

Big data/Hadoop Developer Big data/Hadoop Developer Big data/Hadoop Developer - Bank of America Avenel, NJ Above 8+ years of experience in Analysis, Design, Development, Testing, Implementation, Maintenance and Enhancements on various IT Projects and experience in Big Data in implementing end-to-end Hadoop solutions. Experience in working in environments using Agile (SCRUM), RUP and Test-Driven development methodologies. Good Knowledge in Amazon AWS concepts like EMR and EC2 web services which provides fast and efficient processing Extensive experience in installing, configuring and using ecosystem components like Hadoop MapReduce, HDFS, Sqoop, Pig, Hive, Impala & Spark Expertise in using J2EE application servers such as IBM Web Sphere, JBoss and web servers like Apache Tomcat. Experience in different Hadoop distributions like Cloudera (CDH3 & CDH4) and Horton Works Distributions (HDP). Experience in analyzing data using HIVEQL, PIG Latin and custom Map Reduce programs in JAVA. Extending HIVE and PIG core functionality by using custom UDF's. Experienced in configuring and administering the Hadoop Cluster using major Hadoop Distributions like Apache Hadoop and Cloudera. Diverse experience utilizing Java tools in business, Web, and client-server environments including Java Platform, J2EE, EJB, JSP, Java Servlets, Struts, and Java database Connectivity (JDBC) technologies. Good understanding in integration of various data sources like RDBMS, Spreadsheets, Text files, JSON and XML files. Good working experience on using Sqoop to import data into HDFS from RDBMS and vice-versa. Implemented Service Oriented Architecture (SOA) using Web Services and JMS ( Java Messaging Service). Implemented J2EE Design Patterns such as MVC, Session Fa ade, DAO, DTO, Singleton Pattern, Front Controller and Business Delegate. Experienced in developing web services with XML based protocols such as SOAP, Axis, UDDI and WSDL. Experienced in MVC (Model View Controller) architecture and various J2EE design patterns like singleton and factory design patterns. Extensive experience in loading and analyzing large datasets with Hadoop framework (Map Reduce, HDFS, PIG, HIVE, Flume, Sqoop, SPARK, Impala), NoSQL databases like Mongo DB, HBase, Cassandra. Solid understanding of Hadoop MRV1 and Hadoop MRV2 (or) YARN Architecture. Hands on experience in configuring and administering the Hadoop Cluster using major Hadoop Distributions like Apache

Hadoop and Cloudera. Good knowledge in SQL and PL/SQL to write Stored Procedures and Functions and writing unit test cases using JUnit. Extensive experience in Extraction, Transformation and Loading (ETL) of data from multiple sources into Data Warehouse and Data Mart. Strong knowledge in Object oriented design/analysis, UML modeling, Classic design patterns, and J2EE patterns. Hands on experience working with databases like Oracle 12g, SQL Server 2010 and MySQL. Hands on experience on the entire latest UI stack including HTML, CSS, mobile friendly, responsive design, user-centric design etc. Experience in developing web-based enterprise applications using Java, J2EE, Servlets, JSP, EJB, JDBC, Hibernate, Spring IOC, Spring AOP, Spring MVC, Spring Web Flow, Spring Boot, Spring Security, Spring Batch, Spring Integration, Web Services (SOAP and REST) and ORM frameworks like Hibernate. Strong knowledge on Hadoop ecosystem systems including HDFS, Hive, Oozie, HBase, Pig, Sqoop, Zookeeper etc. Extensive experience with advanced J2EE Frameworks such as spring, Struts, JSF and Hibernate. Expertise in JavaScript, JavaScriptMVC patterns, ObjectOrientedJavaScriptDesign Patterns and AJAX calls. Installation, configuration and administration experience in Big Data platforms Cloudera Manager of Cloudera, MCS of MapR. Expertise in using XML related technologies such as XML, DTD, XSD, XPATH, XSLT, DOM, SAX, JAXP, JSON and JAXB. Experience in using ANT and Maven for building and deploying the projects in servers and also using Junit and log4j for debugging. Work Experience Big data/Hadoop Developer Bank of America August 2018 to Present Description: Bank of America Merrill Lynch" is the marketing name for the global banking and global markets businesses of Bank of America Corporation. Lending, derivatives and other commercial banking activities are performed globally by banking affiliates of Bank of America Corporation, including Bank of America, N.A., and Member FDIC. Responsibilities: Responsible for installation and configuration of Hive, Pig, Hbase and Sqoop on the Hadoop cluster and created hive tables to store the processed results in a tabular format. Configured Spark Streaming to receive real time data from the Apache Kafka and store the stream data to HDFS using Scala. Developed the Sqoop scripts to make the interaction between Hive and vertica Database. Processed data into HDFS by developing solutions and analyzed the data using Map Reduce, PIG, and Hive to produce summary

