Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - New York Life California City, CA   Over 11+ years of experience in Information Technology which includes experience in Big data and HADOOP Ecosystem. In-depth knowledge and hands-on experience in dealing with Apache Hadoop components like HDFS, MapReduce, HiveQL, HBase, Pig, Hive, Sqoop, Oozie, Cassandra, Flume, and Spark.   Very good understanding/knowledge of Hadoop Architecture and various components such as HDFS, JobTracker, TaskTracker, NameNode, DataNode, Secondary Namenode, and MapReduce concepts   Extensively worked on MRV1 and MRV2 Hadoop architectures.   Hands on experience in writing MapReduce programs, Pig & Hive scripts. Designing and creating Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning and buckets.   Extending Hive and Pig core functionality by writing custom UDFs.   Experience in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice-versa.   Extensively used Kafka to load the log data from multiple sources directly into HDFS. Knowledge on RabbitMQ. Loaded streaming log data from various webservers into HDFS using Flume.   Experience in building Pig scripts to extract, transform and load data onto HDFS for processing. Excellent knowledge of data mapping, extract, transform and load from different data source. Experience in writing HiveQL queries to store processed data into Hive tables for analysis.   Excellent understanding and knowledge of NOSQL databases like HBase and Cassandra.   Expertise in database design, creation and management of schemas, writing Stored Procedures, Functions, DDL, DML, SQL queries & Modeling.   Extensive experience in ETL Architecture, Development, enhancement, maintenance, Production support, Data Modeling, Data profiling, Reporting including Business requirement, system requirement gathering.   Hands-on experience in shell scripting. Knowledge on cloud services Amazon web services (AWS) and Azure.   Proficient in using RDMS concepts with Oracle, SQL Server and MySQL.   Experience in Object Oriented Analysis, Design (OOAD) and development of software using UML Methodology, good knowledge of J2EE design patterns and Core Java design patterns.   Very good experience in complete project life cycle (design, development, testing and implementation) of Client Server and Web applications.   Hands on experience in application development using Java, RDBMS, and

Linux shell scripting. Worked extensively with CDH3, CDH4. Skilled in leadership, self-motivated and ability to work in a team effectively. Possess excellent communication and analytical skills along with a can-do attitude. Strong work ethics with desire to succeed and make significant contributions to the organization. Experience in processing different file formats like XML, JSON and sequence file formats. Good Knowledge in Amazon AWS concepts like EMR and EC2 web services which provides fast and efficient processing of Big Data. Good Experience in creating Business Intelligence solutions and designing ETL workflows using Tableau. Experience with Numpy, Matplotlib, Pandas, Seaborn, Plotly, and Cufflinks python libraries. Worked on large datasets by using Pyspark, numpy and pandas. Good Experience in Agile Engineering practices, Scrum methodologies, and Test Driven Development and Waterfall methodologies. Hands-on Experience in Object Oriented Analysis, Design (OOAD) and development of software using UML Methodology. Exposure to Java development projects. Hands on experience in database design using PL/SQL to write Stored Procedures, Functions, Triggers and strong experience in writing complex queries, using Oracle, DB2 and MySQL. Good working experience on different OS like UNIX/Linux, Apple Mac OS-X Windows. Experience working both independently and collaboratively to solve problems and deliver high quality results in a fast-paced, unstructured environment. Exhibited strong written and oral communication skills. Rapidly learn and adapt quickly to emerging new technologies and paradigms. Authorized to work in the US for any employer Work Experience Sr. Hadoop Developer New York Life - Franklin, TN January 2016 to Present Responsibilities: Responsible for building scalable distributed data solutions using Hadoop. Written multiple Map Reduce programs in Java for Data Analysis. Wrote Map Reduce job using Pig Latin and Java API. Performed performance tuning and troubleshooting of Map Reduce jobs by analyzing and reviewing Hadoop log files. Developed pig scripts for analyzing large data sets in the HDFS. Collected the logs from the physical machines and the OpenStack controller and integrated into HDFS using Flume. Designed and presented plan for POC on impala. Experienced in migrating Hive QL into Impala to minimize query response time. Knowledge on handling Hive queries using Spark SQL that integrate with Spark environment.

