Hadoop/spark Developer Hadoop/spark Developer Hadoop/spark Developer - BCBSA Chicago, IL

Over 4 years of IT experience including Big Data technologies, Web Application development and Business Intelligence.   Experience in deploying and managing a multi-node Cloudera Hadoop cluster with different components (MFS, NFS, CLDB, Web server, Spark, Resource Manager, Node Manager, Hive, HBase, Zookeeper, History Server)   Exposure to Spark, Spark Streaming, Scala and Implementing Spark using Scala and utilizing Data frames and Spark SQL API, Data Frames and Pair RDD's for faster processing of data.   Experienced with the Scala, Spark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Pair RDD's, Spark YARN.   Extensive experience in working with various distributions of Hadoop Enterprise versions of Cloudera good knowledge on Amazon's EMR (Elastic MapReduce)   Used AWS S3 and Local Hard Disk as underlying File System (HDFS) for Hadoop   Experience in deploying scalable Hadoop cluster on AWS using S3 as underlying file system for Hadoop.   Hands on experience in writing Python scripts.   Experience in converting SQL queries into Spark Transformations using Spark RDDs, Scala and Performed map-side joins on RDD's.   Extensive Experience on importing and exporting data using stream processing platforms like Flume and Kafka.   Hands on experience in configuring and working with Flume to load the data from multiple sources directly into HDFS.   Good knowledge in job workflow scheduling and monitoring tools like Oozie, Zookeeper.   Good working experience using Sqoop to import data into HDFS from RDBMS and vice-versa.   Experience with distributed systems, large-scale non-relational data stores, MapReduce systems, data modeling, and big data systems.   Experienced in loading data to hive partitions and creating buckets in Hive.   Experience in handling messaging services using Apache Kafka.   Experience in fine-tuning MapReduce jobs for better scalability and performance. Developed various Map Reduce applications to perform ETL workloads on terabytes of data. Experience working with Cloudera & Hortonworks Distribution of Hadoop.   Highly motivated team player with zeal to learn new technologies.   Experience in all Phases of Software Development Life Cycle (Analysis, Design, Development, Testing and Maintenance) using Waterfall and Agile methodologies Authorized to work in the US for any employer Work Experience Hadoop/spark

Developer BCBSA - Chicago, IL January 2019 to Present Project Description: This project is based on web services and FHIR response to get the prior member coverage for members when provided with member id, plan id and product id. It utilizes the Hadoop cluster for data storage and uses. Responsibilities: Work closely with the business and analytics team in gathering the system requirements.   Data ingestion into HDFS from various mainframe Db2 table using Sqoop.   Skilled on migrating the data from different databases to Hadoop HDFS and Hive using Sqoop.   Analyzed large structured datasets using Hive's data warehousing infrastructure.   Extensive Knowledge of creating manages tables and external tables in Eco system.   Implement Hive UDF's for evaluation, filtering, loading, and storing of data.   Involved in writing the Hive scripts to reduce the job execution time.   Importing & exporting historical data from Db2 to HDFS, Hive.   Monitoring Hadoop jobs on performance and productions cluster. Provided Production support for few successful runs.   Design managed and External tables in Hive to optimize performance to improve performance.   Using API interface (HUE) to query data and managed tables.   Installed Oozie workflow engine and scheduled it to run data/time dependent Hive.   Involved in Agile methodologies, daily Scrum meetings, Sprint planning.   Environment: HDFS, Hive, Sqoop Import/Export, Spark, Hue, Oozie, ETL, Datawarehouse, UNIX, Cloudera. spark Developer Chicago, IL December 2017 to November 2018 Project Description: all scripts are specialized in designing and developing the high-performance production software with state-of -art computer vision capabilities. The increase of data made the existing databases un accommodable. So, all scripts Leap wants to move it to Hadoop, where exactly we can handle massive amount of data by means of its cluster nodes and to satisfy the scaling needs of the business operation.   Responsibilities: Extracted the data from the flat files and other RDBMS databases into staging area and ingested to Hadoop.   Installed and configured Hadoop Map-Reduce, HDFS and developed multiple Map-Reduce jobs in Java for data cleansing and preprocessing.   Importing and exporting data into HDFS and Hive using Sqoop.   Responsible for Coding batch pipelines, Restful Service, Map Reduce program, Hive query's, testing, debugging, Peer code review, troubleshooting and maintain status report.   Implemented Map Reduce programs to classified data organizations into different

