Hadoop Developer Hadoop Developer Hadoop Developer - Walmart   6+ Years of experience in Big Data Analytics using Hadoop eco-system components like Map Reduce, Pig, Hive, HBase, Flume, Sqoop, Oozie, Spark and Kafka.   Solid understanding of Distributed Systems Architecture, MapReduce and Sparx execution frameworks for large scale parallel processing.   Experience working with all major Hadoop distributions like Cloudera(CDH), Horton works(HDP) and AWS EMR.   Developed Spark applications using Spark core anddata frames.   Used Spark-SQL and Spark streaming API's in Scala.   Implemented custom business logic and performed join optimization, secondary sorting, custom sorting using Map Reduce programs.   Experienced testing and running of Map Reduce pipelines on Apache Crunch.   Hands on experience with accessing and perform CURD operations against HBase data using Java API.   Performed Spark application troubleshooting and fine tuning.   Worked on data integration using Kafka, Spark streaming and HBase.   Depth understanding of NoSQL databases such as Hbase and its integration with Hadoop cluster.   Strong work experience in extracting, wrangling, ingestion, processing, storing, querying and analyzing structured, semi-structured and unstructured data.   Solid understanding of Hadoop MRv1 and Hadoop MRv2 (YARN architecture).   Experience in creating Maps side join, Reducer side join, shuffle & sort, distributed cache, compression techniques, multiple Hadoop input and output formats.   Worked with various data formats like csv, text, sequential files, parquet, Avro, Jason, RC files.   Designed HIVE queries to perform data analysis, data transfer and table design to load data into Hadoop environment.   Implemented Sqoop for large datasets transfer between HDFS and RDBMS and vice-versa.   Extensive experience on importing and exporting data using stream processing platforms like Flume and Kafka.   Experienced with AWS, where cluster was built using Ec2 instances, store data in S3 and ran EMR jobs.   Sound knowledge over J2EE Design Patterns like MVC Architecture, Singleton, Factory Pattern, Front Controller, Session Facade, Business Delegate and Data Access Object for building J2EE Applications.   Experienced in developing the unit test cases using MRUnit, JUnit and Easy Mock.   Knowledge on Splunk for logging mechanism.   Experience in using Maven and ANT for build automation.   Experience in using version control and configuration management tools like SVN, CVS.   Used Talend ETL tool

to Extract the data from different data sources, transform and load the data into the dataware house systems for reporting and analysis. Strong understanding of SDLC and process methodologies like waterfall and Agile. Involved in sprint planning, project documentation, story creations project related meetings. Work Experience Hadoop Developer Walmart - Bentonville, AR January 2018 to Present Responsibilities: Developed TDCH Scriptsfor improving and exporting data into HDFS and Hive. Used Fair Schedule to allocate resources in Yarn. Responsible to manage data coming from different sources. Involved in creating Hive tables, loading with data and writing Hive queries. Implemented partitioning, dynamic partitions, buckets in HIVE. Implemented the workflows using Apache Oozie framework to automate tasks. Worked on CICD pipeline (chip/smartcard interface devices), integrating code changes to Git repository and build using Jenkins. Read the ORC files(optimized row columnar) and create data frames to use in spark. Experienced working with Spark core and Spark SQL using Pyspark. Performed data transformation and analytics on large dataset using Spark. Integrated spark jobs with MLP platforms. Wrote Kafka procedures to stream the data from external rest APIs to Kafka topics. Wrote Spark streaming application to consume the data from Kafka topics and write the processed streams to Hbase. Used reporting tools like Tableau to connect with impala for generating daily reports of data. Involved in Analysis, Design, Development and Testing process based on the new business requirements. Environment: Hadoop, Hive, S3, Spark, Oozie, Teradata, Yarn, Unix, Hortonworks, TDCH, Kafka, Tableau. Verizon Wireless - Fort Worth, TX 2017 to December 2017 Responsibilities: Responsible for building scalable distributed data solution using Apache Hadoop and Spark. Involved in combining traditional transactional data and event data with social network data. Involved in loading data from relational database into HDFS using Sqoop. Deployed scalable Hadoop cluster on AWS using S3 as underlying file system for Hadoop. Worked with Spark core, Spark streaming and Spark SQL modules of Spark. Involved in developing generic Spark-Scala functions for transformations, aggregations and designing schema for rows. Experienced in working with Spark APIs like RDDs, Datasets, Data Frames and DStreams to perform transportation on the data. Used Spark SQL to perform interactive analysis on the data using SQL and HiveQL. Experienced

