

Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - 84.51 Over 7 years of professional IT experience with 3+ Years of Big Data Hadoop Ecosystems experience in ingestion, storage, querying, processing and analysis of big data. Hands-on experience architecting and implementing Hadoop clusters on Amazon (AWS), using EMR, S2, S3, Redshift, Cassandra, MangoDB, CosmosDB, SimpleDB, AmazonRDS, DynamoDB, Postgresql., SQL, MS SQL. Experience in Hadoop Administration activities such as installation, configuration, and management of clusters in Cloudera (CDH4, CDH5), & Hortonworks (HDP) Distributions using Cloudera Manager & Ambari. Hands on experience in installing, configuring, and using Hadoop ecosystem components like HDFS, MapReduce, Hive, Impala, Sqoop, Pig, Oozie, Zookeeper, Spark, Solr, Hue, Flume, Storm, Kafka and Yarn distributions. Very good Knowledge and experience in Amazon AWS concepts like EMR and EC2 web services which provides fast and efficient processing of Big Data. Experienced in performance tuning of Yarn, Spark, and Hive and experienced in developing MapReduce Programs using Apache Hadoop for analyzing the big data as per the requirement. Extensively worked on Spark using Scala on cluster for computational (analytics), installed it on top of Hadoop performed advanced analytical application by making use of Spark with Hive and SQL/Oracle. Expert in big data ecosystem using Hadoop, Spark, Kafka with column-oriented big data systems on cloud platforms such as Amazon CCloud (AWS), Microsoft Azure and Google Cloud Platform. Experienced in importing & exporting data between HDFS and Relational Database Management systems using Sqoop and troubleshooting for any issues. Extensive experience in working with various distributions of Hadoop like enterprise versions of Cloudera (CDH4/CDH5), Hortonworks and good knowledge on MapR distribution, IBM Big Insights and Amazon's EMR (Elastic MapReduce) Exposure to Data Lake Implementation using Apache Spark and developed Data pipe lines and applied business logics using Spark and used Scala and Python to convert Hive/SQL queries into RDD transformations in Apache Spark. Good Understanding and experience on NameNode HA architecture and experience in monitoring the health of cluster using Ambari, Nagios, Ganglia and Cronjobs. Experienced in Cluster maintenance and Commissioning /Decommissioning of Data Nodes and good understanding/knowledge of Hadoop Architecture and

various components such as HDFS, Job Tracker, and Task Tracker, NameNode, DataNode and MapReduce concepts. Experienced in implementation of security controls using Kerberos principals, ACLs, Data encryptions using DM-Crypt to protect entire Hadoop clusters. Well-versed in spark components like Spark SQL, MLib, Spark streaming and GraphX. Expertise in installation, administration, patches, upgrade, configuration, performance tuning and troubleshooting of Red hat Linux, SUSE, CentOS, AIX, Solaris. Experienced Schedule Recurring Hadoop Jobs with Apache Oozie and experience in Jumpstart, Kickstart, Infrastructure setup and Installation Methods for Linux. Good knowledge in troubleshooting skills, understanding of system's capacity, bottlenecks, basics of memory, CPU, OS, storage, and network. Experience in administration activities of RDBMS data bases, such as MS SQL Server. Experienced in Hadoop Distributed File System and Ecosystem (MapReduce, Pig, Hive, Sqoop, YARN, MongoDB and HBase) and knowledge of NoSQL databases such as HBase, Cassandra and MongoDB. Major strengths are familiarity with multiple software systems, ability to learn quickly new technologies, adapt to new environments, focused adaptive and quick learner with excellent interpersonal, technical and communication skills.

Work Experience Sr. Hadoop Developer 84.51 - Cincinnati, OH June 2018 to Present Roles & Responsibilities: Worked directly with the Big Data Architecture Team which created the foundation of this Enterprise Analytics initiative in a Hadoop-based Data Lake. Created multi-node Hadoop and Spark clusters in AWS instances to generate terabytes of data and stored it in AWS HDFS. Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Zookeeper and Sqoop. Extracted real time feed using Kafka and Spark streaming and convert it to RDD and process data in the form of Data Frame and save the data as Parquet format in HDFS. Developed data pipeline using Flume, Sqoop, Pig and Java map reduce to ingest customer behavioral data and financial histories into HDFS for analysis. Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis. Upgraded the Hadoop cluster from CDH4.7 to CDH5.2 and worked on installing cluster, commissioning & decommissioning of Data Nodes, NameNode recovery, capacity planning, and slots configuration. Developed Spark scripts to import large files

