Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Grace Note Emeryville, CA

Authorized to work in the US for any employer Work Experience Sr. Hadoop Developer Grace Note - Emeryville, CA September 2018 to Present Grace Note's video business has several different products serving our core video guide business, each with their own unique format and delivery mechanism. There is strong demand from our customers to support a consistent data schema, standardized output format and modern data delivery across the different regions in more modern, scalable and dynamic ways Responsibilities: Multiple Spark Jobs were written to perform Data Quality checks on data before files were moved to Data Processing Layer.   Implemented Spark using Scala and Spark SQL for faster testing and processing of data   Designed and Modified Database tables and used HBASE Queries to insert and fetch data from tables.   Involved in moving all log files generated from various sources to HDFS for further processing through Flume1.7.0.   Involved in deploying the applications in AWSand maintains the EC2 (Elastic Computing Cloud) and RDS (Relational Database Services) in amazon web services.   Implemented the file validation framework, UDFs, UDTFs and DAOs   Strong experienced in working with UNIX/LINUX environments, writing Unix shell scripts, Python and Perl.   Build REST web service by building Node.js Server in the back-end to handle requests sent from the front-end JQuery Ajax calls.   Importing and exporting data from different databases like MySQL, RDBMS into HDFS and HBASE using Sqoop.   Involved in creating Hive tables, loading with data and writing hive queries.   Model and Create the consolidated Cassandra, FiloDB and Spark tables based on the data profiling   Used OOZIE1.2.1Operational Services for batch processing and scheduling workflows dynamically and created UDF's to store specialized data structures in HBase and Cassandra.   Developed multiple MapReduce jobs in Java for data cleaning and pre-processing. Used Impala to read, write and query the Hadoop data in HDFS from Cassandra and configured Kafka to read and write messages from external programs.   Optimizing existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's.   Create a complete processing engine, based on Cloudera distribution, enhanced to performance.   Environment: Hadoop, HDFS, MapReduce, Yarn, Hive, Pig, HBase, Oozie, Sqoop, Kafka, Flume, Oracle 11g,

Core Java, FiloDB, Spark, Scala, Cloudera HDFS, Eclipse, Web Services (SOAP, WSDL), Oozie, Node.js, Unix/Linux, Aws, JQuery, Ajax, Python, Perl, Zookeeper. Hadoop/ Bigdata Developer Vanguard - Charlotte, NC April 2016 to August 2018 Vanguard is one of the world's largest investment companies, offering a large selection of low-cost mutual funds, advice, and related services. This project designed by big data analytics technology platform which brought together data from various structured and unstructured data sources into Hadoop platform, enabling the near real-time collection and analysis of customer data. This solution offered efficient ways to draw new customers and improve their user experience.   Responsibilities: Developed efficient MapReduce programs for filtering out the unstructured data and developed multiple MapReduce jobs to perform datacleaning and preprocessing on Hortonworks.   Implemented Data Interface to get information of customers using RestAPIand Pre-Processdata using MapReduce 2.0 and store into HDFS (Hortonworks)   Extracted files from MySQL, Oracle, and Teradata 2 through Sqoop 1.4.6and placed in HDFS Cloudera Distribution and processed.   Worked with various HDFS file formats like Avro1.7.6, Sequence File, Jsonandvarious compression formats like Snappy, bzip2.   Proficient in designing Row keys and Schema Design for NoSQL DatabaseHbaseand knowledge of other NOSQL database Cassandra.   Used Hive to perform data validation on the data ingested using scoop and flume and the cleansed data set is pushed intoHbase.   Developed the Pig 0.15.0UDF's to pre-process the data for analysis and Migrated ETL operations into Hadoopsystem using Pig Latin scripts and Python Scripts3.5.1.   Used Pig as ETL tool to do transformations, event joins, filtering and some pre-aggregations before storing the data into HDFS.   Troubleshooting, debugging & altering Talend issues, while maintaining the health and performance of theETLenvironment.   Loaded data into the cluster from dynamically generated files usingFlume and from relationaldatabase management systems using Sqoop.   Used spark to parse XML files and extract values from tags and load it into multiple hive tables.   Experienced in runningHadoop streaming jobs to process terabytes of formatted data usingPythonscripts.   Developed small distributed applications in our projects using Zookeeper3.4.7and scheduled the workflows using Oozie 4.2.0.   Proficiency in writing the Unix/Linux shell commands.   Developed a SCP Stimulator

