Hadoop Developer Hadoop Developer Hadoop Developer - American Express Phoenix, AZ   Over 7+ years of total professional experience in IT field involving project development, implementation, deployment and maintenance using Hadoop ecosystem related technologies with domain knowledge in Finance, Banking, Communication, Insurance, Retail Industry and Health care.   4+ years of hands on experience in Hadoop Ecosystem technologies like HDFS, MapReduce, Yarn, Spark, Hive, Pig, Oozie, Sqoop, Flume, Zookeeper, HBase.   Over all two years of hands on experience using Spark framework with Scala.   3+ years of Java programming experience in developing web based applications and Client- Server technologies.   In depth understanding of Hadoop Architecture and its various components such as Job Tracker, Task Tracker, Name Node, Data Node, Resource Manager, and MapReduce concepts.   Proficient knowledge on Apache Spark and Apache Storm to process real time data.   Extensive knowledge in programming with Resilient Distributed Datasets (RDDs).   Good exposure to performance tuning hive queries, map-reduce jobs, spark jobs.   Worked with various formats of files like delimited text files, click stream log files, Apache log files,Avro files, JSON files, XML Files   Experience on installation and configuration of spark standalone mode for testing and development environments.   Developed simple to complex MapReduce jobs using Java language.   Worked on live 60 nodes Hadoop cluster running on Cloudera CDH4.   Extensive experience in analyzing data using Hive QL, Pig Latin and custom MapReduce programs in Java   Developed UDF, UDAF, UDTF functions for Hive and Pig.   Good knowledge of Partitions, Bucketing concepts, designed and managed them and created external tables in Hive in order to optimize performance.   Good experience in Avro files, RC files, Combiners, Counters for best practices and performance improvements.   Good knowledge on Joins, group and aggregation concepts and resolved performance issues in Hive and Pig scripts by implementing them.   Experience with Big Data ML toolkits such as Mahout and Spark ML.   Experience in job work flow scheduling and monitoring tools like Oozie and Zookeeper.   Worked on NoSQL databases including HBase, Cassandra and Mongo DB.   Experience in importing data from a Relational database management system (RDBMS) such as MySql and Oracle into HDFS, Hive and exported the processed data back into RDBMS using Sqoop.

Experience in importing data from RDBMS to HBase and exporting data into RDBMS using Sqoop.

Implemented Flume for collecting, aggregating and moving large amount of server logs and streaming data to HDFS. Experience in HBase cluster setup and implementation. Done Administration, installing, upgrading and managing distributions of Cassandra. Good knowledge in performance troubleshooting and tunning Cassandra clusters and understanding of Cassandra Data Modeling based on applications. Experience in setting up Hadoop in Pseudo distributed environment. Experience in setting up Hive, Pig, HBase and Sqoop in Ubuntu operating system. Good knowledge on Software development life cycle(SDLC). Experience as Java Developer in Web, Client Server technologies using Java, J2EE, Servlets, JSP, EJB, Hibernate framework and Spring framwork. Good understanding of Software Development Life Cycle(SDLC) and sound knowledge of project implementation methodologies including Waterfall and Agile. Authorized to work in the US for any employer Work Experience Hadoop Developer American Express - Phoenix, AZ October 2016 to Present Responsibilities: Worked on analyzing Hadoop cluster using different big data analytic tools including Kafka, Pig, Hive and Map Reduce. Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis. Real time streaming the data using Spark with Kafka. Configured Spark streaming to receive real time data from the Kafka and store the stream data to HDFS using Scala. Worked within the Apache Hadoop framework, utilizing Opinion Lab statistics to ingest the data from a streaming application program interface (API), automate processes by creating Oozie workflows, and draw conclusions about consumer sentiment based on data patterns found through the use of Hive for external client use. Wrote the Storm topology with HDFS Bolt and Hive Bolts as destinations. Expertise in writing Storm topology development, maintenance and bug fixes. Developed Hadoop streaming Map/Reduce works using Java. Worked on debugging, performance tuning of Hive & Pig Jobs. Implemented test scripts to support test driven development and continuous integration. Worked on tuning the performance of Pig queries. Involved in loading data from Linux file system to HDFS. Importing and exporting data into HDFS using Sqoop. Good knowledge on building Apache spark applications using Scala. Experience working on processing unstructured data using Pig.

