

SR. BIGDATA DEVELOPER/ENGINEER SR. BIGDATA DEVELOPER/ENGINEER SR. BIGDATA DEVELOPER/ENGINEER - MeganSoft Austin, TX Around 11+ years of overall experience in software development with extensive experience with Big Data technologies Good knowledge of Hadoop ecosystem such as HDFS, YARN and MapReduce Experience in build scripts using Maven and do continuous integrations systems like Jenkins. Expertise in using Kafka as a messaging system to implement real-time Streaming solutions and implemented Sqoop for large data transfers from RDMS to HDFS/HBase/Hive and vice-versa. Experience in migrating ETL process into Hadoop, Designing Hive data model and wrote Pig Latin scripts to load data into Hadoop. Good Knowledge on Cloudera distributions and in Amazon simple storage service (Amazon S3), AWS and Amazon EC2, Amazon EMR. Experience working with structured, semi-structured and unstructured data ingestion technologies such as Sqoop, Flume and Kafka Experience working with relational databases such as MySQL 5.7, Oracle 10g, and Postgres 9.6 and NoSQL databases such as MongoDB, Cassandra and HBase In-depth knowledge of Apache Cassandra architecture and extensive experience designing Cassandra data models and working with Cassandra Query Language (CQL) Experience writing Hive QL queries and Pig Latin scripts for ETL Expertise in processing and analyzing archived and real-time data using Core Spark, Spark-SQL and Spark Streaming Experience writing workflows in Oozie to schedule various jobs on Hadoop and with Expertise in Core Java features such as multithreading, exception handling, Generics, garbage collection, Collections, lambda expressions, serialization and deserialization Experience with writing unit tests using JUnit and Mockito frameworks Good Knowledge in Amazon Web Service (AWS) concepts like EMR and EC2 web services which provides fast and efficient processing of Teradata Big Data Analytics. Expertise in Data Development in Hortonworks HDP platform & Hadoop ecosystem tools like Hadoop, HDFS, Spark, Zeppelin, Hive, HBase, SQUOP, flume, Atlas, SOLR, Pig, Falcon, Oozie, Hue, Tez, ApacheNiFi, Kafka. Have very good experience in Apache Spark, Spark Streaming, Spark SQL and No SQL databases like Cassandra and Hbase Expert in Amazon EMR, Spark, Kinesis, S3, Boto3, Bean Stalk, ECS, Cloud watch, Lambda, ELB, VPC, Elastic Cache, Dynamo DB, Redshit, RDS, Aethna, Zeppelin &

Airflow. Strong knowledge on creating and monitoring Hadoop clusters on Amazon EC2, VM, Hortonworks Data Platform 2.1 & 2.2, CDH3, CDH4 Cloudera Manager on Linux, Ubuntu OS etc. Expertise in Java Script, JavaScript MVC patterns, Object Oriented JavaScript Design Patterns and AJAX and developed core modules in large cross-platform applications using JAVA, JSP, Servlets, JDBC, JavaScript, XML, and HTML. Zookeeper for managing and coordinating the cluster Experience with multiple Hadoop distributions such as Cloudera, Hortonworks and AWS Experience with VMWare, VirtualBox, Docker and Vagrant Experience with Java SE 8 and Java EE frameworks such as Spring MVC 4.0, Spring In-depth understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Spark MLlib. Expertise in writing Spark RDD transformations, actions, Data Frame's, case classes for the required input data and performed the data transformations using Spark-Core. Good knowledge of Hadoop Architecture and various components such as YARN, HDFS, NodeManager, ResourceManager, JobTracker, TaskTracker, NameNode, DataNode and MapReduce concepts. Strong knowledge in NOSQL column oriented databases like HBase, Cassandra, MongoDB and its integration with Hadoop cluster. Experience in installation, configuration, supporting and managing - Hortonworks/Cloudera's Hadoop platform along with CDH3&4 clusters. Solid SQL skills, can write complex SQL queries; functions, triggers and stored procedures for Backend testing, Database Testing and End-to-End testing. Experienced on Hadoop cluster on Azure HD Insight Platform and deployed Data analytic solutions using tools like Spark and BI reporting tools. Authorized to work in the US for any employer Work Experience SR. BIGDATA DEVELOPER/ENGINEER MeganSoft - Dearborn, MI January 2017 to Present RESPONSIBILITIES: Worked on loading disparate data sets coming from different sources to BDpaas (HADOOP) environment using Spark. Developed UNIX scripts in creating Batch load for bringing huge amount of data from Relational databases to BIG DATA platform. Delivery experience on major Hadoop ecosystem Components such as Pig, Hive, Spark Kafka, Elastic Search & HBase and monitoring with Cloudera Manager. Used AWS Data Pipeline to schedule an Amazon EMR cluster to clean and process web server logs stored in Amazon S3 bucket. Involved in gathering and analyzing system requirements and played key role