Implemented Avro and parquet data formats for apache Hive computations to handle custom business requirements. Responsible for creating Hive tables, loading the structured data resulted from Map Reduce jobs into the tables and writing hive queries to further analyze the logs to identify issues and behavioral patterns. Worked on Sequence files, RC files, Map side joins, bucketing, partitioning for Hive performance enhancement and storage improvement. Implemented Daily jobs that automate parallel tasks of loading the data into HDFS using autosys and Oozie coordinator jobs. Performed streaming of data into Apache ignite by setting up cache for efficient data analysis. Responsible for performing extensive data validation using Hive. Sqoop jobs, PIG and Hive scripts were created for data ingestion from relational databases to compare with historical data. Used Kafka to load data in to HDFS and move data into NoSQL databases(Cassandra) Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios. Involved in submitting and tracking Map Reduce jobs using Job Tracker. Involved in creating Oozie workflow and Coordinator jobs to kick off the jobs on time for data availability. Used Pig as ETL tool to do transformations, event joins, filter and some pre-aggregations. Responsible for cleansing the data from source systems using Ab Initio components such as Join, Dedup Sorted, De normalize, Normalize, Reformat, Filter-by-Expression, Rollup. Implemented business logic by writing Pig UDFs in Java and used various UDFs from Piggybanks and other sources. Implemented Hive Generic UDF's to implement business logic. Implemented test scripts to support test driven development and continuous integration. Involved in story-driven agile development methodology and actively participated in daily scrum meetings. Environment: Hadoop, Map Reduce, HDFS, Pig, Hive, Sqoop, Flume, Oozie, Java, Linux, Teradata, Zookeeper, autosys, Hbase, Cassandra, Apache ignite Sr. Hadoop Developer Express Scripts - Nashville, TN November 2014 to January 2016 Responsibilities: Developed Pyspark code to read data from Hive, group the fields and generate XML files.Enhanced the Pyspark code to write the generated XML files to a directory to zip them to CDAs Implemented REST call to submit the generated CDAs to vendor website Implemented Impyla to support JDBC/ODBC connections for Hiveserver2 Enhanced the Pyspark code to replace spark with

Impyla. Performed installation for Impyla on the Edge node   Evaluated performance of Spark application by testing on cluster deployment mode vs local mode   Experimented submissions with Test OIDs to the vendor website   Explored StreamSet Data collector Implemented StreamSets data collector tool for ingestion into Hadoop.   Created a StreamSet pipeline to parse the file in XML format and convert to a format that is fed to Solr   Built a data validation dashboard in Solr to be able to display the message record. Wrote shell script to run Sqoop job for bulk data ingestion from Oracle into Hive   Created tables for the ingested data in Hive. Scheduled Oozie job for data ingestion for the Sqoop job   Worked with JSON file format for StreamSets. Worked with Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs.   Shell scripts to dump the data from MySQL to HDFS.   Analyzing of large volumes of structured data using SparkSQL.   Configured Spark Streaming to receive real time data from the Kafka and store the stream data to HDFS.   Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework.   Worked on Maven 3.3.9 for building and managing Java based projects. Hands-on experience with using Linux and HDFS shell commands. Worked on Kafka for message queuing solutions.   Developing Unit Test Cases for Mapper, Reducer and Driver classes using MRUNIT.   Loaded and transformed large sets of structured, semi structured and unstructured data in various formats like text, zip, XML and JSON. Generated Java APIs for retrieval and analysis on No-SQL database such as HBase and Cassandra.   Written HBASE Client program in Java and web services.   Environment: Sqoop, StreamSets, Impyla, Pyspark, Solr, Oozie, Hive, Impala Hadoop Developer Barnes & Noble - Brentwood, TN July 2013 to September 2014 Responsibilities: Worked on analyzing, writing Hadoop MapReduce jobs using JavaAPI, Pig and Hive. Responsible for building scalable distributed data solutions using Hadoop.   Involved in loading data from edge node to HDFS using shell scripting.   Created HBase tables to store variable data formats of PII data coming from different portfolios.   Exported the analysed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.   Analyze large and critical datasets using Cloudera, HDFS, Hbase, MapReduce, Hive, Hive UDF, Pig, Sqoop, Zookeeper, & Spark.   Developed custom

