Hadoop Developer Hadoop Developer Hadoop Developer Rockford, IL   Experience around 6+ years in IT industry with complete software development of life cycle (SDLC) which includes business requirements gathering, system analysis & design, data modeling, development, testing and implementation of the projects.   Experienced in configuration, deployments and managing of different Hadoop distributions like Cloudera (CDH4 & CDH5) and Hortonworks (HDP).   Experience of import/export data using Sqoop from Hadoop distributed file systems to relational database systems and vice versa.   Experience in handling various file formats like AVRO, Sequential, text, xml, JSON and Parquet with different compression techniques such as gzip, LZO, Snappy etc. Experienced on Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming and Spark MLlib.   Imported the data from source HDFS into Spark Data Frame for in-memory data computation to generate the optimized output response and better visualizations. Expertise in writing Spark RDD transformations, actions, Data Frame's, case classes for the required input data and performed the data transformations using Spark-Core also convert RRD to Data Frame.   Experienced on collection the real time streaming data and creating the pipeline for row data from different source using Kafka and store data into HDFS and NoSQL using Spark. Extending HIVE core functionality by using custom User Defined Function's (UDF) and User Defined Aggregating Functions (UDAF).   Implemented POC for using Impala for data processing on top of HIVE for better utilization of C++ executions engines.   Experience in NoSQL Databases HBase, Cassandra and it's integrated with Hadoop cluster.   Implemented Cluster for NoSQL tools HBase as a part of POC to address HBase limitations.   Exploring with Spark Beta version API to improve the performance, and optimization of the existing algorithms with different modes such as YARN, Mesos and standalone for POC.   Expertise in using ETL Tool Informatica Power Center designer, workflow manager, repository manager, data quality and ETL concepts.   Experienced with NiFi to automate the data movement between different Hadoop systems.   Worked with different Hadoop Security such as Knox and Ranger integrated LDAP store with Kerberos KDC.   Good understanding on security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure.   Experienced on cloud integration with AWS using Elastic Map Reduce

(EMR), Simple Storage Service (S3), EC2, Redshift and Microsoft Azure. Experienced on different Relational Data Base Management Systems like Teradata, PostgreSQL, DB2, Oracle and SQL Server. Experienced in scheduling and monitoring the production jobs using Oozie and Azkaban. Authorized to work in the US for any employer Work Experience Hadoop Developer Quant Systems - Irving, TX October 2015 to December 2016 Responsibilities: Involved in installation, configuration and Design of Hadoop Distributed using Cloudera and Hortonworks of Hadoop. Involved in complete Big Data flow of the application data ingestion from upstream to HDFS, processing the data in HDFS and analyzing the data using several tools. Imported the data from various formats like JSON, Sequential, Text, CSV, AVRO and Parquet to HDFS cluster with compressed for optimization. Experienced on ingesting data from RDBMS sources like - Oracle, SQL Server and Teradata into HDFS using Sqoop. Configured Hive and written Hive UDF's and UDAF's Also, created partitions such as Static and Dynamic with bucketing. Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like spark. Importing and exporting data into HDFS and hive using Sqoop and Kafka with batch and streaming. Experienced with Spark-Streaming APIs to perform transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into HBase. Performance analysis of Spark streaming and batch jobs by using Spark tuning parameters. Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework. Developed Spark scripts by using Python shell commands as per the requirement. Using Hive join queries to join multiple tables of a source system and load them to Elastic search tables. Experience in managing and reviewing huge Hadoop log files. Expertise in designing and creating various analytical reports and Automated Dashboards to help users to identify critical KPIs and facilitate strategic planning in the organization. Involved in Cluster maintenance, Cluster Monitoring and Troubleshooting Created Data Pipelines as per the business requirements and scheduled it using Oozie Coordinators. Maintaining technical documentation for each and every step of development environment and launching Hadoop clusters. Built the automated build and deployment framework using GitHub