in the high-level design for the implementation of this application. Mavenized the existing applications using Maven tool and added the required jar files to the application as dependencies to the pom.XML file and used JSF & Struts frameworks to interact with the front end. Utilized Swing/JFC framework to develop client side components and developed J2EE components on Eclipse IDE. Implemented the Machine learning algorithms using Spark with Python and worked on Spark Storm, Apache and Apex and python. Involved in analyzing data coming from various sources and creating Meta-files and control files to ingest the data in to the Data Lake. Involved in configuring batch job to perform ingestion of the source files in to the Data Lake and developed Pig queries to load data to HBase Leveraged Hive queries to create ORC tables and developed HIVE scripts for analyst requirements for analysis. Implemented Kafka consumers to move data from Kafka partitions into Cassandra for near real-time analysis and worked extensively on Hive to create, alter and drop tables and involved in writing hive queries. Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig and parsed high-level design spec to simple ETL coding and mapping standards. Created and altered HBase tables on top of data residing in Data Lake and Created external Hive tables on the Blobs to showcase the data to the Hive Meta Store. Involved in requirement and design phase to implement Streaming Architecture to use real time streaming using Spark and Kafka. Use Spark API for Machine learning. Translate a predictive model from SAS code to Spark and used Spark API over Cludera Hadoop YARN to perform analytics on data in Hive. Created Reports with different Selection Criteria from Hive Tables on the data residing in Data Lake. Worked on Hadoop Architecture and various components such as YARN, HDFS, NodeManager, Resource Manager ,JobTracker ,TaskTracker, NameNode, DataNode and MapReduce concepts. Deployed Hadoop components on the Cluster like Hive, HBase, Spark, Scala and others with respect to the requirement. Uploaded and processed terabytes of data from various structured and unstructured sources into HDFS (AWS cloud) using Sqoop. Implemented the Business Rules in Spark/ SCALA to get the business logic in place to run the Rating Engine. Used Spark UI to observe the running of a submitted Spark Job at the node level and used Spark to do Property Bag Parsing of the data to get

the required fields of data. Extensively used ETL methodology for supporting Data Extraction, transformations and loading processing, using Hadoop. Used both Hive context as well as SQL context of Spark to do the initial testing of the Spark job and used WINSCP and FTP to view the data storage structure in the server and to upload JARs which were used to do the Spark Submit. Developed code from scratch in Spark using SCALA according to the technical requirements.

Environment: Hadoop, Hive, HDFS, Pig, Sqoop, Python, SparkSQL, Machine Learning, MongoDB, AWS, AWS S3, AWS EC2, AWS EMR, Oozie, ETL, Tableau, Spark, Spark-Streaming, KAFKA, Netezza, Apache Solr, Cassandra, Cloudera Distribution, Java, Impala, Web Server's, Maven Build, MySQL, Grafana, AWS, Agile-Scrum. SR. BIGDATA ENGINEER/ DEVELOPER Prabhav Services Inc - Newark, NJ September 2014 to December 2016

RESPONSIBILITIES: Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark SQL, Data Frame, Spark Yarn. Worked on cloud computing infrastructure (e.g. Amazon Web Services EC2) and considerations for scalable, distributed systems. Worked on Go-cd (ci/cd tool) to deploy application and have experience with Munin frame work for big data Testing. Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS and converted all Hadoop jobs to run in EMR by configuring the cluster according to the data size. Documented and established policies and procedures for GIS technologies including servers, workstations, storage, server virtualization, GIS data security, web based maps, and service. Extensively worked with Avro and Parquet files and converted the data from either format Parsed Semi Structured JSON data and converted to Parquet using Data Frames in Spark. Involved in converting Hive/SQL queries into Spark Transformations using Spark RDDs and Scala and involved in using SGOOP for importing and exporting data between RDBMS and HDFS. Imported the data from different sources like AWS S3, Local file system into Spark RDD. Collected data using Spark Streaming from AWS S3 bucket in near-real-time and performs necessary Transformations and Aggregations on the fly to build the common learner data model and persistence the data in HDFS. Involved in transforming the relational database to legacy labels to HDFS, and HBASE tables using Sqoop and vice versa. Processed the web server logs by