results from Hadoop to downstream systems. Build servers using AWS: Importing volumes, launching EC2, creating security groups, auto-scaling, load balancers, Route 53, SES and SNS in the defined virtual private connection. Written Map Reduce code to process and parsing the data from various sources and storing parsed data into HBase and Hive using HBase-Hive Integration. Streamed AWS log group into Lambda function to create service now incident. Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running Hive queries and Pig scripts. Created Managed tables and External tables in Hive and loaded data from HDFS. Developed Spark code by using Scala and Spark-SQL for faster processing and testing and performed complex HiveQL queries on Hive tables. Scheduled several times based Oozie workflow by developing Python scripts. Developed Pig Latin scripts using operators such as LOAD, STORE, DUMP, FILTER, DISTINCT, FOREACH, GENERATE, GROUP, COGROUP, ORDER, LIMIT, UNION, SPLIT to extract data from data files to load into HDFS. Exporting the data using Sqoop to RDBMS servers and processed that data for ETL operations. Worked on S3 buckets on AWS to store Cloud Formation Templates and worked on AWS to create EC2 instances. Designing ETL Data Pipeline flow to ingest the data from RDBMS source to Hadoop using shell script, sqoop, package and MySQL. End-to-end architecture and implementation of client-server systems using Scala, Akka, Java, JavaScript and related, Linux Optimized the Hive tables using optimization techniques like partitions and bucketing to provide better. Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive, Pig, and Sqoop. Implementing Hadoop with the AWS EC2 system using a few instances in gathering and analyzing data log files. Involved in Spark and Spark Streaming creating RDD's, applying operations -Transformation and Actions. Created partitioned tables and loaded data using both static partition and dynamic partition method. Developed custom Apache Spark programs in Scala to analyze and transform unstructured data. Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from Oracle into HDFS using Sqoop Using Kafka on publish-subscribe messaging as a distributed commit log,

have experienced in its fast, scalable and durability. Test Driven Development (TDD) process and extensive experience with Agile and SCRUM programming methodology. Implemented POC to migrate Map Reduce jobs into Spark RDD transformations using SCALA. Scheduled map reduces jobs in production environment using Oozie scheduler. Involved in Cluster maintenance, Cluster Monitoring and Troubleshooting, Manage and review data backups and log files. Designed and implemented map reduce jobs to support distributed processing using java, Hive and Apache Pig. Analyzing Hadoop cluster and different Big Data analytic tools including Pig, Hive, HBase and Sqoop. Improved the Performance by tuning of HIVE and map reduce. Research, evaluate and utilize modern technologies/tools/frameworks around Hadoop ecosystem. Environment: HDFS, Map Reduce, Hive, Sqoop, Pig, Flume, Vertica, Oozie Scheduler, Java, Shell Scripts, Teradata, Oracle, HBase, MongoDB, Cassandra, Cloudera, AWS, JavaScript, JSP, Kafka, Spark, Scala and ETL, Python.

Big data/Hadoop Developer Sprint - Overland Park, KS May 2017 to July 2018

Description: Sprint offers all the latest "must-have" smartphones and tablets from top industry-leading manufacturers including Apple, Samsung, LG, HTC, exclusively offering the new Essential-Phone and much more. Sprint is also your one-stop-shop for the coolest accessories from fashion-forward designers and tech innovators.

