

Sr.Hadoop Developer Sr.Hadoop Developer Sr.Hadoop Developer - Rsvp hospitality New York, NY  
Having 8+years of overall experience in IT industry includes implementing, developing and maintenance in Big Data Technologies and Web based applications using Java, J2EE technologies.

Hands on experience in implementing various solutions and analysing data using Hadoop Ecosystems like HDFS, Map Reduce, Yarn, Spark, Sqoop, Hive, Pig, Flume, Kafka, Impala, Oozie, Oozie coordinator, Zookeeper and Cassandra, HBase. Hands on experience in various Hadoop distributions like Cloudera (CDH3, CDH4 & CDH5), Horton works Data Platform (HDP) and MapR. Experience in performing in-memory data processing and real time streaming analytics using Apache Spark with Scala, Java and Python. Experience in Sqoop to import data from various external sources into Hadoop ecosystem components like HDFS, HBase and Hive, as well as exports data from Hadoop to other external sources. Experience on Spark Streaming API with Kafka to feed live streaming data into HDFS and optimized it using spark concepts like Data frames, Partitioning, Bucketing, Parallel execution and Map Side Joins using broadcast joins. Experience in advanced procedures like text analytics, processing using in memory computing capabilities like Spark written in Scala. Experience in Spark-SQL to process large amount of data by implementing Spark-RDD transformations, Actions and Data frames to required input data. Experience on data analysis and developing scripts for pig Latin and Hive QL using Java. Experience in using Ambari for provisioning, managing, monitoring and securing apache Hadoop cluster. Experience in analyzing data from Cassandra for quick processing, sorting and grouping through CQL. Experience on Managing and scheduling Jobs in Hadoop Cluster using Oozie and used Zookeeper for cluster coordination services. Experience in configuring different topologies in Storm to import and process information only from multiple sources, collect into central repository Hadoop. Experience in developing and designing POCs using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata. Good Knowledge in Amazon AWS concepts like EMR, EC2, EBS, S3 and RDS web services which provides fast and efficient processing of Big Data. Extensive experience with SQL, PL/SQL, Shell Scripting and database concepts. Experience with front end technologies like HTML, CSS and JavaScript. Experience in

working with Windows, UNIX/LINUX platform with different technologies such as Big Data, SQL, XML, HTML, Core Java, Shell Scripting etc. Good working experience on different file formats like Json, Avro, Parquet, compression techniques like snappy& bzip. Hands on experience in data warehousing with ETL tools like Informatica, Ab initio, IBM Data Stage for various loading and transformation processes. Good working experience and knowledge in NOSQL databases like HBase, Cassandra, Mongo DB, Couch DB. Intensive working experience with Amazon Web Services(AWS)using S3 for storage, EC2 for computing and RDS, EBS. Experience in NIFI work flow scheduler managing Hadoop jobs by Direct Acyclic graph (DAG) of actions with control flows. Experience in using IDEs and source control repositories like Eclipse, IntelliJ and GitHub, Maven, SBT. Experience in preparing Tableau reports for analyzed data by using excel sheets, flat files, CSV files. Experience in different data sources like Oracle, Netezza, SQL and PostgreSQL. Experience in J2EE technologies like Java servlets, JSP, EJB, and JDBC etc. Work Experience Sr.Hadoop Developer Rsvp hospitality - New York, NY March 2018 to Present Description: RSVP Hospitality is a UAE consulting firm specialized in Revenue Management and Business Analytics for Hotels, Spa/Salons and Service Industry projects. RSVP Hospitality provide our clients, extensive expertise in different sectors in hospitality, leisure and wellness industries. Responsibilities: Contributing as a member of a high performing, the agile team focused on next-generation data &analytics Build Big Data Analytics and Visualization platform for handling high-volume batch-oriented and real-time data streams. Utilized Agile Scrum Methodology to help manage and organize a team with regular code review sessions. Used Pig as ETL tool to do transformations, event joins, filter both traffic and some pre-aggregations before storing the data onto HDFS. Experienced 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. Analysed different big data analytics using Hive import data from RDBMS to HDFS. Implemented complex big data with a focus on collecting, parsing, managing, analyzing, and visualizing large sets of data to turn information into business insights using multiple platforms in the Hadooecosystem. Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.

