

Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - CVS Health Los Angeles, CA

Around 8+ years of IT experience in a variety of industries, which includes 4+ years of working as a Hadoop developer designing and implementing complete end-to-end Hadoop infrastructure using MapReduce, Pig, Hive, Sqoop, Oozie, Flume, Spark, HBase and Zookeeper. Excellent knowledge of Hadoop Architecture and its various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MapReduce Programming paradigm. Hands on expertise with SQL, PL/SQL, UNIX shell scripts and commands. Experience in creating Hive tables and queries using HiveQL, with a good understanding of Hive concepts such as Partitioning, Bucketing and Joins. Developed Pig Latin scripts to extract and load the data into HDFS. Experience in migrating the data using Sqoop from Relational Database Systems to HDFS and vice-versa. Experience in loading unstructured data (Log files, Xml data) into HDFS using Flume. Hands on expertise with different file formats like JSON, XML, CSV etc. Good exposure on setting up job streaming and scheduling with Oozie and working on messaging system such as Kafka integrated with Zookeeper.

Experience in monitoring Hadoop clusters on VM, Horton Works Data Platform 2.1 and 2.2, CDH5 Cloudera Manager, HDP on Linux. Using Kafka HDFS connector to export data from Kafka topics to HDFS files in a variety of formats and integrate with Apache Hive to make data immediately available for querying with Hive QL. Good exposure to NoSQL Databases and hands on work experience in using HBase Experience in using Oozie workflow scheduler to manage Hadoop jobs by Directed Acyclic Graph (DAG) of actions with control flows. Experience in designing and developing applications in Spark using Python to compare the performance of Spark with Hive and SQL/Oracle. Flexible with Unix/Linux and Windows Environments working with Operating Systems like Cent OS 5/6, Ubuntu 13/14. Experience in manipulating the streaming data to clusters through Kafka and Apache Spark-Streaming. Experience in Continuous Integration and Continuous Deployment by the tools like Jenkins. Learning to use Amazon AWS EMR and EC2 for cloud big data processing. Experience in all phases of Software Development Life Cycle including analysis, design, implementation, integration, deployment, testing using different software methodologies like Agile, Scrum, Waterfall models. Technically well versed in designing and developing business

solutions in the field of Banking, Insurance, Ecommerce, Manufacturing divisions. Experienced with the Apache Spark improving the performance and optimization of the existing algorithms in Hadoop using Apache Spark Context, Apache Spark-SQL, Data Frame, Pair RDD's, Apache Spark YARN. Good knowledge on Hadoop MRV1 and Hadoop MRV2 (or) YARN Architecture. An effective team player with excellent communication, analytical and interpersonal skills with exceptional planning and execution skills with a systematic approach and quick adaptability. Experience with databases such as Oracle 9i, PostgreSQL, MySQL Server with cluster setup and writing the SQL queries Triggers & Stored Procedures. Authorized to work in the US for any employer Work Experience Sr. Hadoop Developer CVS Health - Northbrook, IL January 2019 to Present Responsibilities: Designed and implemented scalable infrastructure and platform for large amounts of data ingestion, aggregation and integration in Hadoop, MapReduce, Flume, Kafka, Spark, Hive. Loaded large sets of structured, semi-structured, and unstructured data using Sqoop, Flume, Kafka. Written Sqoop scripts to import, export and update the data between HDFS and Relational databases. Created Flume configure files to collect, aggregate and store the web log data into HDFS. Involved in designing and creating data model to load customer data into a NoSQL database like HBase. For data enrichment, we perform a lookup over the table present in HBase to retrieve the data. After enriching the data, we validate it and analyze the data by performing bucketing and partitioning in Hive. Developed Pig Latin scripts to transform the data and load into HDFS. Import data from different sources like HDFS/HBase into Spark RDD. Utilized Kafka to capture and process real time streaming data. Worked with Oozie and Zookeeper to manage job workflow and job coordination in the cluster. Worked closely with the business analysts to convert the Business Requirements into Technical Requirements and prepared low and high-level documentation. Played a key role in discussing about the requirements, analysis of the entire system along with development and testing. Exported data from HDFS into RDBMS using Sqoop to generate reports. Involved in diagnosing different possible ways to optimize and improve the efficiency of the system. Participated in peer-reviews of solution designs and related code. Developed workflow scheduler job scripts in Apache Oozie to extract the outcomes on a day-to-day

