Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Walmart Bentonville, AR   4+ years of experience in developing applications that perform large scale Distributed Data Processing using Big Data ecosystem tools Hadoop, Hive, Pig, Sqoop, Hbase, Cassandra, Spark, Spark Streaming MLLib, Mahout, Oozie, Zoo Keeper, Flume, Yarn and Avro.   Passionate about Big Data Analytics and skilled in exploring data, content and Expert in distributed computing, algorithms, and data analytics.   Hands on experience in using various Hadoop distributions (Apache, Cloudera, Horton works, MapR).   In-Depth knowledge and experience in design, development and deployments of Big Data projects using Hadoop / Data Analytics / NoSQL / Distributed Machine Learning frameworks.   Solid understanding of SQL & NOSQL databases such as Oracle, Postgre SQL, MySQL, Mongo DB, HBase & Cassandra.   Knowledge on Mongo DB NoSQL data modeling, tuning, disaster recovery backup used it for distributed storage and processing using CRUD. Worked on Mongo DB database concepts such as locking, transactions, indexes, Sharing, replication, schema design.   Very good understanding Cassandra cluster mechanism that includes replication strategies, snitch, gossip, consistent hashing and consistency levels.   Working knowledge in installing and maintaining Cassandra by configuring the cassandra.yaml file as per the business requirement and performed reads/writes using Java JDBC connectivity.   Experience with Cloudera Manager for management of Hadoop cluster.   Good conceptual understanding and experience in cloud computing applications using Amazon EC2, S3, EMR.   Experience in analyzing data using HiveQL, Pig Latin, HBase and custom Map Reduce programs in Java.   Extensive Experience on Teradata database, analyzing business needs of clients, developing effective and efficient solutions and ensuring client deliverables with in committed timelines. Expertise in maintaining data quality, data organization and metadata.   Experience in Coding Teradata SQL, Teradata Stored Procedures, Macros and Triggers   Strong experience in Creating Database Objects such as Tables, Views, Functions, Stored Procedures, Indexes, Triggers, Cursors in Teradata   Experience data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka.   Developed a data pipeline using Kafka and Spark Streaming to store data into HDFS and performed the real-time analytics on the incoming data.   Experience in

importing the real-time data to Hadoop using Kafka and implemented the Oozie job for daily imports.

Experienced in working with in-memory processing framework like Spark Transformations, SparkQL, MLib and Spark Streaming.   Expertise in creating Custom Serdes in Hive.   Good working experience on using Sqoop to import data into HDFS from RDBMS and vice-versa. Expertise in job scheduling and monitoring tools like Oozie and Zoo Keeper.   Experience in design and development of Map Reduce Programs using Apache Hadoop for analyzing the big data as per the requirement.   Experience in performing ad-hoc queries on structured data using Hive QL and used Partition and Bucketing techniques and joins with HIVE for faster data access.   Experience in performing ETL operations using Pig Latin scripts.   Implemented Java APIs and created custom Java programs for full-fledged utilization of Hadoop and its related tools.   Implemented work flows that involve Hadoop actions using Oozie coordinators.   Experienced in implementing POC using Spark Sql and Mlib libraries.   Improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, YARN.   Hands on experience in handling Hive tables using Spark SQL.   Created User Defined Functions (UDFs), User Defined Aggregated Functions (UDAFs) in PIG and Hive.   Experience in dealing with log files to extract data and to copy into HDFS using Flume.   Experience with Testing Map Reduce programs using MR Unit, JUnit and Easy Mock.   Implemented distributed searching capabilities using Solr to empower the geospatial search and navigation feature.   Experienced in Java Application Development, Client/Server Applications, Internet/Intranet based applications using Core Java, J2EE patterns, Spring, Hibernate, Struts, JMS, Web Services (SOAP/REST), Oracle, SQL Server and other relational databases.   Experience writing Shell scripts in Linux OS and integrating them with other solutions.   Expert in developing web page interfaces using JSP, Java Swings, and HTML scripting languages.   Excellent understanding on Java beans and Hibernate framework to implement model logic to interact with RDBMS databases.   Experience in using IDEs like Eclipse, NetBeans and Maven.   Hands on experienced working with source control tools such as Rational Clear Case and Clear Quest.   Hands on experience on writing Queries, Stored procedures, Functions and Triggers by using SQL.   Used EMR (Elastic Map Reducing) to perform bigdata

operations in AWS. Proficient using version control tools like GIT, VSS, SVN and PVCS. Involvement in all stages of software development life cycle (SDLC) and follow agile methodologies and continuous delivery. Strong skills in Object Oriented Analysis and Design (OOAD). Well versed in enterprise software development methodologies and practices including TDD, BDD design patterns and performance testing. Authorized to work in the US for any employer Work Experience