and Maven etc.   Worked on IntelliJ IDEA to develop the code and dubbing.   Worked on BI tools as Tableau to create dashboards like weekly, monthly, daily reports using tableau desktop and publish them to HDFS cluster.   Environment: Scala, Hadoop, HDFS, Hive, Oozie, Sqoop, NiFi, Spark, Kafka, Elastic Search, Shell Scripting, HBase, Python, GitHub, Tableau, Oracle, MySQL, Teradata and AWS. Hadoop Developer ReqRoute, Inc - Milpitas, CA August 2013 to October 2015 Responsibilities: Involved in requirement gathered, narrates the stories and worked with complete Software Development Life Cycle (SDLC) methodologies based on Agile.   Involved in installation, configuration, supporting and managing Hadoop Clusters using Hortonworks Distribution (HDP) to Cloudera Distributions Hadoop (CDH).   Worked on Hadoop MapReduce, HDFS, developed multiple MapReduce jobs in java for data cleaning and preprocessing.   Experienced in managing and reviewing Hadoop log files and documenting the issues on daily basis to the resolution portal. Implemented Dynamic Partitions, Buckets in HIVE.   Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Scala and Python.   Experience configuring spouts and bolts in various Storm topologies and validating data in the bolts.   Experienced in running Hadoop streaming jobs to process terabytes of xml format data.   Have an experience to load and transform large sets of structured, semi structured and unstructured data, using Sqoop from Hadoop Distributed File Systems to Relational Database Systems and also Relational Database Systems to Hadoop Distributed File Systems.   Used Spark streaming to receive real time data from the Kafka and store the stream data to HDFS using Scala and databases such as HBase. Established/implemented firewall rules, validated rules with vulnerability scanning tools   Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java MapReduce Hive, Pig, and Sqoop.   Implemented Storm builder topologies to perform cleansing operations before moving data into HBase.   Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way. Developed a custom File System plug in for Hadoop so it can access files on Data Platform.   Used Spark to create API's in Java and Python for Big Data analysis.   Experience in troubleshooting errors in Cassandra, Hive and MapReduce.   This plugin allows Hadoop MapReduce programs,

Cassandra, Pig and Hive to work unmodified and access files directly. Used versions controls tools such as GitHub to pull data from Upstream to local branch, check conflict, cleaning also reviewing the codes of other developers. Involved with development teams to discuss JIRA stories and understand the requirements. Actively, involved in complete life cycle of agile methodology to design, develop, deploy and support solutions. Environment: Hadoop, Hive, Pig, Strom, Cassandra, Sqoop, Impala, Oozie, Java, Python, Shell Scripting, MapReduce, Java Collection, MySQL. Java Developer Netserv Applications, Inc - Alpharetta, GA June 2012 to August 2013 Responsibilities: Responsible and active in the analysis, design, implementation and deployment of full Software Development Lifecycle (SDLC) of the project. Designed and developed user interface using JSP, HTML and JavaScript. Defined the search criteria and pulled out the record of the customer from the database. Make the required changes and save the updated record back to the database. Validated the fields of user registration screen and login screen by writing JavaScript and jQuery validations. Used DAO and JDBC for database access. Developed stored procedures and triggers using PL/SQL in order to calculate and update the tables to implement business logic. Design and develop XML processing components for dynamic menus on the application. Involved in postproduction support and maintenance of the application. Involved in the analysis, design, implementation, and testing of the project modules. Implemented the presentation layer with HTML, XHTML and JavaScript. Developed web components using JSP and JDBC. Implemented database using SQL Server. Designed tables and indexes. Wrote complex SQL queries and stored procedures. Involved in fixing bugs and unit testing with test cases using JUnit. Created user and technical documentation. Environment: Java, Oracle, HTML, XML, SQL, J2EE, JUnit, JDBC, JSP, Tomcat, SQL Server, MongoDB, JavaScript, GitHub, SourceTree, NetBeans. Java Developer DHI Group, Inc - Greenwood Village, CO April 2011 to June 2012 Responsibilities: Analyzing and preparing the requirement Analysis Document. Deploying the Application to the JBOSS Application Server. Requirement gatherings from various stakeholders of the project. Effort-estimation and estimating timelines for development tasks. Used to J2EE and EJB to handle the business flow and Functionality. Interact with Client to get the confirmation on the

functionalities and implementation. Involved in the complete SDLC of the Development with full system dependency. Actively coordinated with deployment manager for application production launch. Provide Support and update for the period under warranty. Produce detailed low-level designs from high level design Specifications for components of low level complexity. Develops, builds and unit tests components of low level Complexity from detailed low-level designs. Developed user and technical documentation. Monitoring of test cases to verify actual results against expected results. Performed Functional, User Interface test and Regression Test. Carrying out Regression testing to track the problem tracking. Implemented Model View Controller (MVC) architecture at the Web tier level to isolate each layer of the application to avoid the complexity of integration and ease of maintenance along with Validation Framework Environment: Java, JEE, CSS, HTML, SVN, EJB, UNIX, XML, Work Flow, MyEclipse JMS, JIRA, Oracle, JBOSS. Education Bachelor of Engineer Texas A&M University Skills Hadoop, Drupal

Name: Jack Owen

Email: jerrygardner@example.org

Phone: 001-518-396-3601x628