Sr. Hadoop/Spark Developer Sr. Hadoop/Spark Developer Oklahoma City, OK Work Experience Sr. Hadoop/Spark Developer IBM December 2016 to Present Responsibilities: Good Experience in designing and deployment of Hadoopcluster and various Big Data components including HDFS, MapReduce, Hive, Sqoop, Pig, Oozie, Zookeeper in bothClouderaas well as Hortonworksdistribution   Involved in loading and transforming large Datasets from relational databases into HDFSand vice-versa using Sqoop imports and export.   Created Partitions, Buckets based on State to further process using Bucket based Hive joins.   Responsible for loading Data pipelines from webservers and Teradata using Sqoop with Kafka and SparkStreamingAPI.   Managing multiple AWS instances, assigning the security groups, ElasticLoadBalancer and AMIs.   Created detailed AWS Security groups which behaved as virtual firewalls that controlled the traffic allowed reaching one or more AWS EC2 instances.   Wrote ETL jobs to read from web APIs using REST and HTTP calls and loaded into HDFS using java and Talend.   Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like ApacheSpark written in Scala.   Worked with Spark to create structured data from the pool of unstructured data received. Processed Multiple Data sources input to same Reducer using Generic Writable and Multi-Input format.   Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandraas per the business requirement and also used Cassandra through Java services.   Experience in NoSQL Column-Oriented Databases like Cassandra and its Integration with Hadoopcluster   Built re-usable HiveUDF libraries which enabled various business analysts to use these UDF's in Hivequerying.   Experienced on creating multiple kind of Report in PowerBI and present it using Story Points.   Handled importing of data from various data sources, performed transformations using HiveMapReduce, loaded data into HDFS and extracted data from MYSQL into HDFS vice-versa using Sqoop.   Support development with application architecture in both real time and batch processing using big data.   Developed MapReduce/ EMR jobs to analyze the data and provide heuristics and reports. The heuristics were used for improving campaign targeting and efficiency.   Written shellscripts that run multiple Hive jobs which helps to automate different hive tables incrementally which are used to generate different reports using Tableau for the

Business use.   Worked with both MapReduce 1 (JobTracker) and MapReduce 2 (YARN) setups Worked totally in agilemethodology and also developedSparkscripts by using Scalashell.   Worked as a Cassandra developer Setting-up configuration and optimized the Cassandracluster. Developed real-time java based application to work along with the Cassandradatabase.   Prepared technical design documents, detailed design documents.   Written complex Hive queries involving external dynamic partitioned on date Hive Tables which stores rolling window time-period user viewing history. Hadoop Developer Bristol-Myers Squibb September 2014 to November 2016 Environment: Hadoop Ecosystem Components, Tableau, EMR, AWS, ETL, EC2, Kafka, Spark, Cassandra, Scala, Maven, Java, JUnit, agile methodologies, MySQL, impala, cloudera, power BI. Hadoop Developer Bristol-Myers Squibb September 2014 to November 2016 Responsibilities: Responsible for planning, organizing, and implementation of complex business solutions, producing deliverables within stipulated time.   Worked with Linux systems and RDBMS database on a regular basis in order to ingest data using Sqoop.   Experience in Cloudbasedservices(AWS) to retrieve the data. Worked and expertise hands on scala programming for processing real time information using SparkAPI's in the cloudenvironment.   Using Kafka and Kafka brokers we initiatedspark context and processed live streaming information with the help of RDD as is.   Experience in supporting multi-region AWS cloud and Created placement groups to maintain cluster of instances.   Installed, Configured TalendETL on single and multi-server environments.   Experience in creating tables, dropping and altered at run time without blocking updates and queries using HBase and Hive. Developed ETL test scripts based on technical specifications/Data design documents and Source to Target mappings.   Hands-on experience with message broker such as ApacheKafka.   Worked on ApacheNifi as ETL tool for batch processing and real time processing.   Developed Solr web apps to query and visualize and solr indexed data from HDFS.   Worked on apache Solr for indexing and load balanced querying to search for specific data in larger datasets.   Extracted files from MongoDB through Sqoop and placed in HDFS and processed.   Worked on MongoDB for distributed storage and processing.   Experienced in defining job flows. Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting. Experienced in managing

