleton Pattern.    Used Spark to create the structured data from large amount of unstructured data from various sources.    Implemented usage of Amazon EMR for processing Big Data across Hadoop Cluster of virtual servers on Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3)    Performed transformations, cleaning and filtering on

imported data using Hive, MapReduce, Impala and loaded final data into HDFS. Developed Python scripts to find vulnerabilities with SQL Queries by doing SQL injection. Experienced in designing and developing POC's in Spark using Scala to compare the performance of Spark with Hive and SQL. Responsible for coding MapReduce program, Hive queries, testing and debugging the MapReduce programs. Extracted Real time feed using Spark streaming and convert it to RDD and process data into Data Frame and load the data into Cassandra. Involved in the process of data acquisition, data pre-processing and data exploration of telecommunication project in Scala. Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and Zookeeper. Specified the cluster size, allocating Resource pool, Distribution of Hadoop by writing the specification texts in JSON File format. Imported weblogs & unstructured data using the Apache Flume and stores the data in Flume channel. Exported event weblogs to HDFS by creating a HDFS sink which directly deposits the weblogs in HDFS. Used RESTful web services with MVC for parsing and processing XML data. Utilized XML and XSL Transformation for dynamic web-content and database connectivity. Involved in loading data from UNIX file system to HDFS. Involved in designing schema, writing CQL's and loading data using Cassandra. Built the automated build and deployment framework using Jenkins, Maven etc. Environment: Hadoop 3.0, Hive 2.3, HBase 1.2, Oozie, Pig 0.17, Zookeeper, Spark, MapReduce, Azure, Java, Agile, J2EE, Cassandra 3.11, Jenkins, Maven Big Data/Hadoop Developer Caterpillar - Peoria, IL March 2015 to December 2017 Responsibilities Responsible for installation and configuration of Hive, Pig, HBase and Sqoop on the Hadoop cluster and created hive tables to store the processed results in a tabular format. Involved in Agile methodologies, daily scrum meetings, spring planning. Integrate visualizations into a Spark application using Data bricks and popular visualization libraries (ggplot, matplotlib). Involved in all phases of Software Development Life Cycle (SDLC) and Worked on all activities related to the development, implementation and support for Hadoop. Configured Spark Streaming to receive real time data from the Apache Kafka and store the stream data to HDFS using Scala. Developed the Sqoop scripts to make the interaction between Hive and vertica Database. Processed data into HDFS by developing solutions and analyzed the data using MapReduce, PIG,

and Hive to produce summary results from Hadoop to downstream systems. Streamed AWS log group into Lambda function to create service now incident. Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running Hive queries and Pig scripts. Created Managed tables and External tables in Hive and loaded data from HDFS. Developed Spark code by using Scala and Spark-SQL for faster processing and testing and performed complex HiveQL queries on Hive tables. Scheduled several times based Oozie workflow by developing Python scripts. Exporting the data using Sqoop to RDBMS servers and processed that data for ETL operations. Worked on S3 buckets on AWS to store Cloud Formation Templates and worked on AWS to create EC2 instances. Designing ETL Data Pipeline flow to ingest the data from RDBMS source to Hadoop using shell script, sqoop, package and MySQL. End-to-end architecture and implementation of client-server systems using Scala, Akka, Java, JavaScript and related, Linux Optimized the Hive tables using optimization techniques like partitions and bucketing to provide better. Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive, Pig, and Sqoop. Implementing Hadoop with the AWS EC2 system using a few instances in gathering and analyzing data log files. Involved in Spark and Spark Streaming creating RDD's, applying operations -Transformation and Actions. Created partitioned tables and loaded data using both static partition and dynamic partition method. Developed custom Apache Spark programs in Scala to analyze and transform unstructured data. Using Kafka on publish-subscribe messaging as a distributed commit log, have experienced in its fast, scalable and durability. Scheduled map reduces jobs in production environment using Oozie scheduler. Involved in Cluster maintenance, Cluster Monitoring and Troubleshooting, Manage and review data backups and log files. Designed and implemented MapReduce jobs to support distributed processing using java, Hive and Apache Pig Analyzing Hadoop cluster and different Big Data analytic tools including Pig, Hive, HBase and Sqoop. Improved the Performance by tuning of Hive and MapReduce. Implemented POC to migrate MapReduce jobs into Spark RDD transformations using Scala. Environment: Hive 2.3, Pig 0.17, HBase 1.2, Sqoop, Hadoop 3.0, Agile, Spark, Scala, AWS, JavaScript Hadoop Developer

