Hadoop Developer Hadoop Developer Hadoop Developer - Dell Austin, TX Hadoop Developer with 8+ Years of IT experience including 4 years in Big Data and Analytics field in Storage, Querying, Processing and Analysis for developing E2E Data pipelines. Expertise in designing scalable Big Data solutions, data warehouse models on large-scale distributed data, performing wide range of analytics.   Expertise in all components of Hadoop/Spark Ecosystems - Spark, Hive, Pig, Flume, Sqoop, HBase, Kafka, Oozie, Impala, Stream sets, Apache NIFI, Hue, AWS.   3+ years of experience working in programming languages Scala/Python.   Extensive knowledge on data serialization techniques like Avro, Sequence Files, Parquet, JSON and ORC.   Acute knowledge on Spark architecture and real-time streaming using Spark.   Hands on experience with Spark Core, Spark SQL and Data Frames/Data Sets/RDD API.   Good knowledge on Amazon Web Services (AWS) cloud services like EC2, S3, EMR and VPC.   Experienced in Data Ingestion, Data Processing, Data Aggregations, Visualization in Spark Environment.   Hands on experience in working with large volume of Structured and Un-Structured data.   Expert in migrating the code components from SVN repository to Bit Bucket repository.   Experienced in building Jenkins pipelines for continuous code integration from Github into Linux machine. Experience in Object Oriented Analysis Design (OOAD) and development.   Good understanding in end-to- end web applications and design patterns.   Hands on experience in application development using Java, RDBMS, and Linux shell scripting.   Experience in implementing by using agile methodology. Well versed in using Software development methodologies like Agile Methodology and Waterfall processes.   Experienced in handling databases: Netezza, Oracle and Teradata.   Strong team player with good communication, analytical, presentation and inter-personal skills. Work Experience Hadoop Developer Dell - Austin, TX, US December 2017 to Present Dell EMC is the undisputed industry leader for unstructured data storage. Tens of thousands of customers have bet their business on UDS products including Isilon Scale-Ou t File Storage and ECS Object Storage. Moreover, Gartner has consistently recognized Dell EMC as the clear Magic Quadrant leader for Distributed File and Object Storage. As a part of Data science team we are responsible for analyzing, processing and visualizing one of our client's information to improve the business.

Responsibilities: Worked on Hortonworks-HDP 2.5 distribution. Involved in review of functional and non-functional requirements. Responsible for designing and implementing the data pipeline using Big Data tools including Hive, Spark, Scala and Stream Sets. Experience in using Apache Storm, Spark Streaming, Apache Spark, Apache NiFi, Kafka and Flume in creating data streaming solutions. Developed and implemented Apache NIFI across various environments, written QA scripts in Python for tracking files. Involved in importing data from Microsoft SQL Server, MySQL, and Teradata into HDFS using Sqoop. Good knowledge in using Apache NIFI to automate the data movement. Developed Sqoop scripts to import data from relational sources and handled incremental loading. Extensively used Stream Sets Data Collector to create ETL pipeline for pulling the data from RDBMS system to HDFS. Implemented the data processing framework using Scala and Spark SQL. Worked on implementing the performance optimization methods to improve the data processing timing. Experienced in creating the shell scripts and made jobs automated. Extensively worked on Data frames and Datasets using Spark and Spark SQL. Responsible for defining the data flow within Hadoop eco system and direct the team in implement them and exported the result set from Hive to MySQL using Shell scripts. Worked on Kafka Streaming using stream sets to process continuous integration of data from Oracle systems to hive tables. Developed a generic utility in Spark for pulling the data from RDBMS system using multiple parallel connections. Integrated existing code logic in HiveQL and implemented in the Spark application for data processing. Extensively used Hive/Spark optimization techniques like Partitioning, Bucketing, Map Join, parallel execution, Broadcast join and Repartitioning. Environment: Spark, Python, Scala, Hive, Hue, UNIX Scripting, Spark SQL, Stream sets, Kafka, Impala, Beeline, Git, Tidal. Hadoop Developer Inter-American Development Bank - Washington, D.C, US January 2016 to November 2017 A currency transaction report (CTR) is a report that U.S. financial institutions are required to file with FinCEN for each deposit, withdrawal, exchange of currency, or other payment and transfer, by, through, or to the financial institution which involves a transaction in currency of more than $10,000. Report is generated daily basis with T1 data with pre-defined rules and alerts. Once report is generated with potential CTRs, analysts will parse alerted CTRs to identify actual vs

