Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Wells Fargo Charlotte, NC

Work Experience Sr. Hadoop Developer Wells Fargo - Charlotte, NC January 2018 to Present

Responsibilities: Multiple Spark Jobs were written to perform Data Quality checks on data before files were moved to Data Processing Layer.   Implemented Spark using Scala and Spark SQL for faster testing and processing of data.   Designed and Modified Database tables and used HBASE Queries to insert and fetch data from tables.   Responsible for creating data pipeline using Kafka, Flume and Spark Streaming for Twitter source to collect the sentiment tweets of Eaton customers about the reviews   Involved in moving all log files generated from various sources to HDFS for further processing through Flume1.7.0.   Involved in deploying the applications in AWS and maintains the EC2 (Elastic Computing Cloud) and RDS (Relational Database Services) in amazon web services.   Implemented the file validation framework, UDFs, UDTFs and DAOs.   Strong experienced in working with UNIX/LINUX environments, writing UNIX shell scripts, Python   Created reporting views in Impala using Sentry Policy files.   different databases like MySQL, RDBMS into HDFS and HBASE using Sqoop.   Advanced knowledge in performance troubleshooting and tuning Cassandra clusters.   Analyzing the source data to know the quality of data by using Talend Data Quality.   Involved in creating Hive tables, loading with data and writing hive queries.   Developed REST APIs using Java, Play framework and Akka.   Model and Create the consolidated Cassandra, Filo DB and Spark tables based on the data profiling.   Used OOZIE1.2.1Operational Services for batch processing and scheduling workflows dynamically and created UDF's to store specialized data structures in HBase and Cassandra.   Developed multiple MapReduce jobs in Java for data cleaning and pre-processing.   Used Impala to read, write and query the Hadoop data in HDFS from Cassandra and configured Kafka to read and write messages from external programs.   Optimizing existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's. Create a complete processing engine, based on Cloudera distribution, enhanced to performance Environment: Hadoop, HDFS, MapReduce, Yarn, Hive, Pig, HBase, Oozie, Sqoop, Kafka, Flume, Oracle 11g, FiloDB, Spark, Akka, Scala, Cloudera HDFS, Talend, Eclipse, Oozie, Node.js, Unix/Linux, Aws, JQuery, Ajax, Python, Zookeeper Hadoop/ Bigdata Developer Target, MN April

2016 to January 2018 Responsibilities: Developed efficient MapReduce programs for filtering out the unstructured data and developed multiple MapReduce jobs to perform data cleaning and preprocessing on Hortonworks. Implemented Data Interface to get information of customers using RestAPIand Pre-Processdata using MapReduce2.0 and store into HDFS (Hortonworks) Extracted files from MySQL, Oracle, and Teradata2 through Sqoop1.4.6and placed in HDFS Cloudera Distribution and processed. Worked with various HDFS file formats like Avro1.7.6, Sequence File, Json and various compression formats like Snappy, bzip2. Successfully written Spark Streaming application to read streaming twitter data and analyze twitter records in real time using kafka and flume to measure performance of Apache spark streaming. Proficient in designing Row keys and Schema Design for NoSQL Database Hbaseand knowledge of other NOSQL database Cassandra. Used Hive to perform data validation on the data ingested using scoop and flume and the cleansed data set is pushed into Hbase. Good understanding of Cassandra Data Modeling based on applications. Wrote ETL jobs to read from web APIs using REST and HTTP calls and loaded into HDFS using java and Talend. Developed the Pig 0.15.0UDF's to pre-process the data for analysis and Migrated ETL operations into Hadoop system using Pig Latin scripts and Python Scripts3.5.1. Used Pig as ETL tool to do transformations, event joins, filtering and some pre-aggregations before storing the data into HDFS. Troubleshooting, debugging & altering Talend issues, while maintaining the health and performance of the ETL environment. Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop. Used spark to parse XML files and extract values from tags and load it into multiple hive tables. Experienced in running Hadoop streaming jobs to process terabytes of formatted data using Pythonscripts. Developed small distributed applications in our projects using Zookeeper3.4.7and scheduled the workflows using Oozie 4.2.0. Proficiency in writing the Unix/Linux shell commands. Developed a SCP Stimulator which emulates the behavior of intelligent networking and Interacts with SSF Environment: Hadoop, HDFS, MapReduce, Yarn, Hive, Pig, HBase, Oozie, Sqoop, Kafka, Flume, Oracle 11g, Spark, Scala, Cloudera HDFS, Talend, Eclipse, Oozie, Unix/Linux, Aws, Python, Perl, Zookeeper. Hadoop/ Bigdata Developer Middle by

