Sr. Big Data Developer Sr. Big Data Developer Sr. Big Data Developer - UHG Hartford, CT years of overall experience in building and developing Hadoop Map Reduce solutions. Strong experience with Big Data and Hadoop technologies with excellent knowledge of Hadoop ecosystem: Hive, Spark, Sgoop, Impala, Pig, HBase, Kafka, Flume, Storm, Zookeeper, Oozie. Experienced in transferring data from different data sources into HDFS systems using Kafka producers, consumers and Kafka brokers. Expert in creating and designing data ingest pipelines using technologies such as Apache Storm-Kafka. Good Knowledge in creating processing data pipelines using Kafka and Spark Streaming. Experienced in data ingestion using Sgoop, Storm, Kafka and Apache Flume. Experience with Python libraries for big data wrangling/manipulation/analysis Experience in Apache Flume for efficiently collecting, aggregating, and moving large amounts of log data. Knowledge in Streaming the Data to HDFS using Flume. Hands on Flume to handle the real time log processing for attribution reports. Experience in Importing and Exporting the data using Sqoop from HDFS to Relational Database systems/mainframe and vice-versa. Experience with Apache Spark ecosystem using Spark-SQL, Data Frames, RDD's and knowledge on Spark MLib. Experience in ingestion, storage, querying, processing and analysis of Big Data with hands on experience in Big Data including Apache Spark, Spark SQL, Spark Streaming. Worked with Spark engine to process large scale data and experience to create Spark RDD. Knowledge on developing Spark Streaming jobs by using RDDs and leverage Spark-Shell. Expertise in Talend Big data tool with involved in architectural designing and development of ingestion and extraction job in Big Data and Spark Streaming. Having experience on RDD architecture and implementing Spark operations on RDD and also optimizing transformations and actions in Spark. Hands on Apache Spark jobs using Scala / Python in test environment for faster data processing and used Spark SQL for querying. Knowledge of Spark code and SparkSQL for testing and processing of data using PySpark. Knowledge on cloud services Amazon web services(AWS). Good in analyzing data using HiveQL, PigLatin and custom MapReduce program in Java. Expertise in MapReduce programs in HIVE and PIG to validate and cleanse the data in HDFS, obtained from heterogeneous data sources, to make it suitable for analysis. Good in Hive and Impala queries to

load and processing data in Hadoop File system (HFS). Good understanding of NoSQL Data bases and hands on work experience in writing applications on NoSQL databases like Cassandra and MongoDB. Hands on Datalake cluster with Hortonworks Ambari on AWS using EC2 and S3. Strong knowledge in MongoDB concepts includes CRUD operations and aggregation framework and in document Schema design. Experience in maintenance/bug-fixing of web based applications in various platforms. Experience in managing life cycle of MongoDB including sizing, automation, monitoring and tuning. Experience in storing, processing unstructured data using NoSQL databases like HBase. Good in developing web-services using REST, HBase Native API Experience in working with MapReduce programs using Hadoop Client to guery data from HBase. for working on Big Data. Knowledge of ETL methods for data extraction, transformation and loading in corporate-wide ETL Solutions and Data Warehouse tools for reporting and data analysis. Experienced in job workflow scheduling, monitoring tools like Oozie and Zookeeper. Experience with Oozie Workflow Engine in running workflow jobs with actions that run Java MapReduce and Pig jobs. Good experience with both Job Tracker (Map reduce 1) and YARN (Map reduce 2). Experience in managing and reviewing Hadoop Log files generated through YARN. Experience in using Apache Solr for search applications. Knowledge of Business Intelligence and Reporting. Experienced in Java, Spring Boot, Apache Tomcat, Maven, Gradle, Hibernate and open source frameworks/ software's. Preparation of Dashboards using Tableau. Experience with all stages of the SDLC and Agile Development model right from the requirement gathering to Deployment and production support. Proficient on Test driven development (TDD), Agile to produce high quality Hands on Agile (Scrum), Waterfall model along with automation and enterprise tools deliverables. like Jenkins, Chef, JIRA, Confluence to develop projects and version control, Git. Excellent Java development skills using Java 6/7/8, J2EE, J2SE, Servlets, Junit, JSP, JDBC. Experience using middleware architecture using Sun Java technologies like J2EE, JSP, Servlets, and application Familiarity working with popular frameworks likes Struts, servers like Web Sphere and Web logic. Hibernate, Spring MVC and AJAX. Exceptional ability to learn and master new technologies and to deliver outputs in short deadlines. Quick Learner, with high degree of passion and commitment in

