

Senior Hadoop Developer Senior Hadoop Developer Senior Hadoop Developer - MoneyGram International, Location Dallas, TX 8+ years of experience in IT, which includes experience in Hadoop ecosystem, Bigdata Technologies, QA Automation Tools, SQL related technologies in Retail, Manufacturing, Financial and Communication sectors 4+ Years of experience as Hadoop Developer using Various Hadoop eco-systems tools and Spark Framework and Currently working on Spark and Spark Streaming frameworks extensively using Scala as the main programming dialect Experience installing/configuring/maintaining Apache Hadoop clusters for application development and Hadoop tools like Sqoop, Hive, PIG, Flume, Hbase, Kafka, Hue, Storm, Zoo Keeper, Oozie, MangoDB, Sqoop, Python Worked with major distributions like Cloudera (CDH 3&4) & Horton works Distributions and AWS. Also worked on Unix and DWH in support for various Distributions Hands on experience in developing and deploying enterprise based applications using major components in Hadoop ecosystem like Hadoop 2.X, YARN, Hive, Pig, MapReduce, Spark, Kafka, Storm, Oozie, HBase, Flume, Sqoop and ZooKeeper Experience in handling large datasets using Partitions, Spark in memory capabilities, Broadcasts in Spark with Scala and python, Effective and efficient Joins, Transformations and other during ingestion process itself Experience in developing data pipeline using Pig, Sqoop, and Flume to extract the data from weblogs and store in HDFS and accomplished developing Pig Latin Scripts and using HiveQL for data analytics Extensively dealt with Spark Streaming and Apache Kafka to fetch live stream data. Experience in converting Hive/SQL queries into Spark transformations using Java and experience in ETL development using Kafka, Flume and Sqoop Good experience in writing Spark applications using Scala and Java and used Scala sbt to develop Scala projects and executed using Spark-Submit Experience working on NoSQL databases including Hbase, MangoDB and experience using Sqoop to import data into HDFS from RDBMS and vice-versa Developed Spark scripts by using Scala shell commands as per the requirement Good experience in writing Sqoop queries for transferring bulk data between Apache Hadoop and structured data stores Substantial experience in writing Map Reduce jobs in Java, PIG, Flume, Zookeeper, Hive and Storm Created multiple MapReduce Jobs using Java API, Pig and Hive for data extraction Strong expertise in troubleshooting and

performance fine-tuning Spark, MapReduce and Hive applications Good experience on working with Amazon EMR framework for processing data on EMR and EC2 instances Created AWS VPC network for the installed Instances and configured security groups and Elastic IP's Accordingly Developed AWS Cloud formation templates to create custom sized VPC, subnets, EC2 instances, ELB and security groups Extensive experience in developing applications that perform Data Processing tasks using Teradata, Oracle, SQL Server and MySQL database Worked on data warehousing and ETL tools like Informatica, Tableau, and Pentaho Experience in understanding the security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure Acquaintance with Agile and Waterfall methodologies. Responsible for handling several clients facing meetings with great communication skills

Work Experience Senior Hadoop Developer MoneyGram International, Location - Dallas, TX May 2018 to Present Responsibilities:

Worked on analysing Hadoop cluster and different big data analytical and processing tools including Pig, Hive, Sqoop and Spark with Scala & java, Spark Streaming Wrote Spark-Streaming applications to consume the data from Kafka topics and wrote processed streams to HBase and steamed data using Spark with Kafka Worked on the large-scale Hadoop YARN cluster for distributed data processing and analysis using Spark, Hive, and HBase Involved in creating data-lake by extracting customer's data from various data sources to HDFS which include data from Excel, databases, and log data from servers Developed Apache Spark applications by using Scala for data processing from various streaming sources Used Scala to convert Hive/SQL queries into RDD transformations in Apache Spark Implemented Spark solutions to generate reports, fetch and load data in Hive Experienced in writing real-time processing and core jobs using Spark Streaming with Kafka as a data pipeline system Written HiveQL to analyse the number of unique visitors and their visit information such as views, most visited pages, etc Configured Spark Streaming to receive real time data from the Apache Kafka and store the stream data to HDFS using Scala Monitored workload, job performance and capacity planning using Cloudera Manager Experienced on working with Amazon EMR framework for processing data on EMR and EC2 instances Designing and implementing complete end-to-end Hadoop Infrastructure including Pig, Hive, Sqoop,

