

Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Mattel, Inc El Segundo, CA

Overall 8+ years of overall experience with strong emphasis on Design, Development, Implementation, Testing and Deployment of Software Applications. Over 4+ years of comprehensive IT experience in BigData and BigData Analytics, Hadoop, HDFS, MapReduce, YARN, Hadoop Ecosystem and Shell Scripting. 5+ years of development experience using Java, J2EE, JSP and Servlets. Highly capable for processing large sets of Structured, Semi-structured and Unstructured datasets and supporting BigData applications. Hands on experience with Hadoop Ecosystem components like Map Reduce (Processing), HDFS (Storage), YARN, Sqoop, Pig, Hive, HBase, Oozie, ZooKeeper and Spark for data storage and analysis. Expertise in transferring data between a Hadoop ecosystem and structured data storage in a RDBMS such as MY SQL, Oracle, Teradata and DB2 using Sqoop. Experience in NoSQL databases like Mongo DB, HBase and Cassandra. Experience in Apache Spark cluster and streams processing using Spark Streaming. Expertise in moving large amounts of log, streaming event data and Transactional data using Flume. Experience in developing MapReduce jobs in Java for data cleaning and preprocessing. Expertise in writing Pig Latin, Hive Scripts and extended their functionality using User Defined Functions (UDF's). Expertise in handling arrangement of data within certain limits (Data Layout's) using Partitions and Bucketing in Hive. Expertise in preparing Interactive Data Visualization's using Tableau Software from different sources. Hands on experience in developing workflows execute MapReduce, Sqoop, Pig, Hive and Shell Scripts using Oozie. Experience working with Cloudera Hue Interface and Impala. Hands on experience developing Solr Indexes using MapReduce Indexer Tool. Expertise in Object-Oriented Analysis and Design (OOAD) like UML and use of various design patterns. Experience in Java, JSP, Servlets, EJB, Web Logic, Web Sphere, Hibernate, Spring, JBoss, JDBC, RMI, Java Script, Ajax, JQuery, XML and HTML. Fluent with the core Java concepts like I/O, Multi-Threading, Exceptions, Reg Ex, Data Structures and Serialization. Performed Unit Testing using Junit Testing Framework and Log4J to monitor the error logs. Experience in process Improvement, Normalization/De-normalization, Data extraction, cleansing and Manipulation. Converting

requirement specification, Source system understanding into Conceptual, Logical and Physical Data Model, Data flow (DFD). Expertise in working with Transactional Databases like Oracle, SQL server, My SQL, and Db2. Expertise in developing SQL queries, Stored Procedures and excellent development experience with Agile Methodology. Ability to adapt to evolving technology, Strong sense of Responsibility and Accomplishment. Excellent leadership, interpersonal, problem solving and time management skills. Excellent communication skills both Written (documentation) and Verbal (presentation). Authorized to work in the US for any employer

Work Experience

Sr. Hadoop Developer Mattel, Inc October 2018 to Present

Description: FFM- Federally Facilitated Marketplace-Healthcare: Health Partners, like other healthcare organizations, has a legacy system, clinicians, and researchers that needed access to the data. The data types include EMR generated data, genomic data, financial data , patient and caregiver data, incremental physiological monitoring, ventilator data, temperature, and humidity data. Any electronically generated data in the healthcare environment can be ingested and stored in the hdfs which will be used for analytic to aid in the delivery of quality care at the lowest possible cost and an environment to enable clinical researchers to examine healthcare data

Key Achievements:

- Performed performance tuning and troubleshooting of MapReduce jobs by analyzing and reviewing Hadoop log files.
- Involved Low level design for MR, Hive, Impala, Shell scripts to process data.
- Involved in complete Big Data flow of the application starting from data ingestion upstream to HDFS, processing the data in HDFS and analyzing the data.
- Knowledge on handling Hive queries using Spark SQL that integrate with Spark environment implemented in Scala.
- Used Spark Streaming API with Kafka to build live dashboards;
- Worked on Transformations & actions in RDD, Spark Streaming, Pair RDD Operations, Check-pointing, and SBT.
- Implemented POC to migrate map reduce jobs into Spark RDD transformation using Scala IDE for Eclipse
- Creating Hive tables to import large data sets from various relational databases using Sqoop and export the analyzed data back for visualization and report generation by the BI team.
- Installing and configuring Hive, Sqoop, Flume, Oozie on the Hadoop clusters.
- Involved in scheduling Oozie workflow engine to run multiple Hive and Pig jobs.
- Developed a process for the Batch ingestion of CSV Files, Sqoop from different sources and also

generating views on the data source using Shell Scripting and Python. Integrated a shell script to create Collections/morphline, SolrIndexes on top of table directories using MapReduce Indexer Tool within Batch Ingestion Framework. Implemented partitioning, dynamic partitions and buckets in HIVE. Developed Hive Scripts to create the views and apply transformation logic in the Target Database. Involved in the design of Data Mart and Data Lake to provide faster insight into the Data. Involved in using Stream Sets Data Collector tool and created Data Flows for one of the streaming application. Experienced in using Kafka as a data pipeline between JMS (Producer) and Spark Streaming Application (Consumer) Involved in the development of Spark Streaming application for one of the data source using Scala, Spark by applying the transformations. Developed a script in Scala to read all the Parquet Tables in a Database and parse them as Json files, another script to parse them as structured tables in Hive. Designed and Maintained Oozie workflows to manage the flow of jobs in the cluster. Configured Zookeeper for Cluster co-ordination services. Developed a unit test script to read a Parquet file for testing PySpark on the cluster. Involved in exploration of new technologies like AWS, Apache Flink, and Apache NIFletc which can increase the business value. Environment: Hadoop, HDFS, Map Reduce, Hive, HBase, Zookeeper, Impala, Java(jdk1.6), Cloudera, Oracle, SQL Server, UNIX Shell Scripting, Flume, Oozie, Scala, Spark, Sqoop, Python, kafka, PySpark.

Sr. Hadoop Developer Intel - Santa Clara, CA January 2018 to September 2018

Description: The purpose of the project is to perform the analysis on the Effectiveness and validity of controls and to store terabytes of log information generated by the source providers as part of the analysis and extract meaningful information out of it. The solution is based on the open source Big Data software Hadoop. The data will be stored in Hadoop file system and processed using Map Reduce jobs, which intern includes getting the raw data, process the data to obtain controls and redesign/change history information, extract various reports out of the controls history and Export the information for further processing.

Key Achievements: Responsible for Writing MapReduce jobs to perform operations like copying data on HDFS and defining job flows on EC2 server, load and transform large sets of structured, semi-structured and unstructured data. Developed a process for Sqooing data from multiple

sources like SQL Server, Oracle and Teradata. Responsible for creation of mapping document from source fields to destination fields mapping. Developed a shell script to create staging, landing tables with the same schema like the source and generate the properties which are used by Oozie jobs. Developed Oozie workflow's for executing Sqoop and Hive actions. Worked with NoSQL databases like Hbase in creating Hbase tables to load large sets of semi structured data coming from various sources. Performance optimizations on Spark/Scala. Diagnose and resolve performance issues. Responsible for developing Python wrapper scripts which will extract specific date range using Sqoop by passing custom properties required for the workflow. Developed scripts to run Oozie workflows, capture the logs of all jobs that run on cluster and create a metadata table which specifies the execution times of each job. Developed Hive scripts for performing transformation logic and also loading the data from staging zone to final landing zone. Worked on Parquet File format to get a better storage and performance for publish tables. Involved in loading transactional data into HDFS using Flume for Fraud Analytics. Developed Python utility to validate HDFS tables with source tables. Designed and developed UDF'S to extend the functionality in both PIG and HIVE. Import and Export of data using Sqoop between MySQL to HDFS on regular basis. Responsible for developing multiple Kafka Producers and Consumers from scratch as per the software requirement specifications.. Involved in using CA7 tool to setup dependencies at each level (Table Data, File and Time). Automated all the jobs for pulling data from FTP server to load data into Hive tables using Oozie workflows. Involved in developing Spark code using Scala and Spark-SQL for faster testing and processing of data and exploring of optimizing it using Spark Context, Spark-SQL, Pair RDD's, Spark YARN. Migrating the needed data from Oracle, MySQL in to HDFS using Sqoop and importing various formats of flat files in to HDFS. Environment: Hadoop, HDFS, Map Reduce, Hive, HBase, Kafka, Zookeeper, Oozie, Impala, Java(jdk1.6), Cludera, Oracle, Teradata SQL Server, UNIX Shell Scripting, Flume, Scala, Spark, Sqoop, Python. Sr. Hadoop Developer Health partners - Bloomington, MN January 2016 to December 2017 Description: The purpose of this project is to acquire Core metrics data, export and process in Hadoop. The Core metrics data files to be loaded in to Hadoop data warehouse daily to enable business reporting and