Proficiency with mentoring and on-boarding new engineers who are not proficient in Hadoop work. and getting them up to speed quickly. Work Experience Sr. Big Data Developer UHG - Hartford, CT, US November 2017 to Present UnitedHealth Group Inc. is an American profit managed health care company based in Minnetonka, Minnesota. It is sixth in the United States on the Fortune 500. UnitedHealth Group offers health care products and insurance services. UnitedHealth Group is the largest healthcare company in the world by revenue \$184 billion in 2016. UnitedHealth Group subsidiaries companies together serve approximately 115 million individuals in 2016. As a Sr. Big Data developer, I am leading a \$15M EDW project to implement combined data management infrastructure for Reporting & Analytics team. The task at hand is to process and analyze huge sets of health insurance data in Datalake using Hadoop and Spark tools. Responsibilities: Act as a liaison for visualizers, business and product teams for their specific data and reporting needs. Manage Agile Software Practice using Rally by creating Product Backlogs, Iterations and Sprints in collaboration with business and product teams. Working with cloud services like Amazon Web Services (AWS) and involved in ETL, Data Integration and Migration. Import data using Sgoop to load data from MySQL to HDFS on regular basis. 3+ years of experience with Python libraries such as Pandas, SciKit-Learn and Numpy Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis. Collect and aggregate large amounts of web log data from different sources such as webservers, mobile and network devices using Apache Kafka and stored the data into HDFS for analysis. Develop multiple Kafka Producers and Consumers from scratch implementing organization's requirements. Responsible for creating, modifying topics (Kafka Queues) as and when required with varying configurations involving replication factors, partitions and TTL. Write and test complex MapReduce jobs for aggregating identified and validated data. Create Managed and External Hive tables with static/dynamic partitioning. Extensively involved in performance tuning of the HiveQL by performing bucketing on large Hive tables. Implement Spark applications using Python/PySpark and Spark SQL for faster testing and processing of data. Develop an equivalent Spark-PySpark for existing SAS code to extract summary insights on the Hive tables. Extensive experience in

writing Pig scripts to transform raw data from several data sources into forming baseline data. Implement workflow using Oozie for running Map Reduce jobs and Hive Queries. Expert in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. Develop and execute shell scripts to automate the jobs. Write complex Hive gueries and UDFs. Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala. Develop multiple POCs using PySpark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata. Analyze the SQL scripts and designed the solution to implement using Involved in loading data from UNIX file system to HDFS. PySpark. Extract the data from Teradata into HDFS using Sqoop. Handle importing of data from various data sources, performed transformations using Hive, Map Reduce, Spark and loaded data into HDFS. Involved in analysis, design, testing phases and responsible for documenting technical specifications. Develop Kafka producer and consumers, HBase clients, Spark and Hadoop MapReduce jobs along with components on HDFS, Hive. Very good understanding of Partitions, Bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance. Work on the core and Spark SQL modules of Spark extensively. Experienced in running Hadoop streaming jobs to process terabytes data. Involved in importing the real-time data to Hadoop using Kafka and implemented the Oozie job for daily imports. Environment: Hadoop, Amazon Web Services (AWS), HDFS, Map Reduce, Hive, Sgoop, Apache Kafka, Zookeeper, Spark, Hbase, Python, Shell Scripting, Oozie, Maven, Hortonworks Hadoop Developer Target - Minneapolis, MN, US October 2015 to November 2017 Target Corporation is the second-largest discount retailer in the United States, behind Walmart, and is a component of the S&P 500 Index. Headquartered in Minneapolis, Minnesota, Target currently operates 1793 locations across the United States, and has annual revenue of \$74 billion. The project is to implement 300 node Hadoop cluster at Target and replace traditional systems like Oracle Exadata, Teradata, SAS, etc. to store and process customer activity and sales data. The goal was to be able to analyze the customer behavior and know more about their buying patterns and come up with recommended products based on their behavior. Target's

