Sr.Hadoop Developer Sr.Hadoop Developer Sr.Hadoop Developer - Utica National Insurance Group

New York, NY   IT Professional with 10+ years of experience in using Java /J2EE technologies along with 6+ years of experience in Hadoop and related Big data technologies.   Experienced in working with Cloudera (CDH), Hortonworks (HDP) and Amazon EMR environments.   Strong experience writing Scala based Spark applications to perform large scale data transformations, data cleansingand other data preparation for advanced analytics.   Good understanding of Spark execution framework and internal memory management.   Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, YARN and MapReduce concepts.   Experience in working with different kind of MapReduce programs using Hadoop for working with Big Data analysis.   Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, YARN and MapReduce concepts.   Experience in working with different kind of MapReduce programs using Hadoop for working with Big Data analysis.   Experience in analyzing data using Hive QL, Pig Latin and custom MapReduce programs in Java.   Experience in importing/exporting data using Sqoop into HDFS from Relational Database Systems and vice-versa.   Extensive knowledge and experience on real time data streaming techniques like Kafka, Storm and Spark Streaming. Working experience on designing and implementing complete end-to-end Hadoop Infrastructure including PIG, HIVE, Sqoop, Oozie, Flume and zookeeper.   Good Knowledge in providing support to data analyst in running Pig and Hive queries.   Experience in writing shell scripts to dump the shared data from MySQL servers to HDFS.   Experience in designing both time driven and data driven automated workflows using Oozie.   Knowledge in performance tuning the Hadoop cluster by gathering and analyzing the existing infrastructure.   Experience in working with various Cloudera distributions (CDH4/CDH5), Hortonworks and Amazon EMR Hadoop Distributions.   Extensively worked on Hive and Sqoop for sourcing and transformations.   Experience in automating the Hadoop Installation, configuration and maintaining the cluster by using the tools like Puppet. Hands on experience knowledge in NoSQL databases like HBase, Cassandra, Mongo db. Experience in working with flume to load the log data from multiple sources directly into HDFS.

Strong debugging and problem solving skills with excellent understanding of system development methodologies, techniques and tools.   Wrote Flume configuration files for importing streaming log data into HBase with Flume.   Processing this data using Spark Streaming API with Scala.   Worked in complete Software Development Life Cycle (analysis, design, development, testing, implementation and support) in different application domain involving different technologies varying from object oriented technology to Internet programming on Windows NT, Linux and UNIX/ Solaris platforms and RUP methodologies.   Familiar with RDBMS concepts and worked on Oracle 8i/9i, SQL Server 7.0., DB2 8.x/7.x   Involved in writing shell scripts, Ant scripts for Unix OS for application deployments to production region.   Used Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive.   Experience in Hadoop Distributions like Cloudera, Hortonworks, Big Insights, MapR Windows Azure, and Impala.   Hands on experience in developing the applications with Java, J2EE, J2EE - Servlets, JSP, EJB, SOAP, Web Services, JNDI, JMS, JDBC2, Hibernate, Struts, Spring, XML, HTML, XSD, XSLT, PL/SQL, Oracle10g and MS-SQL Server RDBMS.   Having very good POC and Development experience on Apache Flume, Kafka, Spark, Storm, and Scala.   Involved in writing shell scripts, Ant scripts for Unix OS for application deployments to production region.   Developed Unit test cases using Junit, Easy Mock and MRUnit testing frameworks.   Experienced with version controller systems like SVN, Clear case. Experience using IDEs tools Eclipse 3.0, My Eclipse, RAD and NetBeans. Authorized to work in the US for any employer Work Experience Sr.Hadoop Developer Utica National Insurance Group - Utica, NY October 2018 to Present Responsibilities: Worked on writing transformer/mapping Map-Reduce pipelines using Java.   Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run Map Reduce jobs in the backend.   Involved in loading data into HBase using HBase Shell, HBase Client API, Pig and Sqoop.   Designed and implemented Incremental Imports into Hive tables.   Deployed an Apache Solr search engine server to help speed up the search of the government cultural asset.   Involved in collecting, aggregating and moving data from servers to HDFS using Apache Flume   Written Hive jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data   Involved in creating

