

Sr. Data Engineer Sr. Data Engineer Sr. Data Engineer - AT&T Plano, TX Proactive IT developer with 9 years of working experience in Java/J2EE Technology and development design of various scalable systems using Hadoop Technologies on various environments. Experience in installation, configuration, supporting and managing Hadoop Clusters using Horton works, and Cloudera (CDH3, CDH4) distributions on Amazon web services (AWS). Extraordinary Understanding of Hadoop building and Hands on involvement with Hadoop segments such as Job Tracker, Task Tracker, Name Node, Data Node and HDFS Framework. Extensive experience in analyzing data using Hadoop Ecosystems including HDFS, Hive, PIG, Sqoop, Flume, MapReduce, Spark, Kafka, HBase, Oozie, Solr and Zookeeper. Extensive knowledge on NoSQL databases like HBase, Cassandra, and Mongo DB. Configured Zookeeper, Cassandra and Flume to the existing Hadoop cluster. Expertise in writing Hadoop Jobs for analyzing data using Hive QL (Queries), Pig Latin (Data flow language), and custom MapReduce programs in Java. Experience in converting Hive queries into Spark transformations using Spark RDDs and Scala. Hands on Experience in troubleshooting errors in HBase Shell, Pig, Hive and MapReduce. Hands-on experience in provisioning and managing multi-tenant Cassandra cluster on public cloud environment - Amazon Web Services (AWS) - EC2, Open Stack. Experience in NoSQL Column-Oriented Databases like HBase, Cassandra and its Integration with Hadoop cluster. Experience in maintaining the big data platform using open source technologies such as Spark and Elastic Search. Planned and created answer for constant information ingestion utilizing Kafka, Storm, Spark spilling and different NoSQL databases. 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. Experience in understanding the security requirements for Hadoop and integrate with Kerberos authentication and authorization infrastructure. Good hands on experience in creating the RDD's, DF's for the required input data and performed the data transformations using Spark Scala. Knowledge in developing a Nifi flow prototype for data ingestion in HDFS. Extensive experience working in Oracle, DB2, SQL Server, PL/SQL and My SQL database and Java Core concepts like OOPS, Multithreading, Collections and IO. Experience in Service Oriented Architecture using Web

Services like SOAP & Restful. Work Experience Sr. Data Engineer AT&T - Plano, TX January 2018 to Present Sr. Bigdata Developer T Mobile - Bellevue, WA May 2018 to January 2019

Responsibilities: Developed various spark applications using Scala to perform various enrichment of these click stream data merged with user profile data. Utilized Spark-SQL to event enrichment and used Spark-SQL to prepare various levels of user behavior summaries. Worked on SQS Queue receiver using Spark Streaming context to consume the data from extended queue and integrated with ETL Functions. Real time streaming the data using Spark with SQS. Responsible for handling Streaming data from web server console logs. Optimize the Hive tables using optimization techniques such as partitions and bucketing to provide better performance with HiveQL queries. Worked on migrating data from traditional RDBMS to HDFS. Used Scala to convert Hive/SQL queries into RDD transformations in Apache Spark. Written Programs in Spark using Scala for Data quality check. Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing. Implemented the use of Amazon EMR for Big Data processing among a Hadoop Cluster of virtual servers on Amazon related EC2 and S3. Optimized Hive QL/ pig scripts by using execution engine like Tez, Spark. Environment: Hadoop, HDFS, Hive, Spark AWS EC2, S3, Kafka, Yarn, Shell Scripting, Scala, Pig, Oozie, Java, Agile methods, Linux, MySQL, Elastic Search, Kibana, Teradata. Hadoop Developer Visa - Austin, TX August 2017 to May 2018

Developed Spark Applications by using Spark, Java and Implemented Apache Spark data processing project to handle data from various RDBMS and Streaming sources. Handled importing of data from various data sources, performed data control checks using Spark and loaded data into HDFS. Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala. Used Spark SQL to Load JSON data and create SchemaRDD and loaded it into Hive Tables and handled structured data using Spark SQ. Imported data from AWS S3 into Spark RDD, Performed transformations and actions on RDD's. Used Spark and Spark SQL to read the parquet data and create the tables in hive using the Scala API. Implemented Spark using Scala and utilizing Data frames and Spark SQL API for faster processing of data. Developed Scala scripts, UDFs using both Data frames/SQL/Data sets and RDD/Map Reduce in Spark 1.6 for Data

