Sr. Hadoop /Spark Developer Sr. Hadoop /Spark Developer Sr. Hadoop /Spark Developer Atlanta, GA   Over all7+years of progressive experience in the IT industry with proven expertise in architecting and implementing Software Solutions using Java&Big Data technologies.   4 years of experience on Batch Analytics using Hadoop working environment includes Map Reduce, HDFS, Hive, Pig, HBase, Oozie, Kafka, SparkandSqoop.   Worked extensively in Real time analytics using Storm and Spark-Streaming.Used ingestion tools like Kafka and Sqoop.   Worked extensively with NoSql databases like Cassandra and HBase.   In depth understanding of Hadoop Architecture and its various components such as Resource Manager,Node Manager, Applications Master, Name Node, Data Node concepts.   Experience in importing and exporting data using Sqoop from Relational DatabaseSystems to HDFS andvice-versa.   Extending HIVE and PIG core functionality by using custom User Defined Function's (UDF), User Defined Table-Generating Functions (UDTF) and User Defined Aggregating Functions (UDAF) for Hive and Pig.   Designed and built user interface using spring and JavaScript& employed collection libraries.   Designed a website for understanding the user requirements and validated the web page using JQuery in Conjunction with Java Spring/ Hibernate/ PHP.   Experience in analyzing data using HiveQL, Pig Latin, and custom MapReduce programs in Java.   Developed Pig Latin scripts for data cleansing and Transformation.   Job workflow scheduling and monitoring using tools like Oozie, IBM Tivoli.   Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from different sources like Storm and Spark.   Good experience in Cloudera, Horton works&MapRApache Hadoop distributions.   Worked with relational database systems (RDBMS) such as MySQL, MSSQL, OracleRelational database systems like HBase and Cassandra.   Assisted with performance tuning and monitoring of Kafka, HBase, Storm, Pig and Hive.   Used Shell Scripting to move log files into HDFS.   Good hands on experience in creating the RDD's, DF's for the required input data and performed the data transformations using Spark.   Developed Scala scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop.   Good understanding in processing of real-time data using Spark.Import the data from different sources like HDFS/HBase into Spark RDD.   Experience in writing

MapReduce jobs in python for some complicated queries.   Experienced with different file formats like Parquet, ORC, CSV, Text, Sequence, XML, JSON and Avro files.   Good knowledge on Data Modelling and Data Mining to model the data as per business requirements.   Involved in unit testing of Map Reduce programs using Apache MRunit.   Good knowledge on python scripting and bash scripting languages.   Expert in Data Visualization development using Tableau to create complex and innovative dashboards.   Extensively used Java and J2EE technologies like Core Java, Java Beans, Servlet, JSP, spring, Hibernate, JDBC, JSON Object, and Design Patterns. Experienced in Application Development using Java, J2EE, JSP, Servlets, RDBMS, Tag Libraries, JDBC, Hibernate and XML.   Worked with different software version control, Jira, bug tracking and code review systems like CVS, Clear Case. Authorized to work in the US for any employer Work Experience Sr. Hadoop /Spark Developer American Express - Phoenix, AZ August 2016 to July 2017 Description: American express is a global services company that provides customers with access to products, insights and experiences that enrich lives and build business success.The company is best known for its credit card, charge card, and traveler's Cheque businesses. Storm, Kafka and Hbase are used here to processing the users credit card, charge card and cheque's business data for processing all the data in the real time. UI layer is developed for retrieving the data from Hbase. Responsibilities: Designed the solution using Storm Spouts (to stream data from Kafka) and Bolts connecting to Java APIs developed independently based on the application logic. Imported bulk data into HBase Using Map Reduce programs.   Written Storm topology to accept the events from Kafka producer and emit into HBase.   Developed a data pipeline using Kafka and Strom to store data into HDFS.   Developed HDFS with huge amounts of data using Apache Kafka.   Implemented a proof of concept (Poc's) using Kafka, Strom, HBase for processing streaming data.   Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.   Development of software using core java with integration of Apache Storm, Apache Kafka.   Integrated Oozie with the rest of the Hadoop stack supporting several types ofHadoop jobs out of the box (such as Map-Reduce, Pig, Hive, and Sqoop) as well as system specific jobs (such as Java programs and shell scripts).   Exploring with

