

Sr. Hadoop Developer Sr. Hadoop Developer Sr. Hadoop Developer - Inovalon Herndon, VA, VA

Around 8+ years of professional IT industry experience encompassing wide range of skill set in Big Data technologies and Java/J2EE technologies. 4+ years of experience in working with Big Data Technologies on systems which comprises of massive amount of data running in highly distributive mode in Cloudera, Hortonworks Hadoop distributions. Hands on experience in using Hadoop ecosystem components like Hadoop, Hive, Pig, Sqoop, HBase, Cassandra, Spark, Spark Streaming, Spark SQL, Oozie, ZooKeeper, Kafka, Flume, MapReduce and Yarn. Strong Knowledge on architecture and components of Spark, and efficient in working with Spark Core, SparkSQL, Spark streaming. Implemented Spark Streaming jobs by developing RDD's (Resilient Distributed Datasets) and used pyspark and spark-shell accordingly. Experience in configuring Spark Streaming to receive real time data from the Apache Kafka and store the stream data to HDFS using Scala. Experience in importing and exporting data using stream processing platforms like Flume and Kafka. Accomplished complex HiveQL queries for required data extraction from Hive tables and written Hive UDF's as required. Pleasant experience of Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance. Involved in converting Hive/SQL queries into Spark transformations using Spark Data frames and Scala. Used Spark Data Frames API over Cloudera platform to perform analytics on Hive data. Used Spark Data Frame Operations to perform required Validations in the data. Experience in integrating Hive queries into Spark environment using Spark SQL. Good understanding and knowledge of NoSQL databases like MongoDB, Hbase and Cassandra. Worked on HBase to load and retrieve data for real time processing using Rest API. Excellent understanding and knowledge of job workflow scheduling and locking tools/services like Oozie and Zookeeper. Experienced in designing different time driven and data driven automated workflows using Oozie. Knowledge of ETL methods for data extraction, transformation and loading in corporate-wide ETL Solutions and Data Warehouse tools for reporting and data analysis. Worked on developing ETL Workflows on the data obtained using Python for processing it in HDFS and HBase using Oozie. Experience in configuring the Zookeeper to coordinate the servers in clusters

and to maintain the data consistency. Experienced in working with Amazon Web Services (AWS) using EC2 for computing and S3 as storage mechanism. Capable of using AWS utilities such as EMR, S3 and cloud watch to run and monitor Hadoop and spark jobs on AWS. Hands on experience with Spark Data frames, Spark-SQL and RDD API of Spark for performing various data transformations and dataset building. Involved in developing Impala scripts for extraction, transformation, loading of data into data warehouse. Good knowledge in using apache NiFi to automate the data movement between different Hadoop systems. Experienced in using Pig scripts to do transformations, event joins, filters and some pre-aggregations before storing the data onto HDFS. Experience in importing and exporting the data using Sqoop from HDFS to Relational Database systems and vice-versa. Good Knowledge in UNIX Shell Scripting for automating deployments and other routine tasks. Experience in relational databases like Oracle, MySQL and SQL Server. Experienced in using Integrated Development environments like Eclipse, NetBeans, IntelliJ, Spring Tool Suite. Used various Project Management services like JIRA for tracking issues, bugs related to code and GitHub for various code reviews and Worked on various version control tools like GIT, SVN. Experienced in developing and implementing web applications using Java, J2EE, JSP, Servlets, JSF, HTML, DHTML, EJB, JavaScript, AJAX, JSON, JQuery, CSS, XML, JDBC and JNDI. Experience in Java development GUI using JFC, Swing, JavaBeans, and AWT. Excellent Java development skills using J2EE, J2SE, Servlets, JSP, EJB, JDBC, SOAP and RESTful web services. Experienced in working in SDLC, Agile and Waterfall Methodologies. Excellent Communication skills, Interpersonal skills, problem solving skills and a team player. Ability to quickly adapt new environment and technologies. Work Experience Sr. Hadoop Developer Inovalon - Herndon, VA December 2017 to Present Inovalon is a leading technology company providing cloud-based platforms empowering data-driven healthcare. We bring to the marketplace a national-scale capability to interconnect with the healthcare ecosystem, aggregate and analyze data in real-time, and empower the application of resulting insights to drive meaningful impact at the point of care. Responsibilities: Working knowledge of Spark RDD, Data Frame API, Data set API, Data Source API, Spark SQL and Spark Streaming. Exploring with the Spark improving the

performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN. Used Spark SQL on data frames to access hive tables into spark for faster processing of data. Performed Spark join optimizations, troubleshooted, monitored and wrote efficient codes using Scala. Experienced with batch processing of data sources using Apache Spark and Elastic search. Used Different Spark Modules like Spark core, Spark SQL, Spark Streaming, Spark Data sets and Data frames. Worked on analyzing Hadoop stack and different big data analytic tools including Pig, Hive, HBase database and Sqoop. Developed Spark Streaming script which consumes topics from distributed messaging source Kafka and periodically pushes batch of data to spark for real time processing. Experienced to implement Hortonworks distribution system. Creating Hive tables and working on them for data analysis to cope up with the requirements. Developed a frame work to handle loading and transform large sets of unstructured data from UNIX system to HIVE tables. Used Spark Data Frames Operations to perform required Validations in the data and to perform analytics on the Hive data. Experienced in working with Elastic MapReduce (EMR) Developed Map Reduce programs for some refined queries on big data. In-depth understanding of classic MapReduce and YARN architecture. Worked with business team in creating Hive queried for ad hoc access. Use Hive to analyze the partitioned and bucketed data and compute various metrics for reporting. Implemented Hive Generic UDF's to implement business logic. Analyzed the data by performing Hive queries, ran Pig scripts, Spark SQL and Spark Streaming. Installed and configured Pig for ETL jobs. Responsible for developing multiple Kafka Producers and Consumers from scratch as per the software requirement specifications. Involved in creating generic Sqoop import script for loading data into Hive tables from RDBMS. Designed Columnar families in Cassandra and Ingested data from RDBMS, performed data transformations, and then exported the transformed data to Cassandra as per the business requirement. Extracted files from Cassandra through Sqoop and placed in HDFS for further processing. Enabled speedy reviews and first mover advantages by using Oozie to automate data loading into the Hadoop Distributed File System and Pig to pre-process the data. Created detailed AWS Security groups which behaved as virtual firewalls

that controlled the traffic allowed reaching one or more AWS EC2 instances. Involved in creating data-lake by extracting customer's data from various data sources to HDFS which include data from Excel, databases, and log data from servers. Performed data integration with a goal of moving more data effectively, efficiently and with high performance to assist in business-critical projects using Talend Data Integration. Design, developed, unit test, and support ETL mapping and scripts for data marts using Talend. Used HUE for running Hive queries. Created partitions according to day using Hive to improve performance. Built a data flow pipeline using flume, Java (MapReduce) and Pig. Involved in Agile methodologies, daily Scrum meetings, Sprint planning. Experience in using version control tools like GITHUB to share the code snippet among the team members.

Environment: Hadoop, Map Reduce, HDFS, Hive, Cassandra, Sqoop, Oozie, SQL, Kafka, Spark, Scala, Java, AWS, GitHub, Talend Big Data Integration, Solr, Impala.

Hadoop Developer Surgere Inc - Green, OH September 2016 to November 2017

Since 2004, Surgere has been re-inventing how packaging assets influence profitability. While raising the bar on packaging supply chain standards, Surgere developed a wide variety of visibility and control technologies to increase data accuracy and manage assets (packaging, parts, tools, finished vehicles) moving through manufacturing, supply chain and logistics verticals.

Responsibilities:

- Good in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.
- Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data.
- Developed Spark scripts by using Scala shell commands as per the requirement.
- Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework.
- Used Spark RDD for faster Data sharing.
- Experienced in querying data using SparkSQL on top of Spark engine for faster data sets processing.
- Worked on Ad hoc queries, Indexing, Replication, Load balancing, Aggregation in MongoDB.
- Extracted and restructured the data into MongoDB using import and export command line utility tool.
- Worked on the large-scale Hadoop YARN cluster for distributed data processing and analysis using Hive and MongoDB.
- Wrote XML scripts to build Oozie functionality.
- Experience in workflow Scheduler Oozie to manage and schedule job on

