Sr.Hadoop Developer Sr.Hadoop Developer Sr.Hadoop Developer - Amtrak Washington, DC Overall 10+ years of IT industry experience in product Development, Implementation and Maintenance of various cloud-based web applications using Java, J2EE technologies and Big Data ecosystems on Linux environment Overall 5 years of experience working with analytics using Big Data technologies. Have hands-on experience in Storing, Querying, Processing and Data Analysis Experience in importing and exporting data using Sqoop between HDFS and Relational Database Management Systems Populated HDFS with huge amounts of data using Apache Kafka and Flume Excellent knowledge of data mapping, extracting, transforming and loading from different data sources Worked with different File Formats like TEXTFILE, SEQUENCE FILE, AVROFILE, ORC, and PARQUET for Hive querying and processing Experience in developing custom MapReduce Programs in Java using Apache Hadoop for analyzing Big Data as per the requirement Comprehensive work experience in implementing Big Data projects using Apache Hadoop, Pig, Hive, HBase, Spark, Sqoop, Flume, Zookeeper, Oozie Experience with distributed systems, large-scale non-relational data stores and multi-terabyte data warehouses Excellent knowledge on Hadoop architecture: Hadoop Distributed File system (HDFS), Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming paradigm Hands-on experience building data pipelines using Hadoop components Sqoop, Hive, Pig, MapReduce, Spark, Spark SQL Hands on experience in various Big Data application phases like Data Ingestion, Data Analytics and Data Visualization Experience working on Hortonworks / Cloudera / MapR distributions Extensively worked on MRV1 and MRV2 Hadoop architectures Experience working on Spark, RDD's, DAG's, Spark SQL and Spark Streaming Well experienced in data transformation using custom MapReduce, Hive and Pig scripts for different types of file formats Expertise in extending Hive and Pig core functionality by writing custom UDFs and UDAF's Designing and creating Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning and buckets Experience building solutions with NoSQL databases, such as HBase, Cassandra, MongoDB Firm grip on data modeling, data mapping, database performance tuning and NoSQL map-reduce systems In-depth understanding of Spark architecture including Spark Core, Spark SQL, Data

Frames, and Spark Streaming Hands on experience migrating complex MapReduce programs into Apache Spark RDD transformations Experienced in Apache Spark for implementing advanced procedures like text analytics and processing using the in-memory computing capabilities written in Scala Experience in Kafka installation & integration with Spark Streaming Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing Good Experience in Linux Bash scripting and following PEP Guidelines in Python Involved in converting Hive/SQL queries into Spark transformations using Spark RDD in Scala and Python Experienced with different scripting language like Python and shell scripts Hands-on experience in using python Scripts to handle data manipulation Worked well with Python, R scripts for statistics analytics for generating reports for Data Quality Experience in designing both time driven and data driven automated workflows using Oozie Good understanding of ZooKeeper for monitoring and managing Hadoop jobs Good understanding of ETL tools and how they can be applied in a Big Data environment Monitoring Map Reduce Jobs and YARN Applications Hands-on experience with Amazon Elastic MapReduce (EMR), Storage S3, EC2 instances and Data Warehousing Experience with RDBMS and writing SQL and PL/SQL scripts used in stored procedures Used Git for source code and version control management Proficient in Java, J2EE, JDBC, Collection Framework, Servlets, JSP, Spring, Hibernate, JSON, XML, REST, SOAP Web Services Strong understanding in Agile and Waterfall SDLC methodologies Excellent problem solving, proactive thinking, analytical, programming and communication skills Ability to learn new technologies and work effectively in cross-functional team environments Authorized to work in the US for any employer Work Experience Sr.Hadoop Developer Amtrak - Washington, DC September 2018 to Present Responsibilities: Involved in complete project life cycle starting from design discussion to production deployment Worked closely with the business team to gather their requirements and new support features Involved in running POC's on different use cases of the application and maintained a standard document for best coding practices Developed a 16-node cluster in designing the Data Lake with the Hortonworks distribution Responsible for building scalable distributed data solutions using Hadoop Installed, configured and implemented high availability

