Sr. Hadoop / Spark Developer Sr. Hadoop / Spark Developer Sr. Hadoop / Spark Developer - American Express Palo Alto, CA   Over 6 years of professional IT experience which includes experience in Big data ecosystem experience in complete project life cycle (design, development, unit testing and implementation) of which over 3+ years of work experience in ingestion, storage, querying, processing and analysis of Big Data with hands on experience in Hadoop Ecosystem (YARN, HDFS) and its components Hive, Pig, HBase, Sqoop, Hue, Kafka, Flume, Oozie, Zookeeper, Spark, Spark SQL and Spark Streaming.   Hands on experience in improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, pair RDD's, Spark YARN and experience in application in Apache Spark using python (Spark).   Working experience on building spark applications using build tools like SBT, Maven and Gradle.   Good experience in dealing with different file formats like text, Sequence, RCFILE, ORC, Parquet, Avro and JSON and different compression formats like GZip, LZO, BZip2 and snappy. Good knowledge on relational databases like MySQL, Oracle and NoSQL databases like HBase, MongoDB.   Working experience in handling semi/un-structured data from different data sources. Working experience in developing Map side join, Reducer side join, Distributed Cache, Compression techniques, Multiple Input & output.   Good working experience in performing ad-hoc analysis on structured data using HiveQL, joins and Hive Generic UDF's good exposure to Counters, Shuffle & Sort parameters, Dynamic Partitions, Bucketing for performance improvement.   Good knowledge in using IDE like Net Beans, Eclipse, Intellij. Authorized to work in the US for any employer Work Experience Sr. Hadoop / Spark Developer American Express - Phoenix, AZ January 2018 to Present Responsibilities: Designed the solution using Storm Spouts (to stream data from Kafka) and Bolts connecting to Java APIs developed independently based on the application logic. Imported bulk data into HBase Using Map Reduce programs.   Written Storm topology to accept the events from Kafka producer and emit into HBase.   Developed a data pipeline using Kafka and Strom to store data into HDFS.   Developed HDFS with huge amounts of data using Apache Kafka.   Implemented a proof of concept (Poc's) using Kafka, Strom, HBase for processing streaming data.   Created HBase tables to load large sets of structured, semi-structured and unstructured data

coming from UNIX, NoSQL and a variety of portfolios.   Development of software using core java with integration of Apache Storm, Apache Kafka.   Integrated Oozie with the rest of the Hadoop stack supporting several types ofHadoop jobs out of the box (such as Map-Reduce, Pig, Hive, and Sqoop) as well as system specific jobs (such as Java programs and shell scripts).   Exploring with Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, pair RDD's, Spark YARN.   Developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.   Experience in deploying data from various sources into HDFS and building reports using Tableau.   Performed real time analysis on the incoming data.   Load the data into Spark RDD and performed in-memory data computation to generate the output response.   Developed Spark scripts by using Python shell commands as per the requirement.   Developed Shell scripts and Python programs to automate tasks   Environment: Hadoop, Map Reduce, HDFS, Spark, Java, Kafka, Hive, HBase, maven, Jenkins, Pig, UNIX, Python, Git, Storm, MapR, Oozie Hadoop/Spark Developer East West Bank - Palo Alto, CA October 2016 to June 2017 Responsibilities: Worked with lambda architecture in handling and processing batch and real-time data.   Using Sqoop, ingested the Data from data warehouse to HDFS.   Using Kafka, collected real-time streaming and log data from web applications and click stream data, analyzing a part of data using spark streaming and rest stored into HDFS for future use.   Worked in creating External and Managed Hive tables.   Worked in writing Hive Queries for analyzing data in Hive warehouse using Hive Query Language (HiveQL) and worked with Hive Tables, Hive queries, Partitioning, Bucketing.   Performed Data Profiling, identify data quality and validating rules regarding data integrity and data quality as it relates to the impact on business requirements.   Build spark applications using SBT builds.   Used Spark SQL to process the huge amount of structured data.   Worked on Python plugin on MySQL workbench to upload CSV files   Migrated python scikit learn machine learning to data frame-based spark machine learning algorithms.   Created Rich dashboards using Tableau Dashboard.   Connected Tableau server to publish dashboard to a central location for portal integration.   Creation of metrics, attributes, filters, reports, and dashboards created advanced chart types, visualizations and complex calculations to manipulate

