

Sr. Hadoop/Scala Developer Sr. Hadoop/Scala Developer Sr. Hadoop/Scala Developer - Safeway Phoenix, AZ Having 8+ years of Experience in IT industry in Designing, Developing and Maintaining using Big data Technologies like Hadoop, Spark Ecosystems and Java/J2EE Technologies. 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. Excellent Programming skills at a higher level of abstraction using Scala, Java and Python. Extensive experience in working with various distributions of Hadoop enterprise versions of Cloudera(CDH4/CDH5), Horton works and good knowledge on MAPR distribution, IBM Big Insights and Amazon's EMR (Elastic Map Reduce). Working knowledge of Amazon's Elastic Cloud Compute(EC2) infrastructure for computational tasks and Simple Storage Service (S3) as Storage mechanism. Experienced in implementing scheduler using Oozie, Airflow, Crontab and Shell scripts. Good working experience in importing data using Sqoop, SFTP from various sources like RDMS, Teradata, Mainframes, Oracle, Netezza to HDFS and performed transformations on it using Hive, Pig and Spark. Extensive experience in importing and exporting streaming data into HDFS using stream processing platforms like Flume and Kafka messaging system. Strong experience and knowledge of real time data analytics using Spark Streaming, Kafka and Flume. Extensively worked on Spark streaming and Apache Kafka to fetch live stream data. Expertise in writing SparkRDD transformations, Actions, Data Frames, Case classes for the required input data and performed the data transformations using Spark-Core. Experience in integrating Hive queries into Spark environment using Spark SQL. Expertise in performing real time analytics on big data using HBase and Cassandra. Developed customized UDFs and UDAFs in java to extend Pig and Hive core functionality. Proficient in NoSQL databases including HBase, Cassandra, MongoDB and its integration with Hadoop cluster. Good experience in optimizing Map Reduce algorithms using Mappers, Reducers, combiners and practitioners to deliver the best results for the large datasets. Extracted data from various data source including OLEDB, Excel, Flat files and XML. Experienced in using build tools like Ant, SBT, Log4j, Maven to build and deploy applications into the server. Had competency in using Chef, Puppet and Ansible configuration and automation

tools. Configured and administered CI tools like Jenkins, Hudson Bambino for automated builds. Proficient in developing, deploying and managing the SOLR from development to production. Experience in Enterprise search using SOLR to implement full text search with advanced text analysis, faceted search, filtering using advanced features like dismax, extended dismax and grouping. Worked on data warehousing and ETL tools like Informatica, Talend, and Pentaho. Designed ETL workflows on Tableau, Deployed data from various sources to HDFS. Working experience on Test Data Management tools HP Quality Center, HPALM, Load Runner, QTP and Selenium. Worked on ELK stack like Elastic search, Logstash, Kibana for log management. Worked on various programming languages using IDEs like Eclipse, NetBeans, and IntelliJ. Experience in Software Design, Development and Implementation of Client/Server Web based Applications using JSTL, jQuery, JavaScript, Java Beans, JDBC, Struts, PL/SQL, SQL, HTML, CSS, PHP, XML, AJAX and had a bird's eye view on React Java Script Library. Used various Project Management services like JIRA for tracking issues, bugs related to code and GitHub for various code reviews and Worked on various version control tools like CVS, GIT, PVCS, SVN. Experience with best practices of Web services development and Integration (both REST and SOAP). Generated various kinds of knowledge reports using Power BI and Qlik based on Business specification. Experience in automated scripts using Unix shell scripting to perform database activities. Experience in complete Software Development Life Cycle (SDLC) in both Waterfall and Agile methodologies. Good understanding of all aspects of Testing such as Unit, Regression, Agile, White-box, Black-box. Authorized to work in the US for any employer Work Experience Sr. Hadoop/Scala Developer Safeway - Phoenix, AZ February 2016 to Present Responsibilities: Experienced in designing and deployment of Hadoop cluster and different Big Data analytic tools including Pig, Hive, Cassandra, Oozie, Sqoop, Kafka, Spark, Impala with Horton works distribution Performed source data transformations using Hive. Supporting infrastructure environment comprising of RHEL and Solaris. Involved in developing a Map Reduce framework that filters bad and unnecessary records. Developed Spark scripts by using Scala shell commands as per the requirement. Used Kafka to transfer data from different data systems to HDFS. Created Spark