which emulates the behavior of intelligent networking and Interacts with SSF.   Environment: Hadoop, HDFS, MapReduce, Yarn, Hive, Pig, HBase, Oozie, Sqoop, Kafka, Flume, Oracle 11g, Core Java, Spark, Scala, Cloudera HDFS, Eclipse,Oozie, Node.js, Unix/Linux, Aws, JQuery,Ajax, Python, Perl, Zookeeper. Hadoop/ Bigdata Developer Middle by Corp - Elgin, IL January 2014 to March 2016 The Middle by Corporation, through its subsidiaries, engages in the design, manufacture, marketing, distribution, and service of commercial foodservice and food processing equipment in the United States, Canada, Asia, Europe, the Middle East, and Latin America. Responsibilities: Developed multiple Map-Reduce jobs in java for data cleaning and pre-processing.   Performed Map Reduce Programs those are running on the cluster.   Involved in loading data from RDBMS and web logs into HDFS using Sqoop and Flume.   Worked on loading the data from MySQL to HBase where necessary using Sqoop.   Configured Hadoop cluster with Namenode and slaves and formatted HDFS.   Performed Importing and exporting data from Oracle to HDFS and Hive using Sqoop   Performed source data ingestion, cleansing, and transformation in Hadoop.   Supported Map-Reduce Programs running on the cluster.   Wrote Pig Scripts to perform ETL procedures on the data in HDFS.   Used Oozie workflow engine to run multiple Hive and Pig jobs.   Analyzed the partitioned and bucketed data and compute various metrics for reporting. Created HBase tables to store various data formats of data coming from different portfolios. Worked on improving the performance of existing Pig and Hive Queries.   Involved in developing HiveUDFs and reused in some other requirements. Worked on performing Join operations. Developed fingerprinting rules on HIVE which help in uniquely identifying a driver profile   Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs.   Exported the result set from Hive to MySQL using Sqoop after processing the data.   Analyzed the data by performing Hive queries and running Pig scripts to study customer behaviour.   Used Hive to partition and bucket data.   Environment: Hadoop, MapReduce, HDFS, HBase, HDP Horton, Sqoop, Data Processing Layer, HUE, AZURE, Erwin, MS Visio, Tableau, SQL, MongoDB, Oozie, UNIX, MySQL, RDBMS, Ambari, Solr Cloud, Lily HBase,Cron. Java/Hadoop Developer Fiserv - Noida, Uttar Pradesh January 2012 to December

2013 ACUMEN is software product that provides solution to credit unions.This project was based on agile methodology. Implemented using core java, Oracle, DWR and spring technology. Responsibilities: Worked on AGITAR tool, which is Junit generating software to increase the code coverage. Code coverage was major quality issue faced in Acumen at that time. It was a critical short term project. Analyze the generated Junit and add proper asserts and make it more code specific along with increasing the code coverage. This helped to boast my product knowledge as well as my Junit writing skills. Which Improved Code Quality to a commendable level. Joined EFT team in ACUMEN. This team basically dealt with the Electronic Fund Transfer, ATM, Credit Cards and online banking. Explored almost all the areas of EFT. Learned DWR. Worked with various challenging aspects like JPOS for ATM and online banking. Various logger applications for the cards. Worked on all the layers of the product, enhancing knowledge on Core Java. Domain knowledge gain was tremendous in this assignment. Environment: Core Java, Oracle, DWR, spring MVC, Agitar, Tomcat, Glassfish, ClearCase, JIRA Java Developer Xoriant - Mumbai, Maharashtra July 2011 to December 2012 Get Insured as the name suggest was an insurance domain project.It basically dealt with contracts with various State of US under Obama Health care for providing online Insurance to the user. This included showing various plans to the user according to their wages, need, age, health, existing illness and other such parameters. This health insurance plans falls in various categories of health care like children must have a dental plan. So if a person has a family member less than 18 he is shown all the dental plans. Responsibilities: Worked on one of the most critical module for project, right from the beginning phase which included requirement gathering, analysis, design, review and development. Module lead located to another location had KT from him about roughly 2 weeks, Lead was absorbed by client. Took initiative in building a new team of more than 6 members with proper knowledge transfer sessions assigning and managing tasks with JIRA. Learned Backbone JS and worked with UI team on UI enhancements. Actively participating in the daily Scrums, understanding new user stories. Implementing new requirements after discussion with Scrum masters. Working with BA,QA to identify and fix bugs, raise new feature and enhancements. Was greatly appreciated by client with

appreciation certificate and client bonus of 10k and 50k respectively.   Environment: Java/J2EE, spring MVC, Hibernate, Oracle, Backbone.js, HTML, Tomcat, WebSphere, SVN, JIRA Education Bachelor's Skills APACHE CASSANDRA, CASSANDRA, HDFS, IMPALA, MAPREDUCE, OOZIE, SQOOP, HBASE, KAFKA, ETL, FLUME, HADOOP, INFORMATICA, MONGODB, NOSQL, Avro, Hadoop, HBase, Hive, HTML Additional Information Technical Skills: Programming Languages Java, J2EE, C, SQL/PLSQL, PIG LATIN, Scala, HTML, XML Hadoop HDFS, MapReduce, HBase, Hive, Pig, Impala, SQOOP, Flume, OOZIE, Spark, SparkQL, and Zookeeper, AWS, Cloudera, Hortonworks, Kafka, Avro.   Web Technologies JDBC, JSP, JavaScript, AJAX, SOAP. Scripting Languages Java Script, Pig Latin, Python 2.7and Scala. RDBMS Languages Oracle, Microsoft SQL Server, MYSQL. NoSQL MongoDB, HBase, Apache Cassandra, FiloDB. SOA Web Services (SOAP, WSDL) IDES MyEclipse, Eclipse, and RAD Operating System Linux, Windows, UNIX, CentOS. Methodologies Agile, Waterfall model. ETL Tools Talend, Informatica Testing Hadoop MR UNIT Testing, Quality Center, Hive Testing. Other Tools SVN, Apache Ant, Junit and Star UML, TOAD, Pl/SQL Developer, JIRA, Visual Source, QC, Agile Methodology

Name: Kimberly Tucker

Email: ccruz@example.org

Phone: 888-583-0772