

Scala/Hadoop Developer Scala/Hadoop Developer Scala/Hadoop Developer - Fidelity Investment Smithfield, RI Experienced Hadoop Developer with over 10+ years of IT experience has a strong background with Big-data arena, File Systems, Data Management/ Analysis and Java Based enterprise application using Java/ J2EE technologies. Expertise with installing, configuring, testing and using Apache Hadoop framework 2.7.x, its ecosystem components like HDFS, MapReduce, Sqoop 1.4.6, Flume 1.7.0, Hive 2.1.1, Pig 0.16.0, Spark 2.1.0, Scala 2.12.0, Kafka 0.10.1, Yarn, Oozie 3.1.3 and Zookeeper 3.4.10. Experience and strong knowledge on implementation of Spark Core -Spark Streaming, Spark SQL. Developed applications in Spark 2.1.0 using Scala 2.12.0 to compare the performance of Spark with Hive. Implemented POC's and developed pipeline using Kafka 0.10.1, Spark Streaming and Spark SQL. Extensive experience in Spark 2.1.0 transformations using Scala 2.12.0 and Spark SQL for faster testing and processing of data files. Hands on real time processing with Spark modules, Spark RDD, Dataset API using Scala 2.12.0. Hands on experience of writing programs in MapReduce to analyze unstructured data. Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems (RDBMS) and from RDBMS to HDFS Well versed. Expert in writing Pig Latin scripts and HiveQL queries to process and analyze data. Experience in Spark, and in-depth knowledge on Spark-SQL, RDD's, Lazy transformation and actions. Worked on Apache Kafka 0.10.1 and Flume for handling megabytes of streaming data. Transformed various formats of data like sequence File, RC, ORC, Parquet, JSON, AVRO, and experience in dealing with Compression techniques such as Gzip, Snappy. Deep knowledge in batch job scheduling workflow using Oozie 4.2.0. and zookeeper 3.4.10. Understanding of Amazon Web Services stack and hands-on experience in using S3, EC2 and EMR. Worked with different NOSQL databases such as Cassandra 3.10, HBase 1.3.0 and MongoDB. Experience with Relational databases including Oracle, SQL Server and MySQL and experience in writing complex SQL queries, PL/SQL Stored Procedures, Triggers, sequences. Well versed with the working of Data Visualization tools such as Tableau 9.3, D3.js. Developed Web applications using HTML5, CSS3, Bootstrap, JavaScript, JQuery, AJAX. Expertise in Core Java, Data Structures, Multithreading, JDBC, J2EE, Algorithms, Object Oriented Design (OOD) and

Exception Handling and frameworks like Spring MVC and Hibernate. Produced and consumed SOAP and RESTful Web Services and experience in developing Hadoop applications on Spark using Scala as a functional and object-oriented programming. Experienced in writing ANT and Maven scripts to build and deploy Java applications. Experienced in TDD (Test-Driven Development) and SDLC methodologies such as Agile(Scrum). Through knowledge of development environments such as Maven, Git, JIRA 6.4, Jenkins and Confluence. Excellent understanding and knowledge of Hadoop architecture and various daemons such as Name Node, Data Node, Job Tracker, Task Tracker, Resource Manager and MapReduce programming paradigm. Knowledge of Cyber Security concepts like cryptography, Access Control, Data Security in Linux, Unix. Expertise in Creating Hive Internal/External Tables/Views using shared Meta store, writing scripts in HiveQL also data transformation & file processing using Pig Latin Scripts. Experience in Object Oriented Analysis Design (OOAD) and development of software using UML Methodology, good knowledge of J2EE design patterns and Core Java design patterns. Possess excellent presentation, documentation, communication skills, detail oriented, zeal to learn new technologies with a cooperative team focused attitude, strong analytical and problem-solving skills.

Work Experience Scala/Hadoop Developer Fidelity Investment - Smithfield, RI June 2018 to Present

Project Description: Fidelity Investment is a financial services firm involved in different sectors including Mutual fund, Brokerage, Retirement & IRA and Wealth Management and in one of the leading companies in 410k plan and Stock Market activity. The Distributed reporting project involves transferring from legacy system to the Hadoop environment for the power query using HDFS for the optimize query performance for the end customer using hive data warehouse. They are using different tools for this project like Apache Hue, Apache Hadoop, Apache Spark, HDFS, Impala and few internal frameworks.

