

Hadoop developer Hadoop developer Hadoop developer - Humac Inc Albany, NY Around 8+ years of experience in Information Technology with Hands on experience in all the stages of system development efforts, including requirement definition, design, implementation, testing and documentation. Experience in Hadoop ecosystem including the state of Art technologies. Worked in a highly dynamic team using agile methodologies like Scrum and waterfall. A quick learner, punctual and trustworthy. Good working experience on Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming. Techno functional responsibilities include interfacing with users, identifying technical and functional gaps, estimates, designing, developing, producing documentation and extensive Production support. Integrated various Big Data Technologies into the overall Experiences in analyzing, developing, testing and optimizing data transformation processes using Hadoop components. Proficiency in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice versa. Extensive Experience in Developing and maintaining Big Data streaming applications using Kafka, Storm, Spark and other Hadoop Components. Experience in column-family based Databases HBase and Accumulo. Skilled in Creating, scheduling and maintaining Workflows in UC4. Procedural knowledge on cleansing and analysing data using Hive, Presto on Hadoop Platform and also on Relational databases such as Oracle, SQL, Teradata and MongoDB. Extensively worked on creating Teradata Bteq Scripts and used Informatica to load data into Teradata. Preparation of Standard Code guidelines, analysis and testing documentations. In depth understanding of data structures and algorithms. Worked on various IDEs Eclipse, IntelliJ and repositories Git and SVN. Hands on Experience in designing and developing applications in Spark using Scala to compare performance of Spark with Hive. Involved in ooptimizing existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's. Good knowledge on installing, configuring, and using Hadoop components like Map Reduce, HDFS, Hive, Sqoop, Pig, Zookeeper and Flume. In depth knowledge of database like SQL, MySQL and extensive experience in writing SQL queries, Stored Procedures, Triggers, Cursors, Functions and Packages. Proficient in writing Shell, JavaScript and Python Scripts.

Good understanding on Installing and maintaining the Linux servers. Experience in Monitoring System Metrics and logs for any problems adding, removing, or updating user account information, resetting passwords, etc. Authorized to work in the US for any employer Work Experience Hadoop developer Humac Inc - Phoenix, AZ December 2016 to Present Responsibilities: Analyzed various Relational Databases such as Oracle, SQL, Teradata, and MongoDB to understand the source systems and develop application to migrate the data to Hadoop Environment. Implemented several Batch Ingestion jobs for Historical data migration from various relational databases and files using Sqoop. Hands on Developing a Near Real-Time Framework using Kafka, storm to ingest data from several source systems like Oracle, SQL, Teradata and MongoDB into Hadoop Environment. Involved in implementing a Real-time framework to capture Streaming data and store in HDFS using Kafka, Spark. Developed Kafka consumer component for near real-time and Real-Time data processing in Java and Scala. Defined multiple Kafka Topics with several Partitions and replication factors across data centers. Part of designing and developing a custom Java daemon to pull data from source systems and publish the resultant to a specific Kafka Topic. Expertise in integrating Kafka with Storm and Spark streaming for near real-time and real-time Frameworks. Migrated streaming data to HDFS, Accumulo by creating Hdfs Bolt and Accumulo Bolt in Storm. Successfully migrated complex SQL transformations in Pentaho that belong to Xfinity Home Security, to Hadoop using Spark and Scala. Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's. Responsible for creating Hive tables and views using Partitions, HQL Scripts in landing layer for analytics. Also, created tables in Accumulo Database and built external tables on top of them in hive using Accumulo Connectors for data analysis. Analyzed massive Data sets in Hadoop ranging in Gigabytes using Hive, Accumulo and Presto. Developed Data archival and data purge in HDFS using Shell and Python scripts, UC4 workflow to automate the process. Defined UC4 workflows for running sequential job flows in Production Responsible for creating Deployment SMOP and LLDs Documents. Created a Wiki page and migrated code to Git repository. Wrote tested and implemented Teradata Bteq Scripts, DML and DDL which is transactional data of Xfinity Mobile.

Performance tuned and optimized various SQL queries. Hands on experience on Informatica to load data to Teradata by making various connections to load and extract data to and from Teradata efficiently. Environment: Hortonworks Hadoop 2.6, HDFS, Hive 2.4.3, Presto, Spark 2.2, Scala 2.11.8, Kafka 0.9, Apache Storm 0.10, Accumulo, Sqoop, UC4, Rest API, Java (1.7 & 1.8), Shell Scripting, Python Scripting, MySQL Oracle 11g, MongoDB, Teradata 15, SQL, Informatica 10.1.1.