Implemented Partitioning, Dynamic Partitions, Buckets in Hive.   Implemented Spark using Scala and SparkSQL for faster testing and processing of data.   Good knowledge with NoSQL databases like HBase, Cassandra   Handled Administration, installing, upgrading and managing distributions of Cassandra.   Advanced knowledge in performance troubleshooting and tuning Cassandra clusters.   Done Scaling Cassandra cluster based on lead patterns.   Good understanding of Cassandra Data Modeling based on applications.   Experience with Cassandra Performance tuning.   Highly involved in development/implementation of Cassandra environment.   Plan, deploy, monitor, and maintain Amazon AWS cloud infrastructure consisting of multiple EC2 nodes and VMWare Vm's as required in the environment.   Expertise in AWS data migration between different database platforms like SQL Server to Amazon Aurora using RDS tool.   Supported Map Reduce Programs those are running on the cluster.   Gained experience in managing and reviewing Hadoop log files.   Involved in scheduling Oozie workflow engine to run multiple pig jobs.   Responsible for developing data pipeline using flume, Sqoop and Pig to extract the data from weblogs and store in HDFS.   Data scrubbing and processing with Oozie.   Developed Pig Latin scripts to extract data from the web server output files to load into HDFS.   Involved in developing Hive DDLs to create, alter and drop tables.   Created and maintained technical documentation for launching Hadoop clusters and for executing Hive queries and Pig Scripts.   Also exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark SQL, Data Frame, Pair RDDs, Storm, Spark YARN.   Used many features like Parallelize, Partitioned, Caching (both in-memory and disk Serialization), Kryo Serialization, etc.   Environment: Hadoop, Cloudera, Big Data, HDFS, MapReduce, Sqoop, Spark, Hive, HBase, Linux, Java, Eclipse, Hadoop Distribution of Cloudera, PL/SQL, Toad 9.6, Windows NT, MongoDB, Cassandra, Tableau, Unix shell scripting, Putty and Eclipse. Hadoop Developer Barclays - New York, NY August 2015 to September 2016 Responsibilities: Developed simple to complex MapReduce streaming jobs using Java language for processing and validating the data.   Developed data pipeline using MapReduce, Flume, Sqoop and Pig to ingest customer behavioral data into HDFS for analysis.   Developed MapReduce and Spark jobs to discover trends in data usage by users.   Implemented Spark using

Python and Spark SQL for faster processing of data.   Implemented algorithms for real time analysis in Spark.   Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data.   Used the Spark -Cassandra Connector to load data to and from Cassandra.   Real time streaming the data using Spark with Kafka.   Handled importing data from different data sources into HDFS using Sqoop and also performing transformations using Hive, MapReduce and then loading data into HDFS.   Exported the analyzed data to the relational databases using Sqoop, to further visualize and generate reports for the BI team.   Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis   Analyzed the data by performing Hive queries (HiveQL) and running Pig scripts (Pig Latin) to study customer behavior.   Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.   Developed Pig Latin scripts to perform Map Reduce jobs.   Developed product profiles using Pig and commodity UDFs.   Developed Hive scripts in HiveQL to De-Normalize and Aggregate the data.   Created HBase tables and column families to store the user event data.   Written automated HBase test cases for data quality checks using HBase command line tools.   Created UDF's to store specialized data structures in HBase and Cassandra.   Scheduled and executed work flows in Oozie to run Hive and Pig jobs.   Used Impala to read, write and query the Hadoop data in HDFS from HBase or Cassandra.   Used Tez framework for building high performance jobs in Pig and Hive.   Configured Kafka to read and write messages from external programs.   Configured Kafka to handle real time data.   Developed end to end data processing pipelines that begin with receiving data using distributed messaging systems Kafka through persistence of data into HBase.   Written Storm topology to emit data into Cassandra DB.   Written Storm topology to accept data from Kafka producer and process the data. Continuous monitoring and managing the Hadoop cluster using Cloudera Manager.   Used JUnit framework to perform Unit testing of the application   Developed interactive shell scripts for scheduling various data cleansing and data loading process.   Performed data validation on the data ingested using MapReduce by building a custom model to filter all the invalid data and cleanse the data.   Experience with data wrangling and creating workable datasets.   Developed schemas

to handle reporting requirements using Jaspersoft.   Environment: Hadoop, MapReduce, Spark, Pig, Hive, Sqoop, Oozie, HBase, Zookeeper, Kafka, Flume, Solr, Storm, Tez, Impala, Mahout, Cassandra, Cloudera manager, MySQL, Jaspersoft, Multi-node cluster with Linux-Ubuntu, Windows, Unix. Hadoop Developer Ally Financial - Auburn Hills, MI November 2014 to July 2015 Responsibilities: Installed and configured Hadoop Ecosystem components and Cloudera manager using CDH distribution.   Frequent interactions with Business partners.   Designed and developed a Medicare-Medicaid claims system using Model-driven architecture on a customized framework built on spring.   Moved data from HDFS to Cassandra using MapReduce and BulkOutputFormat class.