jobs to see trends in data usage by users. Responsible for generating actionable insights from complex data to drive real business results for various application teams. Designed the Column families in Cassandra. Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandra as per the business requirement. Developed Spark code to using Scala and Spark-SQL for faster processing and testing. Experience in NoSQL Column-Oriented Databases like Cassandra and its Integration with Hadoop cluster. Used Spark API over Hadoop YARN as execution engine for data analytics using Hive. Exported the analyzed data to the relational databases using Sqoop to further visualize and generate reports for the BI team. Good experience with Talend open studio for designing ETL Jobs for Processing of data. Experience in processing large volume of data and skills in parallel execution of process using Talend functionality. Worked on different file formats like Text files and Avro. Created various kinds of reports using Power BI and Tableau based on the client's needs. Worked on Agile Methodology projects extensively. Experience designing and executing time driven and data driven Oozie workflows. Setting up Kerberos principals and testing HDFS, Hive, Pig, and MapReduce access for the new users. Experienced in working with Spark eco system using SCALA and HIVE Queries on different data formats like Text file and parquet. Collected the logs data from web servers and integrated in to HDFS using Flume. Developed Hive scripts in Hive QL to de-normalize and aggregate the data. Implemented map-reduce counters to gather metrics of good records and bad records. Work experience with cloud infrastructure like Amazon Web Services (AWS). Developed customized UDF's in java to extend Hive and Pig functionality. Jobs via Zeppelin notebooks; mentored and guided offshore team in troubleshooting and fine-tuning Spark. Experience in importing data from various data sources like Mainframes, Teradata, Oracle and Netezza using Sqoop, SFTP, performed transformations using Hive, Pig and Spark and loaded data into HDFS. Extracted the data from Teradata into HDFS/Databases/Dashboards using SPARK STREAMING. Implemented best income logic using Pig scripts. Worked on different file formats (ORCFIELD, Parquet, Avro) and different Compression Codecs (GZIP, SNAPPY, LZO). Created applications using Kafka, which monitors consumer lag within Apache Kafka clusters. Used

in production by multiple companies. Using Spark-Streaming APIs to perform transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra. Design and document REST/HTTP, SOAP APIs, including JSON data formats and API versioning strategy. Experience in using Apache Kafka for collecting, aggregating, and moving large amounts of data from application servers. Used Hibernate ORM framework with Spring framework for data persistence and transaction management. Performance analysis of Spark streaming and batch jobs by using Spark tuning parameters. Used React Bindings for embracing Redux. Worked towards creating real time data streaming solutions using Apache Spark/Spark Streaming, Kafka. Worked along with the Hadoop Operations team in Hadoop cluster planning, installation, maintenance, monitoring and upgrades. Used File System check (FSCK) to check the health of files in HDFS. Worked in Agile development environment in sprint cycles of two weeks by dividing and organizing tasks. Participated in daily scrum and other design related meetings. Environment: Hadoop, Hive, Map Reduce, Sqoop, Kafka, Spark, Yarn, Pig, Cassandra, Oozie, shell Scripting, Scala, Maven, Java, React JS, JUnit, agile methodologies, Horton works, Soap, NIFI, Teradata, MySQL. Sr. Spark Developer Digital Realty Trust November 2014 to February 2016 Responsibilities: Developed data pipeline using SPARK, Apache Kafka to ingest customer behavioral data and financial histories into HADOOP cluster for analysis. Collected data using SPARK STREAMING from AWS S3 bucket in near-real-time and performed necessary Transformations and Aggregation to build the common learner data model and persist the data in HDFS. Explored the usage of SPARK for improving the performance and optimization of the existing algorithms in HADOOP using SPARK. Worked with SPARK STREAMING to ingest data into Spark Engine. Imported the data from different sources like AWS S3, Local file system into Spark RDD. Experience working with various services in Azure like Data lake to store and analyze the data. Hands on experience configuring Hadoop cluster in professional environment and on Azure using HDInsight. Developed Spark applications using SCALA with SPARK-SQL/STREAMINGAPI for faster testing and processing of data. Involved in converting Hive/SQL queries into Spark Transformations using SPARK RDDs and SCALA.