from Amazon S3 buckets and imported the data from different sources like HDFS/HBase into Spark RDD. Involved in migration of ETL processes from Oracle to Hive to test the easy data manipulation and worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop. Worked on Installing Cloudera Manager, CDH and install the JCE Policy File to Create a Kerberos Principal for the Cloudera Manager Server, enabling Kerberos Using the Wizard. Developed Spark jobs using Scala and Python on top of Yarn/MRv2 for interactive and Batch Analysis. Monitored cluster for performance and, networking and data integrity issues and responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files. Created 25+ Linux Bash scripts for users, groups, data distribution, capacity planning, and system monitoring. Install OS and administrated Hadoop stack with CDH5 (with YARN) Cloudera distribution including configuration management, monitoring, debugging, and performance tuning. Supported MapReduce Programs and distributed applications running on the Hadoop cluster and scripting Hadoop package installation and configuration to support fully-automated deployments. Migrated existing on-premises application to AWS and used AWS services like EC2 and S3 for large data sets processing and storage and worked with ELASTIC MAPREDUCE and setup Hadoop environment in AWS EC2 Instances. Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters. Perform maintenance, monitoring, deployments, and upgrades across infrastructure that supports all our Hadoop clusters and worked on Hive for further analysis and for generating transforming files from different analytical formats to text files. Created Hive External tables and loaded the data in to tables and query data using HQL and worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required. Worked and learned a great deal from AWS Cloud services like EC2, S3, EBS, RDS and VPC. Monitoring Hadoop cluster using tools like Nagios, Ganglia, and Cloudera Manager and maintaining the Cluster by adding and removing of nodes using tools like Ganglia, Nagios, and Cloudera Manager. Worked on Hive for exposing data for further analysis and for generating transforming files from different analytical formats to text files. Worked on migrating MapReduce programs into

Spark transformations using Spark and Scala, initially done using python (PySpark) Environment: Hadoop, MapReduce, Hive, PIG, Sqoop, Python, Spark, Spark-Streaming, Spark SQL, AWS EMR, AWS S3, AWS Redshift, Python, Scala, Pyspark, MapR, Java, Oozie, Flume, HBase, Nagios, Ganglia, Hue, Cloudera Manager, Zookeeper, Cloudera, Oracle, Kerberos and RedHat 6.5. Sr. Hadoop Developer AMEX - Phoenix, AZ May 2018 to May 2018 Roles & Responsibilities:

Collaborate in identifying the current problems, constraints and root causes with data sets to identify the descriptive and predictive solution with support of the Hadoop HDFS, MapReduce, Pig, Hive, and Hbase and further to develop reports in Tableau. Architect the Hadoop cluster in Pseudo distributed Mode working with Zookeeper and Apache and storing and loading the data from HDFS to Amazon AWS S3 and backing up and Created tables in AWS cluster with S3 storage. Evaluated existing infrastructure, systems, and technologies and provided gap analysis, and documented requirements, evaluation, and recommendations of system, upgrades, technologies and created proposed architecture and specifications along with recommendations

Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Zookeeper and Sqoop. Installed and Configured Sqoop to import and export the data into MapR-FS, HBase and Hive from Relational databases. Administering large MapR Hadoop environments build and support cluster set up, performance tuning and monitoring in an enterprise environment. Installed and Configured MapR-zookeeper, MapR-cldb, MapP-jobtracker, MapR-tasktracker, MapR-resourcemanager, MapR-node manager, MapR-fileserver, and MapR-webserver. Installed and configured Knox gateway to secure HIVE through ODBC, WebHcat and Oozie services. Load data from relational databases into MapR-FS filesystem and HBase using Sqoop and setting up MapR metrics with NoSQL database to log metrics data. Close monitoring and analysis of the MapReduce job executions on cluster at task level and optimized Hadoop clusters components to achieve high performance. Developed data pipeline using Flume, Sqoop, Pig and Java MapReduce to ingest customer behavioral data into HDFS for analysis. Integrated HDP clusters with Active Directory and enabled Kerberos for Authentication. Worked on commissioning & decommissioning of Data Nodes, NameNode