Aggregation, queries and writing data back into OLTP system through Sqoop. Processing the schema oriented and non-schema-oriented data using Scala and Spark. Used Spark API over Cloudera Hadoop YARN to perform analytics on data in HDFS. Worked on streaming pipeline that uses Spark to read data from Kafka transform it and write it to HDFS. Analyzed the weblog data using the HiveQL, integrated Oozie with the rest of the Hadoop stack Utilized cluster co-ordination services through Zookeeper. Environment: Scala, Spark, Spark SQL, Spark Streaming, Azkaban, Presto, Hive, Apache Crunch, Elastic Search, GIT Repository, Amazon S3, Amazon AWS Ec2/EMR, Spark cluster, Hadoop Framework, Sqoop, DB2. Data Engineering Walt Disney - Glendale, CA December 2016 to August 2017 Responsibilities: Developed optimal strategies for distributing the web log data over the cluster importing and exporting the stored web log data into HDFS and Hive using Sqoop. Design and develop ELT data pipeline using Spark App to fetch data from Legacy system and third-party API, social media sites. Developed custom mappers in python script and Hive UDFs and UDAFs based on the given requirement. Design and develop DMA (Disney Movies anywhere) dashboard for BI analyst team. Perform data analytics and load data to Amazon s3/Data Lake/ Spark cluster. Involved in querying data using Spark SQL on top of Spark engine. Developed Spark scripts by using Python shell commands as per the requirement. Writing Pig and Hive scripts with UDF in MR and Python to perform ETL on AWS Cloud Services. Worked with file formats text, avro, parquet and sequence files. Involved in migrating HiveQL into Impala to minimize query response time. Created Hive tables, dynamic partitions, buckets for sampling, and working on them using HQL. Defined job flow using Azkaban, scheduler to automate the Hadoop jobs and installed Zookeepers for automatic node failovers. Performed Tableau type conversion functions when connected to relational data sources. Environment: Languages/Technologies: Java (JDK1.6 and higher), Azkaban, Spark SQL, Presto, Hive, Apache Crunch, Elastic Search, Spring boot, Eclipse, GIT Repository, Amazon S3, Amazon AWS Ec2/EMR, Spark cluster, Hadoop Framework, Sqoop. Hadoop Developer Williams-Sonoma, Inc - San Francisco, CA September 2015 to October 2016 Responsibilities: Involved in managing nodes on Hadoop cluster and monitor Hadoop cluster job performance using Cloudera manager. Involved in

loading data from edge node to HDFS using shell scripting. Created Map Reduce programs to handle semi/unstructured data like xml, json, Avro data files and sequence files for log files. Developed Spark scripts by using Python shell commands as per the requirement. Integrated Elastic Search and implemented dynamic faceted-search. Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis. Design and Develop Pig Latin scripts and Pig command line transformations for data joins and custom processing of Map reduce outputs. Implemented Spark using Scala and Spark SQL for faster testing and processing of data. Implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. Implemented Spark RDD transformations to map business analysis and apply actions on top of transformations. Used maven to build and deploy the Jars for MapReduce, Pig and Hive UDFs. Reviewed basic SQL queries and edited inner, left, and right joins in Tableau Desktop by connecting live/dynamic and static datasets. Environment: Hadoop, Scala, Map Reduce, HDFS, Spark, Scala, Kafka, AWS, Apache SOLR, Hive, Cassandra, maven, Jenkins, Pig, UNIX, Python, MRUnit, Git.

Hadoop Developer Veritas Technologies LLC - Mountain View, CA November 2014 to August 2015

Responsibilities Responsible for building scalable distributed data solutions using Hadoop. Worked in joining raw data with the reference data using Pig scripting. Implemented DataStax Enterprise Search with Apache Solr. Created java operators to process data using DAG streams and load data to HDFS. Configured, Designed implemented and monitored Kafka cluster and connectors. Developed ETL jobs using Spark-Scala to migrate data from Oracle to new hive tables. Developed and Deployed applications using Apache Spark, Scala. Developed Oozie workflow for scheduling and orchestrating the ETL process. Implemented Spark using Scala and Spark SQL for faster testing and processing of data. Helped in troubleshooting Scala problems while working with Micro Strategy to produce illustrative reports and dashboards along with ad-hoc analysis. Developed Hive queries for the analysts and I have written scripts using Scala. Created and worked Sqoop jobs with incremental load to populate Hive External tables. Installed and configured Flume, Hive, Pig, Sqoop and Oozie on the Hadoop cluster. Continuous Integration