and reviewing the Hadoop log files. Used OOZIE Operational Services for batch processing and scheduling workflows dynamically. Used Oozie workflow engine to create the workflows and automate the MapReduce, Hive, Pig jobs. Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop. Cluster co-ordination through Zookeeper. Hands on experience on HIVE queries and functions for evaluation, filtering, loading and storing of data. Developed Unixshellscripts to load large number of files into HDFS from LinuxFile System. Experience in creating hive tablesHiveQL. Using HIVE join queries to join multiple tables of a source system and load them into Elastic Search Tables. Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig. Environment: Hadoop ecosystem components, ETL, Spark, Kafka, Shell Scripting, SQL Talend, Elastic search, solr, Linux- Ubuntu, AWS, Hortonworks, MongoDB, Map Reduce. Hadoop Developer January 2013 to September 2014 Responsibilities: Hands on experience in loading data from UNIX file system and Teradata to HDFS Installed and configured Flume, Hive, Pig, Sqoop and Oozie on the Hadoop cluster. Developed PIG scripts for the analysis of semi structured data. Developed JavaMapReduce programs on log data to transform into structured way to find user location, age group, spending time. Collected and aggregated large amounts of web log data from different sources such as webservers, mobile and network devices using ApacheFlume and stored the data into HDFS for analysis Create a complete processing engine, based on Cloudera's distribution, enhanced to performance. Working with Eclipse using Maven plugin for EclipseIDE. Created HBase tables to store variable data formats coming from different portfolios Performed real time analytics on HBase using JavaAPI and RestAPI. Integrated Oozie with the rest of the Hadoop stack supporting several types of Hadoop jobs out of the box (such as Map-Reduce, Pig, Hive, and Sqoop) as well as system specific jobs (such as Java programs and shellscripts). Developed ETL using Hive, Oozie, shellscripts and Sqoop. Used Scala for coding the components, & Utilized Scala pattern matching in coding Supported DataAnalysts in running MapReduce Programs. Implemented Name Node backup using NFS. This was done for High availability Analyzed the weblog data using the HiveQL, integrated Oozie with the rest of the Hadoop stack supporting

several types of Hadoop jobs out of the box (such as MapReduce, Pig, Hive, and Sqoop) as well as system specific jobs (such as Java programs and shellscripts).   Experience in Pigscripts for sorting, joining and grouping the data   Experienced with working on Avro Data files using AvroSerialization system.   Environment: HDFS, Map Reduce, Pig, Hive, Sqoop, Flume, HBase, Java, Maven, Avro, Cloudera, Eclipse and Unix Shell Scripting, Oozie, ETL, Scala. Java Developer Hyderabad, Telangana March 2011 to December 2012 Responsibilities: Developed UseCase diagrams, business flow diagrams, Activity/State diagrams.   Designed and implemented the training and reports modules of the application using Servlets, JSP and ajax   Developed XMLWebServices using SOAP, WSDL, and UDDI.   Interact with Business Users and Develop Custom Reports based on the criteria defined.   Requirement gathering and information collection. Analysis of gathered information so as to prepare a detail work plan and task breakdown structure   Experience in custom JSP tags for the application   Experience in develop of SDLC life cycle and undergo in all the phases in it.   Implemented applications using Java, J2EE, JSP, Servlets, JDBC, RAD, XML, HTML, XHTML, HibernateStruts, spring and JavaScript on Windows environments.   Developed action Servlets and JSPs for presentation in StrutsMVC framework.   Developed PL/SQL View function in Oracle9i database for get available date module.   Used OracleSQL4.0 as the database and wrote SQL queries in the DAO Layer.   Used RESTFUL Services to interact with the Client by providing the RESTFULURL mapping.   Implementing project using Agile SCRUM methodology, involved in daily stand up meetings and sprint showcase and sprint retrospective.   Used SVN and GitHub as version control tool.   Developed presentation layer using HTML, JSP, Ajax, CSS and JQuery.   Worked with StrutsMVC objects like Action Servlet, Controllers, validators, WebApplication Context, HandlerMapping, Message Resource Bundles, Form Controller, and JNDI for look-up for J2EE components.   Used Quartz Scheduler for batch jobs.   Experience in JIRA and tracked the test results and interacted with the developers to resolve issue.   Created the UI tool - using Java, XML, XSLT, DHTMLand JavaScript   Used XSLT to transform my XML data structure into HTML pages.   Developed and maintained elaborate services based architecture utilizing open source technologies like HibernateORM (Data Access Layer) and SpringFramework