  Imported trading and derivatives data in Hadoop Distributed File System and Eco System (MapReduce, Pig, Hive, Sqoop).   Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running Hive queries and Pig scripts.

  Created tables in HBase and loading data into HBase tables.   Developed scripts to load data from HBase to Hive Meta store and perform Map Reduce jobs.   Was part of an activity to setup Hadoop ecosystem at dev & QA Environment.   Managed and reviewed Hadoop Log files.   Responsible writing Pig Script and Hive queries for data processing   Running Sqoop for importing data from Oracle & Other Database.   Creation of shell script to collect raw logs from different machines. Created Partition in a Hive as static and dynamic.   Implemented Pig Latin scripts using operators such as LOAD, STORE, DUMP, FILTER, DISTINCT, FOREACH, GENERATE, GROUP, COGROUP, ORDER, LIMIT, AND UNION.   Optimized the Hive tables using optimization techniques like partitions and bucketing to provide better performance with HiveQL queries. Defined some Pig UDF for some financial functions such as swap, hedging, Speculation and arbitrage   Coded many MapReduce program to process unstructured logs file.   Worked on import and export data into HDFS and Hive using Sqoop.   Used different data formats (Text format and Avro format) while loading the data into HDFS.   Used parameterize Pig script and optimized script using illustrate and explain.   Involved in the process of configuring HA, Kerberos security issues and name node failure restoration activity time to time as a part of zero downtime.   Implemented FAIR Scheduler as well.   Environment: Hadoop, Linux, MapReduce, HDFS, Hbase, Hive, Pig, Shell

Scripting, Sqoop, CDH Distribution, Windows, Linux, Java 6, Eclipse, Ant, Log4j and JUnit. Hadoop Developer Hitachi America - Chula Vista, CA May 2014 to October 2014 Responsibilities: Hands on Experience in joining raw data with the reference data using Pig scripting.   Written custom UDF's in Hive.   Hands on experience in extracting data from different databases and to copy into HDFS file system using Sqoop.   Written Sqoop incremental import job to move new/updated info. From database to HDFS.   Created Oozie coordinated workflow to execute Sqoop incremental job daily.   Used Oozie workflow engine to run multiple Hive and Pig jobs.   Hands on experience in exporting the results into relational databases using Sqoop for visualization and to generate reports for the BI team.   Involved in Installing and configuring Hive, Pig, Sqoop, Flume and Oozie on the Hadoop cluster.   Worked with application teams to install Operating System, Hadoop updates, patches, versions upgrades as required.   Working with clients on requirements based on their business needs.   Communicate deliverables status to user/stakeholders, client and drive periodic review meetings.   On time completion of tasks and the projects per quality goals.   Environment: Hadoop, HDFS, MapReduce, Hive, Pig, Sqoop, HBase, Oozie, MySql, SVN, Putty, Zookeeper, Ganglia, UNIX and Shell scripting. Hadoop Developer Florida Power & Light - Miami, FL October 2013 to April 2014 Responsibilities: Integrated Kafka with Storm for real time data processing and written some storm topologies to store the processed data directly to MongoDB and HDFS. Experience in writing Spark SQL scripts.   Imported data from different sources into Spark RDD for processing.   Developed custom aggregate functions using Spark SQL and performed interactive querying.   Involved in loading data from edge node to HDFS using shell scripting.   Worked on installing cluster, commissioning and decommissioning of Datanode, Namenode high availability, capacity planning and slots configuration.   Completion of unit testing for the new Hadoop jobs in standalone mode designated for Unit region using MR Unit.   Developed Spark scripts by using Scala and Python shell commands as per the requirement.   Experience in managing and reviewing Hadoop log files.   Experience in Hive partitioning, bucketing and perform joins on Hive tables and implementing Hive SerDe like REGEX, JSON and Avro.   Optimized Hive analytics Sql queries, created tables/views, written custom UDF's and Hive based exception processing.   Involved in