Responsibilities: Involved in Installing, Configuring Hadoop Eco-System, Cloudera Manager using CDH3, CDH4 Distributions. Responsible to manage data coming from different sources and involved in HDFS maintenance and loading of structured and unstructured data. Imported data using Sqoop from Teradata using Teradata connector. Created Data Pipeline of Map-Reduce programs using Chained Mappers. Implemented Optimized join base by joining different data sets to get top claims based on state using Map Reduce. Visualize the HDFS data to the customer using BI tool with the help of HiveODBCDriver. Worked on POC of Talend integration with Hadoop where Created Talend Jobs to extract data from Hadoop. Imported data using Sqoop to load data from MySQL to HDFS on a regular basis. Worked on social media (Facebook, Twitter etc) data crawling using Java and R language and MongoDB for unstructured data storage. Integrated Quartz scheduler with Oozieworkflows to get data from multiple data sources parallels using a fork. Created Partitions,

Buckets based on State to further process using Bucket based Hive joins. Experienced with different kind of compression techniques like LZO, GZip, Snappy. Created Hive Generic UDF's to process business logic that varies based on policy. Imported Relational Database data using Sqoop into Hive Dynamic partition tables using staging tables. Worked on custom Pig Loaders and storage classes to work with a variety of data formats such as JSON and XML file formats. Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive, Pig, and Sqoop. Developed Unit test cases using JUnit, Easy Mock, and MRUnit testing frameworks. Experienced in Monitoring Cluster using Cloudera Manager. Environment: Hadoop, HDFS, HBase, Spark, MapReduce, Tera Data, MySQL, Java, Hive, Pig, Sqoop, Flume, Oozie, SQL, Cloudera Manager. Hadoop Developer Center Light Health - Bronx, NY January 2016 to April 2017 Description: CenterLight Healthcare, a member of CenterLight Health System, is a not-for-profit New York State Managed Long Term Care organization. Founded in 1985 to provide home healthcare services to the elderly, disabled and chronically ill in the New York metropolitan area, CenterLight Healthcare today is recognized as an innovator in the managed care field. Responsibilities: Provided application demo to the client by designing and developing a search engine, report analysis trends, application administration prototype screens using AngularJS, and Bootstrap JS. Took the ownership of complete application Design of Java part, Hadoop integration. Apart from the normal requirement gathering, participated in a Business meeting with the client to gather security requirements. Assisted with the architect to analyze the existing system and future system. Prepared design blue prints and application flow documentation. Experienced in managing and reviewing Hadoop log files. Load and transform large sets of structured, semi-structured and unstructured data. Responsible to manage data coming from different sources and application. Supported Map Reduce Programs those are running on the cluster. Responsible for working with Message broker system such as Kafka. Extracted data from mainframes and feed to KAFKA and ingested to HBase to perform Analytics. Written event-driven, link tracking system to capture user events and feed to KAFKA to push it to HBASE. Created MapReduce jobs to extract the contents from HBase and configured in OOZIE workflow to

generate analytical reports. Developed the JAX-RS web services code using Apache CXF framework to fetch data from SOLR when the user performed the search for documents. Participated in SOLR schema, and ingested data into SOLR for data indexing. Written MapReduce programs to organize the data, and ingest the data to suitable for analytics in client specified format. Hands on experience in writing python scripts to optimize the performance. Implemented Storm builder topologies to perform cleansing operations before moving data into Cassandra. Extracted files from Cassandra through Sqoop and placed in HDFS and processed. Implemented Bloom filters in Cassandra using keyspace creation. Involved in writing Cassandra CQL statements. Good hands-on experience in developing concurrency using spark and Cassandra together. Involved in writing spark applications using Scala. Hands on experience in creating RDDs, transformations, and Actions while implementing spark applications. Good knowledge in creating data frames using Spark SQL. Involved in loading data into Cassandra NoSQL Database. Implemented record level atomicity on writes using Cassandra. Written PIG Scripts to query and process the Datasets to figure out the patterns of trends by applying client-specific criteria, and configured OOZIE workflows to run the jobs along with the MR jobs. Stored the derived results in HBase from analysis and make it available to data ingestion for SOLR for indexing data. Involved in integration of java search UI, SOLR and HDFS. Involved in code deployments using continuous integration tool using Jenkins. Documented all the challenges, issues involved to deal with the security system and Implemented best practices. Created Project structures and configurations according to the project architecture and made it available to the junior developer to continue their work. Handled onsite coordinator role to deliver work to offshore. Involved in core reviews and application lead supported activities. Environment: Cassandra, Spring 3.2, Restful services using CXF web services framework, spring data, SOLR 5.2.1, PIG, HIVE, apache AVRO, Map Reduce, Sqoop Zookeeper, SVN, Jenkins, windows AD, windows KDC, Hortonworks distribution of Hadoop 2.3, YARN, Ambari Hadoop developer. Engage Point - Calverton, MD March 2014 to December 2015. Description: Engage Point chosen as Agile Development Pre-qualified Vendor by California Health and Human Services Agency. Responsibilities: Collected and aggregated large amounts of