Imported data from the structured data source into HDFS using Sqoop incremental imports. Created Hive tables, partitions and implemented incremental imports to perform ad-hoc queries on structured data. Build Hive tables using list partitioning and hash partitioning and created Hive Generic UDF's to process business logic with HiveQL. Integrated HBase with MapReduce to move the bulk amount of data into HBase. Developed SQL scripts using Spark for handling different data sets and verifying the performance over Map Reduce jobs. Supported MapReduce Programs those are running on the cluster and Wrote MapReduce jobs using JavaAPI. Extensively used Apache Sqoop for efficiently transferring bulk data between Apache Hadoop and relational databases (Oracle, MySQL) for predictive analytics. Developed storytelling dashboards in Tableau Desktop and published them on to TableauServer and used GitHub version controlling tools to maintain project versions. Environment:Hadoop, Java, MapReduce, HDFS, Hive, Linux, XML, Eclipse, Cloudera, Spark, HBase, MongoDB, Python, GitHub, Sqoop, Oozie, DB2, SQL Server, Oracle 12c, MySQL Hadoop Developer Neurotrack - San Francisco, CA June 2016 to February 2018 Description: Neurotrack is dedicated to the development of non-invasive cognitive health assessment tools that will enable earlier and more effective evaluation of patients who may be at risk for cognitive decline, and help advance research of treatments for cognitive diseases, including Alzheimer's. Neurotrack's mission is to revolutionize the early diagnosis and treatment of Alzheimer's disease with our online assessment and digital therapy. Responsibilities: Analysing the requirement to setup a cluster. Worked on analyzing Hadoop cluster and different big data analytic tools including Map Reduce, Hive and Spark. Involved in loading data from LINUX file system, servers, Java web services using Kafka Producers, partitions. Implemented Kafka Custom encoders for custom input format to load data into Kafka Partitions. Implemented Storm topologies to pre-process data before moving into HDFS system. Implemented Kafka High level consumers to get data from Kafka partitions and move into HDFS. Implemented POC to migrate Map Reduce programs into Spark transformations using Spark and Scala. Migrated complex Map Reduce programs into Spark RDD transformations, actions. Implemented Spark RDD transformations to map business analysis and apply actions on top of transformations. Involved in creating Hive

tables, loading with data and writing hive queries which runs internally in Map Reduce way. Developed the Map Reduce programs to parse the raw data and store the pre Aggregated data in the partitioned tables. Loaded and transformed large sets of structured, semi structured, and unstructured data with Map Reduce, Hive and pig. Developed Map Reduce programs in Java for parsing the raw data and populating staging Tables. Experienced in developing custom input formats and data types to parse and process unstructured and semi structured input data and mapped them into key value pairs to implement business logic in Map Reduce. Involved in using HCATALOG to access Hive table metadata for Map Reduce or Pig code. Experience in implementing custom sterilizer, interceptor, source and sink as per the requirement in flume to ingest data from multiple sources. Experience in setting up Fan-out workflow in flume to design vshaped architecture to take data from many sources and ingest into single sink. Developed Shell, Perl and Python scripts to automate and provide Control flow to Pig scripts. Exporting of result set from HIVE to MySQL using Sqoop export tool for further processing. Evaluated usage of Oozie for Workflow Orchestration. Converted unstructured data to structured data by writing Spark code. Indexed documents using Apache Solr. Set up Solr Clouds for distributing indexing and search. Automation of all the jobs starting from pulling the Data from different Data Sources like MySQL and pushing the result dataset to Hadoop Distributed File System and running MR, PIG, and Hive jobs using Kettle and Oozie (Work Flow management) Worked on No-SQL databases like Cassandra, Mongo DB for POC purpose in storing images and URIs. Integrating bulk data into Cassandra file system using Map Reduce programs. Used Talend ETL tool to develop multiple jobs and in setting workflows. Created Talend jobs to copy the files from one server to another and utilized Talend FTP components. Worked on Mongo DB for distributed storage and processing. Designed and implemented Cassandra and associated Restful web service. Implemented Row Level Updates and Real time analytics using CQL on Cassandra Data. Used Cassandra CQL with Java API's to retrieve data from Cassandra tables. Worked on analysing and examining customer behavioural data using Cassandra. Created partitioned tables in Hive, mentored analyst and SQA team for writing Hive Queries. Developed Pig Latin scripts to extract the data from the web server output

files to load into HDFS. Involved in cluster setup, monitoring, test benchmarks for results. Involved in build/deploy applications using Maven and integrated with CI/CD server Jenkins. Involved in agile methodologies, daily scrum meetings, spring planning's. Environment: Hadoop, Cloudera, HDFS, pig0., Hive, Flume, Sqoop, Oozie, AWS Redshift9, Python, Spark, Scala Mongo DB, Cassandra, Solr, ZooKeeper, MySQL, Talend., Shell Scripting, Linux Red Hat, Java. Hadoop Developer AT&T - Dallas, TX February 2015 to May 2016 Description: AT&T Inc. is an American multinational telecommunications holding company headquartered at Whitacre Tower in Downtown Dallas, Texas. It is the world's largest telecommunications company, the second largest provider of mobile telephone services. The purpose of this project was to eliminate the need for various regional databases, to streamline processes and reduce the address fallouts. This provides consistent data for various stakeholders while having access to real-time data. Responsibilities: Apart from the normal requirement gathering, participated in a Business meeting with the client to gather security requirements. Assisted with the architect to analyze the existing system and future system Prepared design blue prints and application flow documentation Experienced 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 and application Supported Map Reduce Programs those are running on the cluster Created MapReduce jobs to extracts the contents from HBase and configured in OOZIE workflow to generate analytical reports. Extracted files from Cassandra through Sqoop and placed in HDFS and processed. Implemented Bloom filters in Cassandra using keyspace creation Involved in writing Cassandra CQL statements God hands-on experience in developing concurrency using spark and Cassandra together Involved in writing spark applications using Scala Hands on experience in creating RDDs, transformations, and Actions while implementing spark applications Good knowledge in creating data frames using Spark SQL. Involved in loading data into Cassandra NoSQL Database Documented all the challenges, issues involved to deal with the security system and Implemented best practices Created Project structures and configurations according to the project architecture and made it available to the junior developer to continue their work Environment: Java, Python,