Sr. Hadoop Developer Walmart - Bentonville, AR May 2018 to Present Description: Walmart Inc. is an American multinational retail corporation that operates a chain of hypermarkets, discount department stores, and grocery stores.we've been dedicated to making a difference in the lives of our customers. Our business is the result of Sam Walton's visionary leadership, along with generations of associates focused on helping customers and communities save money and live better. Responsibilities: Analyzed large and critical datasets using Cloudera, HDFS, HBase, MapReduce, Hive, Hive UDF, Pig, Sqoop, Zookeeper and Spark. Gathered the business requirements from the Business Partners and Subject Matter Experts Developed environmental search engine using JAVA, Apache SOLR and MYSQL. Managed works including indexing data, tuning relevance, developing custom tokenizers and filters, adding functionality includes playlist, custom sorting and regionalization with SOLR Search Engine. Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandra as per the business requirement. Written multiple Map Reduce programs for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats. Developed automated processes for flattening the upstream data from Cassandra which in JSON format. Used Hive UDFs to flatten the JSON Data. Optimized Map Reduce Jobs to use HDFS efficiently by using various compression mechanisms Developed PIG UDFs to provide Pig capabilities for manipulating the data according to Business Requirements and worked on developing custom PIG Loaders and Implemented various requirements using Pig scripts. Experienced on loading and transforming of large sets of structured, semi structured and unstructured data Created POC using Spark Sql and Mlib libraries. Developed a Spark Streaming module for consumption of Avro messages from Kafka. Implementing different machine

learning techniques in Scala using Scala machine learning library, and created POC using SparkSql and Mlib libraries.   Experienced in Querying data using SparkSQL on top of Spark Engine, implementing Spark RDD's in Scala.   Expertise in writing Scala code using Higher order functions for iterative algorithms in Spark for Performance considerations.   Experienced in managing and reviewing Hadoop log files   Worked with different File Formats like TEXTFILE, AVROFILE, ORC, and PARQUET for HIVE querying and processing.   Loading data by using the Teradata loader connection, writing Teradata utilities scripts (Fast Load, Multiload) and working with loader logs.   To monitor query run times using Teradata Performance Monitor   Involved in loading of data into Teradata from legacy systems and flat files using complex MLOAD scripts and FASTLOAD scripts.   Create and Maintain Teradata Tables, Views, Macros, Triggers and Stored Procedures   Monitored workload, job performance and capacity planning using Cloudera Distribution.   Worked on Data loading into Hive for Data Ingestion history and Data content summary.   Involved in developing Impala scripts for extraction, transformation, loading of data into data warehouse.   Used Hive and Impala to query the data in HBase.   Created Impala tables and SFTP scripts and Shell scripts to import data into Hadoop.   Developed Hbase java client API for CRUD Operations.   Created Hive tables and involved in data loading and writing Hive UDFs. Developed Hive UDFs for rating aggregation   Generated Java APIs for retrieval and analysis on No-SQL database such as HBase and Cassandra   Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig Did various performance optimizations like using distributed cache for small datasets, partition and bucketing in hive, doing map side joins etc   Worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop for analysis, visualization and to generate reports. Worked on Apache spark writing python applications to convert txt, xls files and parse.   Developed Python scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop.   Use different components of Talend (tOracleInput, tOracleOutput, tHiveInput, tHiveOutput, tHiveInputRow, tUniqeRow, tAggregateRow, tRunJob, tPreJob, tPostJob, tMap, tJavaRow, tJavaFlex, tFilterRow etc) to develop standard jobs.   Loading data from different source (database & files) into Hive using Talend tool.

