Bigdata developer Bigdata developer Bigdata developer - Caterpillar LLC-Peoria-IL Peoria, IL

Having 5 years of experience in implementation, Development, Customization, integration, Testing applications using Java/J2EE technologies, Bigdata technologies like Hadoop, Apache Spark using Python, Apache Hive, Pig, Flume, Apache Kafka, Impala and support of business application system.   Worked on Cloudera Environments to Ingest, cleanse and process huge data with different file formats   Hands on with Cloudera Environment Setup, Hadoop Administration, and Configurations.   Hands on with Hadoop file systems services.   POC on Spark Projects using Spark SQL and Spark Streaming with Hadoop File System as Data Lake   Experienced in Scrum and pair programming methodologies.   Experienced in writing test cases using Junit.   Experience in writing HIVE test cases for real-time Data Analytics.   Hand on experience with Big Data Hadoop, Spark, Kafka and Cloudera Administration.   Experience on AWS Lambda services and building Lambda function and Lambda layers using python   Experience on AWS Cloud services along with building Hadoop Cluster which includes: EC2, S3, ESF, Glacier, SWF etc.   Gained knowledge on SQL server, MYSQL and 2012 R2 windows server while developing virtual machines using Hyper-V technology.   Experience in Java while working in my pervious projects.   Experience with various stages of Software Development Life cycle (SDLC) and agile   Enthusiastic to learn new tools and technology with my adaptive learning skills Work Experience Bigdata developer Caterpillar LLC-Peoria-IL February 2019 to Present Description: This project is to build various data pipelines for extraction and transformation of data provided by dealers for client Caterpillar. The pipelines are created around the data sent by dealers which are extracted and apply various transformations to cleanse the data and to provide back with business requirements. The transformed data is stored in various data storage systems which are consumed back by various dealers.   Responsibilities: Working on designing and implementing data pipelines by collaborating with Team leads, Architect and Business Analysts.   Involved in coding, testing, integration, and Deployment of various modules in a software development life cycle (SDLC).   Built and Migrated the existing data processing framework to intuitive and scalable Spark applications using Scala as for dealer data processing.   Designed and implemented high performance and high availability data processing

framework using various Hadoop Tools like Hive, Impala, Spark, Sqoop, Oozie, Snowflake and AWS.   Working on migrating existing systems to AWS services like Lambda, EMR EC2 instance Worked on data Integration and imported data from RDBMS(Oracle) into Cloudera Hadoop, HDFS and AWS.   Assisted in Data Migration from Cloudera Hadoop to AWS instance   Designed, developed and tested reusable framework and libraries using coding best practices and Improve performance.   Upgraded data ingestion workflows to store data in accessible, performant, secure and sustainable way.   Managed the ongoing data warehouse performance and service accessibility and troubleshoot ingestion, jobs and workflow Failures proactively.   Supported Hadoop platform in identifying, communicating and resolving data integration and consumption Issues.   Worked on Root Cause Analysis and Problem Management processes and assisted support team resolving issues in a data ingestion solution developed on Hadoop platform.   Built Smoke and regression test cases and performed Unit Testing on all the developed modules. Environment: Hive, Impala, Sqoop, Oozie, Spark, Scala, SQL, Agile, Java, Cloudera, AWS, Putty, Oracle. Hadoop developer Ameriprise Financial - Minneapolis, MN January 2018 to January 2019 Description: Worked as a Hadoop developer- Data implementation project that consist data migration from source systems to Hadoop environment.   Aimed to achieve high performance for near real time Data Analytics using Spark. Data lake was maintained in Hadoop by ingesting data from Oracle using Sqoop. A data ware house was created in Hive to integrate with Spark and to measure performance using Impala.   Using Spark SQL and Data frames, we have achieved high performance for Business analytics by loading data from Hive. Spark jobs were submitted in Spark Job Server and connected to YARN as resource manager. We have used Avro file format to transfer data between applications and stored data in parquet file formats in Hive.   We have started another POC to verify cluster's performance on streaming data for real time analytics using Spark Streaming and Kafka. Responsibilities: Installed Cloudera's Hadoop cluster, configured Name-Nodes, a secondary Name-Node, and configured following services on Edge node - Spark, Hive and Kafka and Data-Nodes.   Used Xen Center and Cent OS 7   Setup ssh permissions to accept request between each node to transfer huge volumes of data.   Provided proper FQDN to address the