Responsibilities: Experience in creating Hive tables to store the processed results in a tabular format, optimizing Hive tables using optimization techniques like partitions and bucketing to provide better performance with HiveQL queries and creating custom user defined functions. Developed, reviewed and updated architecture and process documentation, server diagrams, requisition documents and other technical documents Involved in

Agile methodology, attended daily scrum meetings and sprint planning meetings      Integrated visualizations into a Spark application using Databricks and visualization libraries (ggplot, Matplotlib)

Implemented different analytical algorithms using MapReduce programs to apply on top of HDFS data      Installed, managed and maintained a cluster of 60 nodes from POC to PROD      Installed Hortonworks clusters and Hadoop ecosystem components through Ambari and command line interface      Responsible for cluster maintenance, commissioning and decommissioning data nodes, cluster monitoring, troubleshooting, management and review data backups, and Hadoop log files

Skilled in Tableau Desktop for various types of data visualization, reporting and analysis including Cross Map, Scatter Plots, Geographic Map, Pie Charts and Bar Charts, Page Trails and Density Chart      Created HBase tables to store various data formats of data coming from MySQL, Oracle, Teradata      Installed and configured Hive, Pig, Sqoop, Flume and Oozie on the Hadoop cluster

Configured Spark streaming to get ongoing information from the Kafka and stored the stream information to HDFS      Developed data pipeline using Flume, Sqoop to ingest business data and purchase histories into HDFS for analysis      Utilized SparkSQL to extract and process data by parsing using Datasets or RDDs in Hive Context, with transformations and actions (map, flat Map, filter, reduce, reduceByKey)      Extended the capabilities of Data Frames using User Defined Functions in Python and Scala      Designed and Implemented Sqoop incremental imports, delta imports on tables without primary keys and dates from Oracle and appends directly into Hive Warehouse.      Import the data from different sources like HDFS/HBase into Spark RDD and perform computations using Scala to generate the output response.      Writing a spark shell code to audit report to bring all the unique field under metadata.      Implemented HQL scripts for daily based data loading & further aggregations.      Scheduling Oozie workflow and job to automate the run of Unix script daily as per the team requirement.      Working on clean up the old data and purge the database for having accurate data in the system.      Involved in building the ETL architecture and Target mapping to load data into Data Lake.      Developed Workflows with the help of Oozie to manage the flow of jobs and wrote Custom Expression Language (EL) functions for complex workflows.      Working on Install process monthly or by weekly new sets of report for new table or

existing table and validating all the data coming from working directory to the HDFS. Involved in the complete Software Development Life Cycle (SDLC) phases of the project as a part of Agile scrum methodology. Loaded datasets from MySQL to HDFS and Hive respectively on daily basis.

Worked on Apache Hue as a central web admin User to Query on Hive or Impala and to Schedule job workflow. Assisted in Installation and Configuration of Apache Hadoop clusters CHD and Hadoop tools for application development includes HDFS, HUE, YARN, Sqoop, Impala and Hive.

Used Cloudera Manager for continuous managing and monitoring the Hadoop cluster. Environment: Spark 2.3.0, Scala 2.13.0, Sqoop 1.4.6, Impala 2.2.0, Oracle, Hive 2.3.3, HDFS, Spark SQL, CHD 5.16.1, HiveQL, Hue 4.2.0, YARN 1.12.3, Oozie 4.3.0, Agile. Big Data Analytics Solutions/Scala Developer Bloomberg L.P - New York, NY May 2015 to May 2018 Project Description: My Latest project was with Bloomberg which is a global firm which specializes in financial, software, media and data. Company mainly involve in the financial software tools which provides the analytics and equality trading platform for data services and news. Where in Bloomberg briefs its deliver news, market data, and commentary directly to inbox which give competitive advantage that's changing faster than ever. Bloomberg Briefs is unique and provide downloadable reports to the subscriber and financial professional to get the actionable ideas and insights into a diverse range of markets and industries. Specifically, I was working with the Bloomberg Media division for the Bloomberg Terminal platform which caters to near half millions of corporate customers in more than 160 countries. Instead of using third party application we were using the Hadoop ecosystem to manage the large scale of data with low-latency which is processed daily on this platform. Responsibilities:

Experience in Transform, Stage and Store data using Spark which includes writing Spark applications in Scala. Developed Spark core and Spark SQL scripts using Scala for faster data processing. Developed a data pipeline using Kafka and Storm to store data into HDFS. Developed Spark code using Scala and Spark - SQL for faster testing and data processing. Setting up and managing Kafka and Zookeeper for Stream processing. Developed Kafka consumer's API in Scala for consuming data from Kafka topics. Configured, deployed and maintained multi-node Dev and Test Kafka Clusters. Developed Spark scripts using Scala, Spark

SQL to access Hive tables in Spark for faster data processing    Stored real-time data from Spark on HBase for future analysis.    Involved in implementing High Availability and automatic failover infrastructure to overcome single point of failure for Name node utilizing Zookeeper services.

