

Hadoop Developer Hadoop Developer Hadoop Developer - Sherwin Williams Cleveland, OH Over 7+ years of extensive hands on experience with Hadoop Ecosystem stack including HDFS, MapReduce, Sqoop, Hive, Pig, HBase, Oozie, Flume, Kafka, Zookeeper, and Spark. Experience in different Hadoop distributions like Cloudera and HortonWorks Distributions (HDP). Comfortable working with various facets of the Hadoop ecosystem, real-time or batch, Structured or Unstructured data processing. Experience with NoSQL databases like HBase as well as other ecosystems like Zookeeper, Oozie, Impala, Storm, Spark- Streaming/SQL, Kafka, Flume. Expertise skills in handling analytics projects using Big Data technologies. Hands on experience in ingesting data from external servers to Hadoop. Experience in moving large amounts of log, streaming event data and Transactional data using Flume. Hands on experience developing workflows that execute Sqoop, Pig, Hive and Shell scripts using Oozie. Developed Spark code and Spark-SQL/Streaming for faster testing and processing of data. Good experience with Hive Data Warehousing concepts like Static/Dynamic Partitioning, Bucketing, Managed, and External Tables, join operations on tables. Proficient in building user defined functions (UDF) in Hive and Pig, to analyze data and extended HiveQL and Pig Latin Functionality. Experience in working with Spark transformations and actions on RDDs and Spark-SQL, Data Frames in Python. Experience in implementing unified data ingestion platform using Kafka producers and consumers. Experience in implementing near real-time event processing and analytics using Spark Streaming Proficient with Flume topologies for data ingestion from streaming sources into Hadoop. Well versed with major Hadoop distributions: Cloudera and HortonWorks Having experience on Eclipse, NetBeans IDEs. Ability to adapt to evolving Technology, Strong sense of responsibility and Accomplishment. Has very good development experience with Agile Methodology. Strong experience in distinct phases of Software Development Life cycle (SDLC) including Planning, Design, Development and Testing during the development of software applications. Excellent leadership, interpersonal, problem solving and time management skills. Excellent communication skills both written (documentation) and verbal (presentation). Very responsible and good team player. Can work independently with minimal supervision. Work Experience Hadoop Developer Sherwin Williams - Cleveland, OH March

2018 to Present Responsibilities: Responsible for understanding the scope of the project and requirements gathering. Loaded log data into HDFS using Flume. Developed data pipeline using Flume, Sqoop, Pig and Python MapReduce to ingest customer behavioral data and financial histories into HDFS for analysis. Developed Python scripts to extract the data from the web server output files to load into HDFS. Involved in HBASE setup and storing data into HBASE, which will be used for further analysis. Used Pig as ETL tool to do transformations, event joins and some pre-aggregations before storing the data on to HDFS. Written the Apache PIG scripts to process the HDFS data. Created Hive tables to store the processed results in a tabular format. Developed the Sqoop scripts to make the interaction between Pig and Oracle. Writing the script files for processing data and loading to HDFS. Worked extensively with Sqoop for importing data from Oracle. Utilized Apache Hadoop ecosystem tools like HDFS, Hive and Pig for large datasets analysis. Developed Pig and Hive UDF to analyze the complex data to find specific user behavior.

Experienced in using Pig for data cleansing and developed Pig Latin scripts to extract the data from web server output files to load into HDFS. Maintaining and monitoring clusters. Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop. Worked on Hive by creating external and internal tables, loading it with data and writing Hive queries. Created HBase tables to store data from various sources. Developed workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive (Data Warehouse). Worked with various Hadoop file formats, including Text, Sequence File, RCFILE and ORC File. Configured Zookeeper for Cluster co-ordination services. Environment: Hadoop, HDFS, Pig, Hive, HBase, Zookeeper, Flume, Kafka, Sqoop, Spark, Elastic Search, Oozie, Java, Cloudera, Oracle, Windows, UNIX Shell Scripting.

Hadoop Developer Cox Automotive - Atlanta, GA October 2016 to February 2018 Responsibilities: Responsible for building scalable distributed data solutions using Hadoop. Job duties include design and development of various modules in Hadoop Big Data platform and processing data using MapReduce, Hive, SQUOP, Kafka and Oozie. Developed job processing scripts using Oozie workflow. Worked with Apache Hadoop and Spark Used Dataframe API for converting the

distributed collection of data organized into named columns. Involved in Hadoop cluster task like commissioning & decommissioning Nodes without any effect to running jobs and data. Worked extensively with Sqoop for importing metadata from Oracle. Real streaming the data using Spark with Kafka. Designed, developed and did maintenance of data integration programs in a Hadoop and RDBMS environment with both traditional and non-traditional source systems as well as RDBMS and NoSQL data stores for data access and analysis. Involved in installing, configuring and managing Hadoop Ecosystem components like Hive, Pig, Sqoop, Kafka and Flume. Assisted in exporting analyzed data to relational databases using Sqoop. Wrote Hive Queries and UDF's. Developed Hive queries to process the data and generate the data cubes for visualizing.