transforming the Teradata to legacy lables to HDFS and HBase tables using Sqoop and vice versa.
Configured Fair Scheduler to provide fair resources to all the applications across the cluster.
Environment: Hortonworks Hadoop, Ambari, Spark, Solr, Kafka, MongoDB, Linux, HDFS, Hive, Pig, Sqoop, Flume, Zookeeper, RDBMS. Java/J2EE Developer Mercury Insurance - Roseville, CA November 2012 to September 2013 Responsibilities: Write design document based on requirements from MMSEA user guide.   Performed requirement gathering, design, coding, testing, implementation and deployment.   Worked on modeling of Dialog process, Business Processes and coding Business Objects, QueryMapper and JUnit files.   Involved in the design and creation of Class diagrams, Sequence diagrams and Activity Diagrams using UML models   Created the Business Objects methods using Java and integrating the activity diagrams.   Involved in developing JSP pages using Struts custom tags, jQuery and Tiles Framework.   Used JavaScript to perform client side validations and Struts-Validator Framework for server-side validation   Worked in web services using SOAP, WSDL.   Wrote Query Mappers and MQ Experience in JUnit Test Cases.   Developed the UI using XSL and JavaScript.   Managed software configuration using ClearCase and SVN.   Design, develop and test features and enhancements.   Performed error rate analysis of production issues and technical errors.   Developed test environment for testing all the Web Service exposed as part of the core module and their integration with partner services in Integration test.   Analyze user requirement document and develop test plan, which includes test objectives, test strategies, test environment, and test priorities.   Responsible for performing end-to-end system testing of application writing JUnit test cases   Perform Functional testing, Performance testing, Integration testing, Regression testing, Smoke testing and User Acceptance Testing (UAT).   Converted Complex SQL queries running at mainframes into pig and Hive as a part of a migration from mainframes into Hadoop cluster.   Environment: Shell Scripting, Java 6, JEE, Spring, Hibernate, Eclipse, Oracle 10g, JavaScript, Servlets, Nodejs, JMS, Ant, Log4j and Junit, Hadoop (Pig & Hive). Java Developer Reliance Energy LTD June 2010 to August 2012 Responsibilities: Involved in the design and implementation of the architecture for the project using OOAD, UML design patterns.   Involved in design and development of server side layer using XML,

JSP, JDBC, JNDI, EJB and DAO patterns using eclipse IDE.   Work involved extensive usage of HTML, CSS, Javascript and Ajax for client side development and validations.   Used parsers for the conversion of XML files to java objects and vice versa.   Developed screens using XML documents and XSL.   Developed Client programs for consuming the Web services published by the Country Defaults Department which keeps in track of the information regarding life span, inflation rates, retirement age, etc. using Apache Axis.   Developed Java Beans and JSP's by using spring and JSTL tag libs for supplements.   Development of EJB's, Servlets and JSP files for implementing Business rules and Security options using IBM Web Sphere.   Involved in creating tables, stored procedures in SQL for data manipulation and retrieval using SQL Server, Oracle and DB2.   Trained end users on developed application.   Environment: Java, JSF Framework, Eclipse IDE, Ajax, Apache Axis, OOAD, Web Logic , Java script, HTML, XML, CSS, SQL Server, Oracle, Web services, Ajax, Spring, OOAD and UML, Windows. Education Bachelor's Skills JAVA (6 years), SQL (6 years), Hadoop (5 years), HADOOP (5 years), Hive (5 years), Hadoop, Cloudera, Big Data, HDFS, MapReduce, Sqoop, Spark, Hive, HBase, Linux, Java, Eclipse, Hadoop Distribution of Cloudera, PL/SQL, Toad 9.6, Windows NT, MongoDB, Cassandra, Tableau, Unix shell scripting, Putty and Eclipse (8 years) Additional Information TECHNICAL SKILL   Hadoop Ecosystem Hadoop, HDFS, YARN, MapReduce, Hive, Pig, Oozie, Zookeeper, Flume, Sqoop, Spark, Mahout, Kafka, Storm, Scala, Spark ML, Impala, HBase, Tez.   Languages C, C++, Core Java, PL/SQL, Scala. Java/J2EE Technologies Servlets, JSP, JDBC, Java Beans, EJB, RMI & Web Services. Frameworks EJB, Struts, Hibernate and Spring. Scripting Languages Python, R, SQL, Unix Shell Scripting, Hive QL, Pig Latin. Databases MySQL, Oracle 10g, SQL Server 2008. NoSQL Databases HBase, Cassandra, MongoDB. Web Technologies HTML, CSS, XML, Javascript, Ajax, Node.js, SOAP. Web/Application Servers WebLogic, WebSphere, Apache Tomcat. Development Tools Eclipse, Ant, Putty. Version Control SVN, GIT. Cluster Management and Monitoring Tools Ganglia, Ambari. Visualization Tools Tableau. Methodologies Agile/Scrum, Rational Unified Process and Waterfall. Environment Win 95/98, Win NT, Win XP, Win 7, Unix, Linux(Ubuntu and CentOS).

Name: Cody Fisher

Email: robertlevine@example.com

Phone: 592-936-2508x5215