classifieds based on different type of records. Implemented complex map reduce programs to perform joins on the Map side using Distributed Cache in Java. Wrote Flume configuration files for importing streaming log data into HBase with Flume. Performed masking on customer sensitive data using Flume interceptors. Involved in migrating tables from RDBMS into Hive tables using SQOOP and later generate visualizations using Tableau. Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way. Installed Oozie workflow engine and scheduled it to run data/time dependent Hive and Pig jobs. Involved in Agile methodologies, daily Scrum meetings, Sprint planning. Environment: HDFS, MapReduce, Cassandra, Hive, Pig, Sqoop, Tableau, NoSQL, Shell Scripting, Maven, Git, HDP Distribution, Eclipse, Log4j, JUnit, Linux. Hadoop Developer Waukegan, IL December 2016 to November 2017 Project Description: AbbVie is a global, research-driven biopharmaceutical company committed to developing innovative advanced therapies for some of the world's most complex and critical conditions. The company's mission is to use its expertise, to markedly improve treatments across four primary therapeutic areas: immunology, oncology, virology and neuroscience. The project aims at importing supply chain data from SAP source onto HDFS. Also providing and supporting a self service portal for business users for their day-to-day operations, analysis and reporting. Responsibilities: Installed and configured Hadoop, YARN, MapReduce, Flume, HDFS (Hadoop Distributed File System), developed multiple MapReduce jobs in Python for data cleaning. Developed data pipeline using Flume, Sqoop, Pig and Python MapReduce to ingest customer behavioral data and financial histories into HDFS for analysis. Extensive experience in working with various distributions of Hadoop Enterprise versions of Cloudera good knowledge on Amazon's EMR (Elastic MapReduce) Used AWS S3 and Local Hard Disk as underlying File System (HDFS) for Hadoop Experience in deploying scalable Hadoop cluster on AWS using S3 as underlying file system for Hadoop. Developed Python scripts to extract the data from the web server output files to load into HDFS. Involved in HBASE setup and storing data into HBASE, which will be used for further analysis. Used Pig as ETL tool to do transformations, event joins and some pre-aggregations before storing the data onto HDFS. Wrote Python MapReduce scripts for

processing the unstructured data.   Load log data into HDFS using Flume. Worked extensively in creating MapReduce jobs to power data for search and aggregation.   Worked on Hive for exposing data for further analysis and for generating transforming files from different analytical formats to text files.   Responsible for creating Hive tables, loading data and writing hive queries.   Used forward engineering to create a Physical Data Model with DDL that best suits the requirements.   Worked with Sqoop to export analyzed data from HDFS environment into RDBMS for report generation and visualization purpose.   Implemented Fair schedulers on the Job tracker to share the resources of the Cluster for the Map Reduce jobs given by the users.   Maintaining and monitoring clusters. Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop. Environment: Cloudera, Cloudera Manager, HDFS, Map Reduce, Hive, Impala, Pig, Python, SQL, Sqoop, Flume, Yarn, Linux, Centos, HBase. Java Developer NebuLogic Technologies - Hyderabad, Telangana January 2014 to December 2015 Project Description: Nebulogic is a provider of business outsourcing solutions. They provide Services such as Human Capital Management, Tax & Compliance, Electronic Payment Solutions, Dealer Services and Medical Practice Services. This project is focused on Human Capital Management and particularly on Payroll Services, Human Resource Management, Talent Management, Benefits Administration, Time & Attendance, and International Solutions. This front-end application has interface with various Java applications, which are used to send and receive information to third party vendors.   Responsibilities: Involved in Full Life Cycle Development in Distributed Environment Using Java and J2EE framework.   Responsible for developing and modifying the existing service layer based on the business requirements.   Involved in designing & developing web-services using SOAP and WSDL.   Involved in database design. Created tables, views, triggers, stored procedures in SQL for data manipulation and retrieval Developed Web Services for Payment Transaction and Payment Release.   Involved in Requirement Analysis, Development and Documentation.   Developed front-end using JSP, HTML, CSS and JavaScript.   Coding for DAO Objects using JDBC (using DAO pattern).   XML and XSDs are used to define data formats.   Implemented J2EE design patterns such as singleton, DAO for

the presentation tier, business tier and Integration Tier layers of the project.   Involved in Bug fixing and functionality enhancements.   Followed coding and documentation standards and best practices.   Participated in project planning discussions and worked with team members to analyze the requirements and translate them into working software modules. Environment: Java, J2EE, JSP, SOAP, WSDL, SQL, PL/SQL, XML, JDBC, Eclipse, Windows XP, Oracle Education Master of Science in Computer Science in Computer Science Virginia international university 2018 Skills APACHE HADOOP HDFS (2 years), database (2 years), Eclipse (2 years), Java (2 years), SQL (2 years) Certifications/Licenses Driver's License Additional Information TECHNICAL SKILLS: Hadoop/MapR Ecosystem: Hive, HBase, MFS, HDFS, MapReduce, YARN, Spark, MapR Control System (MCS), Sqoop, Pig, Cloudera Manager, Zookeeper   Tool(s): Apache Tomcat 7.0, Maven, Git, Hibernate, Microsoft SQL Server Management Studio, SQL Developer, MySQL Workbench, Eclipse   Language(s): Java 7/8, Scala, C#, C/C++, JavaScript, ABAP/4, Ruby, HTML5, CSS3 Database(s): Oracle 10g/11g/12c, MySQL, Microsoft SQL Server 2005/2008 Framework(s): Spring MVC, Spring Boot, Hibernate,

Name: Maria Chandler

Email: gking@example.com

Phone: 665-539-4213x9652