the data.   Environment: Cloudera Manager, Sqoop, Java (jdk1.8 Version), Hive, Spark, Spark-SQL, Scala. Hadoop Developer United Airlines - Chicago, IL May 2015 to September 2016 Responsibilities: Worked using Apache Hadoop ecosystem components like HDFS, Hive, Sqoop and Worked with Spark, Scala and Python.   Involved in data extraction from distributed RDBMS like Teradata and Oracle.   Involved in loading data from UNIX file system to HDFS.   Used Map Reduce JUnit for unit testing.   Good experience in handling data manipulation using python Scripts.   Developed data pipeline using Flume, Sqoop, Pig and Java map reduce and Spark to ingest customer behavioral data and purchase histories into HDFS for analysis.   Troubleshooting the cluster by reviewing Hadoop LOG files. Involved in managing and reviewing Hadoop log files. Installed and configured Pig for ETL jobs.   Experienced in migrating iterative map reduce programs into Spark transformations using Scala.   Import the data from different sources like HDFS/Hbase into Spark RDD.   Used Oozie to manage the Hadoop jobs.   Involved in running Hadoop streaming jobs to process terabytes of text data.   Load and transform large sets of structured, semi structured and unstructured data.   Imported data using Sqoop from Teradata using Teradata connector. Implemented Partitioning, Dynamic Partitioning, and Bucketing in HIVE.   Exported the result set from HIVE to MySQL using Shell scripts.   Used Zookeeper for various types of centralized configurations.   Involved in maintaining various Unix Shell scripts.   Implemented Fair schedulers on the Job tracker to share the resources of the Cluster for the Map Reduce jobs given by the users.   Automated all the jobs starting from pulling the Data from different Data Sources like MySQL to pushing the result set Data to Hadoop Distributed File System using Sqoop.   Used GIT for version control.   Maintain System integrity of all sub-components (primarily HDFS, MR, HBase, and Flume).   Monitor System health and logs and respond accordingly to any warning or failure conditions.   Environment: Hadoop, HDFS, Map Reduce, Hive, Pig, Sqoop, Hue, Impala, NoSQL, Java 1.6, Mongo and Spark. Sr. Java Developer Fresenius Medical Care - Waltham, MA March 2014 to April 2015 Roles and Responsibilities: Involved in various phases of Software Development such as modeling, system analysis and design, code generation and testing using AGILE Methodology   Participated in daily Stand up meetings with Scrum Master.   Responsible for

estimating the time for given task (Service Request)   Designed, developed and deployed application using Eclipse and Tomcat application Server   Developed SQL queries and used JDBC to interact with the Oracle database   Configured and customized logs using Log4J   Involved in installing and configuring Eclipse and Maven for development   Used SVN as a version management tool   Written Test Cases for Unit Level Testing using JUnit   Developed database objects like Stored Procedures, Functions, Triggers, index, views to maintain referential integrity of the database   Responsible to monitor all the tasks assigned to offshore and ensure they are delivered with in SLA   Part of the Maintenance and Enhancement Group which provides on-going support for systems after implementation occurs   Environment: Spring Framework, Hibernate, PL/SQL, Tomcat 7.x, log4j, JUnit, Oracle 10g, Eclipse, AGILE Methodology, SCRUM, SOAP Webservices, Unix Shell Scripting, SVN, Oracle 9i, Maven 4 Java/Web Developer Comcast - Herndon, VA October 2012 to February 2014 Responsibilities: Involved in various phases of Software Development Life Cycle (SDLC) as design development and unit testing.   Developed and deployed UI layer logics of sites using JSP, XML, JavaScript, HTML/DHTML and Ajax.   CSS and JavaScript were used to build rich internet pages.   Agile Scrum Methodology been followed for the development process.   Designed different design specifications for application development that includes front-end, back-end using design patterns.   Developed proto-type test screens in HTML and JavaScript.   Involved in developing JSP for client data presentation and, data validation on the client side within the forms.   Developed the application by using the Spring MVC framework.   Collection framework used to transfer objects between the different layers of the application.   Developed data mapping to create a communication bridge between various application interfaces using XML, and XSL.   Spring IOC being used to inject the parameter values for the Dynamic parameters.   Developed JUnit testing framework for Unit level testing.   Actively involved in code review and bug fixing for improving the performance.   Documented application for its functionality and its enhanced features.   Created connection through JDBC and used JDBC statements to call stored procedures.   Environment: Spring MVC, J2EE, Java, JDBC, Servlets, JSP, XML, Design Patterns, CSS, HTML, JavaScript, Junit, Apache Tomcat, My SQL Server 2008. Education

Bachelor's Skills HDFS, MAPREDUCE, OOZIE, SQOOP, HBASE, FLUME, MONGODB, NOSQL, VISUAL STUDIO, Git, HBase, Hive, HTML, JAVASCRIPT, MapReduce, Pig, PYTHON, XML, ZooKeeper, ECLIPSE Additional Information TECHNICAL SKILLS: Operating Systems: Windows, Ubuntu, UNIX   Big Data: HDFS, MapReduce, PIG, Hive, Zookeeper, Sqoop, Flume, Oozie Programming Languages: Python, Java   Databases: Oracle 9i/10g, MySQL, Sybase ASE 12.5 NoSQL: HBase, MongoDB   Web Technologies: Javascript, HTML, CSS, XML   Version Control: Git   IDEs: Eclipse for Java, Visual Studio

Name: Melissa Hall

Email: lewisjonathan@example.com

Phone: (965)485-1913x31033