recovery, capacity planning and installed Oozie workflow engine to run multiple Hive and Pig Jobs. Worked on creating the Data Model for HBase from the current Oracle Data model. Implemented High Availability and automatic failover infrastructure to overcome single point of failure for Name node utilizing zookeeper services. Leveraged Chef to manage and maintain builds in various environments and planned for hardware and software installation on production cluster and communicated with multiple teams to get it done. Monitoring the Hadoop cluster functioning through MCS and worked on NoSQL databases including HBase. Used Hive and created Hive tables and involved in data loading and writing Hive UDFs and worked with Linux server admin team in administering the server hardware and operating system. Worked closely with data analysts to construct creative solutions for their analysis tasks and managed and reviewed Hadoop and HBase log files. Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports and worked on importing and exporting data from Oracle into HDFS and HIVE using Sqoop. Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades when required. Automated workflows using shell scripts pull data from various databases into Hadoop. Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop. Environment: Hadoop, HDFS, Map Reduce, Hive, HBase, Kafka, Zookeeper, Oozie, Impala, Java, Cloudera, Oracle, Teradata SQL Server, Python, UNIX Shell Scripting, ETL, Flume, Scala, Spark, Sqoop, Python, AWS, S3, EC2, Kafka, Oracle, MySQL, Hortonworks, YARN, Python Hadoop Developer Charles Schwab - Littleton, CO January 2016 to May 2017 Roles & Responsibilities: Evaluated suitability of Hadoop and its ecosystem to the above project and implementing / validating with various proof of concept (POC) applications to eventually adopt them to benefit from the Big Data Hadoop initiative. Architected Hadoop system pulling data from Linux systems and RDBMS database on a regular basis in order to ingest data using Sqoop. Installation, configuration, supporting and managing Hadoop Clusters using Apache, Cloudera (CDH4) distributions and on Amazon web services (AWS) Developed MapReduce programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in the EDW. Worked on Cloudera Hadoop Upgrades and Patches and

Installation of Ecosystem Products through Cloudera manager along with Cloudera Manager Upgrade. Created Hive queries that helped market analysts spot emerging trends by comparing fresh data with EDW reference tables and historical metrics. Setting an Amazon Web Services (AWS) EC2 instance for the Cloudera Manager server. Enabled speedy reviews and first mover advantages by using Oozie to automate data loading into the Hadoop Distributed File System and PIG to pre-process the data. Provided design recommendations and thought leadership to sponsors/stakeholders that improved review processes and resolved technical problems. Identify query duplication, complexity and dependency to minimize migration efforts Technology stack: Oracle, Hortonworks HDP cluster, Attunity Visibility, Cloudera Navigator Optimizer, AWS Cloud and Dynamo DB. Shared responsibility for administration of Hadoop, Hive, and Pig and managed and reviewed Hadoop log files and updating the configuration on each host. Worked with Spark ecosystem using Scala, Python and HIVE Queries on different data formats like Text file and parquet. Tested raw data and executed performance scripts and configuring Cloudera Manager Agent heartbeat interval and timeouts. Worked with teams in setting up AWS EC2 instances by using different AWS services like S3, EBS, Elastic Load Balancer, and Auto scaling groups, VPC subnets and CloudWatch. Implemented CDH3 Hadoop cluster on RedHat Enterprise Linux 6.4, assisted with performance tuning and monitoring. Monitoring Hadoop Cluster through Cloudera Manager and Implementing alerts based on Error messages. Used Spark Streaming API with Kafka to build live dashboards; Worked on Transformations & actions in RDD, Spark Streaming, Pair RDD Operations, Check-pointing, and SBT. Providing reports to management on Cluster Usage Metrics and related HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL, and a variety of portfolios. Involved in different Hadoop distributions like Cloudera (CDH3 & CDH4) and Horton Works Distributions (HDP) and MapR. Performed installation, upgrade and configure tasks for impala on all machines in a cluster and supported code/design analysis, strategy development and project planning. Created reports for the BI team using Sqoop to export data into HDFS and Hive and assisted with data capacity planning and node forecasting. Managing Amazon Web Services (AWS) infrastructure with automation and

configuration. Administrator for Pig, Hive and HBase installing updates, patches, and upgrades and performed both major and minor upgrades to the existing CDH cluster and upgraded the Hadoop cluster from CDH3 to CDH4. Developed a process for the Batch ingestion of CSV Files, Sqoop from different sources and also generating views on the data source using Shell Scripting and Python. Environment:: Hadoop, HDFS, Map Reduce, Hive, HBase, Zookeeper, Impala, Java (jdk1.6), Cloudera, Oracle, SQL Server, UNIX Shell Scripting, Flume, Oozie, Scala, Spark, ETL, Sqoop, Python, kafka, PySpark, AWS, S3, MongoDB, Oracle, SQL, Hortonworks, XML, RedHat Linux 6.4. Java Developer Fresh Desk - Chennai, Tamil Nadu November 2015 to November 2015

