Hadoop/Spark Developer Hadoop/Spark Developer Hadoop/Spark Developer - Ford Southfield, MI

Talented and accomplished Software Engineer with 8 years of IT experience in developing applications using BigData, AWS, Java,SQL and Spark.   3+ years of experience with Big Data tools like MapReduce, YARN, HDFS, Hbase, Impala,Hive, Pig, Oozie,AWS, ,ApacheSpark for ingestion, storage, querying, processing and analysis of data.   Performance tuning in Hive&Impala using multiple methods limited to dynamic partitioning, bucketing, indexing, files compressions. Hands on experience withdata ingestion tools Kafka, Flume and workflow management tools Oozie and Zena.   Hands on experience handling different file formats like JSON, AVRO, ORC, Parquet and compression techniques like snappy, zlib and lzo.   Hands on experience in Hadoop Ecosystem components such as Hadoop, Spark, HDFS, YARN, TEZ, Hive, Sqoop, Flume, MapReduce, SCALA, Pig, OOZIE, Kafka, NIFI, Storm, HBASE.   Experience on analyzing data in NOSQL databases like Hbase and Cassandraand its Integration with Hadoop cluster.   Hands on experience with Spark Core, Spark SQL and Data Frames/Data Sets/RDD API.   Experience in using Kafka and Kafka brokers to initiate spark context and processing live streaming information with the help of RDD.   Developed Java applications using various IDE's like Spring Tool Suite and Eclipse.   Good knowledge in using Hibernate for mapping Java classes with database and using Hibernate Query Language (HQL).   Operated on Java/J2EE systems with different databases, which include Oracle, MySQL and DB2.   Knowledge on implementing Big Data in Amazon Elastic MapReduce (Amazon EMR) for processing, managing Hadoop framework dynamically scalable Amazon EC2 instances.   Capable of processing large sets of structured, semi-structured and unstructureddata and supporting systems application architecture.   Extensive development experience in sparkapplications for datatransformations and loading into HDFS using RDD, DataFrames and Datasets.   Extensive knowledge on performance tuning of Spark applications and converting Hive/SQL queries into Sparktransformations.   Hands-on experience with AWS (AmazonWebServices), using ElasticMapReduce (EMR), creating and storing data in S3buckets and creating ElasticLoadBalancers(ELB) for Hadoop front end WebUI's.   Extensive knowledge on creating Hadoop cluster on multiple EC2 instances in AWS and configuring them through ambari

and using IAM (Identity and AccessManagement) for creating groups, users and assigning permissions.   Extensive programming experience in JavaCore concepts like OOPS, Multithreading, Collections and IO.   Experience using Jira for ticketing issues and Jenkins for continuous integration.   Extensive experience with UNIX commands, shellscripting and setting up CRON jobs.

   Experience in software configuration management using Git.   Good experience in using Relational databases Oracle&MySQL.   Able to assess businessrules, collaborate with stakeholders and perform source-to-target datamapping, design.   Successfully working in fast-paced environment, both independently and in collaborative team environments. Work Experience Hadoop/Spark Developer Ford - Southfield, MI May 2018 to Present Responsibilities   Implemented Hive UDF's and did performance tuning for better results.   Analyzed the data by performing Hive queries and running Pig Scripts.   Implemented optimized map joins to get data from different sources to perform cleaning operations before applying the algorithms.   Monitoring/Maintaining day-to-day batch jobs using Event Engine.   Interacting with SOR team to fix production data issues.   Performing data transformation using Pig and Hive.   Performing Snapshot, full refresh and Load append data refresh.   Providing support for CMDL raw data and ODL ingestion, EFS Big Data applications   Experience in using Sqoop to import and export the data from Netezza and Oracle DB into HDFS and HIVE.   Involved in loading data from UNIX file system to HDFS. Handled importing of data from various data sources, performed transformations using Hive, MapReduce and loaded data into HDFS.   Providing weekly/monthly status reports to customer. Worked on analyzing data with Hive and Pig and real time analytical operations using H-base. Worked on loading and transforming of large sets of structured, semi structured and unstructured data.   Participated in all the meetings with the Database owners for the approvals and extraction of the data.   Retrieving data from external web APIs and indexing it into our distributed SolrCloud platform. Also, maintained and deployed the production and development SolrCloud setup using Ansible. Other responsibilities included building proofs of concept in C# for transforming incoming data into schema compatible with the existing SQL data as well as building monitoring scripts in Python.   Resolve missing fields in DataFrame rows using filtering and imputation.   Integrate