main Hadoop cluster has nearly 300 nodes, and it's populated with 2 PB of data--mostly structured data such as customer transaction, point of sale, and supply chain. Responsibilities: Built a suite of Linux scripts as a framework for easily streaming data feeds from various sources onto HDFS. Written multiple MapReduce programs in java 8 for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV and other compressed file formats. Wrote Interface specifications to ingest structured data into appropriate schemas and tables to support the Involved in migration from Hadoop System to Spark System; rules and analytics. Involved in start to end process of Hadoop cluster setup where in installation, configuration and monitoring the Responsible for Cluster maintenance, Cluster Monitoring, Troubleshooting, Hadoop Cluster. Manage & review Hadoop log files. Installation of various Hadoop Ecosystems. Responsible for Installation and configuration of Hive, Pig, Oozie, HBase and sqoop on the Hadoop cluster. Spark Streaming API with Kafka to build live dashboards; Worked on Transformations & actions in RDD and Spark Streaming, MapReduce, Pair RDD Operations, Partitioner, Check-pointing, and SBT. Refactored formal Hive gueries to Spark SQL. Involved in moving all log files generated from various sources to HDFS for further processing through Flume. Extracted the data from Teradata into HDFS using Sqoop and exported the patterns analyzed back into Teradata using Processed HDFS data and created external tables using Hive and developed scripts to Sqoop. ingest and repair tables that can be reused across the project. Used Talend Open Studio to perform ETL aggregations in Hadoop HIVE & PIG. Wrote Pig Scripts to generate MapReduce jobs and performed ETL procedures on the data in HDFS. Developed Pig Scripts, Pig UDFs and Hive Scripts, Hive UDFs to load data files. Used hive schema to create relations in pig using HCatalog. Developed a Java MapReduce and pig cleansers for data cleansing. Implemented POC to migrate map reduce jobs into Spark RDD transformation using Scala IDE for Eclipse. Implemented Machine Learning Models like K-means clustering using PySpark. Used Spark to create reports for analysis of the data coming from various sources like transaction logs. **Used Oozie Operational**

Services for batch processing and scheduling work flows dynamically. Used Maven extensively for building jar files of MapReduce programs and deployed to cluster. Managed Agile Software

Practice using Rally by creating Product Backlog, Iterations and Sprints in collaboration with the Product Team. Environment: Hadoop, HDFS, Hive, Pig, MapReduce, YARN, Datameer, Sqoop, Flume, Oozie, Linux, Teradata, HCatalog, Java 8, Eclipse IDE, GIT. Hadoop Developer Yum Brands - Louisville, KY, US September 2013 to October 2015 Yum! Brands, Inc is an American fast food company. A Fortune 500 corporation, Yum operates the licensed brands Taco Bell, KFC, Pizza Hut, and WingStreet worldwide. Prior to 2011, Yum! Also owns Long John Silver's and A&W Restaurants. Based in Louisville, Kentucky, it is one of the world's largest fast food restaurant companies in terms of system units - with 42,692 restaurants (including 8,927 that are company-owned, 796 that are unconsolidated affiliates, 30,930 that are franchised, and 2,039 that are licensed) around the world in over 130 countries. Their global annual revenue is US\$13.105 billion. The project objective is to extr act all the sales and the products sold across all the Yum Brands Restaurant's and load it into the data warehouse using the Big Data components. It also involves loading of data from traditional databases to HDFS using various Big Data technologies. Yum Brands wants to use big data analytics to gain lot more insight to improve operations at its various stores and enhance customer experience. This would help it analyze data about various factors such as wait times, information on the menu, the size of the orders, ordering patterns of the customers to optimize the operations of its restaurants at specific locations. My role involved the design, development of various modules in Hadoop Big Data Platform and processing data using MapReduce, Hive, Pig, Scoop and Oozie. Responsibilities: Designed, developed and tested Map Reduce programs on Mobile Offers Redemptions and Sent it to the downstream applications like HAVI. Scheduled this MapReduce job through Oozie workflow. Ingested huge amount of XML files into Hadoop by Utilizing DOM Parsers with in Map Reduce. Extracted Daily Sales, Hourly Sales and Product Mix of the items sold in Yum Brand Restaurant's and loaded them into Global Data Warehouse. Wrote and tested the MapReduce code to do aggregations on identified and validated Scheduled Multiple Map Reduce jobs in Oozie. Involved in extracting the promotions data for data. all stores within USA by writing the map reduce jobs and automating it with UNIX shell script. Gathered business requirements in meetings for successful implementation and POC and moving it to Production. Implemented POC to migrate map reduce jobs into Spark RDD transformations using Scala IDE for Eclipse. Implemented different machine learning techniques in Scala using Developed Spark applications using Scala for easy Hadoop Scala machine learning library. Created RDD's in Spark Scala and Python. Closely worked with Admin team to transitions. gather hardware for Data nodes, edge nodes, and Name nodes. Successfully loaded files to Hive and HDFS from Oracle, Netezza and SQL Server using SQOOP. Used Talend Open Studio to load files into Hadoop HIVE tables and performed ETL aggregations in Hadoop HIVE. Designed & Created ETL Jobs through Talend to load huge volumes of data into Cassandra. Used Sqoop to import data from SQL server to Cassandra. Worked on analyzing, writing Hadoop MapReduce jobs using Java API, Pig and Hive. Developed some machine learning algorithms using Mahout for data mining for the data stored in HDFS. Used Flume extensively in gathering and moving log data files from Application Servers to a central location in Hadoop Distributed File System (HDFS). Worked with Oozie Workflow manager to schedule Hadoop jobs and high intensive jobs. Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, manage and review data backups, manage and review Hadoop log files. Loaded data into HIVE tables, and extensively used Hive/HQL or Hive queries to guery data in Hive Tables. Introduced Tableau Visualization to Hadoop to produce reports for Business and BI team.

