

Senior Spark Developer,Hadoop Developer Senior Spark Developer,Hadoop Developer Senior Spark Developer,Hadoop Developer - Synchrony Financial Bank Atlanta, GA Around 9+years of experience in Hadoop/Big Data technologies such as in Hadoop, Pig, Hive, HBase, Oozie, Zookeeper, Sqoop, Storm, Flink, Flume, Zookeeper, Impala, Tez, Kafka and Spark with hands on experience in writing Map Reduce/YARN and Spark/Scala jobs. Have good IT experience with special emphasis on Analysis, Design and Development and Testing of ETL methodologies in all the phases of the Data Warehousing. Expertise in OLTP/OLAP System Study, Analysis and E-R modeling, developing Database Schemas like star schema and Snowflake schema used in relational, dimensional modeling. Experience in optimizing and performance tuning of Mappings and implementing the complex business rules by creating re-usable Transformations, Mapplets and Tasks. Solid Experience in Cloud with Amazon Web services (AWS - EC2, S3, CloudWatch, RDS,EMR,SNS ) and Google cloud (GCP). Responsible for developing data pipeline using flume, Sqoop and pig to extract the data from weblogs and store in HDFS. Queried Vertica, SQL Server for data validation along with developing validation worksheets in Excel in order to validate the dashboards on Tableau. Have knowledge on GCP cloud Pub/Sub and Microservices Event sourcing, cloud functions and algorithm. Experience in implementing and migrating and deploying workloads on Azure VM Developed Spark code using Scala and Spark-SQL for faster testing and data processing. Extensively used SQL and PL/SQL for development of Procedures, Functions, Packages and Triggers. Experienced on Tableau Desktop, Tableau Server and good understanding of tableau architecture Experienced in integrating Kafka with Spark Streaming for high speed data processing. Experience in Implementing AWS solutions using EC2, S3 and Azure storage. Experienced in developing business reports by writing complex SQL queries using views, macros, volatile and global temporary tables. Working with AWS team in testing our Apache Spark- ETL application on EMR/EC2 using S3. Experience in designing both time driven and data driven automated workflows using Oozie. Experienced with work flow schedulers, data architecture including data ingestion pipeline design and data modelling. Configuration of ElasticSearch on Amazon Web Service with static IP authentication security features Experience

in AWS Cloud platform and its features which includes EC2, AMI, EBS Cloudwatch, AWS Config, Auto-scaling, IAM user management, and AWS S3. Managed AWS EC2 instances utilizing Auto Scaling, Elastic Load Balancing and Glacier for our QA and UAT environments as well as infrastructure servers for GI. Solved performance issues in Hive and Pig scripts with understanding of Joins, Group and Aggregation and how does it translate to MapReduce jobs. Work Experience Senior Spark Developer, Hadoop Developer Synchrony Financial Bank - Atlanta, GA May 2018 to Present Responsibilities: Worked on different tools for Presto to process these large datasets. Worked on Core tables of Revenue DataFeed(RDF) that calculates the revenue of the advertisers of the Facebook. Involved into testing and migration to Presto. Worked extensively with Data migration, Data cleansing, Data profiling, and ETL Processes features for data warehouses. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala. Experienced with the tools in Hadoop Ecosystem including Pig, Hive, HDFS, Sqoop, Spark, Yarn and Oozie Have used Python with the Spark Python API (PySpark) to create and analyze Spark DataFrames. Involved in importing the real time data to Hadoop using kafka and implemented the Oozie job for daily. Experienced in writing complex SQL Queries, Stored Procedures, Triggers, Views, Cursors, Joins, Constraints, DDL, DML and User Defined Functions to implement the business logic. Developed Custom ETL Solution, Batch processing and Real-Time data ingestion pipeline to move data in and out of Hadoop using Python and shell Script. Experience in Large Data processing and transformation using Hadoop-Hive and Sqoop. Real time predictive analytics capabilities using Spark Streaming, Spark SQL and Oracle Data Mining tools. Experience with Tableau for Data Acquisition and visualizations. Working with AWS team in testing our Apache Spark- ETL application on EMR/EC2 using S3. Assisted in data analysis, star schema data modeling and design specific to data warehousing and business intelligence environment. Have been using pyspark with Jupyter in Docker Containers. Expertise in platform related Hadoop Production support tasks by analyzing the job logs. Have been using Spark with Python, working with RDD provided by the library Py4j in pyspark module. Monitored System health and logs and responded accordingly to any warning or failure conditions.

