

Big data Developer Big data Developer Parsippany, NJ Over 5 years of industrial experience in Application development and maintenance, data management, programming, data analysis and data visualization. Experience in dealing with Apache Hadoop components like HDFS, MapReduce, Hive, HBase, Pig, Sqoop, Oozie, Python, Spark, Storm, Cassandra, MongoDB, Big Data and Big Data Analytics. Good understanding/knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name node and MapReduce concepts. Experienced managing No-SQL DB on large Hadoop distribution Systems such as: Cloudera, Hortonworks HDP, Map R M series etc. Experienced developing Hadoop integration for data ingestion, data mapping and data process capabilities. Good knowledge in ETL and hands on experience in ETL, experience in RDBMS like Oracle, MS SQL Server, MySQL and DB2, and NoSQL databases like mongo DB and HBase. Software development in Java Application Development Client/Server Applications, Internet/Intranet based database applications and developing, testing and implementing application environment using J2EE, JDBC, JSP, Servlets, Web Services, Oracle, PL/SQL and Relational Databases. Exceptional ability to quickly master new concepts and capable of working in groups as well as independently. Excellent interpersonal skills and the ability to work as a part of a team. Has good knowledge of virtualization and worked on VMware Virtual Center. Hands on experience in installing, configuring, and using Hadoop ecosystem components like Hadoop MapReduce, HDFS, HBase, Hive, Sqoop, Pig, Zookeeper, Spark, Kafka and Flume. Good Knowledge on Hadoop Cluster architecture and monitoring the cluster. Experience in managing Hadoop clusters using Cloudera Manager. Experience in using the Impala usage for the high-performance SQL queries.

Very good experience in complete project life cycle (design, development, testing and implementation) of Client Server and Web applications. Hands on experience in VPN, Putty, win SCP, VNC viewer, etc. Hands on experience in application development using Java, RDBMS, and Linux shell scripting. Performed data analysis using MySQL, SQL Server Management Studio and Oracle. Good knowledge on cloud computing platforms like Amazon Web Services(AWS). Ability to adapt to evolving technology, strong sense of responsibility and accomplishment. Work

Experience Big data Developer Parsippany, NJ January 2018 to Present Responsibilities: Involved in installing Hadoop Ecosystem components and responsible to manage data coming from various sources. Implemented Flume from relational database management systems using Sqoop. Used Pig as ETL tool to do transformations, event joins, filter both traffic and some pre-aggregations before storing the data onto HDFS. Involved in writing Flume and Hive scripts to extract, transform and load the data into Database cluster capacity planning, performance tuning, cluster Monitoring, Troubleshooting. Developed Spark code using Java and Spark-SQL/Streaming for faster processing of data. Develop ETL Process using SPARK, SCALA, HIVE and HBASE. Extensively involved in writing ETL Specifications for Development and conversion projects. Involved in converting Hive/SQL queries into Spark transformations using Spark RDD'S and Scala. Developed various Map Reduce jobs in Java for data cleaning and preprocessing. Developed Oozie workflow for scheduling & orchestrating the ETL process. Experience on loading and transforming of large sets of structured, semi structured and optimizing of existing algorithms in Hadoop using Spark Context, Hive-SQL, Data Frames. Developed Spark scripts by using Scala shell commands as per the requirement. Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala. Created HBase tables to store variable data formats coming from different portfolios Performed real time analytics on HBase using Java API and Rest API. Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations and other during ingestion process itself. Converted the text files and the csv files to parquet form for the analysis of data. Prepared the mapping document, as in which fields has be used from the HIVE DB and perform the transformations. Implemented HDFS and Cassandra with vast amounts of data using Apache Kafka Performed the analytics over the data mining, data visualization using Hive. Worked in Agile environment and used JIRA as a bug-reporting tool for updating the bug report. Environment: Cloudera Hadoop, Hortonworks, Linux, ETL, HDFS, Hive, Spark, Sqoop, Flume, Zookeeper, HBase. Hadoop Developer Anthem Insurance - Norfolk, VA June 2016 to December 2017 Responsibilities: Contributed to the development of key data integration and

advanced analytics solutions leveraging Apache Hadoop and other big data technologies using major Hadoop Distributions like Hortonworks and Cloudera. Worked on Amazon AWS - EMR, EC2, RDS, S3, RedShift, etc., Tools- Hadoop, Hive, Pig, Sqoop, Oozie, HBase, Flume, Spark. Worked on loading log data directly into HDFS using Flume in Cloudera - CDH. Involved in loading data from LINUX file system to HDFS in Cloudera - CDH. Experienced in running Hadoop streaming jobs to process terabytes of xml format data. Experience in importing and exporting data into HDFS and assisted in exporting analyzed data to RDBMS using SQOOP in Cloudera. Installed and configured MapReduce, HIVE and the HDFS. Developing Spark scripts by using Java per the requirement to read/write JSON files. Developed and implemented Java code according to MapReduce for the business requirements. Worked on Importing and exporting data into HDFS and Hive using Sqoop. Worked on Hadoop Administration, development, NoSQL in Cloudera. Load and transform large sets of structured, semi structured and unstructured data. Experienced in Hadoop Big Data Integration with Talend ETL on performing data extract, loading and transformation process. Experience in development of extracting, transforming and loading (ETL), maintain and support the enterprise data warehouse system and corresponding marts. Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map. Automate all the jobs, for pulling data from FTP server to load data into Hive tables, using Oozie workflows. Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios. Supported code/design analysis, strategy development and project planning. Created reports for the BI team using Sqoop to export data into HDFS and Hive. Design & Implemented Data Warehouse creating facts and dimension tables and loading them using Informatica Power Center Tools fetching data from the OLTP system to the Analytics Data Warehouse. Coordinating with business user to gather the new requirements and working with existing issues, worked on reading multiple data formats on HDFS using Scala. Loading data into parquet files by applying transformation using Impala. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala. Analyzed the SQL scripts and designed the solution to implement using Scala. Developed multiple