Spark improving the performance and optimization of the existing algorithms in Hadoopusing Spark context, Spark-SQL,Data Frame, pair RDD's, Spark YARN.   Developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.   Experience in deploying data from various sources into HDFS and building reports using Tableau.   Performed real time analysis on the incoming data.   Load the data into Spark RDD and performed in-memory data computation to generate the output response.   Developed Spark scripts by using Python shell commands as per the requirement.   Developed Shell scripts and Python programs to automate tasks   Environment: Hadoop,Map Reduce, HDFS, Spark, Java, Kafka, Hive, HBase, maven, Jenkins, Pig, UNIX, Python, , Git, Storm ,MapR, Oozie Sr. Hadoop Developer Celgene Corporation July 2016 to July 2016 Description: Celgene is a global biopharmaceutical company committed to improving the lives of patients worldwide. At Celgene, we seek to deliver truly innovative and life-changing treatments for our patients. Responsibilities: Imported data from different relational data sources like RDBMS, Teradata to HDFS using Sqoop.   Imported bulk data into HBase Using Map Reduce programs. Perform analytics on Time Series Data exists in HBase using HBase API.   Designed and implemented Incremental Imports into Hive tables.   Used Rest ApI to Access HBase data to perform analytics.   Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.   Involved in converting Map Reduce programs intoSpark transformations using Spark RDD's on Scala.   Implemented Spark using Scala and Spark SQL for faster testing and processing of data.   Experienced in working with various kinds of data sources such as Teradata and Oracle. Successfully loaded files to HDFS from Teradata, and load loaded from HDFSto hive and impala.   Experienced in running query using Impala and used BI tools to run ad-hoc queries directly on Hadoop.   Experienced with batch processing of data sources using Apache Spark,Elastic search.   Develop wrapper using shell scripting for Hive, Pig, Sqoop, Scala jobs.   Worked on developing Unix Shell scripts to automate Spark-Sql.   Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala.   Worked in Loading and transforming large sets of structured, semi structured and unstructured data.   Involved in collecting, aggregating and moving data from servers to HDFS

using Apache Flume.   Written Hive jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data.   Involved in creating Hive tables, loading with data and writing hive queries that will run internally in MapReduce way.   Developed java Restfulweb services to upload data from local to Amazon S3, listing S3 objects and file manipulation operations. Configured a 20-30 node (Amazon EC2 spot instance) Hadoop cluster to transfer the data from Amazon S3 to HDFS and HDFS to Amazon S3 and also to direct input and output to the Hadoop MapReduce framework.   Experienced in managing and reviewing the Hadoop log files. Successfully ran all Hadoop MapReduce programs on Amazon Elastic MapReduce framework by using Amazon S3 for input and output.   Involve in Data Asset Inventory to gather, analyze, and document business requirements, functional requirements and data specifications for Member Retention from sources SQL / Hadoop.   Worked on solving performance and limit queries to the workbooks that when it connects to live database by using a data extract option in Tableau. Designed and developed Dashboards for Analytical purposes using Tableau.   Designed and implemented facts, dimensions, measure groups, measures and OLAP cubes using dimensional data modeling standards in SQL Server 2008 that maintained data   Creating and Designing OLAP using SAS OLAP Cube Studio.   Designing Source, Job, Target using SAS OLAP Cube Studio and SAS/DIS.   Analyzing OLAP Using SAS OLAP Viewer and SAS Dataset using SAS/EG.   Migrated ETL jobs to Pig scripts do Transformations, even joins and some pre-aggregations before storing the data onto HDFS.   Worked with Avro Data Serialization system to work with JSON data formats.elastic search   Worked on different file formats like Sequence files, XML files and Map files usingMap ReducePrograms.   Involved in Unit testing and delivered Unit test plans and results documents using Junit and MRUnit.   Exported data from HDFS environment into RDBMS using Sqoop for report generation and visualization purpose.   Worked on Oozie workflow engine for job scheduling.   Created and maintained Technical documentation for launching HADOOP Clusters and for executing Pig Scripts.   Environment: CDH 5.3, Map Reduce, Hive0.14, Spark 1.4.1, Oozie, Sqoop, Pig0.11, Java, Rest API, Maven, MRunit, Junit, Tableau, Cloudera, Python. Java Developer/Hadoop Developer CNAINSURANCE - Chicago, IL January 2014 to March 2015