environments in SCRUM and Agile methodologies. Extracted the data from Teradata into HDFS using the Sqoop. Managed real time data processing and real time Data Ingestion in HBase and Hive using Storm. Environment: Hadoop, HDFS, Pig, Hive, Oozie, HBase, Kafka, Apache SOLR, MapReduce, Apache SOLR, Sqoop, Storm, Spark, Scala, LINUX, Cloudera, Maven, Jenkins, Java, SQL. Client: Well Care, Tampa, Florida Industry: HealthCare Mar'13 - Oct'14 Role: Java/Hadoop Developer Responsibilities Exported data from DB2 to HDFS using Sqoop and Developed MapReduce jobs using Java API. Used Spring AOP to implement Distributed declarative transaction throughout the application. Designed and developed Java batch programs in Spring Batch. Installed and configured Pig and wrote Pig Latin scripts. Created and maintained Technical documentation for launching Cloudera Hadoop Clusters and for executing Hive queries and Pig Scripts. Developed workflow-using Oozie for running MapReduce jobs and Hive Queries. Involved in loading data from UNIX file system to HDFS. Created java operators to process data using DAG streams and load data to HDFS. Assisted in exporting analyzed data to relational databases using Sqoop. Involved in Develop monitoring and performance metrics for Hadoop clusters. Continuous monitoring and managing the Hadoop cluster through Cloudera Manager. Environment: Hadoop, HDFS, Hive, Flume, Sqoop, HBase, PIG, Eclipse, Spark, My SQL and Ubuntu, Zookeeper, Maven, Jenkins, Java (JDK 1.6), Oracle10g. Java Developer Hexacorp Technical services, NJ October 2011 to February 2013 Responsibilities: Effectively interacted with team members and business users for requirements gathering. Coded front end components using HTML, JavaScript and jQuery, Back End components using Java, spring, Hibernate, Services Oriented components using Restful and SOAP based web services, and Rules based components using JBoss Drools. Involved in analysis, design and implementation phases of the software development lifecycle (SDLC). Implementation of spring core J2EE patterns like MVC, Dependency Injection (DI), and Inversion of Control (IOC). Implemented REST Web Services with Jersey API to deal with customer requests. Developed test cases using J Unit and used Log4j as the logging framework. Worked with HQL and Criteria API from retrieving the data elements from database. Developed user interface using HTML, Spring Tags, JavaScript, J Query and CSS.

Developed the application using Eclipse IDE and worked under Agile Environment. Utilized Eclipse IDE as improvement environment to plan, create and convey Spring segments on Web Logic Environment: Java, J2EE, JDBC, EJB, UML, Swing, HTML, JavaScript, CSS, J Query, Spring 3.0, JNDI, Hibernate 3.0, Java Mail, Web Services, REST, Oracle 10g, J Unit, Log4j, Eclipse, Web logic 10.3. Java Developer Choice Software Limited - Hyderabad, Telangana August 2008 to September 2011 Responsibilities Involved in various stages of Enhancements in the Application by doing the required analysis, development, and testing. For analysis and design of application created Use Cases, Class and Sequence Diagrams. Developed web-based user interfaces using struts framework. Developed and maintained Java/J2EE code required for the web application. Handled Client Side Validations used JavaScript and Involved in integration of various Struts actions in the framework. Involved in the development of the User Interfaces using HTML, JSP, CSS and JavaScript. Developed, Tested and Debugged the Java, JSP and EJB components using Eclipse.

Environments: Java (JDK 1.5), J2EE, Servlets, Struts, JSP, HTML, CSS, JavaScript, EJB, Eclipse, WebLogic 8.1, Windows, SOAP, Restful. Skills CASSANDRA, HDFS, IMPALA, MAHOUT, OOZIE Additional Information Technical Skills: Big Data Eco systems HDFS, Map Reduce, Hive, YARN, Pig, Sqoop, Kafka, Storm, Flume, Oozie, and ZooKeeper, Apache Spark, Apache Tez, Impala, Nifi, Apache Solr, Rabbit MQ, Scala. No SQL Databases Hbase, Cassandra, mongoDB Programming Languages C, C++, Java, J2EE, PL/SQL, Pig Latin, Scala, Python Java/J2EE Technologies Applets, Swing, JDBC, JNDI, JSON, JSTL, RMI, JMS, Java Script, JSP, Servlets, EJB, JSF, JQuery, AngularJS Frameworks MVC, Struts, Spring, Hibernate Version control SVN, CVS Business Intelligence Tools Tableau, QlikView, Pentaho, IBM Cognos intelligence Databases Oracle 9i/10g/11g, DB2, SQL Server, MySQL, Teradata Tools and IDE Eclipse, Net Beans, Toad, Maven, ANT, Hudson, Sonar, JDeveloper, Assent PMD, DB Visualizer, IntelliJ. Cloud Technologies Amazon Web Services (AWS), CDH3, CDH4, CDH5, HortonWorks, Mahout, Microsoft Azure Insight, Amazon Redshift

Name: Allison Martin

Email: bradtyler@example.com

Phone: +1-539-486-9430x149