Creating UDF functions in Pig & Hive and applying partitioning and bucketing techniques in Hive for performance improvement. Creating indexes and tuning the SQL gueries in Hive and Involved in database connection by using Sqoop. Involved in Hadoop Namenode metadata backups and load balancing as a part of Cluster Maintenance and Monitoring. Worked on Spark with Python and Scala. Serialized data in Hadoop using Avro serialization system. Used File System Check (FSCK) to check the health of files in HDFS. Monitored Nightly jobs to export data out of HDFS to be stored offsite as part of HDFS backup. Used Pig for analysis of large data sets and brought data back to Hbase by Pig. Scheduled, monitored and debugged various MapReduce, Pig, Hive jobs using Oozie Workflow. Design and deployment of Storm cluster integration with Kafka and Implemented authentication and authorization service using Kerberos authentication HBase.

protocol. Environment: Hadoop 1.2.1, MapReduce, Sgoop 1.4.4, Hive 0.10.0, Flume 1.4.0, Oozie 3.3.0, Pig 0.11.1, Hbase 0.94.11, Scala, Python, Zookeeper 3.4.3, Talend Open Studio, kafka, Storm, Oracle 11g/10g, Apache Cassandra, SQL Server 2008, MySQL 5.6.2, Java 7, SQL, PLSQL, Toad 9.7, Eclipse Kepler IDE, Microsoft Office 2007, MS Outlook 2007, SharePoint Teamsite. Hadoop Developer Bank of the West - San Francisco, CA, US July 2012 to September 2013 Bank of the West is a diversified financial services holding company, headquartered in San Francisco, California. It is a subsidiary of BNP Paribas. It has more than 700 branches in the Midwest and Western United States with annual sales of over \$2.1 billion. Bank of the West uses the Hadoop cluster to construct a new and more accurate score of the risk in its customer portfolios. The Bank wants to use Apache Hadoop to process the customer data that is collected from thousands of banking products and different systems. This project aimed at exploiting Hadoop's ability to store vast volumes of unstructured data allows the company to collect and store web logs, transaction data and social media data. Collecting all the customer related data, which is high in volume and velocity, helps the bank with analysis in a cost effective manner for enhanced insight and decision-making, and to identify any credit and debit card frauds. Responsibilities: Analyzed, Designed and developed the system to meet the requirements of business users. Participated in the design review of the system to perform Object Analysis and provide best possible solutions for the application. Imported and exported terabytes of data using Sqoop from HDFS to Relational Developed MapReduce Jobs using Hive and Pig. Collected and aggregated Database Systems. large amounts of log data using Apache Flume and staging data in HDFS for further analysis. Installed and configured Hadoop Map Reduce, HDFS, developed multiple Map Reduce jobs in Java for data cleaning and preprocessing. Developed Map Reduce (YARN) jobs for accessing and validating the data. Involved in managing and reviewing Hadoop log files. Load and transform large sets of structured, semi structured and unstructured data. Responsible to manage data coming from different sources Involved in loading data from UNIX file system to HDFS. and configured Hive and also written Hive QL scripts. Involved in creating Hive tables, loading with data and writing hive queries which run internally in map reduce way. Implemented Partitioning,