Environment: Amazon Web Service, Vertica, Informatica PowerCenter, Pyspark, Spark, AWS, Kafka, AWS-S3, Apache-Hadoop, Hive, Pig, Shell Script, ETL, tableau, Agile Methodology. Spark Developer, Hadoop developer View Lift - New York, NY September 2016 to April 2018

Responsibilities: Collected data using Spark Streaming from AWS S3 bucket in near-real-time and performs necessary Transformations and Aggregation on the fly to build the common learner data model and persists the data in HDFS. Developed Spark Code using Scala and Spark-SQL/Streaming for faster testing and processing of data. Real-time experience in Hadoop Distributed files system, Hadoop framework and Parallel processing implementation (AWS EMR, Cloudera) with hands on experience in HDFS. Using Glue Data Catalog for storing the schema/metadata of Hive External tables. Used Aws EMR long running clusters for processing the spark jobs, use AWS S3 for storing data in buckets. Created Hive Internal/External tables, metastore for storing the metadata. Used sqoop jobs for ingestion from various sources such as Oracle, Salesforce, SAS etc. Worked on various complex SQL queries on the source side which is Oracle database. Worked on using cloudformation template, CI-CD tools like concourse to automate the data pipeline. Worked on AWS CLI Auto Scaling and Cloud Watch Monitoring creation and update. Designed Data Quality Framework to perform schema validation and data profiling on Spark (Pyspark). Experience in Implementing AWS solutions using EC2, S3 and Azure storage. Responsible for monitor the tableau dashboard for reporting purpose and providing the refined data to end users.

Environment: AWS -EMR, S3, Spark, Spark -sql, Scala, Azure, Hive, Sqoop, AWS-Cli, Shell Script, ETL, Tableau, Agile Methodology. Hadoop Developer Cognizant Technology Solutions - Teaneck, NJ October 2014 to August 2016

Responsibilities: Worked with variables and parameter files and designed ETL framework to create parameter files to make it dynamic. Currently working on the Teradata to HP Vertica Data Migration Project Working extensively on the Copy Command for extracting the data from the files to Vertica. Monitor the ETL process job and validate the data loaded in Vertica DW. Built a Full-Service Catalog System which has a full workflow using ElasticSearch, Logstash, Kibana, Kinesis, CloudWatch. Responsible for data extraction and data ingestion from different data sources into Hadoop Data Lake by creating

ETL pipelines using Pig, and Hive. Experienced in transferring data from different data sources into HDFS systems using kafka producers, consumers and kafka brokers. The logs and semi structured content that are stored on HDFS were preprocessed using PIG and the processed data is imported into Hivewarehouse which enabled business analysts to write Hive queries. Worked with data migration from Hadoop clusters to cloud. Good knowledge of cloud components like AWS S3, EMR, Elastic Cache and EC2. Responsible to write Hive and Pig scripts as ETL tool to do transformations, event joins, filter both traffic and some pre-aggregations before storing into the HDFS. Developed the Vertica UDF's to preprocess the data for analysis. Designed the reporting application that uses the Spark SQL to fetch and generate reports on HBase. Build custom batch aggregation framework for creating reporting aggregates in Hadoop. Experience in working with Hive data warehouse tool-creating tables, data distribution by implementing partitioning and bucketing, writing and optimizing the Hive queries. Built real time pipeline for streaming data using Kafka and SparkStreaming. Experienced with NoSQL databases like HBase, MongoDB and Cassandra and wrote Storm topology to accept the events from Kafka producer and emit into Cassandra DB. Experienced in working with spark ecosystem using spark SQL and scala queries on different formats like Text file, CSV file. Great hands on experience with Pyspark for using Spark libraries by using python scripting for data analysis. Wrote Python Script to access databases and execute scripts and commands. Involved in converting Cassandra/Hive/SQL queries into Spark transformations using Spark RDD's in Scala and Python. Created ODBC connection through Sqoop between Hortonworks and SQL Server. Building, publishing customized interactive reports and dashboards, report scheduling using Tableau server. Creating New Schedule's and checking the task's daily on the server. Environment: Hadoop, Hive, Apache Spark, Apache Kafka, Hortonworks, AWS, ElasticSearch, Lambda, Apache Cassandra, Hbase, MongoDB SQL, Sqoop, Flume, Oozie, Java (jdk 1.6), Eclipse, Informatica Power Center 9.1, Tableau, Teradata 13.x, Teradata SQL Assistant. Java Developer PTC - Boston, MA August 2012 to September 2014 Responsibilities: Involved in developing, testing and implementation of the system using Struts, JSF, and Hibernate. Developing, modifying, fixing, reviewing, testing and

migrating the Java, JSP, XML, Servlet, SQLs, JSF. Updated user-interactive web pages from JSP and CSS to HTML5, CSS, and JavaScript for the best user experience. Developed Servlets, Session and Entity Beans handling business logic and data. Created enterprise deployment strategy and designed the enterprise deployment process to deploy Web Services, J2EE programs on more than 7 different SOA/WebLogic instances across development, test and production environments.