basis. Environment: Hadoop2.6.3, MapReduce, HDFS, Yarn, Sqoop1.4.3, Oozie, Pig0.11, Hive1.1.0, HBase0.98, Spark2.x, Java, Eclipse, UNIX shell scripting, python3.5.1, Cloudera manager 5.9.X, Flume, Kafka. Hadoop/ Java Developer Grand Bank of Texas - Dallas, TX July 2016 to November 2018 Responsibilities: Grand Bank of Texas is a locally-owned and operated independent community bank to serve the financial needs of the community and area. Roles and Responsibilities: Experience with the Hadoop ecosystem (MapReduce, Pig, Hive, HBase) and NoSQL. Analyze and determine the relationship of input keys to output keys in terms of both type and number, identify the number, type, and value of emitted keys and values from the Mappers, Reducers and the number and contents of the output files. Developed MapReduce pipeline jobs to process the data and create necessary HFiles and loading the HFiles into HBase for faster access without taking performance hit. Designed & developed UI Screens with Spring (MVC), HTML5, CSS, JavaScript, AngularJS to provide interactive screens to display data. Written multiple MapReduce programs in java for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV and other file formats. Analyze to determine the correct InputFormat, OutputFormat on order of MapReduce job requirements. Created database tables and wrote T-SQL Queries and stored procedures to create complex join tables and to perform CRUD operations. By using AWS, MapReduce job processed the data stored in the AWS. Using Automated Build and continuous integration systems Jenkins and test-driven Unit Testing framework Junit. Access control to users depending on logins using HTML, jQuery for validations. Created the web application using HTML, CSS, jQuery and JavaScript. Used Eclipse as an IDE for developing the application. Loaded the flat files data using Informatica to the staging area. Used UI-router in AngularJS to make this a single page application. Developed unit/assembly test cases and UNIX shell scripts to run along with daily/weekly/monthly batches to reduce or eliminate manual testing effort. Developed mappings in Informatica to load the data including facts and dimensions from various sources into the Data Warehouse, using different transformations like Source Qualifier, Java, Expression, Lookup, Aggregate, Update Strategy and Joiner. Environment: Windows XP/NT, Java, MapReduce, Pig, Hive, Hbase, NoSQL, AWS, Jenkins, HTML, CSS, T-SQL, AngularJS, UI,

jQuery, Korn Shell, Quality Center 10. Hadoop Developer East- West Bank - Pasadena, CA March 2014 to June 2016 Responsibilities: Installed and configured Apache Hadoop to test the maintenance of log files in Hadoop cluster. Installed and configured Hive, Pig, Sqoop, Flume and Oozie on the Hadoop cluster. Developed Sqoop jobs to import and store massive volumes of data in HDFS and Hive. Designed and developed PIG data transformation scripts to work against unstructured data from various data points and created a base line. Setup and benchmarked Hadoop /HBase clusters for internal use. Developed Java MapReduce programs for the analysis of sample log file stored in cluster. Developed Map Reduce Programs for data analysis and data cleaning. Developed PIG UDF's (like UTAF's & UDAF's) for manipulating the data according to business requirements and worked on developing custom PIG Loaders. Implemented data pipeline by chaining multiple mappers by using Chained Mapper. Used Hive and created Hive tables and involved in data loading and writing Hive UDFs. Migration of processes from Oracle to Hive to test the easy data manipulation. Configured deployed and maintained multi-node Dev and Test Kafka Clusters. Implemented data injection systems by creating Kafka brokers, producers, Consumers, custom encoders. Implemented Partitioning, Dynamic Partitions and Bucketing in Hive for efficient data access. Developed a data pipeline using Kafka and Strom to store data into HDFS. Developed some utility helper classes to get data from HBase tables. Good experience in troubleshooting performance issues and tuning Hadoop cluster. Built components, modules and plugins using Angular JS and Bootstrap. Created Java Interfaces and Abstract classes for different functionalities. Loaded and transformed large sets of structured, semi-structured and unstructured data in various formats like text, sequence, XML, and JSON. Written multiple MapReduce programs to power data for extraction, transformation, and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats. Environment: Apache Hadoop, HDFS, Cloudera Manager, Java, MapReduce, Eclipse, Hive, PIG, Sqoop, Oozie, SQL, Zookeeper, CDH3, Cassandra, Oracle, NoSQL and Unix/Linux, Kafka. Big Data Engineer Caesars - Las Vegas, NV August 2012 to February 2014 Responsibilities: Responsible for creating sample Datasets required for testing various Map Reduce applications from various sources. Developed Hive UDF