Hadoop Clusters with required services (HDFS, Hive, HBase, Spark, ZooKeeper) Implemented Kerberos for authenticating all the services in Hadoop Cluster Configured ZooKeeper to coordinate the servers in clusters to maintain the data consistency Involved in designing the Data pipeline from end-to-end, to ingest data into the Data Lake Wrote scripts to automate application deployments and configurations monitoring YARN Configured and developed Sqoop scripts to migrate the data from relational databases like Oracle, Teradata to HDFS Used Flume for collecting and aggregating large amounts of streaming data into HDFS Wrote MapReduce jobs in Java to parse the raw data populate staging tables and store the refined data Developed Map Reduce programs as a part of predictive analytical model development Built re-usable Hive UDF libraries for business requirements which enabled various business analysts to use these UDF's in Hive querying Created different staging tables like ingestion tables and preparation tables in Hive environment Optimized Hive queries and used Hive on top of Spark engine Worked on Sequence files, Map side joins, Bucketing, Static and Dynamic Partitioning for Hive performance enhancement and storage improvement Tested Apache TEZ, an extensible framework for building high performance batch and interactive data processing applications, on Pig and Hive jobs Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Spark SQL, Scala Worked on the Spark core and Spark SQL modules of Spark extensively Developed Python scripts to find vulnerabilities with SQL Queries by doing SQL injection Wrote Python UDFs to process the RegEx and return the valid names using streaming Created tables in HBase to store the variable data formats of data coming from different upstream sources Leveraged AWS cloud services such as EC2; auto-scaling; and VPC (Virtual Private Cloud) to build secure, highly scalable and flexible systems that handled expected and unexpected load bursts and can quickly evolve during development iterations Developed the batch scripts to fetch the data from AWS S3 storage and do required transformations in Spark framework using Scala Configured various workflows to run on top of Hadoop using Oozie and these workflows comprises of heterogeneous jobs like Pig, Hive, Sqoop and MapReduce Experience in managing and reviewing Hadoop log files Utilized capabilities of Tableau such as Data extracts, Data blending, Forecasting, Dashboard

actions and table calculations to build dashboards Followed Agile Methodologies while working on the project Performed bug fixing and 24X7 production support for running the processes Environment: Java, Scala, Hadoop, Hortonworks, AWS, HDFS, YARN, Map Reduce, Hive, Pig, Spark, Flume, Kafka, Sqoop, Oozie, Zookeeper, Oracle, Teradata, MySQL Sr.Hadoop Developer Tiffany & Co. NJ February 2017 to August 2018 Responsibilities: Worked on Hortonworks cluster, which is responsible for providing open source platform based on Apache Hadoop for analyzing, storing and managing big data Worked with analyst to determine and understand business requirements Load and transform large datasets of structured, semi structured and unstructured data using Hadoop/Big Data concepts Developed data pipeline using Flume, Sqoop, Pig and MapReduce to ingest customer data and financial histories into HDFS for analysis Used MapReduce and Flume to load, aggregate, store and analyze web log data from different web servers Created MapReduce programs to handle semi/unstructured data like XML, JSON, AVRO data files and sequence files for log files Involved in submitting and tracking MapReduce jobs using Job Tracker Experience writing Pig Latin scripts for Data Cleansing, ETL operations and query optimizations of exists scripts Written Hive UDF to sort Structure fields and return complex data types Created Hive tables from JSON data using data serialization framework like AVRO Experience writing reusable custom Hive and Pig UDF's in Java and using existing UDF's from Piggybank and other sources Experience in working with NoSQL database HBase in getting real time data analytics Integrated Hive tables to HBase to perform row level analytics Developed Oozie workflows for daily incremental loads, which Sqoop's data from Teradata, Netezza and then imported into Hive tables Involved in performance tuning by using different service engines like TEZ etc. Performed performance tuning and troubleshooting of MapReduce jobs by analyzing and reviewing Hadoop log files Implemented Daily Cron jobs that automate parallel tasks of loading the data into HDFS using AutoSys and Oozie coordinator jobs Developed suit of Unit Test Cases for Mapper, Reducer and Driver classes using MR Testing library Environment: Hortonworks, Java, Hadoop, HDFS, MapReduce, Tez, Hive, Pig, Oozie, Sqoop, Flume, Teradata, Netezza, Tableau Hadoop Developer Syniverse Technologies - Tampa, FL August 2014 to January 2017