Load and transform data into HDFS from large set of structured data /Oracle/Sql server using Talend Big data studio.   Implemented Spark using Python/Scala and utilizing Spark Core, Spark Streaming and Spark SQL for faster processing of data instead of Map Reduce in Java   Experience in integrating Apache Kafka with Apache Spark for real time processing.   Exposure on usage of Apache Kafka develop data pipeline of logs as a stream of messages using producers and consumers.   Scheduled Oozie workflow engine to run multiple Hive and Pig jobs, which independently run with time and data availability   Worked on custom Pig Loaders and Storage classes to work with a variety of data formats such as JSON, Compressed CSV etc   Involved in running Hadoop Streaming jobs to process Terabytes of data   Used JIRA for bug tracking and CVS for version control.   Environment: Hadoop, Map Reduce, Hive, HDFS, PIG, Sqoop, Oozie, Cloudera, Flume, HBase, SOLR, CDH3, Cassandra, Oracle, Unix/Linux, Hadoop, Hive, PIG, SQOOP, Flume, HDFS, J2EE, Oracle/SQL & DB2 , Unix/Linux, JavaScript, Ajax, Eclipse IDE, CVS, JIRA. Hadoop/ Java developer University of Arkansas January 2017 to May 2017 Global Learning Department: The University of Arkansas, performs analysis on the various data they have. The Global Learning Department is one of the projects that perform Data analytics on the data related to the Student applications received, students admitted and other student activities such as enrollments and, on the courses, offered by each department. The aim of the project is to perform data analytics and generate the reports for every month and semester.   Responsibilities: Involved in all Phases of application development   Followed the Agile Scrum development methodology   Developed the web application using Java/J2EE and Spring framework   Developed the Business logic in the middle-tier using Java Beans and Java classes.   Implemented the Java beans, auto-wiring and annotations in Apache Spring framework   Connecting to the Hive from Java layer using the Hive JDBC driver   Conducted Hibernate and created POJO classes for establishing Object Relation Mapping (ORM) between objects and relational database.   Experienced with JDBC API in establishing connection Between databases and java POJO classes   Developed SQL queries to retrieve and validate Oracle data and executed the same using OJDBC drivers.   Played a key role in designing the Fact and dimension table for the Star Schema   Created Hive tables and

loaded, retrieved the data from tables and developed Hive UDFs using Java   Performing testing on the spark modules developed with the help of sample data sets   Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Hive, Pig, and Sqoop.   Environment: Java/J2EE, Apache Spring, Spark, Hive, PIG, Tableau, Oracle, OJDBC, Pig, Sqoop, Scala, GIT. Hadoop Developer Arteria Technologies Pvt. Ltd August 2014 to November 2016 Responsibilities: Installed and configured Hadoop MapReduce, HDFS and developed multiple MapReduce jobs in Java for data cleansing and preprocessing.   Responsible to manage data coming from different sources.   Involved in gathering the business requirements from the Business Partners and Subject Matter Experts.   Proactively monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup and disaster recovery systems and procedures.   Involved in works including indexing data, tuning relevance, developing custom token izers and filters, adding functionality includes playlist, custom sorting.   Supported Map Reduce Programs those are running on the cluster.   Involved in HDFS maintenance and loading of structured and unstructured data.   Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS   Worked extensively with importing metadata into Hive using Python and migrated existing tables and applications to work on AWS cloud(S3).   Installed and configured Pig and written Pig Latin scripts.   Imported data using Sqoop to load data from MySQL to HDFS on regular basis.   Expert knowledge on Mongo DB NoSQL data modeling, tuning, disaster recovery backup used it for distributed storage and processing using CRUD.   Extracted and restructured the data into Mongo DB using import and export command line utility tool.   Designed and Maintained Tez workflows to manage the flow of jobs in the cluster.   Developed Scripts and Batch Job to schedule various Hadoop Program. Installation of Oozie workflow engine to run multiple Hive and pig jobs.   Writing Hive queries for data analysis to meet the business requirements.   Loading log data into HDFS using Flume and performing ETL Integration.   Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis.   Developed Flume ETL job for handling data from HTTP Source and Sink as HDFS.   Created Hive tables and working on them using Hive QL.

Good Understanding of DAG cycle for entire Spark application flow on Spark application Web UI. Developed Spark SQL scripts and involved in converting Hive UDF's to Spark SQL UDF's. Implemented procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.   Developed multiple Spark jobs in Scala/Python for Data cleaning, pre-processing and Aggregating.   Developed Spark programs using Scala, Involved in Creating SparkSQL Queries and Developed Oozie workflow for Spark jobs.   Push data as delimited files into HDFS using Talend Big data studio.   Analyzed and performed data integration using Talend open integration suite.   Wrote complex SQL queries to take data from various sources and integrated it with Talend.   Used Storm for an automatic mechanism for repeating attempts to download and manipulate the data when there is a hiccup.   Designing and development of technical architecture, requirements and statistical models using R.   Used Storm to analyze large amounts of non-unique data points with low latency and high throughput.   Developed UI application using AngularJS, integrated with Elastic Search to consume REST.   Writing the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.   Utilized Agile and Scrum Methodology to help manage and organize a team of developers with regular code review sessions.   Environment: Hadoop, Hive, Linux, Map Reduce, HDFS, Hive, Pig, HBase, Sqoop, Kafka, Flume, Shell Scripting, Storm, Java (JDK 1.6), Java 6, Eclipse, Oracle 10g, PL/SQL, SQL*PLUS, Toad 9.6, Linux, JIRA 5.1, Storm, CVS, JIRA 5.2. Java Developer Maveric Systems - IN June 2013 to June 2014 Responsibilities: Interacted with Team and Analysis, Design and Develop database using ER Diagram, Normalization and relational database concept.   Involved in Design, Development and testing of the system.   Developed SQL Server Stored Procedures, Tuned SQL Queries (using Indexes and Execution Plan)   Developed User Defined Functions and created Views.   Created Triggers to maintain the Referential Integrity.   Implemented Exceptional Handling.   Involved in Requirement Analysis, Development and Documentation.   Developed front-end using JSP, HTML, CSSand JavaScript.   Coding for DAO Objects using JDBC (using DAO pattern).   XML and XSDs are used to define data formats.   Implemented J2EE design patterns such as singleton, DAO for the presentation tier, business tier