(Application Layer).   Configured Design shipping rate template upload UI using AdobeFlex and Developed Jasper report.   Deployed EJB Components on Tomcat. Used JDBCAPI for interaction with OracleDB.   Wrote build & deployment scripts using shell, Perl and ANTscripts   Extensively used Java multi-threading to implement batch Jobs with JDK 1.5 features   Experience in application using CoreJava, JDBC, JSP, Servlets, spring, Hibernate, WebServices, SOAP, and WSD   Implemented Hibernate in the data access object layer to access and update information in the Oracle10gDatabase   Environment: HTML, Java Script, Ajax, Servlets, JSP, SOAP, SDLC life cycle, Java, Hibernate, Scrum, JIRA, Git Hub, JQuery, CSS, XML, ANT, Tomcat Server, Jasper Reports. Java/ETL Developer Read Mind Info Services - Hyderabad, Telangana March 2009 to January 2011 Responsibilities: Involved in the design, development and deployment of the Application using Java/J2EE Technologies.   Developed web components using JSPServlets, JDBC and Coded JavaScript for AJAX and client side data validation.   Designed and Developed mappings using different transformations like Source Qualifier, Expression, Lookup (Connected & Unconnected), Aggregator, Router, Rank, Filter and SequenceGenerator.   Imported data from various Sources transformed and loaded into DataWarehouseTargets using Informatica Power Center.   Made substantial contributions in simplifying the development and maintenance of ETL by creating re-usable Source, Target, Mapplets, and Transformation objects.   Experience in development of extracting, transforming and loading (ETL), maintain and support the enterprise data warehouse system and corresponding marts   Prepare DRplan and recovery process for GDW application.   Developed JSP pages using Custom tags and Tilesframework and Strutsframework. Used different user interface technologies JSP, HTML, CSS, and JavaScript for developing the GUI of the application.   Skills gained on web-basedRESTAPI, SOAPAPI, Apache for real-time data streaming   Programmed OracleSQL, T-SQL Stored Procedures, Functions, Triggers and Packages as back-end processes to create and update staging tables, log and audit tables, and creating primary keys.   Extensively used Transformations like Aggregator, Router, Joiner, Expression, Lookup, Update Strategy, and Sequence Generator.   Developed mappings, sessions and workflows using InformaticaDesigner and Workflow Manager based on source to target mapping

documents to transform and load data into dimension tables.   Used FTP services to retrieve Flat Files from the external sources.   Environment: Java, Ajax, Informatica Power Center 8.x/9.x, REST API, SOAP API, Apache, Oracle10/11g, SQL Loader, MS SQL SERVER, Flat Files, Targets, Aggregator, Router, Sequence Generator, Education Bachelor's Skills JAVA (8 years), ETL (6 years), EXTRACT, TRANSFORM, AND LOAD (6 years), SQL (5 years), APACHE HADOOP HDFS (5 years), Hadoop Ecosystem Components, Tableau, EMR, AWS, ETL, EC2, Kafka, Spark, Cassandra, Scala, Maven, Java, JUnit, agile methodologies, MySQL, impala, cloudera, power BI. Additional Information SKILLS   ETL (6 years), EXTRACT, TRANSFORM, AND LOAD (6 years), JAVA (5 years), SQL (5 years), APACHE HADOOP HDFS (4 years)   ProfessionalSkills: 8+ years of experience in IT in fields of software design, implementation, and development and also support of business applications for health, insurance and telecom industries.   4+Years of experience in Big data Hadoop, Hadoop Ecosystem components like MapReduce, Sqoop, Flume, Kafka, Pig, Hive, Spark, Storm, HBase, Oozie, and Zookeeper   Having good experience in Hadoop framework and related technologies like HDFS, MapReduce, Pig, Hive, HBase, Sqoop and Oozie Hands of experience on data extraction, transformation and load in Hive, Pig and HBase Experience in the successful implementation of ETL solution between an OLTP and OLAP database in support of Decision Support Systems with expertise in all phases of SDLC.   Experience in creating DStreams from sources like Flume, Kafka and performed different Spark transformations and actions on it.   Experience in integrating Apache Kafka with ApacheStorm and created Storm data pipelines for real time processing.   Worked on improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, DataFrames, RDD's, SparkYARN.   Delivery experience on major Hadoop ecosystem Components such as Pig, Hive, Spark, Kafka, ElasticSearch&HBase and monitoring with Cloudera Manager. Extensive working experience using Sqoop to import data into HDFS from RDBMS and vice-versa.   Procedural knowledge in cleansing and analyzing data using HiveQL, PigLatin and custom MapReduce programs in Java.   Training and Knowledge in Mahout, SparkMLlib for use in data classification, regression analysis, recommendation engines and anomaly detection.   Experienced in Developing