visualizations into a Spark application using Databricks and popular visualization libraries (ggplot, matplotlib). Faster processing and testing of data is achieved by implementing Spark SQL and Spark using Scala. Experienced on adding/installation of new components and removal of them through Ambari. Experience with data wrangling and creating workable datasets. Monitoring systems and services through Ambari dashboard to make the clusters available for the business. Responsible for building scalable distributed data solutions using Hadoop. Working as Hadoop Developer and admin in Hortonworks (HDP 2242) distribution for 10 clusters ranges from POC to PROD. Prepared the Business and high level design documents. Created and maintained Technical documentation for launching HADOOP Clusters and for executing Hive queries and Pig Scripts. Installed Oozie workflow engine to run multiple Hive and Pig jobs which run independently with time and data availability. Assisted application teams in installing Hadoop updates, operating system, patches and version upgrades when required. Implemented Map reduce secondary sorting to get better performance for sorting results in Map Reduce programs. Worked on User Defined Functions in Hive to load the data from HDFS to run aggregation function on multiple rows.

 Involved in moving all log files generated from various sources to HDFS for further processing through Flume. Coordinated with the testing team for bug fixes and created documentation for recorded data, agent usage and release cycle notes. Used Teradata system Priority Schedule in controlling the load of the system. Created different UDF's and UDAF's to analyze the partitioned and bucketed data and compute various metrics for reporting on the dashboard and stored them in different summary tables. Implemented POC to introduce Spark Transformations. Environment: Hadoop, HDFS, hive, Sqoop, Kafka, Spark, Scala, MapReduce, Cloudera, Kafka, Zookeeper, HBase, Shell Scripting, Python, Mongo DB, AWS, UNIX Shell Scripting. Hadoop Developer T-Mobile - Bellevue, WA January 2017 to April 2018 AWS with Spark) Responsibilities Worked on improving the performance of existing Pig and Hive Queries. Developed Oozie workflow engines to automate Hive and Pig jobs. Worked on performing Join operations. 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.   Performed various source data ingestions, cleansing, and transformation in Hadoop. Design and developed many Spark Programs using pyspark   Produce unit tests for Spark transformations and helper methods   Creating RDD's and Pair RDD's for Spark Programming. Implement Joins, Grouping and Aggregations for the Pair RDD's.   Write Scaladoc-style documentation with all code   Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.   Developed Pig Scripts to perform ETL procedures on the data in HDFS.   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 systems.   Advanced knowledge in performance troubleshooting and tuning Cassandra clusters.   Analyzing the source data to know the quality of data by using Talend Data Quality.   Created Scala/Spark jobs for data transformation and aggregation   Involved in creating Hive tables, loading with data and writing hive queries.   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.   Preparation of Technical architecture and Low-level design documents Tested raw data and executed performance scripts.   Environment: eclipse, jdk1.8.0, Hadoop2.8, HDFS, MapReduce,Spark 2.0 Pig0.15.0, Hive2.0, HBase, Apache-Maven3 Java/Hadoop Developer Charles Schwab - Austin, TX January 2015 to December 2016 Responsibilities   Implemented J2EEDesignPatterns like DAO, Singleton, and Factory.   Managed connectivity using JDBC for querying/inserting & data management including triggers and stored procedures.   Used Spring/MVC framework to enable the interactions between JSP/View layer and implemented different design patterns with J2EE and XML technology.   Worked on Python and build the custom ingest framework.   Implemented application using MVC architecture integrating Hibernate and spring frameworks.   Utilized various JavaScript and JQuery libraries Bootstrap, Ajax for form validation and other interactive features.   Extensively worked on Hadoop eco-systems including Hive, Spark Streaming with MapRdistribution.   Upgraded the Hadoop Cluster from CDH3 to CDH4, setting up High Availability Cluster and integrating Hive with existing applications.   Worked on storing data in HDFS either directly or through Hbase.   Worked on NoSQL support enterprise

production and loading data into HBase using Impala and Sqoop.   Performed multiple MapReduce jobs in Pig and Hive for data cleaning and pre-processing.   Build Hadoop solutions for big data problems using MR1 and MR2 in YARN.   Handled importing of data from various data sources, performed transformations using Hive, PIG, and loaded data into HDFS.   Worked on data using Sqoop from HDFS to Relational Database Systems and vice-versa. Maintaining and troubleshooting.   Exploring with Spark to improve the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, pair RDD's.   Created Hive Tables, loaded claims data from Oracle using Sqoop and loaded the processed data into target database.   Worked on Delete printer module using python.   Involved in PL/SQL query optimization to reduce the overall run time of stored procedures.   Exported data from HDFS to RDBMS via Sqoop for Business Intelligence, visualization and user report generation. Implemented the J2EE design patterns Data Access Object (DAO), Session Fa ade and Business Delegate.   Developed Nififlows dealing with various kinds of data formats such as XML, JSON and Avro.   Implemented MapReduce jobs in HIVE by querying the available data.   Proactively involved in ongoing maintenance, support and improvements in Hadoop cluster.   Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.   Used Cloudera Manager for installation and management of Hadoop Cluster.   Collaborated with business users/product owners/ developers to contribute to the analysis of functional requirements.   Involved in converting HiveQL into Spark transformations using Spark RDD and through Scala programming.   Integrated Kafka-Sparkstreaming for high efficiency throughput and reliability Worked in tuning Hive&Pig to improve performance and solved performance issues in both scripts. Environment: Hadoop 3.0, Python,Hive 2.1, J2EE, Hbase, JDBC, Pig 0.16, HBase 1.1, Sqoop, NoSQL, Impala, Java, Spring, MVC, XML, Spark 1.9, PL/SQL, HDFS, JSON, Hibernate, Bootstrap, JQuery, JavaScript, Ajax Java Developer Catalyte - Chicago, IL October 2013 to December 2014 Responsibilities   As a Java Developer involved in back-end and front-end developing team. Involved in the Software Development Life Cycle (SDLC) including Analysis, Design, Implementation   Responsible for use case diagrams, class diagrams and sequence diagrams using Rational Rose

