Hadoop/Spark Developer Hadoop/Spark Developer Hadoop/Spark Developer - ConnectiveRx Over 7 years of diversified IT experience in E2E data analytics platforms (ETL-BI- Java) as Big data, Hadoop, Java/J2EE Development and System Analysis. Worked for over 4 years with Big Data/ Hadoop Ecosystem in the implementation of Data Lake. Hands on experience Hadoop framework and its ecosystem like Distributed file system (HDFS), MapReduce, Pig, Hive, Sqoop, Flume and Spark. Experience in layers of Hadoop Framework - Storage (HDFS), Analysis (Pig and Hive), Engineering (Jobs and Workflows), extending the functionality by writing custom UDFs. Extensive experience in developing Data warehouse applications using Hadoop, Informatica, Oracle, Teradata, MS SQL server on UNIX and Windows platforms and experience in creating complex mappings using various transformations and developing strategies for Extraction, Transformation and Loading (ETL) mechanism by using Informatica 9.x/8.x. Proficient in Hive Query language and experienced in hive performance optimization using Static-Partitioning, Dynamic-Partitioning, Bucketing and Parallel Execution concepts. As ETL developer, designed and maintained high performance ELT/ETL processes. Experience in analyzing data using Hive QL, Pig Latin, and custom MapReduce programs in Java, custom UDF s. Good Understanding of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MapReduce concepts. Knowledge on Cloud computing infrastructure AWS (amazon web services). Created modules for spark streaming in data into Data Lake using Strom and Spark. Experience in Dimensional Data Modeling Star Schema, Snow-Flake Schema, Fact and Dimensional Tables, concepts like Lambda Architecture, and Batch processing, Oozie. Extensively used Informatica client tools Source Analyzer, Warehouse designer, Mapping designer, Mapplet Designer, ETL Transformations, Informatica Repository Manager and Informatica Server Manager, Workflow Manager & Workflow Monitor. Expertise in using core Java, J2EE, Multithreading, JDBC, Shell Scripting and proficient in using Java API's Collections, Servlets, JSP for application development. Worked closely to review pre- and post-processed data to ensure data accuracy and integrity with Dev and QA teams. Experience in Java, J2ee, JDBC, Collections, Servlets, JSP, Struts, Spring, Hibernate, JSON, XML, REST, SOAP Web services, Groovy, MVC, Eclipse,

Weblogic, Websphere, and Apache Tomcat severs. Extensive knowledge of Data Modeling, Data Conversions, Data integration and Data Migration with specialization in Informatica Power Center. Expertise in extraction, transformation and loading data from heterogeneous systems like flat files, excel, Oracle, Teradata, MSSQL Server. Good work experience with UNIX/Linux commands, scripting and deploying the applications on the servers. Maintained tuning, and monitoring Hadoop jobs and clusters in a production environment. Strong skills in algorithms, data structures, Object oriented design, Design patterns, documentation and QA/testing. Excellent domain knowledge in Insurance, Telecom and Banking. Work Experience Hadoop/Spark Developer ConnectiveRx - Bridgeville, PA April 2018 to Present Description: American Family Insurance is a private mutual company that focuses on property, casualty, auto insurance and also offers commercial insurance, life, health and homeowners' coverage. The purpose of the project is to build a complex ETL pipeline which will handle huge trip data that is collected from vendor servers. Transformations which we implemented can calculate different trip level details and events that are based on latitude, longitude details from raw data. Storing all trip level calculated summary data in data warehouse on top of Hadoop. Different machine learning algorithms are used on this summary data to generate the scores, which will be useful to provide better service to the existing customers. Responsibilities: Created and worked on Sqoop jobs with incremental load to populate Hive External tables. Designed and developed Hive tables to store staging and historical data. Created Hive tables as per requirement, internal and external tables are defined with appropriate static and dynamic partitions, intended for efficiency. Experience in using ORC file format with Snappy compression for optimized storage of Hive tables. Solved performance issues in Hive and Pig scripts with understanding of Joins, Group and aggregation and used them using Impala process engine Developed Spark scripts by using Scala shell commands as per the requirement. Created Oozie workflows for sqoop to migrate the data from source to HDFS and then to target tables. Developed Oozie workflow for scheduling and orchestrating the ETL process. Responsible for building scalable distributed data solutions using Hadoop. Experience in Job management using Fair scheduler and Developed job processing scripts using Oozie workflow.