Responsibilities: Experience with complete SDLC process staging code reviews, source code management and build process Implemented Big Data platforms as data storage, retrieval and processing systems Developed data pipeline using Kafka, Sqoop, Hive and Java MapReduce to ingest customer behavioral data and financial histories into HDFS for analysis Wrote Sqoop scripts for importing and exporting data into HDFS and Hive Wrote MapReduce jobs to discover trends in data usage by the users Load and transform large sets of structured, semi structured and unstructured data Pig Experienced working on Pig to do transformations, event joins, filtering and some pre-aggregations before storing the data onto HDFS Involved in developing Hive UDF's for the needed functionality that is not available out of the box from Hive Created Sub-Queries for filtering and faster execution of data Experienced in migrating Hive QL into Impala to minimize query response time Used HCATALOG to access the Hive table metadata from MapReduce and Pig scripts Experience loading and transforming large amounts of structured and unstructured data into HBase and exposure handling Automatic failover in HBase Ran POC's in Spark to take the benchmarking of the implementation Developed Spark jobs using Scala in test environment for faster data processing and querying Worked on migrating MapReduce programs into Spark transformations using Spark and Scala Used Python for pattern matching in build logs to format warnings and errors Configured big data workflows to run on the top of Hadoop using Oozie and these workflows comprises of heterogeneous jobs like Pig, Hive, Sqoop Cluster co-ordination services through ZooKeeper Hands on experience in Tableau for Data Visualization and analysis on large data sets, drawing various conclusions Involved in developing test framework for data profiling and validation using interactive queries and collected all the test results into audit tables for comparing the results over the period Documented all the requirements, code and implementation methodologies for reviewing and analyzation purposes Extensively used GitHub as a code repository and Phabricator for managing day to day development process and to keep track of the issues Environment: Java, Scala, Hadoop, Spark, HDFS, MapReduce, Yarn, Hive, Pig, Impala, Oozie, Sqoop, Flume, Kafka, Teradata, SQL, GitHub, Phabricator, Amazon Web Services Hadoop Developer LakeLand Bank - West Milford, NJ September 2013 to July 2014 Responsibilities:

Installed Cloudera distribution of Hadoop Cluster and services HDFS, Pig, Hive, Sqoop, Flume and MapReduce Responsible for providing open source platform based on Apache Hadoop for analyzing, storing and managing big data Loaded and transformed large sets of structured, semi-structured and unstructured data Responsible for managing data coming from different sources Imported and exported data into HDFS and Hive using Sqoop Wrote Hive queries Involved in loading data from UNIX file system to HDFS Created Hive tables, loaded with data and wrote queries which will run internally in MapReduce and performed data analysis as per the business requirements Worked with analysts to determine and understand business requirements Loaded and transformed large datasets of structured, semi structured and unstructured data using Hadoop/Big Data concepts Developed data pipeline using Flume, Sqoop, Pig and MapReduce to ingest customer data and financial histories into HDFS for analysis Used MapReduce and Flume to load, aggregate, store and analyze web log data from different web servers Created MapReduce programs to handle semi/unstructured data like XML, JSON, AVRO data files and sequence files for log files Involved in submitting and tracking MapReduce jobs using Job Tracker Experience writing Pig Latin scripts for Data Cleansing, ETL operations and query optimizations of exists scripts Written Hive UDF to sort Structure fields and return complex data types Created Hive tables from JSON data using data serialization framework like AVRO Experience writing reusable custom Hive and Pig UDF's in Java and using existing UDF's from Piggybank and other sources Experience in working with NoSQL database HBase in getting real time data analytics Integrated Hive tables to HBase to perform row level analytics Performed performance tuning and troubleshooting of MapReduce jobs by analyzing and reviewing Hadoop log files Developed Unit Test Cases for Mapper, Reducer and Driver classes using MR Testing library Supported operations team in Hadoop cluster maintenance including commissioning and decommissioning nodes and upgrades Provided technical assistance to all development projects Hands-on experience with Qlik Sense for Data Visualization and Analysis on large data sets, drawing various insights Created dashboards using Qlik Sense and performed Data extracts, Data blending, Forecasting, and table calculations Environment: Hortonworks, Java, Hadoop, HDFS, MapReduce,