nodes from client applications   Installed services from Cloudera Manager on Edge nodes   Verified cluster performance and management of the cluster from Cloudera Manager to see the health of the services and nodes.   Used Sqoop to ingest data from Oracle and created Sqoop jobs for ingesting day to day updates.   Created Spark and Hive scripts for data analysis.   Created Test Strategy and Test Design documents and responsible for overall quality of the project.   Created test cases to validate the HIVE tables and data quality in different process areas.   Created use cases for test case creation to validate the business flow process   Ingested data from Sqoop as Avro datafile. Transferred Avro schema file from client using cron jobs.   Created Hive scripts to load data into data ware house, and stored data as parquet files in Hive. Cleansed data using Spark or Hive before inserting into Hive ware house.   Wrote Spark SQL jobs to achieve data analysis by loading data from Hive and writing data back to Hive and relational database.   Maintained and cleaned data in Hive warehouse using Spark Jobs. Environment: Cloudera manager, Spark, Hive, Sqoop, Kafka, YARN, MySQL, Xen Center. Java/Hadoop Developer Kohl's January 2017 to December 2017 Description: This is a service-based application which acts as a standard middleware application to orchestrate between different applications. Our application will act like a communication barrier between frontend user applications and backend applications. Also worked on Hadoop environment with other team to write Sqoop jobs to ingest data to HDFS. Integrated Hive with HBase to overcome the limitation faced by Hive for daily updated. Used Flume to input data streaming faster to HDFS.   Responsibilities: Designed system per J2EE specifications. Servlets were used as a Front Controller gateway into the system. Helper classes were used to limit the business logic in the servlet. EJB was used to talk to the database and JSP along with HTML, XML was used to control the client view   Designed and added new functionality extended existing application using J2EE, XML, Servlets, JSP   Client-side validations are done using JavaScript and server-side validations are done using Struts validator framework   Developed Action Classes, which acts as the controller in Struts framework   Created table and different batch programs to clean up tables in DB2 database   Extensively used Collections and Exceptions in the batch program for database cleanup   Established JDBC connection using database connection pool

Wrote complex SQL statements to retrieve data from the DB2 database   Developed application using Eclipse on Windows XP.   Implemented Log4j to maintain system log.   Used Junit framework for Unit testing of application   Used Clear Case for version control.   Used Python to perform data validation on the data ingested using Sqoop and Flume and the cleansed data set is pushed into HBase.   Used Hive to analyze data ingested into HBase by using Hive-HBase integration and compute various metrics for reporting on the dashboard.   Created test cases to validate the HIVE tables and data quality in different process areas.   Environment: JAVA, J2EE, HTML, XML, CSS, JavaScript, windows/Unix, Struts, Servlets, JSP, EJB, Eclipse, JDBC, DB2, ClearCase, Hive, HBase, Kafka, Flume. Creating templates on Virtual machine Concordia University Wisconsin May 2016 to December 2016 Description: Developed and deployed virtual machines using windows virtualization technology to minimize the infrastructure cost. I developed different VM's using either windows or ubuntu. I created VM's based on customer's requirements. I installed different applications and setup those machines with user credentials.   Responsibilities: Used Windows Virtualization to configure different VM's.   Used MYSQL server and configure user with secured user credentials.   Configured Telnet server to in VM's so that user can connect to VM's using Telnet Client application remotely. Java/J2EE Developer Cura Software Solutions July 2014 to November 2015 Title: IVR system development Description: IVR Processing System: this system was developed based on java. User can call in and select different options to navigate through and complete his transaction. IVR system will send HTTP requests to this application which will process request and respond accordingly.   Responsibilities: Used Maven to build the application.   Used HTTP protocol to receive and respond back with the response.   Used Eclipse as IDE and SVN as source code repository.   Attended agile scrum meetings to manage iterations.   Participated in daily stand ups, sprint reviews, sprint planning sessions and demo the changes to business.   Wrote unit test cases using JUnit.   Supported the application for UAT testing and with the bug fixes. Environment: Java, Servlets, Maven, Junit, Eclipse, SVN, UNIX. Java developer Cura Software Solutions January 2014 to June 2014 Title: Web application development Description: Intranet based application to track User requests received from IVR systems like Billing Data, customer data.

Once the request was solved then our application will generate conformation and send to the user.

Responsibilities: Working on HTML, CSS, JavaScript and JSP for building user interface pages. Using Tomcat to deploy the application in different environments.   Using Maven to build the application artifacts.   Using JDBC to access database.   Using Eclipse as IDE and SVN as source code repository. Education Masters in Information Technology in Information Technology Concordia University Wisconsin Skills Impala, Sqoop, Kafka, Cdh4, Flume, Hadoop, Avro, C++, Hadoop, Hive, Html, Javascript, Jenkins, Pig, Eclipse, Java, Jsp, Servlets, Database, Jdbc Additional Information Technical Skills: Big Data Technologies Apache Hadoop, Spark, Hive, Sqoop, Kafka, Pig, Flume, AVRO, Parquet, Impala   AWS Technologies Lambda services, EMR, Cloudwatch   Hadoop Distributions Cloudera(CDH4/CDH5)   Programming Languages C, C++, Java, HTML, CSS, javascript, Servlets, JSP, JDBC, Junit, log4j, Maven   Databases Oracle Database, SQL server, TSQL, MYSQL   IDE Tools Eclipse, SoapUI, Jenkins, JIRA, Confluence.   Applications MS-Office (Word, Excel, PowerPoint, MS Access), MS outlook, Internet Applications.   Operating Systems Windows 2000/2012 server R2, Windows XP, Pro, Vista, 7/8/10.

Name: Brooke White

Email: steve64@example.net

Phone: 5884333640