

Sr. Hadoop/Big Data Developer Sr. Hadoop/Big Data Developer Sr. Hadoop/Big Data Developer - MedAssets Alpharetta, GA Authorized to work in the US for any employer Work Experience Sr. Hadoop/Big Data Developer MedAssets - Alpharetta, GA April 2018 to Present Responsibilities

Developed Spark applications using Scala utilizing Data frames and Spark SQL API for faster processing of data. Developed highly optimized Spark applications to perform various data cleansing, validation, transformation and summarization activities according to the requirement Data pipeline consists Spark, Hive and Sqoop and Custom build Input Adapters to ingest, transform and analyze operational data. Developed Spark jobs and Hive Jobs to summarize and transform data. Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data. Involved in converting Hive/SQL queries into Spark transformations using Spark Data Frames and Scala. Analyzed the SQL scripts and designed the solution to implement using Scala. Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data. Real time streaming the data using Spark with Kafka Created applications using Kafka, which monitors consumer lag within Apache Kafka clusters. Used in production by multiple report suites. Ingested syslog messages parse them and streams the data to Apache Kafka. Handled importing data from different data sources into HDFS using Sqoop and performing transformations using Hive, Map Reduce and then loading data into HDFS. Exported the analyzed data to the relational databases using Sqoop, to further visualize and generate reports for the BI team. Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis Analyzed the data by performing Hive queries (Hive QL) to study customer behavior. Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting. Developed Hive scripts in Hive QL to de-normalize and aggregate the data. Created HBase tables and column families to store the user event data. Scheduled and executed workflows in Oozie to run various jobs.

Environment: HDFS, Spark, Scala, Tomcat, Netezza, EMR, Oracle, Sqoop, AWS, Terraform, Scylla DB, Cassandra, MySql, Oozie Sr. Hadoop Developer Gilead - Foster City, CA February 2017 to March 2018 Responsibilities: Implemented AWS solutions using EC2, S3 and load balancers.

Installed application on AWS EC2 instances and also configured the storage on S3 buckets. Storing and loading the data from HDFS to Amazon S3 and backing up the Namespace data. Working with data delivery teams to setup new Hadoop users. This job includes setting up Linux users, setting up Kerberos principals and testing HDFS, Hive. Involved in creating Hadoop streaming jobs using Python. Created concurrent access for hive tables with shared and exclusive locking that can be enabled in hive with the help of Zookeeper implementation in the cluster. Worked on 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. Implemented PySpark and Spark SQL for faster testing and processing of data. Developed multiple MapReduce jobs in Java for data cleaning. Developed Hive UDF to parse the staged raw data to get the Hit Times of the claims from a specific branch for a particular insurance type code. Expert in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. Used Scala to write the code for all the use cases in Spark and Spark SQL. Expertise in implementing Spark and Scala application using higher order functions for both batch and interactive analysis requirement. Implemented SPARK batch jobs. Worked with Spark core, Spark Streaming and Spark SQL module of Spark. Worked on reading multiple data formats on HDFS using PySpark. Ran many performance tests using the Cassandra-stress tool in order to measure and improve the read and write performance of the cluster. Created data model for structuring and storing the data efficiently. Implemented partitioning and bucketing of tables in Cassandra. Worked on migrating MapReduce programs into PySpark transformation. Built wrapper shell scripts to hold Oozie workflow. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala. Worked on Distributed/Cloud Computing (Map Reduce/Hadoop, Pig, Hbase, AVRO, Zookeeper, etc.), Amazon Web Services (S3, EC2, EMR etc.) Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig. Worked on MRJ in querying multiple semi-structured data as per analytic needs. Used Hive to analyze the partitioned and bucketed

data and compute various metrics for reporting. Used Hive and created Hive tables and involved in data loading and writing Hive UDFs. Developed POC for Apache Kafka and Implementing real-time streaming ETL pipeline using Kafka Streams API. Involved in using Solr Cloud implementation to provide real time search capabilities on the repository with tera bytes of data. Populated HDFS and Cassandra with huge amounts of data using Apache Kafka. Integrated Apache Storm with Kafka to perform web analytics. Uploaded click stream data from Kafka to HDFS, Hbase and Hive by integrating with Storm. Configured Spark Streaming to receive real time data from the Apache Kafka and store the stream data to HDFS using Scala. Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions. Familiarity with NoSQL databases such as Cassandra. Wrote shell scripts for rolling day-to-day processes and it is automated. Involved in story-driven agile development methodology and actively participated in daily scrum meetings. Environment: Cloudera, Java, Scala, Hadoop, Spark, HDFS, MapReduce, Yarn, Hive, Pig, Zookeeper, Impala, Oozie, Sqoop, Flume, Kafka, Teradata, SQL, GitHub, Phabricator, Amazon Web Service Hadoop Developer Stens Corporation - Jasper, IN September 2015 to January 2017 Responsibilities: Developed data pipeline using Flume, Sqoop, Pig and MapReduce to ingest customer behavioral data and financial histories into HDFS for analysis. Worked on analyzing Hadoop cluster and different Big Data analytic tools including Pig, Hive, HBase database and Sqoop. Implemented real time analytics pipeline using Confluent Kafka, storm, elastic search, splunk and green plum. Design and develop Informatica BDE Application and Hive Queries to ingest Landing Raw zone and transform the data with business logic to refined zone and to Green plum data marts for reporting layer for consumption through Tableau. Installed, configured, and maintained big data technologies and systems. Maintained documentation and troubleshooting playbooks. Automated the installation and maintenance of Kafka, storm, zookeeper and elastic search using salt stack technology. Developed connectors for elastic search and green plum for data transfer from a Kafka topic. Performed Data Ingestion from multiple internal clients using Apache Kafka. Developed k-streams using java for real time data processing. Responded to and resolved access and