Description: CNA insurance provided B2B insurance to their customers. As part of enhancements, we developed ELS (Enterprise Logging Service) to provide statistics to support team and implemented Processor to send alerts to Support Teams. Responsibilities: Participated in requirement gathering and converting the requirements into technical specifications.   Created UML diagrams like use cases, class diagrams, interaction diagrams, and activity diagrams.   Developed the application using Spring Framework that leverages classical Model View Controller (MVC) architecture.   Created Business Logic using Servlets, POJO's and deployed them on Web logic server.   Wrote complex SQL queries and stored procedures.   Developed the XML Schema and Web services for the data maintenance and structures.   Implemented the Web Service client for the login authentication, credit reports and applicant information using Apache Axis 2 Web Service. Developed and implemented custom data validation stored procedures for metadata summarization for the data warehouse tables, for aggregating telephone subscribers switching data, for identifying winning and losing carriers, and for identifying value subscribers.   Identified issue and developed a procedure for correcting the problem which resulted in the improved quality of critical tables by eliminating the possibility of entering duplicate data in a Data Warehouse.   Designed and implemented SQL based tools, stored procedures and functions for daily data volume and aggregation status   Responsible to manage data coming from different sources.   Developed map reduce algorithms.   Got good experience with NOSQL database.   Involved in loading data from UNIX file system to HDFS.   Installed and configured Hive and also written Hive UDFs.   Worked with cloud services like Amazon web services (AWS)   Designed the logical and physical data model, generated DDL scripts, and wrote DML scripts for Oracle 10g database.   Used Hibernate ORM framework with Spring framework for data persistence and transaction management.   Wrote test cases in JUnit for unit testing of classes.   Involved in creating templates and screens in HTML and JavaScript.   Involved in integrating Web Services using SOAP.   Environment: Hive 0.7.1, Apache Solr - 3.x, HBase-0.90.x/0.20.x, JDK, Spring MVC, WebSphere 6.1, HTML, XML, JavaScript, JUnit 3.8, Oracle 10g, Amazon Web Services. Java Developer INTEQ SOFTWARE PVT LTD - Hyderabad, Telangana January 2012 to December 2013 Description: This project deals with

the development of automotive protocol for vehicles based on serial interfaces of ECU. The protocol development has to finally test by the automotive tools to confirm the message integrity. Responsibilities: Using JAVA developed a website for e-Recruitment consists of many modules. Followed MVC Architecture for implementing the functionality.   Designed and reviewed the test scenarios and scripts for given functional requirements.   Implemented Services using Core Java. Involved in development of classes using java.   Designed and built user interface using spring and JavaScript& employed collection libraries.   Designed and involved in preparing activity diagrams, usecasediagrams, sequence diagrams as per the business requirement.   Used JavaScript for Client validation.   Designed a website for understanding the user requirements and validated the web page using JQuery in Conjunction with Java Spring/ Hibernate/ PHP.   Developed user interfaces using Servlets, CSS, XSLT, XML, HTML and JavaScript.   Good proficiency in developing algorithms for serial interfaces.   Involved in testing of CAN protocols.   Developed the flow of algorithm in UML.   Developed verification and validation scripts in java.   Environment: Java, JSP, Servlets, JDBC, JavaScript, MySQL, JUnit, Eclipse IDE, Windows 7/XP/Vista, UNIX, LINUX. Education Bachelor's Skills Html (3 years), java (4 years), Javascript. (3 years), Model view controller (3 years), Xml (3 years)

Name: Chris Booth

Email: debralopez@example.net

Phone: 595-395-7297