outbound feeds. We use Hadoop to store vast volumes of unstructured data allows the company to collect web logs, transaction data and social media data. Key Achievements: Responsible for Managing, Analyzing and Transforming petabyte's of data and also quick validation check on FTP file arrival from S3 Bucket to HDFS. Responsible for analyzing large data sets and derive customer usage patterns by developing new MapReduce programs. Experienced in creation of Hive tables and loading data incrementally into the tables using Dynamic Partitioning and Worked on Avro Files, JSON Records. Experienced in using Pig for data cleansing and developed Pig Latin scripts to extract the data from web server output files to load into HDFS. Worked on Hive by creating external and internal tables, loading it with data and writing Hive queries. Involved in development and usage of UDTF's and UDAF's for decoding Log Record Fields and Conversion's, Generating Minute Buckets for the specified Time Interval's and JSON Field Extractor. Developed Pig and Hive UDF's to analyze the complex data to find specific user behavior. Responsible for Debug, Optimization of Hive Scripts and also implementing Deduplication Logic in Hive using a Rank Key Function (UDF). Experienced in writing Hive Validation Scripts which are used in validation framework (for daily analysis through graphs and presented to business users). Developed workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive. Involved for Cassandra Database Schema design. Using BULK LOAD Utility data pushed to Cassandra databases. Responsible for creating Dashboards on Tableau Server. Generated reports for hive tables in different scenarios using Tableau Responsible for Scheduling using Active Batchjobs and Cron jobs. Experienced in Jar builds that can be triggered by commits to Github using Jenkins. Exploring new tools for data tagging like Tealium (POC Report). Actively updated the upper management with daily updates on the progress of project that include the classification levels that were achieved on the data. Environment: Hadoop, Map Reduce, HDFS, Pig, Hive, HBase, Zookeeper, Oozie, Impala, Cassandra, Java(jdk1.6), Cloudera, Oracle 11g/10g, Windows NT, UNIX Shell Scripting, Tableau, Tealium. Sr. Java Developer Well Point INC - Cincinnati, OH January 2013 to December 2015 Description: Well Point Health Insurance portal provides ability for end customers/users to request a quote, provide quote details, purchase health

insurance and payment of insurance amount. This platform is built using a Struts framework in the presentation layer. The back end functionality is built using Oracle as data persistence layer and Hibernate as data access layer

Key Achievements: Responsible for understanding the scope of the project and requirements gathering

Used MapReduce to Index the large amount of data to easily access specific records.

Supported MapReduce Programs which are running on the cluster.

Developed MapReduce programs to perform data filtering for unstructured data.

Designed the application by implementing Struts Framework based on MVC Architecture.

Designed and developed the front end using JSP, HTML and JavaScript and JQuery.

Developed framework for data processing using Design patterns, Java, XML.

Implemented J2EE standards, MVC2 architecture using Struts Framework.

