

Hadoop/Spark Developer Hadoop/Spark Developer Hadoop/Spark Developer - American Family Insurance Jersey City, NJ Having over 8 years of professional IT experience in Analysis, Development, Integration and Maintenance of Web based and Client/Server applications using Java and Big Data technologies. 4 years of relevant experience in Hadoop Ecosystem and architecture (HDFS, MapReduce, YARN, Pig, Hive, HBase, Sqoop, Flume, Oozie). Experience in real time analytics with Apache Spark (RDD, DataFrames and Streaming API). Used Spark DataFrames API over Cloudera platform to perform analytics on Hive data. Experience in integrating Hadoop with Apache Storm and Kafka. Expertise in uploading Click stream data from Kafka to HDFS, HBase and Hive by integrating with Storm. Developed producers for Kafka which compress and bind many small files into a larger Avro and Sequence files before writing to HDFS to make best use of Hadoop block size. Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice-versa. Configured Flume to extract the data from the web server output files to load into HDFS. Extensive hands on experience in writing MapReduce jobs in Java. Performed data analysis using Hive and Pig. Experience in analyzing large datasets using HiveQL and PigLatin. Experience in using Partitioning and Bucketing concepts in Hive and designed both Managed and External tables in Hive for optimized performance. Experience in developing custom UDFs for Pig and Hive to incorporate methods and functionality of Python/ Java into PigLatin and HQL (HiveQL) and Used UDFs from Piggybank UDF Repository. Good understanding and knowledge of NoSQL databases like MongoDB, Cassandra and HBase. Having Experience in monitoring and managing the Hadoop cluster using Cloudera Manager. Experience in job work-flow scheduling and monitoring using Oozie Worked extensively on different Hadoop distributions like Cloudera's CDH and Hortonworks HDP. Good working knowledge in cloud integration with Amazon Web Services (AWS) components like Redshift, DynamoDB, EMR, S3 and EC2 instances. Worked with Apache NiFi to develop Custom Processors for processing and distributing data among cloud systems. Having good knowledge of Scala programming concepts. Expertise in distributed and web environments focused in Core Java technologies like Collections, Multithreading, IO, Exception Handling and Memory

Management. Expertise in development of Web applications using J2EE technologies like Servlets, JSP, Web Services, Spring, Hibernate, HTML5, JavaScript, jQuery, AJAX etc., Knowledge of standard build and deployment tools such as Eclipse, Scala IDE, Maven, Subversion, SBT. Extensive knowledge in Software Development Lifecycle (SDLC) using Waterfall, Agile methodologies. Facilitate Sprint planning, daily scrums, retrospectives, stakeholder meetings, and software demonstrations. Excellent communication skills with the ability to communicate complex issues to technical and non-technical audiences that includes peers, partners, and Senior IT and Business management. Authorized to work in the US for any employer Work Experience Hadoop/Spark Developer American Family Insurance - Madison, WI October 2017 to Present Description: American Family Insurance is a private mutual company that focuses on property, casualty, auto insurance and also offers commercial insurance, life, health and homeowners coverage. The purpose of the project is to build a complex ETL pipeline which will handle huge trip data that is collected from vendor servers. Transformations which we implemented can calculate different trip level details and events that are based on latitude, longitude details from raw data. Storing all trip level calculated summary data in data warehouse on top of Hadoop. Different machine learning algorithms are used on this summary data to generate the scores, which will be useful to provide better service to the existing customers. Responsibilities: Created and worked on Sqoop jobs with incremental load to populate Hive External tables. Designed and developed Hive tables to store staging and historical data. Created Hive tables as per requirement, internal and external tables are defined with appropriate static and dynamic partitions, intended for efficiency. Experience in using ORC file format with Snappy compression for optimized storage of Hive tables. Solved performance issues in Hive and Pig scripts with understanding of Joins, Group and aggregation and used them using Impala process engine Developed Spark scripts by using Scala shell commands as per the requirement. Created Oozie workflows for sqoop to migrate the data from source to HDFS and then to target tables. Developed Oozie workflow for scheduling and orchestrating the ETL process. Responsible for building scalable distributed data solutions using Hadoop. Experience in Job management using Fair scheduler and Developed job processing

scripts using Oozie workflow. Involved in migrating MapReduce jobs into Spark jobs and used Spark SQL and DataFrames API to load structured and semi-structured data into Spark clusters. Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing. Worked extensively with Sqoop for importing metadata from Oracle. Developed Oozie workflow jobs to execute Hive, Pig, Sqoop and Mapreduce actions. Configured Flume to transport web server logs into HDFS. Experience on Amazon Web Services(AWS), Amazon Cloud Services like Elastic Compute Cloud(EC2), Simple Storage Service(S3), Elastic Map Reduce(EMR), Amazon Simple DB and Amazon Cloud Watch. Implemented Spark using Scala and SparkSQL for faster testing and processing of data. Used Apache Kafka for importing real time network log data into HDFS. Worked on numerous POCs to prove if Big Data is the right fit for a business case. Experience data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka. Created web-based User interface for creating, monitoring and controlling data flows using Apache Nifi. Environment: Apache Hadoop, CDH 4.7, HDFS, MapReduce, Sqoop, Flume, Pig, Hive, HBase, Oozie, Scala, Spark, Spark Streaming, Kafka, Linux Hadoop Developer

Travelport - Englewood, CO March 2015 to September 2017 Description: Travelport operates one of the largest financial communities in the world, providing communications platforms, network and cloud based connectivity solutions for the global financial market I have worked on a platform where we import data from different sources to a central repository. Manipulations are performed before and after storing the data so that analytics can be performed over it. The application uses various Hadoop ecosystem components like Sqoop, Flume for importing various types of data and HBase, Hive and MapReduce to store and apply transformations over it and Oozie as a workflow engine.