Cassandra, HTML5, CSS, PIG, HIVE, Hortonworks distribution of Hadoop 2.3, YARN, Ambari

Hadoop developer Standard & Poor's - New York, NY March 2013 to January 2015 Description: Standard & Poor's Financial Services LLC (S&P) is an American financial services company. It is a division of S&P Global that publishes financial research and analysis on stocks, bonds, and commodities. S&P is known for its stock market indices such as the U.S.-based S&P 500, the Canadian S&P/TSX, and the Australian S&P/ASX 200. Responsibilities: Worked on analyzing Hadoop stack and different big data analytic tools including Pig, Hive, HBase database and Sqoop. Experienced to implement Hortonworks distribution system (HDP 2.1, HDP 2.2 and HDP 2.3). Developed Map Reduce programs for some refined queries on big data. Developed interactive shell scripts for scheduling various data cleansing and data loading process. Experienced in managing and reviewing the Hadooplog files and importing log files from various sources into HDFS using Flume. Used HIVE queries to import data into MicrosoftAzure cloud and analyzed the data using HIVE scripts. Creating Hive tables and working on them for data analysis to cope up with the requirements. Developed a framework to handle loading and transform large sets of unstructured data from the UNIX system to HIVE tables. Worked with a business team in creating Hive queries for ad hoc access. In-depth understanding of Classic MapReduce and YARN architectures. Implemented Hive Generic UDF's to implement business logic. Used Hive to analyse the partitioned and bucketed data and compute various metrics for reporting. Analysed the data by performing Hive queries, ran Pig scripts, SparkSQL and Spark Streaming. Extracted files from Cassandra through Sqoop and placed in HDFS for further processing. Involved in creating generic Sqoop import script for loading data into Hive tables from RDBMS. Environment: Horton works, Hadoop, Map Reduce, HDFS, Hive, Pig, Sqoop, AZURE, Oozie, SQL, Spark, HBase, Cassandra, and GitHub.

Java Developer Comcast - Philadelphia, PA December 2011 to February 2013 Description: Comcast NBCUniversal creates incredible technology and entertainment that connects millions of people to the moments and experiences that matter most. Responsibilities: Involved in each phase of Software Development Life Cycle(SDLC) models like Requirement gathering and analysis, Design, Implementation, Testing, Deployment and Maintenance. Developed Login,

Policy and Claims Screens for customers using HTML 5, CSS3, JavaScript, AJAX, JSP, and jQuery. Used Core Java to develop Business Logic. Involved in the development of business module applications using J2EE technologies like Servlets, JSP. Designed and developed the web-tier using JSP's, Servlets framework. Used various Core Java concepts such as Multi-Threading, Exception Handling, Collection APIs to implement various features and enhancements. Strong experience in design & development of applications using Java/J2EE components such as Java Server Pages (JSP). Developed EJB MDB's and message Queue's using JMS technology. EJB Session Beans were used to process requests from the user interface and CMP entity beans were used to interact with the persistence layer. Developed stored procedures, triggers, and queries using PLSQL in SQL Server. Use Spring MVC as framework and JavaScript for client-side view, used frameworks for client-side data validation, creating dynamic web pages-Ajax, jQuery. Developed model classes based on the forms to be displayed on the UI. Implemented various design patterns in the project such as Business Delegate, Data Transfer Object, Data Access Object, Service Locator and Singleton. Used SQL statements and procedures to fetch the data from the database. Developed test cases and performed unit test using JUnit Framework. Used CVS as version control and ANT scripts to fetch, build, and deploy application to development environment. Environment: Java, HTML, CSS, JavaScript, MySQL, Struts, EJB, Spring MVC.

Java Developer Unisys - Blue Bell, PA April 2010 to November 2011 Description: Unisys Corporation is an American global information technology company that provides IT services, software, and technology. It is the legacy proprietor of the Burroughs and UNIVAC line of computers, formed when the former bought the latter Responsibilities: Performed Analysis, Design, Development, Integration and Testing of application modules. Implemented application prototype using JSP, Servlets, JDBC, and to give the presentation. Developed new web page designs and development of the project presentation layer by using HTML, JavaScript, JSF, Ajax and implemented CSS for User Interface and better appearance. Implemented data base queries by using SQL and PL/SQL to perform data analysis, extraction and various functions. Implemented an application by using Struts Framework which leverages classical Model View Layer Architecture(MVC). Involved in

fixing bugs and unit testing with test cases using Junit. Worked with various software development methodologies like Agile, Waterfall to increase the development of the project. Environment: Java, HTML, CSS, JavaScript, JSON, JSP, JDBC and SQL, PL/SQL. Education Bachelor's ..... Skills Cassandra, Hdfs, Impala, Oozie, Sqoop Certifications/Licenses Driver's License

Name: Travis Glass

Email: william32@example.net

Phone: 968.659.2328x0453