weblog data from different sources such as web servers, mobile and network devices using Apache Flume and stored the data into HDFS for analysis. Collecting data from various Flume agents that are imported on various servers using Multi-hop flow. Ingest real-time and near-real-time (NRT) streaming data into HDFS using Flume. Extensively involved in Installation and configuration of Cloudera distribution Hadoop, Name node, Secondary Name Node, Job Tracker, Task Trackers, and Data Nodes. Developed Map Reduce programs in Java and Sqoop the data from ORACLE database. Responsible for building Scalable distributed data solutions using Hadoop. Written various Hive and Pig scripts. Used various HBase commands and generated different Datasets as per requirements and provided access to the data when required using grant and revoke. Created HBase tables to store variable data formats of input data coming from different portfolios. Worked on HBase for support enterprise production and loading data into HBase using SQOOP. Installed Oozie workflow engine to run multiple Hive and Pig jobs which run independently with time and data availability. Expertise in understanding Partitions, Bucketing concepts in Hive. Experience working with Apache SOLR for indexing and querying. Created custom SOLR Query segments to optimize ideal search matching. Used Oozie Scheduler system to automate the pipeline workflow and orchestrate the Map Reduce Jobs that extract the data in a timely manner. Responsible for loading data from the UNIX file system to HDFS. Developed suit of Unit Test Cases for Mapper, Reducer and Driver classes using MR Testing library. Analyzed the weblog data using the HiveQL, integrated Oozie with the rest of the Hadoop stack Utilized cluster co-ordination services through Zookeeper. Worked on the Ingestion of Files into HDFS from remote systems using MFT. Got good experience with various NoSQL databases and Comprehensive knowledge in process improvement, normalization/de-normalization, data extraction, data cleansing, and data manipulation. Developed Pig scripts to convert the data from Text file to Avro format. Created Partitioned Hive tables and worked on them using HiveQL. Developed Shell scripts to automate routine DBA tasks. Used Maven extensively for building jar files of Map Reduce programs and deployed to Cluster. Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, managing and reviewing data backups and Hadoop log files.

Environment: HDFS, MapReduce, Pig, Hive, Oozie, Sqoop, Flume, HBase, Java, Maven, Avro, Cloudera, Eclipse and Shell Scripting. Java Developer Financial Technologies Limited - Chennai, Tamil Nadu December 2012 to February 2014 Description: Financial Technologies (India) Ltd. is a world leader in providing next-generation technology ventures, innovations, platforms, and solutions for creating digital markets and marketplaces that enable price discovery and transaction efficiencies across industry segments. Responsibilities: Involved in gathering system requirements for the application and worked with the business team to review the requirements, and went through the Software Requirement Specification document and Architecture document Involved in intense User Interface (UI) operations and client-side validations using AJAX toolkit. Used SOAP to expose company applications as a Web Service to outside clients. Log package is used for the debugging. Used Clear Case for version control. Ensuring adherence to delivery schedules and quality process on projects. Used Web Services for creating rate summary and used WSDL and SOAP messages for getting insurance plans from the different module and used XML parsers for data retrieval. Developed business components and integrated those using Spring features such as Dependency Injection, Auto wiring components such as DAO layers and service proxy layers. Used Spring AOP to implement Distributed declarative transaction throughout the application. Wrote Hibernate configuration XML files to manage data persistence. Used TOAD to generate SQL queries for the applications, and to see the reports from log tables. Involved in the migration of Data from Excel, Flat file, Oracle, XML files to SQL Server by using BCP and DTS utility.

