Sr Hadoop Developer Sr Hadoop Developer Sr Hadoop Developer - Provish Consulting Atlanta, GA

Over 9+ years of professional IT experience that includes of BigData/Hadoop and of web and Windows application development with .net and Java. Hands on experience with the Hadoop stack (MapReduce, HDFS, Sqoop, Pig, Hive, HBase, SPARK, Kafka,Control-M, Oozie , ZooKeeper and Talend). Have experience in configuring and administrating the Hadoop Cluster using major Hadoop Distributions like Hortonworks and Cloudera. Excellent experience in developing different components using Apache Hadoop ecosystem components like, Map Reduce, Hive, HBase, PIG, Sqoop, Spark, Kafka, Flume, Zookeeper, Oozie, and Storm. Expertise in depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, MRv1 andMRv2 (YARN). Experienced with data architecture including data ingestion pipeline design, Hadoop information architecture, data modeling and data mining, machine learning and advanced data processing. Expertise in writing Apache Spark streaming API on Big Data distribution in the active cluster environment. Experienced on implementation of a log producer in Scala that watches for application logs, transform incremental log and sends them to a Kafka and Zookeeper based log collection platform. Experienced in working with Flume to load the log data from multiple sources directly intoHDFS. Excellent knowledge in building and scheduling Big Data workflows with the help of OOZIEand Auto-sys. Experienced in importing and exporting data from the different Data sources like (Teradata and DB2) using Sqoop from HDFS to Relational Database Systems (RDBMS) and vice-versa and load into partitioned Hive tables. Proven Expertise in performing analytics on Big Data using Map Reduce, Hive, Pig and Talend. Experienced with performing real time analytics on NoSQL databases like HBase and Cassandra. Experienced with ETL to load data into Hadoop/NoSQL. Experienced with Dimensional modeling, Data migration, Data Masking, Data cleansing, Data profiling, and ETL Processes features for data warehouses. Experience in integration of various data sources like SQL Server, Oracle, Vertica, MySQL, Flat files. Authorized to work in the US for any employer Work Experience Sr Hadoop Developer Provish Consulting - Charlotte, NC September 2018 to Present Responsibilities: Worked on WEB and Windows

Application development. Worked on the application which talks to different technological platforms such as Share point , SAP , Oracle and SQL. Worked with BI teams in generating the reports and designing ETL workflows on SAP BI. Worked on 3-Tier architecture of application development. Used OOPS concepts in application development. Worked as co-Ordinator between On-shore and Off-shore teams. Worked closely with business and gather the requirements. Worked on importing data from various sources and performed transformations using MapReduce, Hive to load data into HDFS. Involved in complete implementation lifecycle, specialized in writing custom MapReduce, Pig and Hive programs. Handled importing of data from RDBMS into HDFS using Sqoop. Managing data flow into Pivotal HAWQ (Internal / External tables). Experienced in data cleansing processing using Pig Latin operations and UDFs. Experienced in writing Hive Queries for analyzing data in Hive warehouse using Hive Query Language. Involved in creating Hive tables, loading with data and writing hive queries to process the data. Created scripts to automate the process of Data Ingestion. Developed PIG scripts for source data validation and transformation. Performed various performance optimizations like using distributed cache for small datasets, Partitioning, Bucketing in Hive and Map Side Joins. Managing and scheduling jobs to remove the duplicate log data files in HDFS using Oozie. Extensively used Hive/HQL or Hive queries to query or search for particular string in Hive tables in HDFS. Experience in developing customized UDF's in java to extend Hive and Pig Latin functionality. Created HBase tables to store various data formats for data coming from different portfolios. Environment: Horton works, Map Reduce, HBase, HDFS, Hive, Pig, Java, SQL, Cloudera Manager, Sqoop, Flume, Oozie. Sr Hadoop Developer Trimble Navigation Limited - Chantilly, VA May 2016 to August 2018 Responsibilities: Worked as a part R&D team in Zeta, experimenting with emerging technologies. Involved in Installing, Configuring Hadoop Eco System, Cloudera Manager using CDH5.4 and Horton works Distribution. Worked on POC to upgrade the tradition relational data process to Hadoop. Used Talend as ETL tool with Hadoop and other database components. Used Control-M was workflow job scheduler for all the bigdata Talend jobs. Played a key role in installation and configuration of the various Hadoop ecosystem tools such asSolr, Kafka, Pig, HBase