and Integration Tier layers of the project. Worked on client requirement and wrote Complex SQL Queries to generate Crystal Reports. Creating and automating the regular Jobs. Tuned and Optimized SQL Queries using Execution Plan and Profiler. Developed user interface (view component of MVC architecture) with JSP, Struts Custom Tag libraries, HTML5 and JavaScript. Used DOJO toolkit to construct Ajax requests and build dynamic web pages using JSP's, DHTML and JavaScript's, extensively used jQuery in web based applications Developed the controller component with Servlets and action classes. Business Components are developed (model components) using Enterprise Java Beans (EJB) Established schedule and resource requirements by planning, analyzing and documenting development effort to include time lines, risks, test requirements and performance targets Analyzing System Requirements and preparing System Design document Developing dynamic User Interface with HTML and JavaScript using JSP and Servlet Technology Designed and developed a sub system where Java Messaging Service (JMS) applications are developed to communicate with MQ in data exchange between different systems Used JMS elements for sending and receiving messages Used hibernate for mapping from Java classes to database tables Created and executed Test Plans using Quality Center by Test Director Mapped requirements with the Test cases in the Quality Center Supporting System Test and User acceptance test Rebuilding Indexes and Tables as part of Performance Tuning Exercise. Involved in performing database Backup and Recovery. Worked on Documentation using MS word. Environment: SQL Server 7.0/2000, SQL, T-SQL, BCP, Visual Basic 6.0/5.0, Crystal Reports 7/4.5, Java, J2ee, JDBC, EJB, JSP, EL, JSTL, JUNIT, XML, SOAP, WSDL, Eclipse, Windows XP, Oracle Education Master's Additional Information TECHNICAL SKILL SET: Languages/Tools Java, C, C++, C#, Scala, VB, XML, HTML/XHTML, HDML, DHTML. Big Data HDFS, Map Reduce, HIVE, PIG, HBase, SQOOP, Oozie, Zookeeper, Spark, Mahout, Kafka, Storm, Cassandra, Solr, Impala, Green plum, Mongo DB Web/Distributed Technologies J2EE ,Servlets, JSP, Struts, Hibernate, JSF, JSTL,EJB,RMI,JNI, XML,JAXP,XSL,XSLT, UML, MVC, STRUTS, Spring, Corba, Java Threads. J2EE Technologies J2EE, Servlets, JSP 2.1/2.2, EJB2.1/3.0, JDBC, MVC Architecture, Java Beans, JNDI, RMI, JMS, Java, ANT 1.8, JavaScript, Spring. Browser

Languages/Scripting HTML, XHTML, CSS, XML, XSL, XSD, XSLT, Java script, HTML DOM, DHTML, AJAX. App/Web Servers IBM Web sphere [ ] BEA Web logic, Jdeveloper, Apache Tomcat, JBoss. GUI Environment Swing, AWT, Applets. Messaging & Web Services Technology SOAP, WSDL, UDDI, XML, SOA, JAX-RPC, IBM Web Sphere MQ v5.3, JMS. Testing &Case Tools JUnit, Log4j, Rational Clear case, CVS, ANT, Maven, JBuilder. Configuration Management Chef, Puppet, Ansible, Docker. Build Tools CVS, Subversion, GIT, Ant, Maven, Gradle, Hudson, TeamCity, Jenkins, Chef, Puppet, Ansible, Docker. CI Tools Jenkins, Bamboo. Scripting Languages Python, Shell (Bash), Perl, PowerShell, Ruby, Groovy, PowerShell. Monitoring Tools Nagios, Cloud Watch, JIIRA, Bugzilla and Remedy. Databases NO SQL Oracle, MS SQL Server 2000, DB2, MS Access & My SQL, Teradata. Cassandra, Greenplum and Mongo DB Operating systems Windows, Solaris, Unix, Linux (Red Hat 'SUSE Linux), Sun Solaris, Ubuntu, CentOS.

Name: Mia Mcguire

Email: jessica37@example.net

Phone: 304-428-5348x97013