Environment: Java/J2EE, MVC Arch with CICS interaction, HTML, Axis, SOAP, Servlets, Web services, Restful Web Services, Sybase, Spring, DB2, RAD, Rational Clear case, WCF, AJAX, Toad. Java Developer Rockwell Collins - Hyderabad, Telangana January 2011 to November 2012 Description: Rockwell Collins (NYSE: COL) is a leader in aviation and high-integrity solutions for commercial and military customers around the world. Every day we help pilots safely and reliably navigate to the far corners of the earth; keep warfighters aware and informed in battle; deliver millions of messages for airlines and airports; and help passengers stay connected and comfortable throughout their journey. Responsibilities: Responsible for developing various modules, front-end



and back-end components using several design patterns based on the client's business requirements. Designed and Developed application modules using spring and Hibernate frameworks. Designed and developed the front-end with Swings and Spring MVC framework, Tag libraries and Custom Tag Libraries and development of Presentation Tier using JSP pages integrating AJAX, Custom Tag's, JSP Tag Lists, HTML, JavaScript, and JQuery. Used Hibernate to develop persistent classes following ORM principles. Deployed spring configuration files such as application context, application resources, and application files. Used Java-J2EE patterns like Model View Controller (MVC), Business Delegate, Session facade, Service Locator, Data Transfer Objects, Data Access Objects, Singleton and factory patterns. Used JUnit for Testing Java Classes. Used Waterfall methodology. Worked with Maven for build scripts and Setup the Log4J Logging framework. Involved in the Integration of the Application with other services. Involved in Units integration, bug fixing, and testing with test cases. Fixed the bugs reported in User Testing and deployed the changes to the server. Managing the version control for the deliverables by streamlining and re-basing the development streams of the SVN. Environment: Java/JDK, J2EE, Spring 2.5, Spring MVC, Hibernate, Eclipse, Tomcat, XML, JSTL, JavaScript, Maven2, Web Services, JQuery, SVN, JUnit, Log4J, Windows, Oracle. Development Methodologies 2010 to 2012 M, Ozii, Hue, SOAP UI Reporting Tools MS Office (Word/Excel/PowerPoint/ Visio/Outlook), Crystal Reports XI, SSRS, Cognos 7.0/6.0. Databases Microsoft SQL Server 2008,2010/2012, MySQL 4.x/5.x, Oracle 11g, 12c, DB2, Teradata, Netezza Operating Systems All versions of Windows, UNIX, LINUX, Macintosh HD, Sun Solaris Education Bachelor's Skills Cassandra, Hdfs, Impala, Mapreduce, Oozie, Sqoop, Hbase, Kafka, Flume, Hadoop, MongoDB, Splunk, Tableau server, C++, Hadoop, Hbase, Hive, Json, Mapreduce, Pig Additional Information TECHNICAL SKILLS: Hadoop Technologies Hadoop, HDFS, YARN, MapReduce, Hive, Pig, Impala, Sqoop, Flume, Spark, Kafka, Storm, Drill, Zookeeper, and Oozie Languages HTML5, DHTML, WSDL, CSS3, C, C++, XML, R/R Studio, SAS, Schemas, JSON, Ajax, Java, Scala, Python, Shell Scripting Big Data Platforms Hortonworks, Cloudera. NO SQL Databases Cassandra, HBase, MongoDB, MariaDB Business Intelligence Tools Tableau server, Tableau Reader, Tableau, Splunk, SAP

Business Objects, QlikView, Amazon Redshift, or Azure Data Warehouse Development Tools  
Microsoft SQL Studio, IntelliJ, Eclipse, NetBeans.

Name: Hannah Mason

Email: diana90@example.org

Phone: 301.533.6934x663