Hadoop cluster for generating reports on Day and weekly basis. Used Flume to collect, aggregate, and store the web log data from various sources like web servers, mobile and network devices and pushed to HDFS. Implemented custom serializer, interceptor, source and sink in Flume to ingest data from multiple sources. Involved in writing query using Impala for better and faster processing of data. Responsible for analyzing and cleansing raw data by performing Hive/Impala queries and running Pig scripts on data. Involved in moving log files generated from various sources to HDFS for further processing through Flume. Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and AWS cloud. Worked on partitioning the HIVE table and running the scripts in parallel to reduce the run time of the scripts. Analyzed the data by performing Hive queries and running Pig scripts to know user behavior. Programmed pig scripts with complex joins like replicated and skewed to achieve better performance. Developed data pipeline expending Pig and Java MapReduce to consume customer behavioral data and financial antiquities into HDFS for analysis. Designing & creating ETL jobs through Talend to load huge volumes of data into MongoDB, Hadoop Ecosystem and relational databases. Created Talend jobs to connect to Quality Stage using FTP connection and process data received from Quality Stage. Migrated data from MySQL server to Hadoop using Sqoop for processing data. Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team. Experienced in developing Shell scripts and Python scripts for system management. Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required. Worked with SCRUM team in delivering agreed user stories on time for every Sprint. Environment: CDH 3.x and 4.x, Java, Hadoop, Python, Map Reduce, Hive, Pig, Impala, Flume, MongoDB, Sqoop, Talend, Spark, MySQL, AWS. Hadoop Developer Archimedes Inc - San Francisco, CA November 2015 to July 2016 Archimedes is a healthcare modelling and analytics organization. The company enables people to combine real-world healthcare data and simulation data to create compelling and actionable evidence used in individual healthcare decision-making, as well as population, with applications in health and economic outcomes research, policy creation, and clinical trial design and operations.

Responsibilities: Worked on migrating MapReduce programs into Spark transformations using Scala. Responsible for building scalable distributed data solutions using Hadoop. Responsible for building scalable distributed data solutions using Hadoop. Responsible for cluster maintenance by adding and removing cluster nodes. Cluster monitoring, troubleshooting, managing and reviewing data backups and log files. Wrote complex MapReduce jobs in Java to perform operations by extracting, transforming and aggregating to process terabytes of data. Collected and aggregated large amounts of stream data into HDFS using Flume and defined channel selectors to multiplex data into different sinks. Analyzed data using Hadoop components Hive and Pig. Scripted complex HiveQL queries on Hive tables to analyze large datasets and wrote complex Hive UDFs to work with sequence files. Scheduled workflows using Oozie to automate multiple Hive and Pig jobs, which run independently with time and data availability. Responsible for creating Hive tables, loading data and writing Hive queries to analyze data. Generated reports using QlikView. Wrote several Hive queries to get valuable information from the hidden large datasets. Loaded and transformed large sets of structured, semi-structured and unstructured data using Hadoop/Big Data concepts. Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS. Imported data from Teradatadatabase into HDFS and exported the analyzed patterns data back to Teradata using Sqoop. Worked with Talend Open Studio to perform ETL jobs. Environment: Hadoop (Hortonworks), HDFS, Hive, Pig, Sqoop, Map Reduce, HBase, Shell Scripting, QlikView, Teradata 14, Oozie, Java 7, Maven 3.x.

Java/Hadoop Developer Evidera - Bethesda, MD August 2014 to October 2015 Evidera provides health economics, outcomes research, market access, data analytics and epidemiology researchand consulting services. We partner with life science organization worldwide to optimize the market access and commercial success of their products. Responsibilities: Involved in Installation and configuration of JDK, Hadoop, Pig, Sqoop, Hive, HBase on Linux environment. Assisted with performance tuning and monitoring. Imported data using Sqoop to load data from MySQL to HDFS on regular basis. Created reports for the BI team using Sqoop to export data into HDFS and Hive. Worked on creating MapReduceprograms to parse the data for claim report generation and

running the Jars in Hadoop. Co-ordinated with Java team in creating MapReduce programs. Worked on creating Pig scripts for most modules to give a comparison effort estimation on code development. Collaborated with BI teams to ensure data quality and availability with live visualization. Created HIVE Queries to process large sets of structured, semi-structured and unstructured data and store in Managed and External tables. Created HBase tables to store variable data formats coming from different portfolios Performed real-time analytics on HBase using Java API and Rest API. Performed test run of the module components to understand the productivity. Written Java program to retrieve data from HDFS and providing REST services. Shared responsibility and assistance for administration of Hadoop, Hive, Sqoop, HBase and Pig in team. Shared the knowledge of Hadoop concepts with team members. Used JUnit for unit testing and Continuum for integration testing. Environment: Cloudera, Hadoop, Pig, Sqoop, Hive, HBase, Java, Eclipse, MySQL, MapReduce. Java Developer Metanoia software Solutions January 2012 to May 2014 Metanoia offers an Ecommerce platform designed and developed for rapid deployment of Ecommerce services by small and medium size organizations. Ecloud is a strong and vibrant deployment option and business model for e-commerce solution. Our cloud system is a cost effective alternative designed to address the diverse needs of companies of varying sizes across industries. Responsibilities: Responsible for the analyzing, documenting the requirements, designing and developing the application based on J2EE standards. Strictly Followed Test Driven Development. Used Microsoft Visio for designing use cases like Class Diagrams, Sequence Diagrams, and Data Models. Extensively developed user interface using HTML, JavaScript, jQuery, AJAX and CSS on the front end. Designed Rich Internet Application by implementing jQuery based accordion styles. Used JavaScript for the client-side web page validation. Used Spring MVC and Dependency Injection for handling presentation and business logic. Integrated Spring DAO for data access using Hibernate. Developed Struts web forms and actions for validation of user request data and application functionality. Developed programs for accessing the database using JDBC thin driver to execute queries, prepared statements, Stored Procedures and to manipulate the data in the database. Created tile definitions, Struts