in the Design phase.   Developed ANT scripts that checkout code from SVN repository, build EAR files.   Used XML Web Services using SOAP to transfer information to the supply chain and domain expertise Monitoring Systems.   Use Eclipse and Tomcat web server for developing & deploying the applications.   Developed REST Web Services clients to consume those Web Services as well other enterprise wide Web Services.   Used JavaScript and AJAXtechnologies for front end user input validations and Spring validation framework for backend validation for the User Interface.   Used both annotation based configuration and XML based.   Developed application service components and configured beans using (applicationContext.xml) Spring IOC.   Implemented persistence mechanism using Hibernate (ORM Mapping).   Developed the DAO layer for the application using Spring Hibernate Template support.   Used WebLogic workshop, Eclipse IDE to develop the application.   Performed the code build and deployment using Maven.   Implementation of Spring Restful web services which produces JSON.   Responsible for maintaining the code quality, coding and implementation standards by code reviews.   Developed the front end of the application using HTML, CSS, JSP and JavaScript.   Created RESTFULL APIs using Spring MVC.   Used SVN version controller to maintain the code versions.   Worked on web applications using open source MVCframeworks.   Developed Web interface using JSP, Standard Tag Libraries (JSTL), and SpringFramework.   Implemented logger for debugging and testing purposes using Log4j.   Environment: JSON, HTML 4, CSS, XML, Hibernate 3.6, Eclipse, Maven, JUnit, JDBC, ANT, SOAP, Log4j Java Developer Paychex - Rochester, NY October 2011 to September 2013 Responsibilities   Individually worked on all the stages of a SoftwareDevelopmentLifeCycle (SDLC).

   Responsible for design and implementation of various modules of the application using Struts-Spring-Hibernate architecture.   Created user-friendly GUI interface and Web pages using HTML, CSS and JSP.   Developed web components using MVC pattern under Struts framework. Wrote JSPs, Servlets and deployed them on Weblogic Application server.   Used JSP's, HTML on front end, Servlets as Front Controllers and JavaScript for client side validations.   Wrote the Hibernate-mapping XML files to define java classes-database tables mapping.   Developed the UI using JSP, HTML, CSS and AJAX and learned how to implement JQuery, JSP and client &server

validations using JavaScript.   Implemented MVC architecture by using spring to send and receive the data from front-end to business layer.   Designed, developed and maintained the data layer using JDBC and performed configuration of JavaApplication Framework.   Extensively used Hibernate in data access layer to access and update information in the database.   Migrated the Servlets to the Spring Controllers and developed Spring Interceptors, worked on JSPs, JSTL, and JSP Custom Tags.   Used Jenkins for continuous integration purpose in using SVN, JUnit and Mockito as version control and Unit testing by Creating design documents and test cases for development work.   Worked on Eclipse IDE for front end development environment for insertions, updating and retrieval operations of data from oracle database by writing stored procedures. Responsible for writing Struts action classes, Hibernate POJO classes and integrating Struts and Hibernate with spring for processing business needs.   Developed the application using Servlets and JSP for the presentation layer along with JavaScript for the client side validations.   Wrote Hibernate classes, DAO's to retrieve & store data, configured Hibernate files.   Used Web Logic for application deployment and Log4J used for Logging/debugging.   Used CVSversion controlling tool and project build tool using ANT.   Used various Core Java concepts such as multi-threading, Exception Handling, Collection APIs to implement various features and enhancements.   Wrote and debugged the MavenScripts for building the entire web application.   Designed and developed Ajax calls to populate screens parts on demand.   Environment: Struts, HTML, CSS, JSP, MVC, Hibernate, JSP, AJAX, JQuery, Java, Jenkins, ANT, Maven Education Bachelor's Skills Hdfs, Impala, Mapreduce, Oozie, Sqoop

Name: Patrick Dyer

Email: debrajenkins@example.org

Phone: 2126894575