Corp - Elgin, IL October 2014 to March 2016 Responsibilities: Developed multiple Map-Reduce jobs in java for data cleaning and pre-processing. Performed Map Reduce Programs those are running on the cluster. Involved in loading data from RDBMS and web logs into HDFS using Sqoop and Flume. Worked on loading the data from MySQL to HBase where necessary using Sqoop. Configured Hadoop cluster with Name node and slaves and formatted HDFS. Performed Importing and exporting data from Oracle to HDFS and Hive using Sqoop Performed source data ingestion, cleansing, and transformation in Hadoop. Supported Map-Reduce Programs running on the cluster. Wrote Pig Scripts to perform ETL procedures on the data in HDFS. Used Oozie workflow engine to run multiple Hive and Pig jobs. Analyzed the partitioned and bucketed data and compute various metrics for reporting. Created HBase tables to store various data formats of data coming from different portfolios. Worked on improving the performance of existing Pig and Hive Queries. Involved in developing HiveUDFs and reused in some other requirements. Worked on performing Join operations. Developed fingerprinting rules on HIVE which help in uniquely identifying a driver profile Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs. Exported the result set from Hive to MySQL using Sqoop after processing the data. Analyzed the data by performing Hive queries and running Pig scripts to study customer behavior. Used Hive to partition and bucket data Environment: Hadoop. MapReduce, HDFS, HBase, HDP Horton, Sqoop, Data Processing Layer, HUE, AZURE, Erwin, MS Visio, Tableau, SQL, MongoDB, Oozie, UNIX, MySQL, RDBMS, Ambari, Solr Cloud, Lily HBase, Cron. Java developer YahSat Abu Dhabi June 2011 to September 2014 Responsibilities: Involved in complete requirement analysis, design, coding and testing phases of the project. Developed the application using , which is based on Model View Controller design pattern. Extensively used Hibernate in data access layer to perform database operations. Used Spring Framework for Dependency Injection and integrated it with the Struts Framework and Hibernate. Developed front end using Struts framework. Configured Struts DynaAction Forms, Message Resources, Action Messages, Action Errors, Validation.xml, and Validator-rules.xml. Designed and developed front-end using struts framework. Used JSP, JavaScript, JSTL, EL,

Custom Tag libraries and Validations provided by struts framework. Used Web services - WSDL and SOAP for getting credit card information from third party. Worked on advanced Hibernate associations with multiple levels of Caching, lazy loading. Created Use case, Sequence diagrams, functional specifications and User Interface diagrams using Star UML. Worked in Agile development environment in sprint cycles of two weeks by dividing and organizing tasks. Participated in daily scrum and other design related meetings. Designed various tables required for the project in Oracle 9i database and used Stored Procedures and Triggers in the application. Involved in consuming RESTful Web services to render the data to the front page. Performed unit testing using JUnit framework. Environment: HTML, JSP, Servlets, JDBC, JavaScript, Java API, Spring 3.0, Spring MVC, JDBC, Maven, SVN, Servlets, Struts, Amazon WS, RESTful Web Services, Bootstrap Java developer ST Electronics - Singapore July 2010 to May 2011 Responsibilities: Created Use case, Sequence diagrams, functional specifications and User Interface diagrams using Star UML. Involved in complete requirement analysis, design, coding and testing phases of the project. Participated in JAD meetings to gather the requirements and understand the End Users System. Developed user interfaces using JSP, HTML, XML and JavaScript. Created Stored Procedures & Functions. Used JDBC to process database calls for DB2/AS400 and SQL Server databases. Developed the code which will create XML files and Flat files with the data retrieved from Databases and XML files. Created Data sources and Helper classes which will be utilized by all the interfaces to access the data and manipulate the data. Used Servlets to implement business components. Designed and Developed required service classes for database operation. Developed web application called iHUB (integration hub) to initiate all the interface processes using Struts Framework, JSP and HTML. Used Java Script validation in JSP pages. Developed the interfaces using Eclipse 3.1.1 and JBoss 4.1 Involved in integrated testing, Bug fixing and in Production Support. Environment: HTML, JSP, Servlets, JDBC, JavaScript, Tomcat, Eclipse IDE, XML, XSL, Tomcat 5 Education Bachelor's Skills Apache cassandra, Cassandra, Hdfs, Impala, Mapreduce

Name: Kurt Martinez

Email: wardsamantha@example.net

Phone: 804-634-2859