Involved in analyzing data coming from various sources and creating Meta-files and control files to ingest the data in to the Data Lake.    Generate reports for analytics development through Zeppelin.

Developed Scala code with Spark Streaming for faster testing and processing of data.    Worked with Spark RDD and Dataframes for sessionization and other transformations.    Used Apache Avro to transform data between different format.    Streamed log files using Flume into HDFS and load into Hive tables to query data.    Created multiple Hive tables, implemented partitioning, dynamic partitioning and bucketing in Hive for efficient data access.    Performed daily ad-hoc data analysis and pulled data from Hadoop using Hive (HiveQL).    Worked closely with team members, managers and other teams in Agile Scrum environment.    Environment: Kafka 0.10.1, Spark 2.1.0, Scala 2.12.0, Sqoop 1.4.6, Avro, MySql5.x, HBase 1.3.0, Zeppelin, Hive 2.1.1, HDFS, Spark SQL, Flume 1.7.0, HiveQL, Zookeeper, Agile.

Hadoop Developer IBM - New York, NY October 2012 to April 2015  
Project: Explore Solutions    Project Description: IBM Watson is a distinctively creating a business unit around the World. Company headquartered in the United States, where it specializes in the sector of Discovery Advisor, Engagement Advisor and Explorer and building the application of business, smart cities, consumer applications and life in general.    Using Hadoop Eco-system, we have been developing predictive models to identify customer Insight for Banking by making use of real-time and historical data of bank user like mortgage, credit history, income source, spending and reviews. The project is about analysis of data by proper cleaning and normalization process that combine predictive and cognitive capabilities. We need to collect data from various resources and enables dynamic solution behavioral segmentation to uncover actionable customer insights allowing banks to create personalized sales offerings and marketing campaigns. Capability to store large unstructured data sets in NoSQL databases and using spark to analyze this data.    Responsibilities:

Worked on performance and optimization of existing algorithms in Hadoop using Spark context, Spark-SQL and Spark YARN using Scala.    Experience in installation, upgrading, configuration,

monitoring supporting and managing in Hadoop clusters using Cloudera CDH 4, CDH 5, Hortonworks HDP 2.x, and 3x on Ubuntu, RedHat, CentOS systems    Worked on components of CDH and HDP including HDFS, MapReduce, Job tracker, Task tracker, Sqoop, Zookeeper, YARN, Oozie, Hive, Hue, Flume, HBase, Fair Scheduler, Spark and Kafka    Deployed Hadoop clusters on public and private cloud environments like AWS and OpenStack    Administered the Linux systems to deploy Hadoop cluster and monitoring the cluster using Nagios and Ganglia    Experienced in performing backup, recovery, failover and DR practices on multiple platforms    Implemented Kerberos and LDAP authentication of all the services across Hadoop clusters using Apache Sentry

Worked with highly transactional merchandise and investment in SQL databases with PCI, HIPAA compliance involving data encryption with certificates and security keys at various levels

Experienced in automating the provisioning processes and system resources using Puppet

Implemented Hadoop-based solutions to store archives and backups from multiple sources.

Involved in importing and exporting data with Sqoop from RDBMS MySQL, Oracle, Teradata and used fast loaders and connectors; and processed data for commercial analytics    Built ingestion framework using flume for streaming logs and aggregating the data into HDFS    Experienced in upgrading SQL Server software, patches and service packs    Practiced Agile scrum to provide operational support, installation updates, patches and version upgrades    Experienced in collaborative platforms including Jira, Rally, SharePoint and Discovery    Installed, monitored and performance tuned standalone multi-node clusters of Kafka    Successfully loaded files to Hive and HDFS from MongoDB, Cassandra, and HBase    Experienced in understanding and managing Hadoop log files with Flume    Experienced on SQL DBA in HA and DDR like replication, log shipping, mirroring and clustering and database security and permissions    Responsible for Importing and exporting data into HDFS and Hive using Sqoop from Oracle 11g and MySQL.