Involved in migrating MapReduce jobs into Spark jobs and used Spark SQL and Data Frames API to load structured and semi-structured data into Spark clusters. Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing. Worked extensively with Sqoop for importing metadata from Oracle. Developed Oozie workflow jobs to execute Hive, Pig, Sqoop and MapReduce actions. Configured Flume to transport web server logs into HDFS. Experience on Amazon Web Services (AWS), Amazon Cloud Services like Elastic Compute Cloud (EC2), Simple Storage Service(S3), Elastic Map Reduce (EMR), Amazon Simple DB and Amazon Cloud Watch. Implemented Spark using Scala and SparkSQL for faster testing and processing of data. Used Apache Kafka for importing real time network log data into HDFS. Worked on numerous POCs to prove if Big Data is the right fit for a business case. Experience data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka. Created web-based User interface for creating, monitoring and controlling data flows using Apache Nifi. Environment: Apache Hadoop, CDH 4.7, HDFS, MapReduce, Sqoop, Flume, Pig, Hive, HBase, Oozie, Scala, Spark, Spark Streaming, Kafka, Linux Hadoop Developer Verizon - Bel Air, MD August 2016 to March 2018 With over 100 million customers and annual revenue of $30 billion, Sears generates a huge amount of data on the transactions made by the customers. The scope of the project is to use the customer transaction data in store and online over a period to recommend items to customer, that they will find engaging. Another concept is a recommendation engine that introduces new products to a customer, which they might have not, came across before. To achieve these goals, we perform data exploration to learn about user behaviour. We perform future engineering to create new features from existing features that truly reflects the signals in the data and avoid noise. Responsibilities: Worked with the source team to understand the format & delimiters of the data files. Responsible for generating actionable insights from complex data to drive significant business results for various application teams. Developed and implemented API services using Python in spark. Troubleshoot and resolve data quality issues and maintain important level of data accuracy in the data being reported. Extensively implemented POC's on migrating to Spark-Streaming to process the live data. Ingested data from RDBMS and performed

data transformations, and then export the transformed data to Cassandra as per the business requirement. Re-writing existing map-reduce jobs to use new features and improvements for achieving faster results. Analyses large amount of data sets to determine optimal way to aggregate and report on it. Performance tuned slow running resource intensive jobs. Worked on Data serialization formats for converting complex objects into sequence bits by using Avro, Parquet, JSON, CSV formats. Hands on experience working on in-memory based Apache Spark application for ETL transformations. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python. Developed multiple POCs using Spark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata. Developed Flume configuration to extract log data from different resources and transfer data with different file formats (JSON, XML, and Parquet) to hive tables. Setup Oozie workflow /sub workflow jobs for Hive/SQOOP/HDFS actions. Experience in accessing Kafka cluster to consume data into Hadoop. Involved in importing the real-time data to Hadoop using Kafka and implemented the Oozie job for daily imports. Worked with business and functional requirement gathering team, updated user comments in JIRA and documented in confluence Handled tasks like maintaining accurate roadmap for project or certain product. Monitoring the sprints, burndown charts and completing the monthly reports. Environment: Hive, SQL, Pig, Flume, Kafka, Map reduce, SQOOP, Spark, Python, Java, Shell Scripting, Teradata, Oracle, Oozie, Cassandra Hadoop Developer SSI Group-Mobile, AL June 2015 to July 20116 Description: SSI Group is a Clinical Data Services provides healthcare organizations with a proven technology to share clinical information. The solution enables authorized users for covered entities to see a complete picture of patient-exhibited conditions and the treatments received. SSI solutions increase the accuracy and velocity of data exchange among healthcare providers and payers, with the highest levels of security. Responsibilities: Gathered business requirements in meetings for successful implementation and POC (Proof-of-Concept) of Hadoop Cluster. Importing data in regular basis using Sqoop into the Hive partition and controlled work flow by using apache Oozie. Developed Sqoop Jobs to both import data into HDFS from Relational Database Management System like Oracle & DB2 and export