aggregate functions using Spark SQL and performed interactive querying. Used Pig to store the data into HBase.   Creating Hive tables, dynamic partitions, buckets for sampling, and working on them using HiveQL.   Used Pig to parse the data and Store in Avro format.   Stored the data in tabular formats using Hive tables and Hive SerDes.   Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.   Worked with NoSQL databases like Hbase in creating Hbase tables to load large sets of semi structured data coming from various sources.   Implemented a script to transmit information from Oracle to Hbase using Sqoop.   Worked on tuning the performance Pig queries.   Involved in writing the shell scripts for exporting log files to Hadoop cluster through automated process.   Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team. Implemented MapReduce programs to handle semi/unstructured data like XML, JSON, and sequence files for log files.   Installed Oozie workflow engine to run multiple Hive and pig jobs. Analyzed large amounts of data sets to determine optimal way to aggregate and report on it. Environment: Hadoop, HDFS, Pig, Sqoop, Spark, MapReduce, Cloudera, Snappy, Zookeeper, NoSQL, HBase, Shell Scripting, Ubuntu, Linux Red Hat. Hadoop Developer Teva Pharmaceuticals - Nashville, TN January 2012 to June 2013 Responsibilities: Evaluated Spark's performance vs Impala on transactional data. Used Spark transformations and aggregations to perform min, max and average on transactional data.   Experienced in migrating data from HiveQL to SparkSQL. Knowledge in using Spark Data-frames to load data in Spark Data-frames. Knowledge on handling Hive queries using Spark SQL that integrate with Spark environment.   Used java to develop Restful API for database Utility Project. Responsible for performing extensive data validation using Hive. Designed a data model in Cassandra(POC) for storing server performance data.   Implemented a Data service as a rest API project to retrieve server utilization data from this Cassandra Table. Implemented Python script to call the Cassandra Rest API, performed transformations and loaded the data into Hive.   Designed data model to ingest transactional data with and without URIs into Cassandra.   Implemented shell script to call python script to perform min, max and average on utilization data of 1000s hosts and compared the performance on various levels of summarization.

Involved in creating Oozie workflow and Coordinator jobs for Hive jobs to kick off the jobs on time for data availability.   Generated reports from this hive table for visualization purpose.   Migrated HiveQL to SparkSQL to validate Spark's performance with Hive's. Implemented Proof of concept for Dynamo DB, Redshift and EMR   Proactively researched on Microsoft Azure. Presented Demo on Microsoft Azure, an overview of cloud computing with Azure. Environment: Hadoop, Azure, AWS, HDFS, Hive, Hue, Oozie, Java, Linux, Cassandra, Python, Open TSDB Hadoop Developer TDS Telecom - Mount Juliet, TN February 2010 to December 2011 Responsibilities: Worked on writing transformer/mapping Map-Reduce pipelines using Java.   Handling structured and unstructured data and applying ETL processes.   Collected the logs data from web servers and integrated in to HDFS using Flume.   Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run Map Reduce jobs in the backend.   Involved in loading data into HBase using HBase Shell, HBase Client API, Pig and Sqoop.   Designed and implemented Incremental Imports into Hive tables.   Worked in Loading and transforming large sets of structured, semi structured and unstructured data.   Extensively used Pig for data cleansing.   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.   Worked extensively with Sqoop for importing and exporting the data from HDFS to Relational Database system and vice-versa. Loading data into HDFS.   Involved in creating Hive tables, loading with data and writing hive queries that will run internally in map reduce way.   Facilitated the Production move ups of ETL components from Acceptance to Production environment Experienced in managing and reviewing the Hadoop log files.   Migrated ETL jobs to Pig scripts do Transformations, even joins and some pre-aggregations before storing the data onto HDFS. Implemented the workflows using Apache Oozie framework to automate tasks.   Worked with Avro Data Serialization system to work with JSON data formats.   Worked on different file formats like Sequence files, XML files and Map files using Map Reduce Programs.   Developed scripts and automated data management from end to end and sync up between all the clusters.   Involved in Setup and benchmark of Hadoop /HBase clusters for internal use.   Created and maintained