Designed user interface HTML, Swing, CSS, XML, Java Script and JSP. Implemented the presentation using a combination of Java Server Pages (JSP) to render the HTML and well-defined API interface to allow access to the application services layer. Used Enterprise Java Beans (EJBs) extensively in the application. Developed and deployed Session Beans to perform user authentication. Involved in Requirement Analysis, Design, Code Testing and debugging, Implementation activities. Involved in the Performance Tuning of Database and Informatica. Improved performance by identifying and rectifying the performance bottle necks. Understanding how to apply technologies to solve big data problems and to develop innovative big data solutions. Designed and developed Job flows using Oozie. Developed Sqoop commands to pull the data from Teradata. The data is collected from distributed sources into Avro models. Applied transformations and standardizations and loaded into HBase for further data processing. Wrote PL/SQL Packages and Stored procedures to implement business rules and validations.

Environment: Java, J2EE, Java Server Pages (JSP), JavaScript, Hadoop, Oozie, Hive, Teradata, Servlets, JDBC, PL/SQL, ODBC, Struts Framework, XML, CSS, HTML, DHTML, XSL, XSLT and MySQL.

Java Developer New York, NY February 2011 to July 2012 Responsibilities: Involved in Requirements study, Functional analysis, detailed design including entity relations and various table design. Ability to support application deployments, building new systems and upgrading and patching existing ones through ATG methodologies. Designed and implemented a GUI framework for Swing. Involved in Creation of Adobe Flex Families in Content Server and associated the, JDBC and XSD, XML pages to the assets. Involved in executing all Selenium test scripts on the different browsers and checked for compatibility, regression test cases were automated using Selenium Web Driver, and Web Driver Backed Selenium. Used validation frameworks for specifying the

validations rules. Extensive work on Web services (SOAP) and Restful application. Developed Java Script for Client Side validations. Uses coding methods in JNI to initiate or enhance in-house custom developed RF optimization drive test software in support of 1Xused SDLC concepts. Involved in applying SDLC (Agile, Scrum, RUP, Waterfall) concepts. Designed and developed the Java bean components and O/R Mapping using Hibernate, designed roles and groups for users and resources using AWS Identity Access Management (IAM). Involved in writing the screen classes and Action classes for implementing the business logic of Pilot and object oriented programming and monitored and responsible for troubleshooting the WebSphere Application Server with JVM logs, Process Logs, Service logs. Involved in Design, Development and Support phases of Software Development Life Cycle (SDLC), developed Custom Tags to simplify the JSP code. Designed UI screens using JSP, CSS, XML and HTML. Used JavaScript for client side validation. Expertise in creating DevOps strategy in a mix environment of Linux servers, responsible for the implementation of application system with core java and Spring framework, uses Rational Rose for model driven development and UML modeling. Used Spring Core Annotations for Dependency Injection and used Apache Camel to integrate Spring framework, use Apache Camel to route and transform messages and designed and implemented new customer flow using Apache Velocity template. Created PHP/HTML5/CSS3 Web pages to support Comcast Business Voice Xpress VOIP phone support portal using Agile practices and Rally management software Extensively used JSTL tags and Struts tag libraries. Used Struts tiles as well in the presentation tier, participated in coding Spring AOP components for the Transactional Model to handle many requests. Involved in writing JSP and Servlet components. Actively involved in designing and implementing the application using various design patterns such as Singleton, DAO, Front Controller, Service Locator, Business Delegate, Fa ade and Data Access Object, used Java Message Service (JMS) for reliable and asynchronous exchange of important information such as loan status report. Involved in the JMS Connection Pool and the implementation of publish and subscribe using Spring JMS. Used JMS Template to publish and Message Driven Bean (MDB) to subscribe from the JMS provider. Used Hibernate, object/relational-mapping (ORM) solution, technique of mapping data

representation from MVC model to Oracle Relational data model with SQL-based schema.