Responsibilities: Extracted the data from Teradata into HDFS using Sqoop. Used Flume to collect, aggregate and store the web log data from different sources like web servers, mobile and network devices and pushed into HDFS. Implemented MapReduce programs on log data to transform into structured way to find user information. Extensive experience in writing Pig scripts to transform raw data from several data sources into forming baseline data. Analyzed the web log data using the HiveQL to extract number of unique visitors per day, page views and visit duration.

Utilized Flume to filter out the JSON input data read from the web servers to retrieve only the required data needed to perform analytics. Developed UDF functions for Hive and wrote complex queries in Hive for data analysis. Developed a well-structured and efficient ad-hoc environment for functional users. Export the analyzed data to relational databases using Sqoop for visualizations and to generate reports for the BI team. Loaded cache data into HBase using Sqoop. Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.

Wrote ETLs using Hive and processed the data as per business logic. Hands on experience on Amazon EC2 Spot integration & and Amazon S3 integration. Optimizing the EMRFS for Hadoop to directly read and write in parallel to AWS S3 performantly. Extensive work in ETL process consisting of data transformation, data sourcing, mapping, conversion and loading using Informatica. Extensively used ETL processes to load data from flat files into the target database by applying business logic on transformation mapping for inserting and updating records when loaded.

Created Talend ETL jobs to read the data from Oracle Database and import in HDFS. Worked on data serialization formats for converting complex objects into sequence bits by using Avro, RC and ORC file formats. Environment: Apache Hadoop, Hortonworks HDP 2.0, HDFS, MapReduce, Sqoop, Flume, Pig, Hive, HBase, Oozie, Teradata, Talend, Avro, Java, Linux Hadoop Developer
First National Bank - Omaha, NE March 2013 to February 2015 Description: First National Bank offers a wide range of banking, lending, credit card, investing and financial services for consumers and businesses. It is recognized as the largest privately held bank in the United States.

Responsibilities: Worked on live 8 node Hadoop cluster running CDH 4. Used Sqoop to import the data from RDBMS to Hadoop Distributed File System (HDFS). Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports by Business Intelligence tools. Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis. Developed several MapReduce programs to analyze and transform the data to uncover insights into the customer usage patterns. Used Pig as ETL tool to do transformations, event joins and some pre-aggregations before storing the data into HDFS. Responsible for creating Hive External tables and loaded the data into tables and query

data using HiveQL. Used Hive data warehouse tool to analyze the unified historic data in HDFS to identify issues and behavioral patterns. Created concurrent access for Hive tables with shared and exclusive locking that can be enabled in Hive with the help of Zookeeper implementation in the cluster. Integrated Oozie with the rest of Hadoop stack supporting several types of Hadoop jobs as well as the system specific jobs (such as Java programs and shell scripts). Created HBase tables to store various data formats coming from different portfolios, worked on NoSQL databases including HBase, Cassandra and MongoDB. Used Jenkins for build and continuous integration for software development. Worked with application teams to install Operating systems, Hadoop updates, patches and version upgrades as required. Environment: Apache Hadoop, CDH 4, Sqoop, Flume, MapReduce, Pig, Hive, HBase, Cassandra, MongoDB, Oozie, Zookeeper, Jenkins

Java Developer Intersect Group - Hyderabad, Telangana April 2011 to February 2013 Description: The Intersect Group is a privately held, Atlanta-based firm offering flexible IT, financing & accounting expertise to progressive, high-growth companies. Responsibilities: Involved in development of business domain concepts into Use cases, Sequence Diagrams, Class Diagrams, Component Diagrams and Implementation Diagrams. Implemented various J2EE Design patterns such as Model-View-Controller (MVC), Data Access Object, Business Delegate and Transfer Object. Involved in designing and development of project using Java/J2EE technologies by following MVC architecture of which JSPs are views and Servers as controllers. Involved in configuring Struts, Tiles and developing the configuration files. Developed Struts Action classes and Validation classes using Struts controller component and Struts validation framework. Developed and deployed UI layer logics using JSP, XML, JavaScript, HTML/DHTML. Using Star UML designed network and use case diagrams to monitor the workflow. Wrote Server side programs to handle requests coming from different types of devices using RESTful Web Services. Designed a light weight model for the product using Inversion of Control principal and implemented it successfully using Spring IOC Container. Used Hibernate ORM tool to store and retrieve the data from PostgreSQL database. Provided connections using JDBC to the database and developed SQL queries to manipulate the data. Environment: Java J2EE, Struts MVC, Tiles, JSP, XML, JavaScript,

Spring IOC, Websphere Application Server, PostgreSQL Java Developer Kraft Foods Group - Hyderabad, Telangana June 2009 to March 2011 Description: Kraft Foods Group, Inc. is an American manufacturing and processing conglomerate headquartered in the Chicago. It is focused mainly on the grocery products for the North American market. Responsibilities: Work involved providing support to the production environment for various applications and actively work on incidents and issues raised by users. This also involved in on call support during off hours. Developed service layer logic for core modules using JSPs and Servlets and involved in integration with presentation layer. Involved in complete project such as Business Delegate, Data Transfer Object, Service Locator, Data Access Object and Singleton. Developed XML configuration and data description using Hibernate. Hibernate Transaction Manager is used to maintain the transaction persistence. Developed the user interface using JSP and DHTML lifecycle of the project from gathering business requirements to creating an architecture and build applications on Java/J2EE with Spring MVC framework. Involved in fixing bugs and minor enhancements for the front-end module. Environment: Java, Servlets, JSP, Spring, Hibernate, XML, XPath, jQuery, JavaScript, WebSphere Application Server Education Bachelor's

Name: Matthew Wade

Email: angelameyers@example.net

Phone: 001-566-690-9088x926