POCs using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata. Designed and Development the Integration APIs using various Data Structure concepts, Java Collection Framework along with exception handling mechanism to return response within 500ms. Usage of Java Thread concept to handle concurrent request. Environment: Hadoop 1.x/2.x MR1, Cloudera CDH, Hortonworks, HDFS, Spark, Scala, Impala, HBase 0.90.x, Flume 0.9.3, Java, Sqoop 2.x, Hive 0.7.1, Tableau. Bigdata Developer Prime key software solutions, Hyd April 2014 to June 2015 Responsibilities: Worked on analyzing Hadoop cluster and different big data analytic tools including Pig, HBase and Sqoop. Responsible for building scalable distributed data solutions using Hadoop. Configured a cluster by editing config files such as core-site.xml, mapred-site.xml, hdfs-site.xml and masters/ slaves. Installed and configured Flume, Hive, Pig, Sqoop, and HBase on the Hadoop Cluster. Responsible to monitor block Scanner Reports on data nodes. Implemented 30 nodes CDH3/CDH4 Hadoop cluster on CentOS. Worked on installing cluster, commissioning & decommissioning of data node, name node recovery, capacity planning, and slots configuration. Involved in loading data from file system to HDFS and Implemented best offer logic using Pig scripts and Pig UDFs. Implemented test scripts to support test driven development and continuous integration. Responsible to manage data coming from different sources. Installed and configured Hive and written Hive UDFs. Developed simple and complex MapReduce programs in Java for Data Analysis on different data formats Experience in managing and reviewing Hadoop log files. Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team. Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. Imported data using Sqoop to load data from MySQL to HDFS on regular basis. Developing Scripts and Batch Job to schedule various Hadoop Program. Responsible for writing Hive queries for data analysis to meet the business requirements. Responsible for importing and exporting data into HDFS and Hive using Sqoop. Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way. Designed and implemented MapReduce based large-scale parallel relation-learning system. Environment: Hadoop, HDFS, Pig, Sqoop, Storm, VPN, MapReduce,

CentOS. Java / J2ee Developer Logic solutions Pvt Ltd - Chennai, Tamil Nadu September 2012 to March 2014 Responsibilities: Responsible for developing Use Case, Class diagrams and Sequence diagrams for the modules using UML and VISIO. Used Spring Framework for dependency injection with the help of spring configuration files. Developed the presentation layer using JSP, HTML, CSS and client validations using JavaScript. Used HTML, CSS, JQUERY and JSP to created and UI. Involved in Installation and configuration of Tomcat Server. Involved in Dynamic form generation, Auto completion of forms and user validation functionalities using AJAX. Designed developed and maintained the data layer using Hibernate and performed configuration of Struts, Application Framework. Created stored procedures using SQL for data access layer. Worked on tuning of back-end stored procedures using TOAD. Involved in the configuration management using CVS. Developed various test cases and performed unit testing using JUnit. Developed Unit test cases for the classes using JUnit/Eclipse. Developed stored procedures to extract data from Oracle database. Application developed with strict adherence to J2EE best practices. Environment: Java, J2EE, JSP, Servlets, Hibernate, SQL, Web Services, SOAP, WSDL, JUnit, Tomcat, MySQL, CVS and Windows. Education Bachelor of Technology in Technology JNTU - Hyderabad, Telangana Skills Cassandra, Hdfs, Sqoop, Hbase, Db2, Etl, Flume, Hadoop, Informatica, Jms, Map reduce, Nosql, C++, Hadoop, Hbase, Hive, Html, Javascript, Perl, Pig Additional Information Technical Skills: Scripting Languages Python, Perl, Shell. Big Data Technologies Hadoop, HDFS, Hive, Map Reduce, Pig, Sqoop, Flume, Zookeeper, AWS. Programming Languages C++, Java. Web Technologies HTML, J2EE, CSS, JavaScript, AJAX, Servlets, JSP, DOM, XML, XSLT, XPATH. Java Frameworks Struts, Spring, Hibernate. DB Languages SQL, PL/SQL Messaging Systems JMS, Active-MQ, IBM MQ Databases /ETL Oracle 9i/10g/11g, MySQL 5.2, DB2, Informatica v 8.x. NoSQL Databases Hbase, Cassandra, Mango DB. Operating Systems Windows, Unix and Linux

Name: Kevin Brown

Email: rogerszachary@example.org

Phone: 789.853.2451x4658