Spark application using SparkCore, SparkSQL and SparkStreaming API's. Knowledge in HIVEQL, PIGLatin, MapReduce design patterns and scala Experience in extensive usage of Struts, HTML, CSS, JSP, JQuery, AJAX and JavaScript for interactive pages. Involved in configuring and working with Flume to load the data from multiple sources directly into HDFS. Hands-on experience with Hortonworks&ClouderaDistributed Hadoop (CDH) Experience in tools like Maven, Log4j, Junit and Ant Experience in understanding security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure. Experience on cloud infrastructure like AmazonWebServices (AWS). Experience on predictive intelligence and smooth maintenance in spark streaming is done using Conviva and MLlib from Spark. Involved in installing cloudera distributionof Hadoop on amazonEC2 Instances. Worked on analyzing Hadoop cluster using different big data analytic tools including Pig, Hive and MapReduce on EC2. Hand on experience in storm for configuring various topologies to ingest and process data on fly from multiple sources and aggregate into the central repository system. Experience onHadoop ecosystems like Sqoop2 and YARN. Imported data using Sqoop to load data from MySQL to S3Buckets on regular basis Hands on experience in using BI tools like Splunk/Hunk, Tableau. Experience in Installing, upgrading and configuring RedHatLinux 4.x, 5.x, and 6.x using KickstartServers Worked on AmazonAWS concepts like EMR and EC2 web services for fast and efficient processing of Big Data. Experience in deployment of BigData solutions and the underlying infrastructure of HadoopCluster using Cloudera, MapR and Hortonworks distributions. Experience in importing and exporting data using Sqoop from HDFS to RelationalDatabaseSystems and vice-versa Experience of MPP databases such as HPVertica and Impala. Experience in understanding the security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure. Hands on experience on implementation projects like Agile and Waterfallmethodologies. Strong Experience on Data WarehousingETL concepts using InformaticaPower Center, OLAP, OLTP and AutoSys. Experienced the integration of various data sources like Java, RDBMS, ShellScripting, Spreadsheets, and Text files. Working knowledge of database such as Oracle 8i/9i/10g, MicrosoftSQLServer, DB2, Netezza. Experience in NoSQLdatabases like HBase, Cassandra,

Redis and MongoDB.   Experience in using design pattern, Java, JSP, Servlets, JavaScript, HTML, JQuery, Angular JS, Mobile JQuery, JBOSS 4.2.3, XML, Web Logic, SQL, PL/SQL, JUnit, and Apache-Tomcat, Linux.   Experience in Object Oriented Analysis, Design (OOAD) and development of software using UMLMethodology, good knowledge of J2EE design patterns and Core Java design patterns.   Experience using various HadoopDistributions (Cloudera, Hortonworks, and MapR) to fully implement and leverage new Hadoop features.   Admin task includesSettingup Linux users, setting up Kerberos principals and testing HDFS, Hive, Pig and MapReduce access for the new users.   Technical skills: Big Data Ecosystem Hadoop, MapReduce, Pig, Hive, YARN, Kafka, Flume, Sqoop, Impala, Oozie, Zookeeper, Spark, Ambari, Mahout, MongoDB, Cassandra, Avro, Storm, Parquet and Snappy.   Hadoop Distributions Cloudera (CDH3, CDH4, and CDH5), Hortonworks, MapR and Apache Languages Java, Python, Jruby, SQL, HTML, DHTML, Scala, JavaScript, XML and C/C++ No SQL Databases Cassandra, MongoDB and HBase Java Technologies Servlets, JavaBeans, JSP, JDBC, JNDI, EJB and struts XML Technologies XML, XSD, DTD, JAXP (SAX, DOM), JAXB Development Methodology Agile, waterfall Web Design Tools HTML, DHTML, AJAX, JavaScript, JQuery and CSS, AngularJs, ExtJS and JSON Development / Build Tools Eclipse, Ant, Maven, IntelliJ, JUNIT and log4J Frameworks Struts, spring and Hibernate App/Web servers WebSphere, WebLogic, JBoss and Tomcat DB Languages MySQL, PL/SQL, PostgreSQL and Oracle RDBMS Teradata, Oracle [ ] MS SQL Server, MySQL and DB2 Operating systems UNIX, LINUX, Mac os and Windows Variants Data analytical tools R and MATLAB ETL Tools Talend, Informatica, Pentaho

Name: Michael Haney

Email: brandi45@example.org

Phone: 5267563880