Facebook - Menlo Park, CA October 2013 to February 2015 Responsibilities    Developed Spark code and Spark-SQL/Streaming for faster testing and processing of data using Lambda Architecture.    Experience in deploying data from various sources into HDFS and building reports using Tableau.    Developed a data pipeline using Kafka and Strom to store data into HDFS. Developed REST APIs using Scala and Play framework to retrieve processed data from Cassandra database.    Re-engineered n-tiered architecture involving technologies like EJB, XML and Java into distributed applications.    Explored the possibilities of using technologies like JMX for better monitoring of the system.    Configured deployed and maintained multi-node Dev and Test Kafka Clusters.    Performed transformations, cleaning and filtering on imported data using Hive, MapReduce, and loaded final data into HDFS.    Load the data into Spark RDD and performed in-memory data computation to generate the output response.    Responsible for Cluster Maintenance, Monitoring, Managing, Commissioning and decommissioning Data nodes, Troubleshooting, and review data backups, Manage & review log files for Hortonworks.    Installed Oozie workflow engine to run multiple Hive and Pig jobs.    Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase. Installed and configured HA of Hue to point Hadoop Cluster in cloudera Manager.    Installed and configured MapReduce, HDFS and developed multiple MapReduce jobs in java for data cleaning and pre-processing.    Working with applications teams to install operating system, Hadoop updates, patches, version upgrades as required.    Responsible for developing data pipeline using HDInsight, Flume, Sqoop and Pig to extract the data from weblogs and store in HDFS.    Performed transformations, cleaning and filtering on imported data using Hive, MapReduce, and loaded final data into HDFS.    Developed and maintained the continuous integration and deployment systems using Jenkins, ANT, Akka and MAVEN.    Effectively used GIT (version control) to collaborate with the Akka team members.    Developed HDFS with huge amounts of data using Apache Kafka. Collected the log data from web servers and integrated into HDFS using Flume. Environment: Spark, HDFS, Kafka, Scala, Cassandra, Java, Hive, MapReduce, Hortonworks, Oozie, Hadoop, HBase, Flume, Sqoop, Pig, MAVEN, ANT Java/J2ee Developer Key Bank - San Francisco, CA June

2012 to September 2013 Responsibilities    Worked closely with the Requirements team and analyzed the Use cases. Elaborated on the Use cases based on business requirements and was responsible for creation of class diagrams, sequence diagrams.    Adopted J2EE best Practices, using Core J2EE patterns. Developed in Eclipse environment using Struts based MVC framework.    Designed and developed presentation layer using JSP, HTML and JavaScript.    Created JSPs using JSTL and Spring tag libraries. Developed Struts Action and Action Form classes.    Deployed J2EE components (EJB, Servlets) in Tomcat Application server.    Involved in understanding business requirements and provide technical designs and necessary documentation.    Implemented Microservices architecture to make application smaller and independent.    Developed new Spring Boot application with microservices and added functionality to existing applications using Java/ J2EE technologies.    Used Hibernate as Object relational mapping tool for mapping Java Objects to database tables.    Used Hibernate Query Language (HQL), annotations and Criteria for access and updating data.    Implemented REST Web Services and performed the HTTP operations.    Implemented multithreading to process multiple tasks concurrently in order to perform the read/write operations.    Worked extensively with Core Java concepts like Collections, Exception Handling, Java I/O, and Generics to implement business logic.    Developed web pages using with HTML, CSS, JavaScript, jQuery and Ajax.    Used JavaScript for client-side validations in the JSP and HTML pages.    Created SQL queries and stored procedures to create, retrieve and update data from database.    Developed Maven Scripts to build and deploy EAR files.    GitHub was used for the version control and source code management.    Followed Test Driven Development (TDD), responsible for testing, debugging and bug fixing of the application.    Used log4j to capture the logs that included runtime exceptions and debug information.    The application was deployed in LINUX environment.  Environment: J2EE, MVC, HTML, JavaScript, Java, Hibernate, jQuery, Ajax, Maven Java Developer Commvault - Hyderabad, Telangana March 2010 to May 2012 Responsibilities Involved in the complete Software Development Life Cycle including Requirement Analysis, Design, Implementation, Testing, and Maintenance.    Designed the front end using JSP, jQuery, CSS, and HTML as per the requirements that are provided.    Developed Responsive web application for the