Dynamic Partitions, Buckets in HIVE. Monitor System health and logs and respond accordingly to any warning or failure conditions. Used ClearCase for version control. Environment: Hadoop, Map Reduce, HDFS, Hive, Java, Hadoop distribution of Hortonworks, Cloudera, Map, Flat files, Oracle 11g/10g, UNIX Shell Scripting, ClearCase, Junit Java Developer Penn Mutual - Horsham, PA, US February 2010 to July 2012 The Penn Mutual Life Insurance Company, commonly referred to as Penn Mutual, was founded in Philadelphia, Pennsylvania, in 1847. It was the seventh mutual life insurance company chartered in the United States. The project was for an insurance portal which would enable Penn Mutual's customers to get a quote, Login to their policy, Report a Claim, Pay Bills, Update their Policy information, and make updates to their policies. Responsibilities: Involved in the analysis, Design, Coding, Modification and implementation of User Requirements in the Electronic Credit File Management system. Designed the application using Front Controller. Service Controller, MVC, Session Facade Design Patterns. The application is designed using MVC Architecture. Implemented the required functionality using Hibernate for persistence & Spring Frame work. Used Spring Framework for Dependency Injection. Designed and implemented the Hibernate Domain Model for the services. Developed UI using HTML, JavaScript and JSP and developed Business Logic and Interfacing components using Business Objects, XML, and JDBC. Designed user-interface and checking validations using JavaScript. Involved in design of JSP's and Servlets for navigation among the modules. Developed various EJBs for handling business logic and data manipulations from database. Managed connectivity using JDBC for querying/inserting & data management including triggers and stored procedures. Developed SQL queries and Stored Procedures using PL/SQL to retrieve and insert into multiple database schemas.

Developed the XML Schema and Web services for the data maintenance and structures Wrote test cases in JUnit for unit testing of classes. Used DOM and DOM Functions using Firefox and IE Developer Tool bar for IE. Debugged the application using Firebug to traverse the documents. Involved in developing web pages using HTML and JSP. Provided Technical support for production environments resolving the issues, analysing the defects, providing and implementing the solution defects. Built and deployed Java applications into multiple UNIX based environments and

produced both unit and functional test results along with release notes. Developed the presentation layer using CSS and HTML taken from bootstrap to develop for browsers. Environment: Java 6, spring, JSP, Hibernate, XML, HTML, JavaScript, JDBC, CSS, SOAP Web services. Jr. Java Developer Apollo Hospitals August 2008 to February 2010 Apollo Hospitals is an Indian hospital chain based in Chennai, India. It was founded in 1983 and has hospitals in India, Bangladesh, Kuwait and Qatar. Several of the group's hospitals have been among the first in India to receive international healthcare accreditation by America-based Joint Commission International (JCI). The project is to develop a web-based application containing all the features for Hospital Management. It contains the different modules like Clinical Modules, Supporting Modules & Admin Modules. It is focused on delivering services to residents. Services include inpatient and outpatient care, primary care, community health and wellness, workplace health, home health, community mental health, rehabilitation and long-term care. Responsibilities: Involved in Development of master screens like Service Requests, Change Requests Screens. Design architecture following J2EE MVC framework. Developed interfaces using HTML, JSP pages and Struts -Presentation View. Developed Struts Framework and configuring web.xml and struts-config.xml according to the struts framework. Developed and implemented Servlets running under JBoss. Used J2EE design patterns and Data Access Object (DAO) for the business tier and integration Tier layer of the project; Developed Java UI using swing. Used Java Message Service (JMS) for reliable and asynchronous exchange of important information between the clients and the customer and developed Message driven beans that consumed the messages from the Java message queue.