to parse the staged raw data to get the item details from a specific store. Built re-usable Hive UDF libraries for business requirements, which enabled users to use these UDF's in Hive querying. Designed workflow by scheduling Hive processes for Log file data, which is streamed into HDFS using Flume. Developed SQL scripts to compare all the records for every field and table at each phase of the data movement process from the original source system to the final target. Implemented map reduce jobs to process standard data in Hadoop cluster. Involved in the performance enhancement by analyzing the workflows, joins, configuration parameters etc. Collaborating with business users/product owners/ developers to contribute to the analysis of functional requirements. Design & Develop workflow using Oozie for business requirements, which includes automating the extraction of data from MySQL database into HDFS using Sqoop scripts. Worked on migration of Informatica Power center to Hadoop eco system. Assisted with the admin department with issues related to migration. Performed data analytics in Hive and then exported this metrics back to Oracle Database using Sqoop. Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig. Developed Map Reduce Programs in Java for applying business rules on the data. Implemented Partitioning, Dynamic Partitions, Bucketing in Hive. Document and manage failure/recovery steps for any production issues. Involved in Minor and Major Release work activities. Proactively involved in ongoing maintenance, support and improvements in Hadoop cluster. Import the data from various sources like HDFS/HBase into Kafka. Wrote MapReduce jobs using java and Pig Latin. Worked on NoSQL databases including HBase and Cassandra. Configured SQL Database to store Hive Teradata. Participated in development/implementation of Cloudera impala Hadoop environment Imported data using Sqoop to load data from MySQL to HDFS on regular basis. Used Pig as ETL (Informatica) tool to do Transformations, even joins and some pre-aggregations before storing the data onto HDFS. Developed Application components and API's using Scala. Created ETL (Informatica) jobs to generate and distribute reports from MySQL database Involved in loading data from LINUX file system to HDFS using Sqoop and exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the Business intelligence (BI) team. Installed

and configured Flume, Hive, Pig, Sqoop and Oozie on the Hadoop cluster. Extracted the Teradata from Oracle into Hive using the Sqoop. Worked on Agile Methodology. Environment: Hadoop, HDFS, Pig, Zookeeper, Sqoop, HBase, Shell Scripting, Ubuntu, Linux Red Hat, Kafka, Cassandra.

Senior Hadoop Developer AFLAC - Columbus, GA March 2011 to July 2012 Responsibilities:

Worked closely with the business analysts to convert the Business Requirements into Technical Requirements and preparing low and high-level documentation Performing transformations using Hive, MapReduce, hands on experience in copying .log, snappy files into HDFS from Greenplum using Flume & Kafka, loaded data into HDFS and extracted the data into HDFS from MYSQL using Sqoop Involved in preparing the S2TM document as per the business requirement and worked with Source system SME's in understanding the source data behavior Imported required tables from RDBMS to HDFS using Sqoop and used Storm/ Spark streaming and Kafka to get real time streaming of data into HBase Experience in Writing Map Reduce jobs for text mining and worked with predictive analysis team and Experience in working with Hadoop components such as HBase, Spark, Yarn, Kafka, Zookeeper, PIG, HIVE, Sqoop, Oozie, Impala and Flume Wrote HIVE UDF's as per requirements and to handle different schema's and xml data Designing and developing MapReduce jobs to process data coming in different file formats like XML, CSV, JSON Involved in Apache SPARK testing Implemented ETL code to load data from multiple sources into HDFS using Pig Scripts Implemented MapReduce programs to handle semi/ unstructured data like XML, JSON, Avro data files and sequence files for log files Responsible to review the test cases in HP ALM Developed Spark applications using Scala for easy Hadoop transitions. And Hands on experienced in writing Spark jobs and Spark streaming API using Scala and Python Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive developed Spark code and Spark-SQL/Streaming for faster testing and processing of data Installed Oozie workflow engine to run multiple Hive and Pig jobs Designed and developed User Defined Function (UDF) for Hive and Developed the Pig UDF'S to pre-process the data for analysis as well as experience in (UDAFs) for custom data specific processing Assisted in problem solving with Big Data technologies for integration of Hive with HBase and Sqoop with HBase Designed and developed the core data

pipeline code, involving work in Java and Python and built on Kafka and Storm    Good knowledge on Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive for optimized performance    Performance tuning using Partitioning, bucketing of IMPALA tables