backend system using AngularJS with HTML and CSS.    Used Spring AOP to enable the log interfaces and cross-cutting concerns.    Developed payment flow using AJAX partial page refresh, validation, and dynamic drop-down list.    Responsible for use case diagrams, class diagrams and sequence diagrams using Rational Rose in the Design phase.    Used Spring Core for dependency injection/Inversion of Control (IOC) to have loose-coupling.    Implemented application using MVC architecture integrating Hibernate and spring frameworks.    Implemented the Enterprise JavaBeans (EJBs) to handle various transactions and incorporated the validation framework for the project. Designed and developed all UI Screens using Java Server Pages (JSP), Static Content, HTML, CSS and JavaScript.    Worked on server-side web applications using Node.js and involved in Construction of UI using jQuery, Bootstrap and JavaScript.    Developed Database access components using Spring DAO integrated with Hibernate for accessing the data.    Integrate the dynamic pages with AngularJS to make the pages dynamic.    Developed Custom Tags and JSTL to support custom user interfaces.    Used CSS style sheets for presenting data from XML documents and data from databases to render on HTML web pages.    Used Spring Framework for integrating Hibernate for dependency injection.    Extensively used Eclipse for writing code.    Used Maven as a build tool and deployed on WebSphere Application Server.    Developed test cases on JUnit.    Used Log4J for logging and tracing the messages.    Environment: HTML, CSS, JavaScript, jQuery, AngularJS, AJAX, Bootstrap, XML, Maven, Eclipse, JUnit Education Bachelor's in Computer Science & Engineering Jawaharlal Nehru Technological University Skills Cassandra, Hdfs, Impala, Mapreduce, Oozie, Sqoop, Hbase, Kafka, Flume, Hadoop, Jboss, Jms, Mongodb, Nosql, Analysis services, Application server, Git, Hadoop, Hbase, Hive Additional Information Technical Skills Hadoop/Big Data Technologies: Hadoop 3.0, HDFS, MapReduce, HBase 1.4, Apache Pig 0.17, Hive 2.3, Sqoop 1.4, Apache Impala 2.1, Oozie 4.3, Yarn, Apache Flume 1.8, Kafka 1.1, Zookeeper    Cloud Platform: Amazon AWS, EC2, EC3, MS Azure, Azure SQL Database, Azure SQL Data Warehouse, Azure Analysis Services, HDInsight, Azure Data Lake, Data Factory    Hadoop Distributions: Cloudera , Hortonworks, MapR    Programming Language: Java, Scala, Python 3.6, SQL, PL/SQL, Shell Scripting, Storm 1.0, JSP, Servlets    Frameworks: Spring 5.0.5, Hibernate 5.2,

Struts 1.3, JSF, EJB, JMS    Web Technologies: HTML, CSS, JavaScript, JQuery 3.3, Bootstrap 4.1, XML, JSON, AJAX    Operating Systems: Linux, Unix, Windows 10/8/7    IDE and Tools: Eclipse 4.7, NetBeans 8.2, IntelliJ, Maven      NoSQL Databases: HBase 1.4, Cassandra 3.11, MongoDB Web/Application Server: Apache Tomcat 9.0.7, JBoss, Web Logic, Web Sphere      SDLC Methodologies: Agile, Waterfall    Version Control: GIT, SVN, CVS

Name: Tracy Gutierrez

Email: robertsjennifer@example.net

Phone: +1-739-562-9977