Environment: MapReduce, Spark, HDFS, Pig, HBase, Oozie, Zookeeper, Sqoop, Linux, Kafka, Hadoop, Maven, NoSQL, MySQL, Hive, Java, Eclipse, Python. Hadoop Consultant AAA Insurance - Phoenix, AZ October 2014 to September 2016 Responsibilities: Automation of data pulls into HDFS from MySQL server and Oracle DB using Sqoop. Analyzing source data tables for best possible loading strategies. Involved in various stages of this project like planning, estimation the hardware and software, installing (SDLC). Develop Shell scripts to perform various ETL jobs like creating staging and final tables. Implemented 2 level staging process for Data Validation. Extracted data from staging tables and analyzed data using Impala. Implement ad-hoc queries using Impala, create tables with partitioning and bucketing to load data. Created a Spark application to process and stream data from Kafka to MySQL. Implement Hive Incremental updates using four-step strategy to load incremental data from RDBM systems. Implement, configure optimization techniques like Bucketing, Partitioning and File Formats. Used Spark to analyze data in HIVE, HBase and HDFS. Involved in Hadoop Cluster Administration that includes adding and removing Cluster Nodes, Cluster Capacity Planning, and Performance Tuning. Worked on Hadoop clusters capacity Planning and Management. Monitoring and Debugging Hadoop jobs Applications running in production. Written a PIG Scripts to read data from HDFS and write into Hive Table. Experience of performance tuning Hive ETL Scripts, Pig Scripts, MR Jobs in production environment by altering job parameters. Providing various hourly/weekly/monthly

aggregation reports required by clients through Spark. Worked on data processing part mainly to make the Unstructured Data to Semi-Structured Data and loaded into Hive tables, HBase tables and integration. Load log data into HDFS using Flume. Written the Apache PIG scripts to process the HDFS Data. Developed Spark SQL scripts with Python for analysis and Demo purposes.

Environment: MapReduce, Spark, HDFS, Pig, HBase, Oozie, Zookeeper, Sqoop, Linux, Kafka, Hadoop, Maven, NoSQL, MySQL, Hive, Java, Eclipse, Python. Hadoop Developer Eventcy - IN October 2013 to August 2014 Responsibilities: Worked on Hadoop Ecosystem using different big data analytic tools including Hive, Pig. Involved in loading data from LINUX file system to HDFS. Importing and exporting data into HDFS and Hive using Sqoop. Implemented Partitioning, Bucketing in Hive. Worked on different file formats (ORCFILE, TEXTFILE) and different Compression Codecs (GZIP, SNAPPY, LZO). Worked with multiple Input Formats such as Text File, Key Value, and Sequence File Input Format. Experienced in running Hadoop Streaming jobs to process terabytes of Json format data. Involved in scheduling Oozie workflow engine to run multiple Hive and Pig jobs Executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business requirements. Created HBase tables to store various data formats of incoming data from different portfolios. Created Pig Latin scripts to sort, group, join and filter the enterprise wise data. Developed the verification and control process for daily load. Experience in Daily production support to monitor and trouble shoots Hadoop/Hive jobs. Worked collaboratively with different teams to smoothly slide the project to production. Environment: HDFS, Pig, Hive, Sqoop, Shell Scripting, HBase, Zoo Keeper, MySQL. Software Developer Aetins - IN May 2011 to September 2013 Responsibilities: Performed analysis for the client requirements based on the developed detailed design documents. Developed Use Cases, Class Diagrams, Sequence Diagrams and Data Models using Microsoft Visio. Developed STRUTS forms and actions for validation of user request data and application functionality. Developed a web service using SOAP, WSDL, XML and SOAP-UI. Developed JSP's with STRUTS custom tags and implemented JavaScript validation of data. Involved in developing business tier using stateless session bean. Used JavaScript for the web page validation and Struts Valuator for server side validation

Designing the database and coding of SQL, PL/SQL, Triggers and Views using IBMDB2. Design patterns of Delegates, Data Transfer Objects and Data Access Objects. Developed Message Driven Beans for asynchronous processing of alerts. Used Clear case for source code control and JUnit for unit testing. The networks are simulated in real-time using an ns3 network simulator modified for multithreading across multiple cores, which is implemented on generic Linux machine. Involved in peer code reviews and performed integration testing of the modules. Environment: STRUTS, JSP's with STRUTS, JDBC, Struts Valuator, SQL, PL/SQL, IBMDB2, JUNIT, Java / J2ee, JSP, servlets, EJB 2.0, SQL Server, Oracle 9i, Jboss & Web Logic Server 6, JavaScript. Skills Mapreduce, Oozie, Sqoop, Hbase, Kafka Additional Information Technical Expertise: Programming Languages C, C++, Java (core), J2EE, UNIX Shell Scripting, Python Web Languages HTML, JAVA SCRIPT, CSS Hadoop Ecosystem MapReduce, HBASE, HIVE, PIG, SQOOP, Zookeeper, OOZIE, Flume, HUE, Kafka, AWS EMR, SPARK, SPARK-SQL Database Languages MySQL, NOSQL Database Oracle, SQL Virtualization & Cloud Tools Amazon AWS, VMware, Virtualbox Visualization tools Power Bi, Tableau Web/Application Servers Apache Tomcat Version Control Tools GIT and SVN Operating Systems Windows, Linux (Ubuntu, Red Hat, Cent OS) IDE Platforms Eclipse, Net Beans, Visual Studio Methodologies Agile, SDLC

Name: Jamie Lee

Email: heatherstone@example.org

Phone: 390.551.0743x945