Technical documentation for launching HADOOP Clusters and for executing pig Scripts. Environment: Hadoop, Big Data, HDFS, Map Reduce, Sqoop, Oozie, Pig, Hive, Hbase, Flume, LINUX, Java, Eclipse, Cassandra, Hadoop Distribution of Cloudera., PL/SQL, Windows, UNIX Shell Scripting, and Eclipse Java Developer Rice University - Houston, TX March 2008 to June 2009 Responsibilities: Involved in design and development phases of Software Development Life Cycle (SDLC).   Involved in designing UML Use case diagrams, Class diagrams, and Sequence diagrams using Rational Rose.   Followed agile methodology and SCRUM meetings to track, optimize and tailored features to customer needs.   Developed user interface using JSP, JSP Tag libraries, and Java Script to simplify the complexities of the application.   Developed a Dojo based front end including forms and controls and programmed event handling.   Created Action Classes which route submittals to appropriate EJB components and render retrieved information.   Used Core java and object oriented concepts.   Used JDBC to connect to backend databases, Oracle and SQL Server 2005.   Proficient in writing SQL queries, stored procedures for multiple databases, Oracle and SQL Server 2005.   Wrote Stored Procedures using PL/SQL. Performed query optimization to achieve faster indexing and making the system more scalable.   Deployed application on windows using IBM Web Sphere Application Server.   Used Java Messaging Services (JMS) for reliable and asynchronous exchange of important information such as payment status report.   Used Web Services - WSDL and REST for getting credit card information from third party.   Used ANT scripts to build the application and deployed on Web Sphere Application Server. Environment: Core Java, J2EE, Oracle, SQL Server, JSP, JDK, JavaScript, HTML, CSS, Web Services, Windows. Java Developer CUNA Mutual - Knoxville, TN September 2006 to February 2008 Responsibilities: Involving in Analysis, Design, Implementation and Bug Fixing Activities. Involving in Functional & Technical Specification documents review.   Created and configured domains in production, development and testing environments using configuration wizard. Involved in creating and configuring the clusters in production environment and deploying the applications on clusters.   Deployed and tested the application using Tomcat web server. Analysis of the specifications provided by the clients.   Involved to Design of the Application.

Ability to understand Functional Requirements and Design Documents. Developed Use Case Diagrams, Class Diagrams, Sequence Diagram, Data Flow Diagram Coordinated with other functional consultants. Web related development with JSP, AJAX, HTML, XML, XSLT, and CSS. Create and enhance the stored procedures, PL/SQL, SQL for Oracle 9i RDBMS. Designed and implemented a generic parser framework using SAX parser to parse XML documents which stores SQL. Deployed the application on WebLogic Application Server 9.0. Extensively used UNIX /FTP for shell Scripting and pulling the Logs from the Server. Provided further Maintenance and support, this involves working with the Client and solving their problems which include major Bug fixing. Environment: Java 1.4, Web logic Server 9.0, Oracle 10g, Web services Monitoring, Web Drive, UNIX/ LINUX, Web Logic Server, JavaScript, HTML, CSS, XML Education University of Memphis - Memphis, TN July 2009 to January 2010 Bachelor's Skills APACHE HADOOP HDFS (8 years), APACHE HADOOP OOZIE (8 years), Java (9 years), Linux (9 years), SQL (10+ years) Additional Information Skills APACHE HADOOP HDFS (7 years), APACHE HADOOP OOZIE (7 years), Java (9 years), Linux (9 years), SQL (10+ years) TECHNICAL SKILLS Hadoop/Big Data Technologies HDFS, MapReduce, Hive, Pig, Sqoop, Flume, Hbase, Oozie, Zookeeper, Apache Kafka, Cassandra, StreamSets, Impyla, Solr Programming Languages Java , (JDK 5/JDK 6), C, HTML, SQL, PL/SQL, Python Client Technologies JQUERY, Java Script, AJAX, CSS, HTML 5, XHTML, D3, Angular JS Operating Systems UNIX, WINDOWS, LINUX Application Servers IBM Web sphere, Tomcat, Web Logic, Web Sphere Web technologies JSP, Servlets, JNDI, JDBC, Java Beans, JavaScript, Web Services (JAX-WS) Databases Oracle 8i/9i/10g & MySQL 4.x/5.x Java IDE Eclipse 3.x, IBM Web Sphere Application Developer, IBM RAD 7.0

Name: Valerie Davidson

Email: ryan96@example.org

Phone: 7715336055