Oozie, Flume, and Zookeeper Further used pig to do transformations, event joins, elephant bird API and pre -aggregations performed before loading JSON files format onto HDFS Involved in resolving performance issues in Pig and Hive with understanding of Map Reduce physical plan execution and using debugging commands to run code in optimized way Used Spark to perform analytics on data in Hive and experienced with ETL working with Hive and Map-Reduce Environment(s): Hdp 2.6.0, HDFS, MapReduce, Spark Streaming, Spark-Core, Spark SQL, Scala, Pig 0.14, Hive 1.2.1, Sqoop 1.4.4, Flume 1.6.0, Kafka, JSON, HBase. Hadoop Developer New York Times - New York, NY August 2016 to May 2018 Responsibilities: Installed and configured Hadoop MapReduce, HDFS and developed multiple MapReduce jobs in Java for data cleansing and prep Worked on Spark streaming to collect TB's of data for every hour from connected cars Worked on Spark Batch Processing to load the processed data into MongoDB Manage migration of on-perm servers to AWS by creating golden images for upload and deployment Manage multiple AWS accounts with multiple VPC's for both production and non-production where primary objectives are automation, build out, integration and cost control Moving log data periodically into HDFS using Flume and building multi-hop flows, fan-out flows, and failover mechanism Developed MapReduce jobs to automate transfer of data from Hbase and to read data files and scrub the data Experience in Spark programming using Scala. Experienced in performing ETL using Spark, Spark SQL Transferring data between MySQL and HDFS using Sqoop with connectors Creating and populating Hive tables and writing Hive queries for data analysis to meet the business requirements Transformed Kafka loaded data using Spark-streaming with Scala Installed and configured Pig and written Pig Latin scripts Migrating data from MySQL database to HBase. Running MapReduce jobs to access HBase data from application using Java Client API's Installed and configured Hive, Pig, Sqoop, Flume and Oozie on the Hadoop cluster and automating the jobs using Oozie Actively participated in software development lifecycle, including design and code reviews, test development, test automation Involved in solution-driven agile development methodology and actively participated in daily scrum meetings Monitoring Hadoop cluster using tools like Cloudera Manager Automation script to monitor HDFS and HBase through Cron jobs

Create a complete processing engine, based on Cloudera's distribution, enhanced to performance

Environment(s): Hadoop, MapReduce, HDFS, Sqoop, Hbase, Oozie, SQL, Pig, Flume, Hive, Java.

Big Data Developer Wells Fargo, Location - Dallas, TX December 2014 to July 2016

Responsibilities: Installed, configured and job creation in Hadoop Map-Reduce, Pig, Hive, HBase, Spark RDD, Pair RDD, Flume, Oozie, Sqoop environment. Involved in application migration from Hadoop to Spark for the fast processing. Extracted data from Oracle database into HDFS using Sqoop. Developed Oozie workflows to schedule and manage Sqoop, Hive, Pig jobs to Extract-Transform-Load process. Used Flume, and configured it to use multiplexing, replicating, multi-source, interceptors, selectors to import log files from Web Servers in to HDFS/Hive. Managed and scheduled Jobs on a Hadoop cluster using Shell Scripts. Maintained Cluster co-ordination services through Zookeeper for system. Involved in filter the partition data based on different year range different format using Hive functions. Defining a schema, creating new relations, performing Pig-Join, sorting and filtering using Pig-Group on large data sets. Performed Map-Side joins and Reduce-Side joins for large tables. Involved in filtering the partition data based on different year range different format using with Hive functions. Creating HBase tables for random read/writes by the map reduce programs. Designed and developed entire pipeline from data ingestion to reporting tables. Performed data cleaning, integration, transformation, reduction by developing Map-Reduce jobs in java for data mining. Creating Hive tables, loading data into it and customizing hive queries, internally operating in Map-Reduce way. Performed Map-Side joins and Reduce-Side joins for large tables. Used Cloudera Manager to monitor and manage Hadoop Cluster