Develop and maintain several batch jobs to run automatically depending on business requirements.

Responsible for building scalable distributed data solutions using Hadoop.    Involved in loading data from edge node to HDFS using shell scripting.    Implemented Partitioning, Dynamic Partitioning, Buckets in Hive.    Developed PIG scripts using Pig Latin and worked on tuning the

performance of Pig queries. Involved in managing and reviewing Hadoop log files. Created HBase tables to store variable data formats of PII data coming from different portfolios. Implemented test scripts to support test driven development and continuous integration. Exported the analyzed data to the relational databases like MySQL using Sqoop for visualization and to generate reports for the BI team. Created MapReduce programs using Java API that filter un-necessary records and find out unique records based on different criteria. Experienced in implementing POC's to migrate iterative map reduce programs into Spark transformations, actions using Scala. Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop. Worked in Test Driven Deployment environment and used Confluence for documentation. Performed unit testing using MRunit. Installed Oozie workflow engine to run multiple Hive and pig jobs. Used Jira for project tracking, Bug tracking and Project Management. Environment: Spark 2.1.0, Apache Hadoop, MapReduce, HDFS, Hive 2.1.1, Java, Pig 0.16.0, Sqoop 1.4.6, MRunit, Oozie 3.1.3, HBase1.3.0, TDD, MySQL5.x, Oracle 11g Java Developer Axis Bank - Gujarat, IN October 2010 to September 2012 Project: Customer Assistance Axis is one of the leading global banks. Customer Assistance is Banking Application for Axis Bank. The Customer Assistance Application plays a crucial role in daily-banking operations. This Application makes its users to create bank accounts, transfer money, update contact history, and update checking and saving accounts, credit card approvals of the customers. Responsibilities: Gathered user requirements, analyzed and wrote functional and technical specifications, we use SharePoint to maintain all of our design docs. Followed Agile software methodology for software development (3 week Sprint). Worked on one of the UI based application and client focus more on look and feel of the UI. We use lots of customs components to design the UI. Chase standards and HTML, CSS, JavaScript, AJAX, EXTJS is being used intensively. Used Spring MVC framework on the server side for creating RESTFul web services by giving JSON out and modifying the DOM object on UI, by making HTTP calls and used GET and PUT. Developed multiple Controller, Service & DAO classes to interact with data layer and developed Entity classes based on the table structure. Involved in core integration to pick the file from FTP location and bring them into our

staging tables and did all the validation on the java side. Created multiple mid-tier services to interact with multiple validations and worked on entitlements services to do user validations. Interaction and also worked on applying security systems. Worked with Oracle database and used Hibernate (ORM). Created POJO/Data Objects in mid-tier service. Hands on experience on implementing lazy loading, first and second level of caching. Leading on-shore offshore model. Coordinating with Offshore team in India and being flexible on gathering updates from team. Work closely with Database team and testing team. Worked on Jasper reports using iReport tool and integrated that JRXML into spring framework. Wrote SQL commands and Stored Procedures to retrieve data from Oracle database. Worked to plug this procedure in Java classes. Used Spring web-flow for MVC pattern. Used Apache-tiles for JSP page fragments for various flows. Used SOAP Web-Services for sending data to Published New Services for given WSDL file. GUI developing using custom JSTL tag library and used AJAX calls for client side Http-Requests. Followed Agile Methodology and regular SCRUM meetings. Involved in creating the Hibernate 3.0 POJO Objects and mapped using Hibernate Annotations. Used Hibernate, object/relational-mapping (ORM) solution, technique of mapping data representation from MVC model to Oracle Relational data model with a SQL-based schema. Provide support to the users for all the service components and help them in production issues. Involved in designing test plans, test cases and overall Unit testing of the system using JUnit. Environment: Java 1.7, JSP, Eclipse, JUnit, Hibernate 3.0, Oracle, Maven, Restful, Git, Scrum, Spring-Web Flow, SQL Java Developer.

elInfochip - Gujarat, IN July 2009 to September 2010 Project: Online Invoice System (OVS)

elInfochips is a Product Engineering and Software R&D Services firm based that offers solutions in software, hardware and VLSI services. The company provides a range of products and services spanning multiple industries including silicon engineering, embedded systems, software engineering, extended services. Where I was involved in software engineering which offers custom software application development and maintenance services with focus on shorter development cycles, enhanced features, better user experience. We also help our clients to migrate applications to open source technologies, while providing a benefit analysis. Responsibilities: Software



Development Life-Cycle (SDLC) phases of design, development, implementation, deployment, testing and maintenance as per quality standards using Agile, Scrum and waterfall methodologies.