Implementing Servlets, JSP and Ajax to design the user interface.

Used JSP, Java Script, HTML5, and CSS for manipulating, validating, customizing, error messages to the User Interface.

Used the light weight container of the Spring Framework to provide architectural flexibility for Inversion of Controller (IOC).

Used SpringIOC for dependency injection to Hibernate and Spring Frameworks.

Designed and developed Session beans to implement the Business logic.

Developed EJB components that are deployed on Web logic Application Server.

Written unit tests using Junit Framework and Logging is done using Log4J Framework.

Used Html, CSS, JavaScript and JQuery to develop front end pages.

Designed and developed various configuration files for Hibernate mappings.

Designed and Developed SQL queries and Stored Procedures.

Used XML, XSLT, XPATH to extract data from Web Services output XML

Extensively used JavaScript, JQuery and AJAX for client-side validation.

Used ANT scripts to fetch, build, and deploy application to development environment.

Developed Web Services for sending and getting data from different applications using SOAP messages.

Actively involved in code reviews and bug fixing.

Applied CSS (Cascading style Sheets) for entire site for standardization of the site.

Offshore co-ordination and User acceptance testing support.

Environment: Java 5.0, Struts, Spring 2.0, Hibernate 3.2, WebLogic 7.0, Eclipse 3.3, Oracle 10g, Junit 4.2, Maven, Windows XP, J2EE, JSP, JDBC, Hibernate, spring, HTML, XMLCSS, JavaScript and JQuery.

Software Programmer Blue Pal Solutions Pvt. Ltd April 2011 to December 2012

Description: Created a home page for the use of internal employees, This home page enables employees to punch in and punch out so the activity tracking report is generated beginning of the week for past week and sent to employee's managers. Description Key Achievements: Involved in the analysis & design of the application using Rational Rose. Developed the various action classes to handle the requests and responses. Designed and created Java Objects, JSP pages, JSF, JavaBeans and Servlets to achieve various business functionalities and created validation methods using JavaScript and Backing Beans. Involved in writing client side validations using JavaScript, CSS. Involved in the design of the Referential Data Service module to interface with various databases using JDBC. Used Hibernate framework to persist the employee work hours to the database. Developed classes and interface with underlying web services layer. Prepared documentation and participated in preparing user's manual for the application. Prepared Use Cases, Business Process Models and Data flow diagrams, User Interface models. Gathered & analyzed requirements for EAuto, designed process flow diagrams. Defined business processes related to the project and provided technical direction to development workgroup. Analyzed the legacy and the Financial Data Warehouse. Participated in Data base design sessions, Database normalization meetings. Managed Change Request Management and Defect Management. Managed UAT testing and developed test strategies, test plans, reviewed QA test plans for appropriate test coverage. Involved in Developing JSP's, action classes, form beans, response beans, EJB's. Extensively used XML to code configuration files. Developed PL/SQL stored procedures, triggers. Performed functional, integration, system and validation testing.

Environment: Java, J2EE, JSP, JCL, DB2, Struts, SQL, PL/DSQL, Eclipse, Oracle, Windows XP, HTML, CSS, JavaScript, and XML. Education Bachelors in Computers Science in Computers Science JNTU Skills APACHE HADOOP HDFS (3 years), APACHE HADOOP IMPALA (3 years), APACHE HADOOP MAPREDUCE (6 years), APACHE HADOOP OOZIE (3 years), APACHE HBASE (3 years), databases (5 years), Hadoop (3 years), HADOOP DISTRIBUTED FILE SYSTEM (3 years), HBase (3 years), HTML (4 years), J2EE (4 years), JavaScript (4 years), JSP (4 years), MapReduce (6 years), Oracle (8 years), Servlets (4 years), SQL (6 years), Struts (4 years), Web

Services (4 years), XML. (4 years)

Name: Jordan Martinez

Email: figueroamatthew@example.org

Phone: (423)665-3075x7418