and Cassandra.   Implemented multiple Map Reduce Jobs in java for data cleansing and pre-processing.   Wrote complex Hive queries and UDFs in Java and Python.   Involved in implementing an HDInsight version 3.3 clusters, which is based on spark version 1.5.1. Responsible for data extraction and data ingestion from different data sources into Hadoop Data Lake by creating ETL pipelines using Pig, and Hive   Job duties involved the design, development of various modules in Hadoop Big Data Platform and processing data using Map Reduce, Hive, Pig, Sqoop and Oozie.   Design, developed and tested Map Reduce programs on Mobile Offers Redemptions and Send it to the downstream applications like HAVI.   Extract, transform, and load (ETL) data from multiple federated data sources (JSON, relational database, etc.) with DataFrames in Spark.   Used Sqoop to import data from Vertica and Oracle.   Worked extensively on Talend , Pig and Hive for ETL processing.   Responsible for importing log files from various sources into HDFS using Kafka.   Used Hbase and Cassandra for data store for Microstrategy and Jasper IReport BI reporting.   Did Experiment with HAWQ to speed up the operation between microstrategy and Hbase.   Optimizing the Hive queries using Partitioning and Bucketing techniques, for controlling the data.   Developed Unit test cases using Junit testing frameworks.   Used Importtsv to create dynamin columns in Hbase.   Experienced in Monitoring Cluster using Cloudera Manager and Ambari   Environment: Hadoop, HDFS, HBase, Talend, MapReduce, Kafka, Java, Hive, Pig, Sqoop, Oozie, SQL, ETL, Cloudera Manager, Ambari, MySQL, Oracle, Vertica. Senior Software Engineer LendingTree - Charlotte, NC June 2012 to April 2016 Responsibilities: Responsible to manage data coming from different sources and involved in HDFS maintenance and loading of structured and unstructured data.   Integrated Quartz scheduler with Oozie work flows to get data from multiple data sources parallel using fork.   Processed Multiple Data sources input to same Reducer using Generic Writable and Multi Input format.   Created Data Pipeline of Map Reduce programs using Chained Mappers.   Visualize the HDFS data to customer using BI tool with the help of Hive ODBC Driver.   Worked on Big Data processing of clinical and non-clinical data using MapReduce.   Implemented complex MapReduce programs to perform joins on the Map side using Distributed Cache in Java.   Responsible for importing log files from various sources into HDFS

using Flume.   Created customized BI tool for manager team that perform Query analytics using HiveQL.   Used Hive and Pig to generate BI reports.   Imported data using Sqoop to load data from MySQL to HDFS on regular basis.   Created Partitions, Buckets based on State to further process using Bucket based Hive joins.   Created Hive Generic UDF's to process business logic that varies based on policy.   Moved Relational Data base data using Sqoop into Hive Dynamic partition tables using staging tables.   Optimizing the Hive queries using Partitioning and Bucketing techniques, for controlling the data.   Worked on custom Pig Loaders and storage classes to work with variety of data formats such as JSON and XML file formats.   Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive, Pig, and Sqoop.   Developed Unit test cases using Junit testing frameworks. Experienced in Monitoring Cluster using Cloudera Manager. Software Engineer Matlen Silver - Charlotte, NC February 2010 to May 2012 Responsibilities: Used Rational Rose for Use Case Diagram, Class Diagrams, Sequence diagrams and Object diagrams in design phase.   Involved in creation of UML diagrams like Class, Activity, and Sequence Diagrams using modeling tools of IBM Rational Rose.   Involved in the full life cycle development of the modules for the project.   Created EC2 instances and implemented large multi node Hadoop clusters in AWS cloud from scratch. Configured AWS IAM and Security Groups.   Responsible for implementing Kerberos, creating service principals, user accounts, keytabs, & syncing with AD.   Developed terraform template to deploy Cloudera Manager on AWS.   Configured different Notifications on AWS Services. Installed, configured Hadoop Cluster using Puppet.   MR2 Batch job was written to fetch required data from DB and store the same in CSV (static file)   Spark job to process the files from Vision EMS and AMN Cache to identify the violations and sending the same to Smarts as SNMP traps. Automated workflows using shell scripting to schedule(crontab) Spark jobs.   Installed Oozie workflow engine to run multiple Map Reduce, Hive HQL and Pig jobs.   Implemented best income logic using Pig scripts and UDFs.   Used Eclipse and Visual Studio IDE for application development.   Used Spring and .Net framework for dependency injection.   Worked with back end database such as Oracle and MS SQL.   Developed Web application and services using C# and ASP .net.

Used Struts (MVC) for developing presentation layer. Used IIS application server for deploying applications. Used SOAP XML , WCF Web services for transferring data between different applications. Used MVC design pattern for designing application. Persistence layer was implemented using Hibernate Framework. Integrated Hibernate with Spring framework. Worked with complex SQL queries, SQL Joins and Stored Procedures using TOAD for data retrieval and update. Used JUnit and NUnit for performing Unit Testing. Used Log4J to capture the logs that included runtime exceptions. Environment: Eclipse,Microsoft Visual Studio, Web Services, UML, MVC, NHibernate , JSP, WSDL, JMS, AJAX, JavaScript, Junit,NUnit, PL/SQL, Oracle 10G, SVN, TFS Education Bachelor's Skills Apache hadoop oozie (9 years), Hadoop (9 years), Hive (9 years), Mapreduce (9 years), Oozie. (9 years) Additional Information Technical Skill: Hadoop Technologies Apache Hadoop, Cloud era Hadoop Distribution (HDFS and Map Reduce) Technologies HDFS, YARN, Map Reduce, Hive, Pig, Sqoop, Flume, Spark, Kafka, Zookeeper, and Oozie, CDH 4, CDH 5 & HDP 2.4.2 Hadoop Ecosystem Hive, Pig, Sqoop, Flume, Zookeeper, Oozie. Streaming Technologies Spark, Kafka, Storm. AWS S3, EC2. Java/J2EE Technologies Core Java, Data Structures, Multithreading. NOSQL Databases Hbase, Cassandra, MongoDB Programming Languages Java, Linux shell scripting, Scala, Python. Web Technologies HTML, CSS, JavaScript, AJAX, JSP, DOM, XML Databases MySQL, SQL, Oracle, SQL Server, DB2, PL/SQL. Application Servers Web Logic, Web Sphere, JBoss. Software Engineering Scrum, Agile methodologies ETL Talend. Operating Systems Windows MAC OS, UNIX, LINUX. IDE Tools Eclipse, IntelliJ IDEA

Name: James Mosley

Email: darrellkane@example.net

Phone: (788)384-0871