Hands on experience on fetching the live stream data from DB2 to HBase table using Spark Streaming and Apache Kafka    Experience in job workflow scheduling and monitoring tools like Oozie and Zookeeper    Worked on NoSQL databases including HBase and Cassandra    Populated HDFS and Cassandra with huge amounts of data using Apache Kafka    Created POC (Proof of Concept) to store Server Log data into Cassandra to identify System Alert Metrics.    Environment: Map Reduce, HDFS, Hive, Pig, HBase, Python, SQL, Sqoop, Flume, Oozie, Impala, Scala, Spark, Apache Kafka, Zookeeper, J2EE, Linux Red Hat, HP-ALM, Eclipse, Cassandra, Talend, Informatica.

Hadoop Developer    Wellcare Health Plans - Tampa, FL    June 2010 to February 2011

Responsibilities:    WellCare focuses exclusively on providing government-sponsored managed care services, primarily through Medicaid, Medicare Advantage and Medicare Prescription Drug Plans, to families, children, seniors and individuals with complex medical needs.    Roles and Responsibilities:

Experience in creating integration between Hive and HBase for effective usage and performed MR Unit testing for the Map Reduce jobs.    Created BI reports (Tableau) and dashboards from HDFS data using Hive.    Experience in importing and exporting the data from Relational Database Systems to HDFS by using Sqoop.    Developed a common framework to import the data from Teradata to HDFS and to export to Teradata using Sqoop.    Imported the log data from different servers into HDFS using Flume and developed MapReduce programs for analyzing the data.    Used Flume to handle the real time log processing for attribution reports.    Worked on tuning the performance of Pig queries.    Involved in loading data from UNIX file system to HDFS.    Performed operation using Partitioning pattern in MapReduce to move records into different categories.    Implemented Fair schedulers on the Job tracker to share the resources of the Cluster for the MapReduce jobs given by the users.    Involved in templates and screens in HTML and JavaScript.    Created HBase tables to load large sets of data coming from UNIX and NoSQL.    Implemented the Web Service client for the login authentication, credit reports and applicant information using

Apache Axis 2 Web Service. Design, develop, test, implement and support of Data Warehousing ETL using Talend and Hadoop Technologies. Built and deployed Java applications into multiple Unix based environments and produced both unit and functional test results along with release notes Environment: WebSphere 6.1, HTML, XML, ANT 1.6, MapReduce, Sqoop, UNIX, NoSQL, Java, JavaScript, MR Unit, Teradata, Node.js, JUnit 3.8, ETL, Talend, HDFS, Hive, HBase. Education Bachelor's Skills APACHE HADOOP MAPREDUCE (9 years), APACHE HBASE (9 years), Hadoop (9 years), HADOOP (9 years), HBase (9 years) Additional Information TECHNICAL SKILLS Big Data Technologies Hadoop, HDFS, MapReduce, Spark, Hive, Sqoop, Oozie, Flume, HBase, Pig, Zookeeper, Kafka. Programming Languages Linux Programming, Basics of Python language, Hive, SQL, JAVA, Pig Latin. Web Development JavaScript, JQuery, HTML 5.0, CSS 3.0, AJAX, JSON Tools CDH 5.9.1, Cloudera Navigator, Cloudera Manager Basics, VMware. Web Servers Apache Tomcat Methodologies Agile, V-model, Waterfall model Operating Systems Linux (CentOS), Windows, Databases MySQL, Oracle 11g/10g, HBase, Mongo DB, Cassandra, Couch, MS SQL server

Name: Miguel Rosales

Email: bradleyunderwood@example.org

Phone: 670.273.8589