Good Experience in Application Software Development and Design, Object Oriented, Technical Documentation, Software Testing and Debugging. Excellent implementation knowledge of JDK 1.6/1.7 and 1.8, Spring, Hibernate, RESTFUL Web Services, AOP, Struts, JDBC, EJB. Consumed and Exposed both REST and SOAP based webservices very good experience with SOA model.

Experience in RDBMS using MySQL, Oracle, SQL Server, PostgreSQL. Involved in configuring, deploying applications on IBM WebSphere Application Server, BEA WebLogic Application Server, Apache Tomcat on UNIX, Linux and Windows platforms. Extensive experience in Design, Development and implementation of Model-View-Controller (MVC) using Spring. Good experience in Database Design, writing stored procedure, functions, triggers, SQL queries. Developed the code for front end using EXTJS, AngularJS JQuery, JavaScript, AJAX, HTML, CSS and JSON.

Good knowledge on TCP/IP tunneling and port management on cloud environments, installation of servers on cloud unix. Good experience on production support and client interaction. Experience in creating build scripts using Ant and Maven also have experience in Jenkins. Strong TDD (test driven development) and continuous integration experience using JUnit, Mock Framework. Worked on Cucumber framework Implemented this application based on MVC Architecture using open source Struts. Used UML to create class diagrams, Sequence diagrams, and State diagrams and implemented these diagrams in Microsoft Visio. Developed front end web pages using Servlets 3.0, HTML, JSP, JavaScript, Swing. Responsible for presentation layers using JSP custom tags and JavaScript. Developed Server-Side Validations using Struts Validation Framework. Created a Transaction History Web Service using SOAP that is used for internal communication in the workflow process. Used Core java concepts in application such as multithreaded programming, Synchronization of threads used thread wait, notify, join methods. Developed Data Access Objects for accessing Relational Database. Designed test cases for unit testing with the help of JUnit. Created database connection using JDBC classes for interacting with Oracle 9i database.

Environment: Java 6/7, JSP, Servlets 3.0, JDBC, Swing, Java Beans 3.0, JavaScript, HTML, Struts,

Oracle 9i, Multithreading, Data Access Objects, SOAP, UML, Junit Education Bachelor's Skills  
DYNAMODB, CASSANDRA, AMBARI, MAPREDUCE, OOZIE, SQOOP, HBASE, KAFKA, FLUME,  
HADOOP, MONGODB, DATA ANALYSIS, DATABASE, MYSQL, ORACLE, PL/SQL, SQL, C++, Git,  
Hadoop Additional Information TECHNICAL SKILLS: Hadoop Ecosystem Database Hadoop  
2.7.X, Spark 2.1.0, Mapreduce, Hive Spark SQL, Cassandra 3.10, HBase 1.3.0, 2.1.1, Sqoop  
1.99.7, Kafka 2.1.0, Oozie 3.1.3, MongoDB, Oracle 12c/11g/8i, MySQL 5.x, Yarn 0.21.3, Pig 0.16.0,  
Flume 1.7.0, Ambari PL/SQL Programming Language AWS Java, SQL, JavaScript, Scala, R, C++  
EC2, S3, EMR, Lambda, Elastic Beanstalk, ELB, EBS, DynamoDB, ElastiCache, SQS, SNS, SWF,  
Testing CloudWatch, CloudFormation, IAM, CloudTrail JUnit, MRUnit, Pytest, ScalaTest Glacier,  
Storage Gateway Scripting Language Web Framework Scala 2.12.0, Unix Shell, Html5, Xml,  
CSS3, JavaScript, jQuery, HTML5, CSS3, J2EE, Spring, JSP, SQL MVC, Hibernate 3.0 Data  
Analysis & Visualization Web Services Tableau 9.3, D3.js SOAP, Restful Environment  
Code/Build/Deployment WinSCP, Putty, SQL Developer, Agile Git, Svn, Maven, Jenkins, Jira 6.4,  
Confluence Operating Systems Tools Mac OS, Ubuntu, Centos, Windows Hive, MongoDB, Spark,  
Tableau, Cloudera, Hue

Name: Michael Jones

Email: lisaramirez@example.com

Phone: 001-285-430-1522x186