Hive, Pig, Oozie, Sqoop, Flume, Netezza, Qlik Sense Java Developer Trimble Navigation Limited - Centreville, VA March 2011 to August 2013 Responsibilities: Built the application based on Rational Unified Process (RUP) Analyzed and developed UML's with Rational Rose including development of class diagrams, sequence diagrams, use case diagrams and activity diagrams Implemented the Middle-Tier employing design patterns like MVC, Business Delegate, Service Locator, Session Fa ade, Data Access Objects (DAO's) Developed using MVC architecture and employed the Struts Framework and used Validator Framework and Tiles Framework as a plug-in with struts Developed user interface using JSP, JSP Tag libraries (JSTL) and Struts Tag Libraries Used Hibernate in data access layer to access and update the information in database Used JSON to pass objects between web pages and server-side application Used XSL-FO to generate PDF reports Extensively worked on XML parsers (SAX/DOM) Used WSDL and SOAP protocol for Web Services implementation Used JDBC to access DB2 UDB database for accessing customer information Developed application level logging using Log4J Used CVS for version controlling and Junit for unit testing Involved in development of Tables, Indices, Stored procedures, Database Triggers and Functions Environment: J2EE 1.7, WebSphere Application Server v8.0, RAD, JSP 2.0, EJB 3.1, Struts 2.0, JMS, JSON, JDBC, JNDI, XML, XSL, XSLT, XSL-FO, WSDL, SOAP, Hibernate 4.0, RUP, Rational Rose (2000), Log4J, Junit, CVS, IBM DB2 v8.2, Red Hat LINUX, RESTful web services Java Developer United Health Group - Sacramento, CA May 2009 to February 2011 Responsibilities: Designed and Developed application using EJB 2.0 and Struts framework Developed POJO's for Data Model to map the Java Objects with Relational Database tables Designed and developed Service layer using Struts framework Used MVC based Struts framework to develop the multi-tier web application presentation layer components Implemented Struts tag libraries like HTML, logic, tab, bean etc. in the JSP pages Used Struts tiles library for layout of web page and performed struts validations using Struts validation framework Implemented Oracle database and JDBC drivers to access the data Involved in design, analysis and architectural meetings, created Architecture Diagrams, and Flow Charts using Rational Rose Followed Agile software development practice paired with

programming, test driven development and scrum status meetings Developed use case diagrams, class diagrams, database tables, and mapping between relational database tables Maintained the application configuration information in various properties file Performed unit testing, system testing and integration testing Environment: Java, Struts Framework, log4j, JBoss, Servlets, JSP, JSTL, I18N, JDBC, HTML, JavaScript, CSS, Rational Rose, UML, Junit, Oracle, Windows NT Education Bachelor's Skills Cassandra, Hdfs, Mapreduce, Oozie, Sqoop, Hbase, Kafka, Flume, Map reduce, Mongodb, Nosql, C++, Git, Hbase, Hive, Html, Javascript, Bootstrap, Mapreduce, Pig Certifications/Licenses Driver's License

Name: Aaron Morgan

Email: qrobbins@example.org

Phone: 001-619-244-3210