in processing live data streams using Spark streaming with high level functions like map, reduce, join and window.   Experienced in minimizing data transfers over Hadoop clusters using Spark optimization like broadcasts variables and accumulators.   Worked on troubleshooting Spark application to make them more error tolerant.   Involved in loading the processed data into Hive warehouse.   Stored the data in table formats using Hive tables and Hive Serdes.   Implemented static partitions, dynamic partitions and buckets in Hive.   Used Oozie operational services for scheduling workflows dynamically.   Used reporting tools like Tableau to connect with Impala for generating daily reports of data.   Designed, documented operational problems by following standards using JIRA.   Collaborated with the infrastructure, network, database, application and other teams to ensure data quality and availability.   Environment: Hadoop 2.x, Spark core, Spark SQL, Spark streaming, Scala, Hive, Sqoop, Oozie, Amazon EMR, Tableau, Impala, JIRA. INDIA.Hadoop developer Infotech Enterprises - Hyderabad, Telangana April 2014 to June 2015 Responsibilities: Worked on Spark SQL, reading/writing data from JSON file, text file, parquet file, schema RDD.   Identifying data sources and create appropriate data ingestion procedures. Transformed the data using Spark, Hive, Pig for BI team to perform visual analytics according to the client requirement.   Developed the service to run the Map-reduce jobs as per the requirement basis.   Importing and exporting data into HDFS and HIVE, PIG using Sqoop.   Populated big data customer marketing data structures.   Developed Spark scripts by using Python as per requirements.   Performed joins on tables in Hive with various optimization techniques. Implemented lateral view in conjunction with UDFs in Hive according to the client requirements. Created Hive tables as per internal requirements in static and dynamic partitions.   Implemented the workflows using Apache Oozie framework to automate the tasks.   Developing design documents considering all possible approaches and identifying best of them.   Developed scripts and automated data management from end to end sync up between the clusters.   Imported the data from different sources like HDFS/Hbase into Spark RDD.   Experienced with Spark context, Spark-SQL, Data frame, Pair RDD's, Spark Yarn.   Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala.   Automated the jobs for pulling data from FTP

server to load data into Hive tables using Oozie workflows.   Environment:Hive, Sqoop, Python, Shell scripting, Spark, Oozie, Scala, Java. Java developer Virinchi Technologies - Hyderabad, Telangana July 2012 to March 2014 Responsibilities: Designed the application using J2EE design patterns based on MVC architecture.   Designed the user interface using HTML, CSS, java script and jQuery.   Used JDBC for accessing data from Oracle database.   Validated web forms for client-side validation as per the requirements.   Developed custom tags, JSTL to support custom user interfaces.   Handled business logic as a model using helper classes and servlets to control the flow of application as controller as server-side validations.   Developed code to convert JSON data to customize JavaScript objects.   Developed servlets and JSPs based on MVC pattern using Struts framework.   Developed EJB-session bean acts as Fa ade.   Used Maven for building application EAR for deploying on Web logic application server.   Used Eclipse development, testing and code review.   Performed unit testing on application to verify and identify various scenarios. Environment:J2EE, Java, Eclipse, EJB, JDBC, Web logic, Pl/SQL, Junit, Maven, Log4j, UML. Education Science & Engineering JNT University Skills Apache (2 years), APACHE HADOOP HDFS (2 years), APACHE HADOOP OOZIE (2 years), Java. (2 years), SQL (4 years)

Name: Christopher Fisher

Email: alicia35@example.org

Phone: 755.620.6246x6671