Hive tables, loading with data and writing hive queries that will run internally in map reduce way. Involved in loading data from UNIX file system to HDFS. Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way. Develop MapReduce jobs for the users. Maintain, update and schedule the periodic jobs which range from updates on periodic MapReduce jobs to creating ad-hoc jobs for the business users. Importing and exporting data into HDFS and Hive using Sqoop. Experienced in defining job flows. Experienced in managing and reviewing Hadoop log files. Extracted files from Couch DB through Sqoop and placed in HDFS and processed. Experienced in running Hadoop streaming jobs to process terabytes of xml format data. Load and transform large sets of structured, semi structured and unstructured data. Responsible to manage data coming from different sources. Got good experience with NOSQL database. Developed a custom File System plug in for Hadoop so it can access files on Data Platform. This plugin allows Hadoop MapReduce programs, HBase, Pig and Hive to work unmodified and access files directly. Designed and implemented Mapreduce-based large-scale parallel relation-learning system Extracted feeds form social media sites such as Facebook, Twitter using Python scripts. Setup and benchmarked Hadoop/HBase clusters for internal use. Gained very good business knowledge on health insurance, claim processing, fraud suspect identification, appeals process etc. Involved in review of functional and non-functional requirements. Facilitated knowledge transfer sessions. Environment: Java , Eclipse, Oracle , Sub Version, Hadoop, Hive, HBase, Linux, MapReduce, HDFS, Hive, Java (JDK), Hadoop Distribution of Horton Works, Cloudera, MapReduce, DataStax, IBM DataStage, Oracle, PL/SQL, SQL*PLUS, UNIX Shell Scripting. Sr. Hadoop Developer TEK Connexion - Pittsburgh, PA June 2016 to September 2018 Responsibilities: Performed fine-tuning of spark applications/jobs to improve the efficiency and overall processing time for the pipelines. Created Spark applications using Spark Data frames and Spark SQL API extensively. Created Java based Kafka producer API to capture live-stream data into various Kafka topics. Developed Spark-Streaming applications to consume the data from Kafka topics and to insert the processed streams to HBase. Used Broadcast variables in Spark, effective & efficient Joins, transformations and other capabilities for

data processing. Used Spark-SQL to perform event enrichment and to prepare various levels of user behavioral summaries Involved in Configuring Hadoop cluster and load balancing across the nodes. Developed MapReduce programs in Java for parsing the raw data and populating staging Tables Created Hive queries to compare the raw data with EDW reference tables and performing aggregates. Experienced in developing custom input formats and data types to parse and process unstructured and semi structured input data and mapped them into key value pairs to implement business logic in Map-Reduce. Experience in implementing custom sterilizer, interceptor, source and sink as per the requirement in Flume to ingest data from multiple sources. Experience in setting up Fan-out workflow in flume to design v shaped architecture to take data from many sources and ingest into single sink. Analyzing the requirement to setup a cluster. Importing and exporting data into HDFS and Hive using Sqoop Experienced in analyzing data with Hive and Pig Experienced knowledge over designing Restful services using java based API's like JERSEY. Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS. Integrating bulk data into Cassandra file system using MapReduce programs. Got good experience with NoSQL databases HBase, Cassandra. Involved in HBase setup and storing data into HBase, which will be used for further analysis. Expertise in designing, data modelling for Cassandra NoSQL database. Experienced in managing and reviewing Hadoop log files. Experienced in defining job flows using Oozie workflow Involved in working with Spark on top of Yarn/MRv2 for interactive and Batch Analysis Worked closely with AWS EC2 infrastructure teams to troubleshoot complex issues Expertise in writing the Scala code using higher order functions for the iterative algorithms in spark for performance consideration Experienced in analyzing and Optimizing RDD's by controlling partitions for the given data Good understanding on DAG cycle for entire spark application flow on Spark application WebUI Experienced in writing live Real-time Processing using Spark Streaming with Kafka Developed custom mappers in python script and Hive UDFs and UDAFs based on the given requirement Used HiveQL to analyze the partitioned and bucketed data and compute various metrics for reporting Experienced in querying data using SparkSQL on top of Spark engine Experience in managing and monitoring Hadoop cluster using

Cloudera Manager   Supported in setting up QA environment and updating configurations for implementing scripts with Pig, Hive and Sqoop   Environment: Hadoop, HDFS, Hive, Cloudera Manager, Sqoop, Flume, Oozie, CDH5, MongoDB, Cassandra, HBase, Hue, Kerberos and Unix/Linux Hadoop Developer ComPsych - Chicago, IL November 2013 to May 2016 Responsibilities: Worked on Spark and Cassandra for the User behavior analysis and lightning speed execution   Installed and configured Hadoop Map-Reduce, HDFS and developed multiple Map-Reduce jobs in Java for data cleansing and preprocessing.   Importing and exporting data into HDFS and Hive using Sqoop.   Used UDF's to implement business logic in Hadoop   Extracted files from Oracle and DB2through Sqoop and placed in HDFS and processed.   Load and transform large sets of structured, semi structured and unstructured data.   Responsible to manage data coming from different sources.   Supported Map-Reduce Programs those are running on the cluster.   Involved in loading data from UNIX file system to HDFS.   Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.   Developed mapping parameters and variables to support SQL override.   Used existing ETL standards to develop these mappings.   Extracted the data from the flat files and other RDBMS databases into staging area and populated onto Data warehouse.   Worked on JVM performance tuning to improve Map-Reduce jobs performance Environment: Hadoop, MapReduce, HDFS, Hive, Oracle, Java, Struts, Servlets, HTML, XML, SQL, J2EE, JUnit, Tomcat. Java Developer SNI Technology - Jacksonville, FL February 2010 to October 2013 Responsibilities: Used Spring framework for implementing IOC/JDBC/ORM, AOP and Spring Security to implement business layer.   Developed and Consumed Web services securely using JAX-WS API and tested using SOAP UI.   Extensively used Action, Dispatch Action, Action Forms, Struts Tag libraries, Struts Configuration from Struts. Extensively used the Hibernate Query Language for data retrieval from the database and process the data in the business methods.   Developed pages using JSP, JSTL, Spring tags, JQuery, Java Script & Used JQuery to make AJAX calls.   Used Jenkins continuous integration tool to do the deployments.   Worked on JDBC for database connections.   Worked on multithreaded middleware using socket programming to introduce whole set of new business rules implementing OOPS design