Worked on the SPARK SQL and Spark Streaming modules of SPARK and used SCALA to write code for all Spark use cases. Developed Spark code using Scala and Spark -SQL for faster testing and data processing. Worked on converting PL/SQL code into Scala code and converted PL/SQL queries into Hive queries. Used IMPALA to give parallel processing database technology on top of Hadoop eco-system. Monitored cluster Health status and assisted in planned Hadoop cluster maintenance activities such as upgrades, expansions, configurations etc. Worked closely with infrastructure, network, database, and business intelligence and application teams to ensure business applications are highly available and performing within agreed on service levels. Configure Flume to ingest log file data into HDFS. Involved in using SQOOP for importing and exporting data between RDBMS and HDFS. Handle data exchange between HDFS and RDBMS. Write Spark applications in Scala to interact with MYSQL database using Spark SQL. Extensively used Hive for ETL Transformations and optimized Hive Queries. Experience in Hive partitioning, bucketing and collections, perform different types of joins on Hive tables. Performed hive performance tuning aspects like Map Join, Cost Based Optimization and Columnar Level Statistics.

Good Experience in loading data from LINUX file system to HDFS. Used ECLIPSE and Sbt to build the application. Environment: Spark, Hive, Hive UDFs, SparkSql, Spark Streaming, Spark Yarn, Spark Sql, HBase, Sqoop, Kafka, Aws Ec2, S3, Cloudera, Scala, IDE(Eclipse), Linux Shell Scripting, Hdfs, Azure Hadoop/ETL Developer Novartis Corporation October 2013 to November 2014 Responsibilities: 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. Moved data from HDFS to Cassandra using Map Reduce and Bulk Output Format class. Experienced with different scripting language like Python and shell scripts. Developed various Python scripts to find vulnerabilities with SQL Queries by doing SQL injection, permission checks and performance analysis. Installed Oozie workflow engine to run multiple Hive and Pig jobs which run independently with time and data availability. Experienced with handling

administration activations using Cloudera manager. Expertise in understanding Partitions, Bucketing concepts in Hive. Used Oozie Scheduler system to automate the pipeline workflow and orchestrate the Map Reduces jobs that extract the data on a timely manner. Responsible for loading data from UNIX file system to HDFS. Analyzed the weblog data using the HiveQL, integrated Oozie with the rest of the Hadoop stack Utilized cluster co-ordination services through Zookeeper.

Got good experience with various NoSQL databases and Comprehensive knowledge in process improvement, normalization/de-normalization, data extraction, data cleansing, data manipulation. Experience with creating script for data modeling and data import and export. Extensive experience in deploying, managing and developing MongoDB clusters. 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 MapReduce 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, Map Reduce, Pig, Hive, Oozie, Sqoop, Flume, HBase, Java, Maven, Avro, Cloudera, Eclipse and Shell Scripting. JAVA/ETL Developer ICICI Bank October 2011 to August 2013 Responsibilities: Developed Maven scripts to build and deploy the application. Developed Spring MVC controllers for all the modules. SAS scripts on UNIX are run, and the output datasets are exported into SAS.

