Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - AT&T Inc Middletown, NJ

Around 9 years of programming experience involved in all phases of Software Development Life Cycle (SDLC)   Over 5+ Years of Big Data experience in building highly scalable data analytics applications.   Strong experience working with Hadoop ecosystem components like HDFS, Map Reduce, Spark, HBase, Oozie, Hive, Sqoop, Pig, Flume and Kafka   Good handson experiencing working with various Hadoop distributions mainly Cloudera (CDH), Hortonworks (HDP) and Amazon EMR.   Good understanding of Distributed Systems architecture and design principles behind Parallel Computing.   Expertise in developing production ready Spark applications utilizing Spark-Core, Dataframes, Spark-SQL, Spark-ML and Spark-Streaming API's, SciKitLearn, SparkML(MLlib) and Tensorflow.   Strong experience troubleshooting failures in spark applications and fine-tuning spark applications and hive queries for better performance.   Worked extensively on Hive for building complex data analytical applications.   Strong experience writing complex map-reduce jobs including development of custom Input Formats and custom Record Readers.   Sound Knowledge in map side join, reduce side join, shuffle & sort, distributed cache, compression techniques, multiple hadoop Input & output formats.   Worked with Apache NiFi to automate the data flow between the systems and managed flow of information between system.   Good experience working with AWS Cloud services like S3, EMR, Redshift, Athena, Dynamo DB etc.,   Deep understanding of performance tuning, partitioning for optimizing spark applications.   Worked on building real time data workflows using Kafka, Spark streaming and HBase.   Extensive knowledge on NoSQL databases like HBase, Cassandra and Mongo DB.   Solid experience in working with csv, text, sequential, avro, parquet, orc, json formats of data.   Extensive experience in performing ETL on structured, semi-structured data using Pig Latin Scripts.   Designed and implemented Hive and Pig UDF's using Java for evaluation, filtering, loading and storing of data.   Experience in using Hadoop ecosystem and processing data using Tableau.   Experience with Apache Phoenix to access the data stored in HBase.   Good knowledge in the core concepts of programming such as algorithms, data structures, collections.   Developed core modules in large cross-platform applications using JAVA, JSP, Servlets, Hibernate, RESTful, JDBC, JavaScript, XML,

and HTML. Extensive experience in developing and deploying applications using Web Logic, Apache Tomcat and JBOSS. Development experience with RDBMS, including writing SQL queries, views, stored procedure, triggers, etc. Strong understanding of Software Development Lifecycle (SDLC) and various methodologies (Waterfall, Agile). Authorized to work in the US for any employer Work Experience Sr. Hadoop Developer AT&T Inc - Middletown, NJ July 2018 to Present Roles & Responsibilities: Developed Spark applications using PySpark utilizing Data frames and Spark SQL API for faster processing of data. Developed highly optimized Spark applications to perform various data cleansing, validation, transformation and summarization activities according to the requirement Data pipeline consists Spark, Hive and Sqoop and custom build Input Adapters to ingest, transform and analyze operational data. Developed Spark jobs and Hive Jobs to summarize and transform data. Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data. Involved in converting Hive/SQL queries into Spark transformations using Spark DataFrames and Scala. Used different tools for data integration with different databases and Hadoop. Analyzed the SQL scripts and designed the solution to implement using Pyspark Involved in installation of Tez and improved the query performance.. Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data. Built real time data pipelines by developing kafka producers and spark streaming applications for consuming. Ingested syslog messages, parses them and streams the data to Kafka. Handled importing data from different data sources into HDFS using Sqoop and performing transformations using Hive, Map Reduce and then loading data into HDFS. Exported the analyzed data to the relational databases using Sqoop, to further visualize and generate reports for the BI team. Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis Analyzed the data by performing Hive queries (Hive QL) to study customer behavior. Helped Dev ops Engineers for deploying code and debug issues. Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting. Developed Hive scripts in Hive QL to de-normalize and aggregate the data. Scheduled and executed workflows in Oozie to run various jobs. Experience