Environment(s): HDFS, CDH, BigInsights, Apache Spark, Flume, Hive, Pig, Scala, Java, Sqoop, SQL, Perl, Shell scripting, C, C++, Java, Oracle, WebSphere Application Server, Spring, Hibernate, Struts, JMS

SQL/Java Developer Staples, Location - Framingham, MA September 2012 to November 2014

Responsibilities: Responsible for the analysis, documenting the requirements and architecting the application based on J2EE standards. Attended Scrum meetings daily as a part of Agile Methodology. Involved in complete Software Development Life Cycle (SDLC) with Object Oriented Approach of client's business process and continuous client feedback.

Implementing MVC Architecture using Spring Framework, customized user interfaces. Used Core Java, and Spring Aspect Oriented programming concepts for logging, security, error handling mechanism. Developed application modules using Spring MVC, Spring Annotations, Spring Beans, Dependency Injection, with database interface using Hibernate. Used the Java Collections API extensively in the application as security protection for XML, SOAP, REST and JSON to make a secure Web Deployment. Developed server-side services using Java, Spring, Web Services (SOAP, Restful, WSDL, JAXB, JAX-RPC). Built Web pages that are more user-interactive using JQuery plugins for Drag and Drop, AutoComplete, AJAX, JSON, Angular JS, JavaScript and Bootstrap. Used XSL to transform XML data structure into HTML pages. Used Struts as the framework in this project and developed struts action classes, form beans. Created dispatch Action classes, and Validation plug-in using Struts framework. DB2 was used as the database and wrote queries to extract data from the database. Developed SQL queries and stored procedures. Designed Developed white-box test cases using JUnit, Git, JMeter, Mockito Framework.

Environment: Core Java, Agile, Scrum, XML, HTML, Jmeter, SOAP, REST, JDK, JSP, Servlets, JDBC, HTML, CSS, JUnit, SQL, MySQL, Windows, Oracle, Eclipse Python Developer Autodesk - San Rafael, CA February 2011 to August 2012 Responsibilities: Involved in the analysis and development of Software Development Life Cycle (SDLC). Contributed in developing a web services middle-tier in Python to integrate with an existing MySQL backend. Wrote several internal API utilities and micro services to carry out specific tasks. Wrote and executed exhaustive SQL queries using Python with help from various query builders in Python to frame the queries. Migrated MySQL to NoSQL data store using ETL processes using Python. Tasks included CRUD, elaborate cleansing of data and packing it into the expected format - JSON. Developed Python scripts to perform auditing tasks and generating Excel reports to support engineering a logistics, sales and inventory management system. Worked as assistant to Web Programmer in creating Django/Flask Web apps. Generated sample JSON format queries for testing REST endpoints. Frequently performed ad hoc File IO tasks in python. Dealt with CSV, JSON, Text, XML and XLSX files. Used shell scripting to automate repetitive tasks. Followed a test-driven approach closely.

Used Python's Unit Testing extensively. Environment: Python, MySQL, PostgreSQL, Jira, Flask, REST, JSON, CSV, Excel, Eclipse Education Bachelor's Skills HDFS, IMPALA, MAPREDUCE, OOZIE, SQOOP

Name: Diana Mckenzie

Email: brandy76@example.org

Phone: (470)588-7081x1027