performance issues. Used Spark API over Hadoop to perform analytics on data in Hive   Exploring with Spark improving the performance and optimization of the existing algorithms Hadoop using Spark context, Spark-SQL, Data Frame, Spark YARN   Imported and exporting data into HDFS and Hive using SQOOP & Developed POC on Apache-Spark and Kafka. Proactively monitored performance, Assisted in capacity planning.   Worked on Oozie workflow engine for job scheduling Imported and exported data into MapReduce and Hive using Sqoop   Performed transformations, cleaning and filtering on imported data using Hive, Map Reduce, and loaded final data into HDFS Good understanding of performance tuning with NoSQL, Kafka, Storm and SQL Technologies Design/Develop framework to leverage platform capabilities using MapReduce, Hive UDFs Worked on data transformation pipelines like Storm. Worked with operational analytics and log management using ELK and Splunk. Assisted teams with SQL and MPP databases such as Green plum.   Worked on Salt Stack automation tools. Helped teams working with batch-processing and tools in Hadoop technology stack (MapReduce, Yarm, Pig, Hive, HDFS) Environment: Hadoop, MapReduce, HDFS, Hive, HBase, Sqoop, Pig, Flume, Oracle 11/10g, DB2, Teradata, MySQL, Eclipse, PL/SQL, Java, Linux, Shell Scripting, SQL Developer, SOLR. Hadoop/ Java Developer Innitech - Hyderabad, Telangana May 2013 to August 2015 Responsibilities:   Worked with several clients and performed jobs for a day to day requests and responsibilities. Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Zookeeper, and Sqoop.   Involved in analyzing system failures, identifying root causes and recommended a course of actions.   Worked on Hive for exposing data for further analysis and for generating transformation files from different analytical formats to text files.   Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.   Managed and scheduled Jobs on a Hadoop cluster.   Implemented and maintained various projects in Java.   Processed the Design of ERD (Entity Relationship Diagrams) for the Relational database.   Extensively used SQL, PL/SQL, Triggers, and Views using IBM DB2.   Utilized Java and MySQL from day to day to debug and fix issues with client processes.   Developed, tested, and implemented a financial-services application

to bring multiple clients into standard database format. Assisted in designing, building, and maintaining a database to analyze life cycle of checking and debit transactions. Excellent JAVA, J2EE application development skills with strong experience in Object Oriented Analysis. Extensively involved throughout Software Development Life Cycle (SDLC). Strong experience of J2SE, XML, Web Services, WSDL, SOAP and, TCP/IP. Strong experience in software and system development using JSP, Servlet, JSF, EJB, JDBC, Struts, Maven, Subversion, Trac, JUnit and, SQL language. Rich experience of database design and hands-on experience with large database systems: Oracle 8i and Oracle 9i, DB2 and languages like SQL and PL/SQL. Worked with Weblogic Application Server, Websphere Application Server, and J2EE application deployment technology. Environment: Apache Hadoop, HDFS, Cassandra, MapReduce, HBase, Impala, Java (jdk1.6), Kafka, MySQL, Amazon, DB Visualizer, Linux, Sqoop, Apache Hive, Apache Pig, Infosphere Python, Scala, NoSQL, Flume, Oozie Java Developer NBS - Hyderabad, Telangana July 2011 to April 2013 Responsibilities: Involved in all the phases of the life cycle of the project from requirements gathering to quality assurance testing. Developed Class diagrams, Sequence diagrams using Rational Rose. Responsible in developing Rich Web Interface modules with Struts tags, JSP, JSTL, CSS, JavaScript, Ajax, GWT. Developed presentation layer using Struts framework, and performed validations using Struts Validator plugin. Created SQL script for the Oracle database. Implemented the Business logic using Java Spring Transaction Spring AOP. Implemented persistence layer using Spring JDBC to store and update data in database. Produced web service using WSDL/SOAP standard. Implemented J2EE design patterns like Singleton Pattern with Factory Pattern. Extensively involved in the creation of the Session Beans and MDB, using EJB 3.0. Used Hibernate framework for Persistence layer. Extensively involved in writing Stored Procedures for data retrieval and data storage and updates in Oracle database using Hibernate. Deployed and built the application using Maven. Performed testing using JUnit. Used JIRA to track bugs. Extensively used Log4j for logging throughout the application. Produced a Web service using REST with Jersey implementation for providing customer information. Used SVN for source code versioning and code repository. Environment:

Java/J2EE, JDK 1.7/1.8, LINUX, Spring MVC, Eclipse, JUnit, Servlets, DB2, Oracle 11g/12c, GIT, GitHub, JSON, RESTful, HTML5, CSS3, JavaScript, Rally, Agile/Scrum Education Bachelor's Skills HDFS, SQOOP, APACHE KAFKA, KAFKA, DATA SCIENCE, DATA SOURCES, DB2, ETL, FLUME, HADOOP, HADOOP CLUSTER, JBOSS, MONGODB, DATABASE, MS SQL SERVER, MS SQL SERVER 2005, SQL SERVER, SQL SERVER 2005, MYSQL, ORACLE

Name: Jeff Stevens

Email: jasminesmith@example.com

Phone: (276)911-4558x58227