data from HDFS to Oracle. Developing HQL queries to implement the select, insert, update and operations to the database by creating HQL named queries. Involved in data extraction that may include analysing, reviewing, modelling based on requirements using higher level tools such as Hive and Impala. Experience in migrating HiveQL into Impala to minimize query response time. Involving in creating Hive tables, loading with data and writing hive queries. Developed Pig functions to pre-process the data for analysis. Created HBase tables to store all data. Deployed the Hbase cluster in cloud (AWS) environment with scalable nodes as per the business requirement. Analysed identified defects and its root cause and recommended course of actions. Loaded data into Hive Tables from Hadoop Distributed File System (HDFS) to provide SQL-like access on Hadoop data. Worked on streaming the analysed data to the existing relational databases using Sqoop for making it available for visualization and report generation by the BI team. Generated reports and did predictions using BI Tool called Tableau, Integrated data by using Talend. Deployed the Hbase cluster in cloud (Amazon AWS) environment with scalable nodes as per the business requirement. Environment: HDFS, Hive, MapReduce, Sqoop, Impala, Java, Pig, SQL Server, HBase, Oracle and Tableau, AWS. Hadoop Developer Net Cracker Technology Solutions-Hyderabad March 2014 to May 2015 Responsibilities: Integrated Kafka with Storm for real time data processing and written some storm topologies to store the processed data directly to MongoDB and HDFS. Experience in writing Spark SQL scripts. Imported data from different sources into Spark RDD for processing. Developed custom aggregate functions using Spark SQL and performed interactive querying. Involved in loading data from edge node to HDFS using shell scripting. Worked on installing cluster, commissioning and decommissioning of Datanode, Namenode high availability, capacity planning and slots configuration. Completion of unit testing for the new Hadoop jobs in standalone mode designated for Unit region using MR Unit. Developed Spark scripts by using Scala and Python shell commands as per the requirement. Experience in managing and reviewing Hadoop log files. Experience in Hive partitioning, bucketing and perform joins on Hive tables and implementing Hive SerDe like REGEX, JSON and Avro. Optimized Hive analytics Sql queries, created tables/views, written custom UDF's and Hive based exception

processing. Involved in transforming the Teradata to legacy lables to HDFS and HBase tables using Sqoop and vice versa. Configured Fair Scheduler to provide fair resources to all the applications across the cluster. Environment: Hortonworks Hadoop, Ambari, Spark, Solr, Kafka, MongoDB, Linux, HDFS, Hive, Pig, Sqoop, Flume, Zookeeper, RDBMS. Java Developer Infor, Grenville - Hyderabad, Telangana March 2012 to February 2014 Description: The workstation Project automates assignment of Workstation and Keys to each employee, capable of identifies the unassigned workstation and Keys, based on that we can easily assign that workstation and keys (pedestal/storage/cabin) to new employee. Gathered the Employee Details, Key Details, Workstation Details based on Floor and Zone and all these are entered by using Bulk Import Concept using this project employees can be viewed/categorized based on their position. Here we can identify the employee location based on Floor/Zone. Responsibilities: Functional and UI design has been prepared. Implementation at BIO level. Creation of Record sets and BIOs for the database schema. Created Relationships for data Integrity. Created Lookups and attribute domains. Implementation at UI level ie. Menus for Navigation, Forms for various Perspectives Implemented shells like List Shell, Detail Shell, Tab Group Shell, Toggle Shell to Provide better look and feel toolbars to allow UI actions for buttons. Used Form Slots by considering the BIO schema.

Attachments of documents has been provided for work orders/invoices. Authentication and authorization have been achieved by creating users and profiles in platadmin. Implemented object-permissions at widget, menu, and form levels. Developed Form level extensions to achieve UI level validations and BIO level extensions to fulfil Functional requirements and validations. All required data is entered by using Bulk Import. Involved in Development process and have knowledge in usage of Tracker Tools like JIRA. Having good Knowledge in Epiphany Platform (Open Architecture). Having Extensive Hands on Experience on Complex PL/SQL Programming.

Environments: CRB Studio, Web logic server 8.1, LDAP, Core Java, SQL Server. Skills Hdfs, Impala, Mapreduce, Oozie, Sqoop, Hbase, Kafka, Flume, Hadoop, Jboss, Mongodb, Nosql, Teradata, Visual studio, Apache spark, Application server, Git, Hadoop, Hbase, Hive Additional Information Skills: Big Data Technologies HDFS, MapReduce, Hive, Pig, Sqoop, Flume, Oozie,

Zookeeper, Kafka, MongoDB, Apache Spark, Spark Streaming, HBase, Flume, Impala Hadoop Distribution Cloudera, Horton Works, Apache, AWS Languages Java, SQL, PL/SQL, Pig Latin, HiveQL, Scala, Regular Expressions Operating Systems Windows(xp/7/8/10), UNIX, LINUX, UBUNTU, CENTOS. Portals/Application servers WebLogic, WebSphere Application server, WebSphere Portal server, JBOSS Build Automation tools SBT, Ant, Maven Version Control GIT IDE & Build Tools, Design Eclipse, Visual Studio, Net Beans, Rational Application Developer, Junit Databases Oracle, SQL Server, MySQL, MS Access, NoSQL Database (HBase, MongoDB), Teradata.

Name: Cody Moreno

Email: courtney18@example.com

Phone: 001-412-973-3520x703