Hadoop Developer State Street - Quincy, MA October 2015 to December 2016 Responsibilities:

Involved in requirement gathering to setup a cluster. Part of Configuring Hadoop cluster and load balancing across the nodes. Developed MapReduce programs in Java for parsing the raw data and populating staging Tables. Created Hive queries to compare the raw data with EDW reference tables and performing aggregates Experienced in developing custom input formats and data types to parse and process unstructured and semi structured input data and mapped them into key value pairs to implement business logic in Map-Reduce. Experience in implementing custom sterilizer, interceptor, source and sink as per the requirement in Flume to ingest data from multiple sources. Experience in setting up Fan-out workflow in flume to design v shaped architecture to take data from many sources and ingest into single sink. Importing and exporting data into HDFS and Hive using Sqoop Experienced in analyzing data with Hive and Pig Experienced knowledge over designing Restful services using java based API's like JERSEY. Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS Integrating bulk data into Cassandra file system using MapReduce programs Expertise in designing, data modeling for Cassandra NoSQL database Experienced in managing and reviewing Hadoop log files Defined multiple job flows using Oozie workflow. Involved in working with Spark on top of Yarn/MRv2 for interactive and Batch Analysis Worked closely with AWS EC2 infrastructure teams to troubleshoot complex issues

Expertise in writing the Scala code using higher order functions for the iterative algorithms in spark for performance consideration Experience in managing and monitoring Hadoop cluster using Cloudera Manager. Experienced in analyzing and Optimizing RDD's by controlling partitions for the given data Good understanding on DAG cycle for entire spark application flow on Spark application WebUI Experienced in writing live Real-time Processing using Spark Streaming with

Kafka    Developed custom mappers in python script and Hive UDFs and UDAFs based on the given requirement    Used HiveQL to analyze the partitioned and bucketed data and compute various metrics for reporting    Supported in setting up QA environment and updating configurations for implementing scripts with Pig, Hive and Sqoop    Unit tested a sample of raw data and improved performance and turned over to production    Environment: CDH, Java(JDK1.7), Hadoop, MapReduce, HDFS, Hive, Sqoop, Flume, Cassandra, Pig, Oozie, Kerberos, Scala, Spark SQL, Spark Streaming, Kafka, Linux, AWS, Shell Scripting, MySQL Oracle 11g, SQL\*PLUS Hadoop Developer Geometric Ltd August 2013 to October 2015 Responsibilities:    Installed Name node, Secondary name node, (Resource Manager, Node manager, Application master), Data node using Cloudera.    Installed and configured Hortonworks Ambari for easy management of existing Hadoop cluster, Installed and Configured HDP.    Installed and configured multi-nodes fully distributed Hadoop cluster of large number of nodes.    Provided Hadoop, OS, Hardware optimizations. Setting up the machines with Network Control, Static IP, Disabled Firewalls, Swap memory. Understanding the performance bottlenecks by analyzing the existing hadoop cluster and provided performance tuning accordingly.    Regular Commissioning and Decommissioning of nodes depending upon the amount of data.    Installed and configured Hadoop components Hdfs, Hive, HBase.    Communicating with the development teams and attending daily meetings.    Addressing and Troubleshooting issues on a daily basis.    Working with data delivery teams to setup new Hadoop users. This job includes setting up Linux users, setting up Kerberos principals and testing HDFS, Hive.    Cluster maintenance as well as creation and removal of nodes.    Monitor Hadoop cluster connectivity and security.    Manage and review Hadoop log files.    Configured the cluster to achieve the optimal results by fine-tuning the cluster.    Dumped the data from one cluster to other cluster by using DISTCP, and automated the dumping procedure using shell scripts.    Designed the shell script for backing up of important metadata and rotating the logs on a monthly basis. Implemented open source monitoring tool GANGLIA for monitoring the various services across the cluster.    Testing, evaluation and troubleshooting of different NoSQL database systems and cluster configurations to ensure high-availability in various crash scenarios.    Performance tuning and

stress testing of NoSQL database environments to ensure acceptable database performance in production mode. Designed the cluster so that only one secondary name node daemon could be run at any given time. Implemented commissioning and decommissioning of data nodes, killing the unresponsive task tracker and dealing with blacklisted task trackers. Dumped the data from HDFS to MYSQL database and vice-versa using SGOOP. Provided the necessary support to the ETL team when required. Integrated Nagios in the Hadoop cluster for alerts. Performed both major and minor upgrades to the existing cluster and rolling back to the previous version. Environment: LINUX, HDFS, MapReduce, KDC, NAGIOS, GANGLIA, OOZIE, SGOOP, Cloudera Manager.