Develop Isolated tests in JRuby with Gherkin/cucumber using Spring Beans config and mocks and execute the tests in an isolated environment and Implemented in Ruby, Java, and JRuby and used a number of AWS Services Also help set up some of the AWS account peered with Shared Services so some users can use their CORP login creds for logging into AWS accounts, help set AWS federation with on prem Shared Services. Used Junit framework for unit testing of application and Log4j to capture the log that includes runtime exceptions. Worked on Agile, SCRUM development methodology and built the application with Test Driven Development (TDD), deployed the application on Web Sphere Application Server. Used ANT as a build tool and IVY as dependency tool. Used CVS version control for implementing the application. Work involved extensive usage of HTML, DHTML, CSS, JQuery, JavaScript and Ajax for client side development and validations Re-wrote several pieces to make them compliant with the emerging JSF standard, experience in working with relational database MySQL, developed complex SQL queries for extracting data from the database. Environment: Core Java, J2EE, JSP 2.0, Struts 1.2, EJB 2.0, JMS, JNDI, Oracle, DHTML XML, DOM, SAX, Rationale Rose, Groovy gails, UNIX, IBM Web Sphere Application Server 5.1, Hibernate 2.0, spring, LOG4J, CVS. Java Developer Boeing, Everett - San Francisco, CA March 2009 to January 2011 Responsibilities: Worked with business teams on requirements analysis building use cases and estimations. Generating high level and low level design documentation. Developed Java/J2EE code, business logic using Spring, Hibernate framework and OOP concepts, involved in Peer code reviews. Created WSDL & Generated data objects using WSDL, Java, Spring, JAX-WS, Axis, apache CXF and developed mapping code for several Web Services interfaces for various profile management endpoints. Developed Java multi threaded batch offline bulk upload tool, web applications using Spring, Servlets and UI layer using JSPs, JavaScript, HTML, CSS, Angular JS. Worked on implementation of new & complex implementations, critical/quick deliveries. Developed and build Ant scripts, Maven for packaging the application code. Developed database scripts and procedures using PL/SQL. Deployed code on Tomcat web application server. Validated requirement deliverables, unit testing using SOAP UI,

set up & executed system endurance, performance tests using JMeter. Set up build automation & deployment to DEV,QA,PRD servers using TeamCity continuous integration platform. Built Regression suits using SOAP UI for automated regression test on CI platform. Coordinated with Architects and Security teams on defining SOAP, RESTful web services architecture and generating/managing artifacts documentation Confluence. Performed use case analysis and design SOAP, REST APIs on social integration of web and native mobile applications implementing OAuth (Ping Federate). Creating/Managing release plan, sprint deliveries, responsible for version control (SVN) and configuration of the project, keeping track of project activities - Sprint execution, planning in JIRA. Coordinated with cross functional teams on resolving integrations issues, bug fixes, RCAs, RFCs. Worked in Agile model, Conducting Daily Scrum/Stand ups, Backlog Grooming, Sprint Planning Sprint Review, Sprint Retrospective Meeting, supporting Product owner in refining and grooming product backlog. Lead development team, Motivating team and helping them work in self organized manner. Acted as Liaison Between Product Owner and Development Team, resolving impediments in order to achieve Sprint Goals. Worked as SME and SPOC for the project within Adidas CRM area. Managing knowledge sharing & resolved problems on critical issues in live system. Master Data management , Data schemas, web forms configuration , workflow /campaign maintenance and set up using Adobe Campaign (formerly Neolane)

Environment: Java, J2EE, Spring, Hibernate,core java, JSP, HTML, XML, CSS, JavaScript, Subversion(SVN), SVN, Oracle PL/SQL,Jms, WSDL, SOAP, XML, JAX-WS, RESTful, JSON, TomCat, Eclipse, SQL Developer, Toad for Oracle, MS Visio, JIRA, Confluence, Maven, Ant, Beyond Compare, Team City (CI), JUnit, SOAP UI, Apache Jmeter, Unix. Education Bachelor's ,,,,,

Skills DATA MODELING, DB2, JDBC, MS SQL SERVER, SQL SERVER, MYSQL, OLTP, ORACLE, SQL, CASSANDRA, MAHOUT, OOZIE, SQOOP, HBASE, KAFKA, DATA ARCHITECTURE, ETL, HADOOP, MAP REDUCE, OLAP Additional Information TECHNICAL SKILLS: Specialties Data warehousing/ETL/BI Concepts, Data Architecture, Software Development methodologies, Data Modeling Business Tools Tableau 10.X Business Objects XI R2, InformaticaPowercenter 8.x, OLAP/OLTP, Talend, Teradata 13.x, Teradata SQL Assistant Big Data hadoop, map reduce



1.0/2.0, pig, hive, hbase, sqoop, oozie, zookeeper, kafka, spark, flume, storm, impala, mahout, hue, tez, hcatalog, storm, Cassandra, pyspark Cloud Technologies AWS-EMR, AWS S3, Glue Data Catalog, Kinesis, Lambda, ELK(Elastic, Logstash, Kibana) Stack, cloudwatch metric, Azure Databases DB2, MySQL, MS SQL server, Vertica, Mongo DB, Oracle, SQL 2008, Hortonworks, Cloudera Languages Python, Java / J2EE, Scala, HTML, SQL, JDBC, JavaScript, PHP Operating System Mac OS, Unix, Linux (Various Versions), Windows 2003/7/8/8.1/XP Web Development HTML, Java Script, XML, PHP, JSP, Servlets, JavaScript Application Server Apache Tomcat, WebLogic, WebSphere Tools Eclipse, NetBeans

Name: Mary Jackson

Email: shannonmiller@example.net

Phone: 3013596805