configuration files, validation files and resource bundles for all modules using Struts framework. Involved in the coding and integration of several business-critical modules using Java, JSF, and Hibernate. Developed SOAP-based web services for communication between its upstream applications. Implemented different Design patterns like DAO, Singleton Pattern and MVC architectural design pattern of spring. Implemented Service Oriented Architecture (SOA) on Enterprise Service Bus (ESB) Developed Message-Driven Beans for asynchronous processing of alerts using JMS. Implemented Rational Rose tool for application development. Used Clear case for source code control and JUnit for unit testing. Performed integration testing of the modules. Used putty for UNIX login to run the batch jobs and check server logs. Deployed application on to Glassfish Server. Involved in peer code reviews. Environment: Java 6, 7, J2EE, Struts 2, Glassfish, JSP, JDBC, EJB, ANT, XML, IBM Web Sphere, JUnit, IBM DB2, Rational Rose 7, CVS, UNIX, SOAP, SQL, PL/SQL. Java Developer Thirdware Solutions Limited July 2009 to December 2011 Thirdware delivers industry-specific technological expertise through a range of services spanning business applications consulting, design, implementation and support. It is involved in developing Smart Data solutions - yielding clean, organized, actionable data to extract information and insight. Responsibilities: Documented functional and technical requirements, wrote Technical Design Documents. Developed analysis level documentation such as Use Case, Business Domain Model, Activity & Sequence and Class Diagrams. Developed presentation layer components comprising of JSP, AJAX, Servlets and JavaBeans using the Struts framework. Implemented MVC (Model View Controller) architecture. Developed XML configuration and data description using Hibernate. Developed Web services using CXF to interact with Mainframe applications. Responsible for the deployment of the application in the development environment using BEA WebLogic 9.0 application server. Participated in the configuration of BEA WebLogic application server. Designed and developed front end user interface using HTML and Java Server Pages (JSP) for customer profile setup. Developed ANT Script to compile the Java files and to build the jars and wars. Responsible for Analysis, Coding and Unit Testing and Production Support. Used JUnit for testing Modules. Environment: Java 1.6, J2EE, JDBC, Struts Framework, Hibernate,



Servlets, MVC, JSP, Web Services, CXF, SOAP, BEA WebLogic 9, Oracle 9i, JavaScript, XML, HTML, Ant, JUnit, SVN, My Eclipse. Skills JAVA (8 years), JUNIT (6 years), AJAX (4 years), HTML (4 years), JAVASCRIPT (4 years) Additional Information Technical Skills: Hadoop Distributions Cloudera (CDH3, CDH4, and CDH5), Hortonworks, MapR and Apache Big Data Ecosystem Hadoop, MapReduce, Pig, Hive, YARN, Kafka, Flume, Sqoop, Impala, Oozie, Zookeeper, Spark, Ambari, Mahout, MongoDB, Cassandra, Avro, Storm, Parquet and Snappy. Languages: Java, Python, Scala Java Technologies Servlets, JavaBeans, JSP, JDBC, and Spring MVC Web Design Tools HTML, DHTML, AJAX, JavaScript, JQuery and CSS, AngularJS, ExtJS and JSON No SQL Databases Cassandra, MongoDB and HBase Development / Build Tools Eclipse, Ant, Maven, IntelliJ, JUNIT and log4J ETL Tools Talend, Informatica DB Languages MySQL, PL/SQL, PostgreSQL and Oracle RDBMS Teradata, Oracle 9i, 10g, 11i, MS SQL Server, MySQL and DB2 Operating systems UNIX, LINUX, Mac os and Windows Variants Data analytical tools R and MATLAB ETL Tools Talend, Informatica, Pentaho

Name: Kim Mueller

Email: william67@example.org

Phone: 348.726.3790x0234