Teradata Developer Enquero - Milpitas, CA June 2011 to August 2013 Responsibilities:

Extensively used ETL to load data from Oracle and Flat files to Data Warehouse Extensively worked in data Extraction, Transformation and Loading from source to target system using power center of Informatica. Developed complex mappings in Informatica to load the data from various sources. Implemented performance tuning logic on targets, sources, mappings, sessions to provide maximum efficiency and performance. Parameterized the mappings and increased the re-usability. Used Informatica Power Center Workflow manager to create sessions, workflows and batches to run with the logic embedded in the mappings. Created procedures to truncate data in the target before the session run. Extensively used Toad utility for executing SQL scripts and worked on SQL for enhancing the performance of the conversion mapping. Used the PL/SQL procedures for Informatica mappings for truncating the data in target tables at run time. Worked on Teradata RDBMS using Fast load, Multi load, Tump, Fast export, Multi load Export, Teradata Sql and Bteq Teradata utilities. Involved in Performance Tuning at various levels including Target, Source, Mapping, and Session for large data files. Extracted data from various source systems like Oracle, SQL Server and flat files as per the requirements. Performed bulk data load from multiple data source (ORACLE 8i, legacy systems) to Teradata RDBMS using BTEQ, Multi Load and Fast Load. Created, optimized, reviewed, and executed Teradata SQL test queries to validate transformation rules used in source to target mappings/source views, and to verify data in target tables Performed tuning and optimization of complex SQL queries using Teradata Explain.

Responsible for Collect Statics on FACT tables. Design and development of the complete Decision Support System using Business Objects. Worked on Migration Strategies between Development, Test and Production Repositories. Supported the Quality Assurance team in testing and validating the Informatica workflows. Extensively involved in development of mappings using various transformations of Informatica according to business logic. Created and Scheduled Sessions and Batches using Server Manager. Created and Monitor the sessions using workflow manager and workflow monitor. Conducting unit testing. Environment: Informatica Power Centre 8.6, UNIX, Teradata, Oracle 8i, TOAD. Java Developer Aroghia - Aurora, CO April 2009 to June 2011 Responsibilities: Developed the spring AOP programming to configure logging for the application. Expertise in developing enterprise applications using Struts Frameworks. Developed the front end using JSF and Portlet. Developed Scalable applications using Stateless session EJBs. Developed the UI panels using JSF, XHTML, CSS, DOJO and JQuery. MySQL to access data in the database at different Levels. Making a connection to backend MySQL database. Design and Developed using Web Service using Apache Axis. wrote numerous session and message driven beans for operation on JBoss and WebLogic. Used VSS (Visual Source Safe) as configuration management tool. Created automated test cases using Selenium. Worked with SDLC process like water fall model, AGILE methodology. JSP interfaces were developed. Custom tags were used. Developed Servlets and Worked extensively on Sql. Used ANT for building the application and deployed on BEA WebLogic Application Server. Was responsible for Developing XML Parsing logic using SAX/DOM Parsers. Good network at EMC Documentum Support Teams who help solve product issues and bugs. Worked on tickets from service-now and Jira on daily basis. Designed the front end using Swing. Used IBM MQ Series in the project. Apache Tomcat Server was used to deploy the application. Involving in Building the modules in Linux environment with ant script. Used Resource Manager to schedule the job in UNIX server. Used web services (REST) to bridge the gap between our MS and Drupal/Word press technology. Design online stores using ASP & JavaScript: develop custom storefront applications, and custom user-interfaces for client sites. J2EE to communicate legacy COBOL based mainframe implementations.

Worked on PL/SQL and SQL queries    Developed Java Script and Action Script, VB Script macros for Client Side validations.    Environment: Spring, Struts, JSF, EJBs, JQuery, MySQL, DB2, Net Beans, JBoss, CVS, VSS, water fall model, UML, JSP, Servlets, ANT, XML, EMC, Jira, IBM MQ, Tomcat Server, Linux, Unix server Education Bachelor of Technology in Technology Texas A&M University Skills MYSQL (6 years), SQL (6 years), LINUX (5 years), APACHE HADOOP HDFS (4 years), APACHE HADOOP SQOOP (4 years) Additional Information Technical Skills:    Big Data Technologies Hadoop, MapReduce, HDFS, Hive, Pig, Sqoop, Flume, solr, Kafka, Spark, Storm Reporting Tools Jaspersoft, Qlik Sense, Tableau    Scripting Languages Python, Shell, R Programming Languages C, C++, Java    Web Technologies HTML, J2EE, CSS, JavaScript, AJAX, Servlets, JSP, DOM, XML, XSLT.    Application Server WebLogic Server, Apache Tomcat.    DB Languages SQL, PL/SQL, Postgres.    NoSQL Databases HBase, Cassandra, Accumulo    Databases /ETL Oracle 10g/11g, MySQL 5.2, DB2, Informatica v 8.x, Talend    Operating Systems Linux, UNIX, Windows 2003 Server    IDE's Eclipse, NetBeans JDeveloper, IntelliJ IDEA.    Version Control CVS, SVN, Git

Name: Jessica Simpson

Email: elizabethgolden@example.org

Phone: 736-367-0215