Implemented jQuery validator components. Extracted data from Oracle as one of the source databases. Using Data stage ETL tool to copy data from Teradata to Netezza Created ETL Data mapping spreadsheets, describing column level transformation details to load data from Teradata Landing zone tables to the tables in Party and Policy subject area of EDW based on SAS Insurance model. Used JSON and XML documents with Mark logic NoSQL Database extensively. REST API calls are made using NodeJS and Java API. SAS data sets were constantly created and updated using the SET and UPDATE statements Built data transformation with SSIS including importing data from files. Loaded the flat files data using Informatica to the staging area. Created SHELL SCRIPTS for generic use. Environment: Java, Spring, MPP, Windows XP/NT, Informatica Power center 9.1/8.6, UNIX, Teradata, Oracle Designer, Autosys, Shell, Quality Center 10. Java Developer

Apollo Hospitals September 2010 to August 2011 Responsibilities: Involved in the analysis, design, implementation, and testing of the project. Implemented the presentation layer with HTML, XHTML and JavaScript. Developed web components using JSP, Servlets and JDBC. Implemented database using SQL Server. Implemented Spring IoC framework Developed Spring REST services for all the modules. Developed custom SAML and SOAP integration for healthcare. Validated the fields of user registration screen and login screen by writing JavaScript validations. Used DAO and JDBC for database access. Built responsive Web pages using Kendo UI mobile. Designed dynamic and multi-browser compatible pages using HTML, CSS, jQuery, JavaScript, Require JS and Kendo UI. Environment: Oracle 11g, Java 1.5, Struts, Servlets, HTML, XML, SQL, J2EE, JUnit, Tomcat 6, Java, JSP, JDBC, JavaScript, MySQL, Eclipse IDE, Rest.

Jr. Java Developer HR Portal June 2009 to July 2010 Responsibilities: Development of two modules Leave Management and Performance Appraisal Management Developed the project using Spring MVC and Hibernate Framework Developed User management and different modules on requirement using HTML, JavaScript, AngularJS, JSON and validations fixing the bugs and handling lot of Exceptions alone. Participated in the enhancement and maintenance of Personal information Management Developed Excel based reports (Leaves Status, Appraisal status) for the consumption of HR Manager Responsible for Implemented key functionalities in Leaves Management such as Auto Approval process if the Manager does not respond to the leave within 3 days of completion of the leave dates. Implemented a workflow-based solution to pass the appraisal from the Employee to Manager until HR Manager for final discussion. Participation in Unit Testing and support for QA Testing Cycles Environment: Java, J2EE, HTML, Servlets, Spring, JSON, Hibernate, AngularJS, JavaScript, Oracle, Web logic, JMS, Eclipse IDE, Maven, UNIX, Junit.

Education Bachelor's Skills JAVA (7 years), ORACLE (7 years), ETL (6 years), EXTRACT, TRANSFORM, AND LOAD (6 years), SQL (5 years) Additional Information Technical Skills: Big Data Ecosystem Hadoop, MapReduce, Pig, Hive, YARN, Kafka, Flume, Sqoop, Impala, Oozie, Zookeeper, Spark, Ambari, Mahout, MongoDB, Cassandra, Avro, Storm, Parquet and Snappy. Hadoop Distributions Cloudera (CDH3, CDH4, and CDH5), Horton works, Map R and Apache

Languages Java, Python, J ruby, SQL, HTML, DHTML, Scala, JavaScript, XML and C/C++ No SQL
Databases Cassandra, MongoDB and HBase Java Technologies Servlets, JavaBeans, JSP, JDBC,
JNDI, EJB and struts Methodology Agile, waterfall Development / Build Tools Eclipse, Ant, Maven,
IntelliJ, JUNIT and log4J. DB Languages MySQL, PL/SQL, PostgreSQL and Oracle RDBMS
Teradata, Oracle 9i,10g,11i, MS SQL Server, MySQL and DB2 Operating systems UNIX, LINUX,
Mac OS and Windows Variants ETL Tools Talend, Informatica, Pentaho

Name: Linda Chung

Email: mglenn@example.com

Phone: 441-876-5107x39469