fractious CTRs vial Web interface. Final consolidated report will be regenerated every night and sent to FINCEN in pre-defined FINCEN format.   Responsibilities: Worked on Hortonworks-HDP 2.5distribution.   Experience in implementing Scala framework code using IntelliJ and UNIX scripting to implement the workflow for the jobs.   Involved in gathering business requirement, analyze the use case and implement the use case end to end.   Worked closely with the Architect; enhanced and optimized product Spark and Scala code to aggregate, group and run data mining tasks using Spark framework.   Experienced in loading the raw data into RDDs and validate the data.   Experienced in converting the validated RDDs into Data frames for further processing. Implemented the Spark SQL code logic to join multiple data frames to generate application specific aggregated results.   Experienced in fine tuning the jobs for better performance in the production cluster space.   Worked totally in agile methodologies, used Rally scrum tool to track the User stories and Team performance.   Worked extensively in Impala Hue to analyze the processed data and to generate the end reports.   Experienced working with hive database through beeline. Worked on analyzing and resolving the production job failures in several scenarios.   Implemented UNIXscripts to define the use case workflow and also to process the data files, and automate the jobs..   Environment: Spark, Scala, Hive, Sqoop, UNIX Scripting, Spark SQL, IntelliJ, Hbase, Kafka, Impala, Hue, Beeline, Git. Hadoop Developer Atlanta Gas Light - Atlanta, GA, US October 2014 to January 2016 Responsibilities: Worked on Cloudera CDH distribution.   Hand on experience on cloud services like Amazon Web Services (AWS)   Created data pipelines for different events to load the data from DynamoDB to AWS S3 bucket and then into HDFS location.   Involved in complete SDLC - Requirement Analysis, Development, Testing and Deployment into Cluster. Worked hand-in-hand with the Architect; enhanced and optimized product Spark code to aggregate, group and run data mining tasks using Spark framework.   Extracted data from various SQL database sources into HDFS using Sqoopand also ran Hive scripts on the huge chunks of data. Implemented a prototype for the complete requirements using Splunk, python and Machine learning concepts.   Design and Implementation of Map reduce code logic for Natural Language Processing of Free Form Text.   Deployed the project on Amazon EMR with S3 Connectivity.   Implemented

usage of Amazon EMR for processing Big Data across a Hadoop Cluster of virtual servers on Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service(S3).   Loaded the data into Simple Storage Service (S3) in the AWS Cloud.   Good Knowledge in using of Amazon Load Balancer for Auto scaling in EC2 servers.   Implemented Spark scripts to migrate map reduce jobs into Spark RDD transformations, streaming data using Apache Kafka.   Implemented Spark SQL queries which intermix the Hive queries with the programmatic data manipulations supported by RDDs and data frames in scala and python.   Involved in Deployment of Code Logic and UDFsacross the cluster.   Communicate deliverables status to user/stakeholders, client and drive periodic review meetings.   Worked on Data Processing using Hive queries in HDFS and the shell Scripts to wrap the HQL scripts.   Developed and Deployed Oozie Workflows for recurring operations on Clusters.   Experienced in performance tuning of hadoop jobs for setting right Batch Interval time, correct level of Parallelism and memory tuning.   Worked extensively with Sqoop for importing metadata from Oracle.   Used Tableau reporting tool to generate reports from the outputs stored in HDFS.   Environment: Hadoop, Spark, HDFS, Hive, Map Reduce, Sqoop, Oozie, Tableau.

Hadoop Developer Toyota Motors July 2012 to September 2014 Responsibilities: Worked on Cloudera CDH distribution.   Design and Implement historical and incremental data ingestion techniques from multiple external systems using Hive, pig and sqoop ingestion tools.   Design physical data models for structured and semi-structured to validate the raw data into HDFS.   Design map/reduce logic and HIVES queries for generating aggregated metrics.   Involved in Design, implementation, development and testing phases in the project.   Responsible to monitor the jobs in production cluster while and trace the error logs when the jobs fails.   Design and Develop data migration logic for exporting data from MySQL to Hive.   Design and Develop complex workflow in Oozie for recurrent job execution.   Used SSRS reporting tool for the generation of data analysis reports.   Environment: Hadoop, MapReduce, HDFS, Pig, Hive, Oozie, Eclipse, Cloudera, Sqoop, SSRS Software Developer Garima Software Solutions November 2010 to June 2012 Responsibilities: Involved in complete SDLC - Requirement Analysis, Development, Testing and Deployments.   Involved in resolving critical Errors.   Responsible to deploy the deliverables of

sprints successfully. Involved in capturing the client's requirements and enhancements on the application document the requirements and populate to the associated teams. Design and Implementation of REST Full services and WSDL in VORDEL. Implemented complex SQL quires to get the analysis reports. Created Desktop applications using J2EE, Swings. Involved in developing applications using Java, JSP, Servlets, Swings. Developed UI using HTML, CSS, Ajax, JQuery and developed Business logic and Interfacing Components using Business Objects, XML and JDBC. Created applications, connection pools, deployment of JSP & Servlets. Used Oracle, MySQL database for storing user information. Developed backed for application using PHP for web applications. Experienced with the Agile Methodologies. Environment: SOAP, REST, HTML, WSDL, 22'Vordel, SQL Developer Education Bachelor's Skills DYNAMODB, CASSANDRA, AMBARI, HDFS, OOZIE Additional Information Technical Skills: Bigdata Technologies HDFS, Map Reduce, Pig, Hive, Sqoop, Oozie, Scala, Spark, Kafka, Flume, Ambari, Hue Hadoop Frameworks: Cloudera CDHs, Hortonworks HDPs, MAPR. Database Oracle 10g/11g, PL/SQL, MySQL, MS SQL Server 2012, DB2 Language C , C++, Java, Scala, Python AWS Components IAH, S3, EMR, EC2,Lambda, Route 53, Cloud Watch, SNS, Methodologies Agile, Waterfall Build Tools Maven, Gradle, Jenkins. NO-SQL Databases HBase, Cassandra, MongoDB, DynamoDB IDE Tools Eclipse, Net Beans, Intellij Modelling Tools Rational Rose, Star UML, Visual paradigm for UML Architecture Relational DBMS, Client-Server Architecture Cloud Platforms AWS Cloud BI Tools Tableau Operating System Windows 7/8/10, Vista, UNIX, Linux, Ubuntu, Mac OS X

Name: Selena Wright

Email: janetfarmer@example.org

Phone: 8199764652