in using Hadoop ecosystem and processing data using Amazon AWS.   Environment: Hadoop, HDFS, HBase, Spark, Scala, Hive, MapReduce, Sqoop, ETL, Java, PL/SQL, Oracle 11g, Unix/Linux. Sr. Hadoop Developer ON SOLVE - Daytona Beach, FL April 2017 to July 2018 Roles & Responsibilities: Build a framework Spark with Scala and migrated existing PySpark applications to improve the runtime and performance.   Developed highly optimized Spark applications to perform various data cleansing, validation, transformation and summarization activities according to the requirement   Performed Transformations like De-normalizing, Cleansing of data sets, Date Transformations, parsing some complex columns.   Worked with different compression codecs like GZIP, SNAPPY and BZIP2 in MapReduce, Pig and Hive for better performance.   Worked with Apache NiFi to automate the data flow between the systems and managed flow of information between systems   Have used Ansible for automation of frameworks.   Handled Avro, JSON and Apache Log data in Hive using custom Hive SerDes.   Worked on batch processing and scheduled workflows using Oozie.   Implemented installation and configuration of multi-node cluster on the cloud using Amazon Web Services (AWS) on EC2.   Worked in agile sprint methodology environment.   Have used the Knox gateway for having Hadoop security between the users and operators.   Used cloud computing on the multi-node cluster and deployed Hadoop application on cloud S3 and used Elastic Map Reduce (EMR) to run Map-reduce.   Used Hive-QL to create partitioned RC, ORC tables, used compression techniques to optimize data process and faster retrieval.   Implemented Partitioning, Dynamic Partitioning and Buckets in Hive for efficient data access.   Environment: Apache Hadoop, HDFS, Cloudera Manager, Java, MapReduce, Eclipse Indigo, Hive, HBASE, PIG, Sqoop, Oozie, SQL, Spring. Hadoop Developer Maximus Inc - Reston, VA September 2013 to March 2016 Roles & Responsibilities: Involved in requirement analysis, design, coding and implementation phases of the project.   Used Sqoop to load structured data from relational databases into HDFS.   Loaded transactional data from Teradata using Sqoop and created Hive Tables.   Worked on automation of delta feeds from Teradata using Sqoop and from FTP Servers to Hive.   Performed Transformations like De-normalizing, Cleansing of data sets, Date Transformations, parsing some complex columns.   Worked with different compression codecs

like GZIP, SNAPPY and BZIP2 in MapReduce, Pig and Hive for better performance. Worked with Apache NiFi to automate the data flow between the systems and managed flow of information between systems Have used Ansible for automation of frameworks. Handled Avro, JSON and Apache Log data in Hive using custom Hive SerDes. Worked on batch processing and scheduled workflows using Oozie. Implemented installation and configuration of multi-node cluster on the cloud using Amazon Web Services (AWS) on EC2. Worked in agile sprint methodology environment. Have used the Knox gateway for having Hadoop security between the users and operators. Used cloud computing on the multi-node cluster and deployed Hadoop application on cloud S3 and used Elastic Map Reduce (EMR) to run Map-reduce. Used Hive-QL to create partitioned RC, ORC tables, used compression techniques to optimize data process and faster retrieval. Implemented Partitioning, Dynamic Partitioning and Buckets in Hive for efficient data access. Environment: Apache Hadoop, HDFS, Cloudera Manager, Java, MapReduce, Eclipse Indigo, Hive, HBASE, PIG, Sqoop, Oozie , SQL, Spring. Java Developer Dataguise Inc - Fremont, CA November 2011 to August 2013 Roles & Responsibilities: Involved in all the phases of the project development - requirements gathering, analysis, design, development, coding, testing and debugging Implemented MVC architecture by using Struts to send and receive the data from front-end to business layer. Integrated the Struts and Hibernate to achieve Object relational mapping. Used apache struts to develop the web-based components and implemented DAO. Implemented Struts framework in the presentation tier for all the essential control flow, business level validations and for communicating with the business layer. Integrated the Struts application with Hibernate for querying/inserting & data management for SQL server database. Responsible for design and development of Web Application in J2EE using Struts MVC Framework. Involved in creating & consuming SOAP based & Restful web services. Used Web Services for communication between the different internal applications. Used SOAP for the communication between the different internal applications. Used GitHub for version control management and consistently produced high quality code through disciplined and rigorous unit testing. Used Maven script for building and deploying the application. Developed the XML schema and Web Services