Development of database interaction code to JDBCAPI making extensive use of SQL Query Statements and advanced prepared statement. Taken care of complete Java multi-threading part in back end components. Inspection/Review of quality deliverables such as Design Documents. Wrote SQL Scripts, Stored procedures and SQL Loader to load reference data. Used Web Services for interacting with a remote client to access data; Performed Unit Testing and Regression testing; Fixed the bugs identified in test phase; Used Junit for testing Java classes. Used Ant building tool to build the application. Environment: J2EE (Java Servlets, JSP, Struts), MVC

Framework, Apache Tomcat, Oracle8i, JMS, SQL, HTML, JDBC, EJB, ANT, Junit, Education MBA in Big Data & Analytics University of St. Thomas - Minneapolis, MN, US 2016 Skills APACHE HADOOP HDFS (6 years), java (9 years), SQL (9 years), UNIX (6 years), XML (8 years) Additional Information TECHNICAL SKILLS Big Data Technology: HDFS, Mapreduce, HBase, Pig, Hive, SOLR, Sgoop, Flume, MongoDB, Cassandra, Puppet, Oozie, Zookeeper, Spark, Kafka, Talend Java/J2EE Technology: JSP, JSF, Servlets, EJB, JDBC, Struts, Spring, Spring MVC, Spring Portlet, Spring Web Flow, Hibernate, iBATIS, JMS, MQ, JCA, JNDI, Java Beans, JAX-RPC, JAX-WS, RMI, RMI-IIOP, EAD4J, Axis, Castor, SOAP, WSDL, UDDI, JiBX, JAXB, DOM. MyFaces(Tomahawk), Facelets, JPA, Portal, Portlet, JSR 168/286, LifeRay, WebLogic Portal, LDAP, JUnit.NET Hadoop Distribution: Cloudera, Horton works, IBM Big Insights Cloud Computing Service: AWS (Amazon Web Services) Languages: Python, Java (5/6/7/8), C/C++, Swing, SQL, HTML, CSS, i18n, I10n, DHTML, XML, XSD, XHTML, XSL, XSLT, XPath, XQuery, SQL, PL/SQL, UML, JavaScript, AJAX(DWR), jQuery, Dojo, ExtJS, Shell Scripts, Perl Development Framework/IDE: RAD 8.x/7.x/6.0, IBM WebSphere Integration Developer 6.1, WSAD 5.x, Eclipse Galileo/Europa/3.x/2.x, MyEclipse 3.x/2.x, NetBeans 7.x/6.x, IntelliJ 7.x, Workshop 8.1/6.1, Adobe Photoshop, Adobe Dreamweaver, Adobe Flash, Ant, Maven, Rational Rose, RSA, MS Visio, OpenMake Meister Web/Application Servers: WebSphere Application Server 8.x/7.0/6.1/5.1/5.0, WebSphere Portal Server 7.0/6.1, WebSphere Process Server 6.1, WebLogic Application Server 8.1/6.1, JBoss 5.x/3.x, Apache 2.x, Tomcat 7.x/6.x/5.x/4.x, MS IIS, IBM HTTP Server Databases: NoSQL, Oracle 11q/10q/9i/8i, DB2 9.x/8.x, MS SQL Server 2008/2005/2000, MySQL NoSQL: HBase, Cassandra, MongoDB Reporting Tools: Tableau, Datameer Operating Systems: Windows XP, 2K, MS-DOS, Linux (Red Hat), Unix (Solaris), HP UX, IBM AIX Version Control: CVS, SourceSafe, ClearCase, Subversion Monitoring Tools: Embarcadero J Optimizer 2009, TPTP, IBM Heap Analyzer, Wily Introscope, Jmeter Other: JBoss Drools 4.x, REST, IBM Lotus WCM, MS ISA,CA Site Minder, BMC WAM, Mingle

Name: Robert James

Email: thompsonrobert@example.org

Phone: 001-482-665-4253x391