**Roles & Responsibilities:** Implemented Multi-Threaded Environment and used most of the interfaces under the collection framework by using Core Java Concepts. Developed Graphical User Interfaces by using JSF, JSP, HTML, DHTML, Angularjs, CSS, and JavaScript and developed scripts in python for Financial Data coming from SQL Developer based on the requirements specified. Implemented several Java/J2EE design patterns like Spring MVC, Singleton, Spring Dependency Injection and Data Transfer Object. Used JAX-WS (SOAP) for producing web services and involved in writing programs to consume the web services using SOA with CXF framework and developed few web pages using JSP, JSTL, HTML, CSS, Java script, Ajax and JSON. Implemented business logic, data exchange, XML processing and created graphics using Python and Django. Wrote code to fetch data from Web services using JQUERY AJAX via JSON response and updating the HTML pages and developed high traffic web applications using HTML, CSS, and JavaScript, jQuery, Bootstrap, Ext JS, AngularJS, Node.js and react.js. Write SQL queries and create PL/SQL functions/procedures/packages that are optimized for APEX and improve performance and response times of APEX pages and reports Used JQuery library, NodeJS and AngularJS for creation of powerful dynamic WebPages and web applications by using its advanced and cross browser functionality. Used Java Server Pages for content layout and presentation with Python and Extracted and loaded data using Python scripts and PL/SQL packages

Worked with various frameworks of JavaScript like BackboneJS, AngularJS, and EmberJS etc. Written with object-oriented Python, Flask, SQL, Beautiful Soup, httpLib2, Jinja2, HTML/CSS,

Bootstrap, jQuery, Linux, Sublime Text, GIT. Developed GUI using JSP, Struts, HTML3, CSS3, XHTML, JQuery, Swing and JavaScript to simplify the complexities of the application. Wrote and executed various MYSQL database queries from python using Python-MYSQL connector and MySQLdb package and generated Python Django forms to record data of online users and used PyTest for writing test cases. Developed and coordinated complex high quality solutions to clients using J2SE, J2EE, Servlets, JSP, HTML, Struts, Spring MVC, SOAP, JavaScript, JQuery, JSON and XML. Exposed business functionality to external systems (Interoperable clients) using Web Services (WSDL-SOAP) ApacheAxis. Used Hibernate, object/relational-mapping (ORM) solution, technique of mapping data representation from MVC model to Oracle10g Relational data model with a SQL-based schema and mapped using Hibernate Annotations. Skilled in using collections in Python for manipulating and looping through different user defined objects. Designed and developed intranet web applications using Ext JS, React.js, JavaScript and CSS and developed Merge jobs in Python to extract and load data into MySQL and Mango Db database. Worked on Oracle & SQL Server as the backend databases and integrated with Hibernate to retrieve Data Access Objects. Developed Python batch processors to consume and produce various feeds and developed entire frontend and backend modules using Python on Django Web Framework and developed Business Logic using Python on Django Web Framework. Designed and documented REST/HTTP APIs, including JSON data formats and API versioning strategy and wrote Apex triggers, apex classes, developing Visual force pages, batch classes. Involved in AJAX driven application by invoking web services/API and parsing the JSON response and involved in writing application level code to interact with APIs, Web Services using JSON. Environment:Python, Django, Java, JSF MVC, Spring IOC, APEX, Ruby on Rails, Spring JDBC, Hibernate, ActiveMQ, Log4j, Ant, MySQL, JDK 1.6, J2EE, JSP, Servlets, HTML, LDAP, Salesforce, ESB Mule, JDBC, MongoDB, DAO, EJB 3.0, PL/SQL, react.js, Web Sphere, Eclipse, Angular.JS, and CVS. Education Bachelor's Skills JAVA (2 years), LINUX (2 years), ORACLE (2 years), PYTHON (2 years), SQL (2 years) Additional Information Technical Competencies: Hadoop Ecosystem Tools:MapReduce, HDFS, Pig, Hive, HBase, Sqoop, Zookeeper, Oozie, Hue, Storm, Kafka, Spark, Flume Languages:



Java, core java, HTML, Programming C, C++ Databases: MySQL, Oracle [ ] SQL Server, MongoDB  
Platforms: Linux (RHEL, Ubuntu, ), open Solaris, AIX Scripting Languages: Shell Scripting, HTML  
scripting, Python, Puppet Web Servers: Apache Tomcat, JBOSS, windows server2003, 2008, and  
2012 Cluster Management Tools: HDP Ambari, Cloudera Manager, Hue, Solr Cloud

Name: Sara Fischer

Email: rachelschultz@example.net

Phone: 224-282-8181