for the data maintenance and structures.   Involved in designing test plans, test cases and overall Unit testing of the system.   Object Oriented Analysis and Design using UML include development of class diagrams, Sequence diagrams and state diagrams and implemented these diagrams in Microsoft Visio.   Worked in agile sprint methodology environment.   Implemented MVC, DAOJ2EE design patterns as a part of application development.   Used Spring IOC and MVC for enhanced modules.   Developed the Persistence Layer using Hibernate.   Used DB2 as the database and wrote SQL & PL-SQL.   Designed and developed message driven beans that consumed the messages from the Java message queue.   Design and development of Web pages using HTML, CSS including Ajax controls and XML.   Written controllers based on Spring MVC and made calls to JSP pages.   Environment: Struts, Spring, HTML, CSS, Java, J2ee, JSP, XML, Eclipse, WebLogic, JavaScript. Java Mail API, Hibernate, SQL Server, JBoss, GitHub, Maven, Agile, Junit. Java Developer Media3 Pvt Ltd - Hyderabad, Telangana June 2010 to October 2011 Roles & Responsibilities: Implemented the presentation layer with HTML, CSS and JavaScript Developed web components using JSP, Servlets and JDBC   Implemented secured cookies using Servlets.   Wrote complex SQL queries and stored procedures.   Implemented Persistent layer using Hibernate API   Implemented Search queries using Hibernate Criteria interface.   Provided support for loans reports for CB&T   Designed and developed Loans reports for Evans bank using Jasper and iReport.   Involved in fixing bugs and unit testing with test cases using Junit.   Object Oriented Analysis and Design using UML include development of class diagrams, Sequence diagrams and state diagrams and implemented these diagrams in Microsoft Visio.   Maintained Jasper server on client server and resolved issues   Actively involved in system testing.   Fine tuning SQL queries for maximum efficiency to improve the performance   Designed Tables and indexes by following normalizations.   Involved in Unit testing, Integration testing and User Acceptance testing   Utilizes Java and SQL day to day to debug and fix issues with client processes.   Environment: Java, Servlets, HTML, Java Script, JSP, Hibernate, Junit Testing, Oracle DB, SQL, Jasper Reports, iReport, Maven, Jenkins. Education Bachelors in Computer Science& Engineering in Computer Science& Engineering Indian Institute of Technology (IIT) Skills ECLIPSE,

EJB, J2EE, JAVA, SPRING, JBOSS, JSP, SERVLETS, STRUTS, DB2, TERADATA, JDBC, MYSQL, SQL, HTML, JAVASCRIPT, PHP, PYTHON, WEB UI, XML Additional Information TECHNICAL SKILLS   Programming Skills Java/J2EE, JSP, Servlets, AJAX, EJB, Struts, Spring, JDBC, JavaScript, PHP and Python. Databases MYSQL, SQL, DB2 and Teradata Web services REST, AWS, SOAP, WSDL, Servers Apache Tomcat, WebSphere, JBoss Operating Systems Unix, Linux, Windows, Solaris IDE tools My Eclipse, Eclipse, NetBeans QA Tools Crashlytics or Fabrics Web UI HTML, JavaScript, XML, SOAP, WSDL

Name: Diana Blackburn

Email: carlos60@example.org

Phone: (765)799-7055x296