and principles.   Involved in implementing Java multithreading concepts.   Developed several REST web services supporting both XML and JSON to perform task such as demand response management.   Used Servlet, Java and Spring for server side business logic.   Implemented the log functionality by using Log4j and internal logging API's.   Used Junit for server side testing.   Used Maven build tools and SVN for version control.   Developed frontend of application using Bootstrap, AngularJS and Node.JS frameworks.   Implemented SOA architecture using Enterprise Service Bus (ESB).   Designed front-end, data driven GUI using JSF, HTML4, JavaScript and CSS   Used IBM MQ Series as the JMS provider.   Responsible for writing SQL Queries and Procedures using DB2.
   Connection with Oracle, MySQL Database is implemented using Hibernate ORM. Configured hibernate, entities using annotations from scratch.   Environment: Core Java, EJB, Hibernate , AWS, JSF, Struts, Spring, JPA, REST, JBoss, DB2, Oracle, XML, JUnit, HTML4, CSS, JavaScript, Apache Tomcat 5x, Log4j Java Developer Object Frontier - Alpharetta, GA June 2009 to January 2010 Responsibilities: Used Struts framework to generate Forms and actions for validating the user request data.   Developed Server side validation checks using Struts validators and Java Script validations.   With JSP's and Struts custom tags, developed and implemented validations of data.   Developed applications, which access the database with JDBC to execute queries, prepared statements, and procedures.   Developed programs to manipulate the data and perform CRUD operations on request to the database.   Used message driven beans for asynchronous processing alerts to the customer.   Worked on developing Use Cases, Class Diagrams, Sequence diagrams, and Data Models.   Developed and Deployed SOAP Based Web Services on Tomcat Server Coding of SQL, PL/SQL, and Views using IBMDB2 for the database.   Working on issues while converting JAVA to AJAX.   Supported in developing business tier using the stateless session bean.
   Extensively used JDBC to access the database objects.   Using Clear case for source code control and JUNIT testing tool for unit testing.   Extensively used Java Collection API to improve application quality and performance.   Vastly used Java 5 features like Generics, enhanced for loop, type safe etc.   Providing production support and enhancements design to the existing product. Environments: Java 1.5, SOA, Spring, ExtJS, Struts 2.0, Servlets, JSP, GWT, JQuery, JavaScript,

CSS, Web Services, XML, Oracle,AngularJS, Web logic Application Server, Eclipse, UML, Microsoft Vision Education Bachelor's Skills Database, Db2, Mysql, Oracle, Pl/sql, Postgresql, Sql, Cassandra, Ambari, Hdfs, Impala, Oozie, Sqoop, Hbase, Kafka, Cdh4, Flume, Hadoop, Jboss, Map reduce Additional Information Technical Skills: Big Data Technologies HDFS, Map Reduce, Hive, Hcat, Pig, Sqoop, Flume, Oozie, Avro, Hadoop Streaming, Zookeeper, Kafka, Impala, Apache Spark, hue, Ambari. Apache ignite.   Hadoop Distributions Cloudera (CDH4/CDH5), Horton Works, AWS EMR Languages Java, SQL, PYTHON, PL/SQL, HQL, Scala IDE Tools Eclipse, NetBeans, RAD Framework Hibernate, Spring, Struts, Junit Web Technologies HTML5, CSS3, JavaScript, jQuery, AJAX, Servlets, JSP, JSON, XML, XHTML, JSF, Angular JS Web Services SOAP, REST, WSDL, JAXB, and JAXP Operating Systems Windows (XP,7,8), UNIX, LINUX, Ubuntu, CentOS Application Servers JBoss, Tomcat, Web Logic, Web Sphere Databases Oracle, MySQL, DB2, Derby, PostgreSQL, No-SQL Database (HBase, Cassandra)

Name: Nicholas Long

Email: cruzmatthew@example.com

Phone: (659)587-6017x932