developing Multi-hop flume agents by using Avro Sink and loaded into MongoDB for further analysis and worked on MongoDB NoSQL data modeling, tuning, disaster recovery and backup. Developed data pipeline using Spark, Hive and HBase to ingest customer behavioral data and financial histories into Hadoop cluster for analysis. Developed a Python Script to load the CSV files into the S3 buckets and created AWS S3 buckets, performed folder management in each bucket, managed logs and objects within each bucket. Worked with different file formats like JSON, AVRO and parquet and compression techniques like snappy and developed python code for different tasks, dependencies, SLA watcher and time sensor for each job for workflow management and automation using Airflow tool. Developed shell scripts for dynamic partitions adding to hive stage table, verifying JSON schema change of source files, and verifying duplicate files in source location. Monitor and Troubleshoot Hadoop jobs using Yarn Resource Manager and EMR job logs using Genie and Kibana. Worked on CICD Automation using tools like Jenkins, Salt stack, Git, Vagrant, Docker, Elastic Search, Grafana. Data Management, Data Access, Data Governance and Integration, Security, and Operations performed by using Hortonworks Data Platform (HDP). Worked with importing metadata into Hive using Python and migrated existing tables and applications to work on AWS cloud (S3). Involved with writing scripts in Oracle, SQL Server and Netezza databases to extract data for reporting and analysis and Worked in importing and cleansing of data from various sources like DB2, Oracle, flat files onto SQL Server with high volume data. Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and AWS cloud and making the data available in Athena and Snowflake. Extensively used Stash Git-Bucket for Code Control and Worked on AWS Components such as Airflow, Elastic Map Reduce (EMR), Athena and Snow-Flake. ENVIRONMENT: FileNet, IBM RAD 6.0, Scala, Java 1.5, JSP, Servlets, Core Java, Spring, Swing, Hibernate, JSF, ICE Faces, Hibernate, HTML, CSS, Jenkins, JavaScript, NodeJs, UNIX, Web Services- SOAP, WAS 6.1, XML, IBM WebSphere 6.1, Rational Clear Case, Log 4j, IBM DB2. HADOOP DEVELOPER Infinite Computer Solutions - Irving, TX, US July 2012 to August 2014 RESPONSIBILITIES: Demonstrated Hadoop practices and broad knowledge of technical solutions, design patterns, and

code for medium/complex applications deployed in Hadoop production. Wrote Spark applications for Data validation, cleansing, transformations and custom aggregations and imported data from different sources into Spark RDD for processing and developed custom aggregate functions using Spark SQL and performed interactive querying Responsible for developing data pipeline with Amazon AWS to extract the data from weblogs and store in HDFS. Worked on data pipeline creation to convert incoming data to a common format, prepare data for analysis and visualization, Migrate between databases, share data processing logic across web apps, batch jobs, and APIs, Consume large XML, CSV, and fixed-width files and created data pipelines in kafka to Replace batch jobs with real-time data. Involved in data pipeline using Pig, Sqoop to ingest cargo data and customer histories into HDFS for analysis. Developed Pig scripts to help perform analytics on JSON and XML data and created Hive tables (external, internal) with static and dynamic partitions and performed bucketing on the tables to provide efficiency. Used Hive QL to analyze the partitioned and bucketed data and compute various metrics for reporting and performed data transformations by writing MapReduce and Pig jobs as per business requirements Used Apache Kafka to aggregate web log data from multiple servers and make them available in Downstream systems for analysis and used Kafka Streams to Configure Spark streaming to get information and then store it in HDFS Design Architecture of data pipeline/ingestion as well as optimization of ETL workflows and developed syllabus/Curriculum data pipelines from Syllabus/Curriculum Web Services to HBASE and Hive tables. Performed data analysis, feature selection, feature extraction using Apache Spark Machine Learning streaming libraries in Python. Leading, managing, planning, the development and implementation of the wide Geographic Information Systems (GIS) program. Worked on setting up and configuring AWS's EMR Clusters and Used Amazon IAM to grant fine-grained access to AWS resources to users Enable and configure Hadoop services such as HDFS, YARN, Hive, Ranger, Hbase, Kafka, Sqoop, Zeppeline Notebook and Spark/Spark2 and involved in analyzing log data to predict the errors by using Apache Spark. Evaluate deep learning algorithms for text summarization using Python, Keras, TensorFlow and Theano on Cloudera Hadoop system Designed Database Schema and created Data Model to store realtime Tick Data

with NoSQL store. Extracting real time data using Kafka and spark streaming by Creating DStreams and converting them into RDD, processing it and stored it into Cassandra. Used AWS Data Pipeline to schedule an Amazon EMR cluster to clean and process web server logs stored in Amazon S3 bucket. Used DataStax Spark-Cassandra connector to load data into Cassandra and used CQL to analyze data from Cassandra tables for quick searching, sorting and grouping and involved in analyzing log data to predict the errors by using Apache Spark. Worked on migrating MapReduce programs into Spark transformations using Spark and Scala, initially done using python (PySpark). Worked and learned a great deal from Amazon Web Services (AWS) Cloud services like EC2, S3, EBS, RDS and VPC. Integrated MapReduce with HBase to import bulk amount of data into HBase using MapReduce programs. Used Impala and Written Queries for fetching Data from Hive tables and developed Several MapReduce jobs using Java API. Worked with Apache SOLR to implement indexing and wrote Custom SOLR query segments to optimize the search. Created kafka spark streaming data pipelines for consuming the data from external source and performing the transformations in scala and contributed towards developing a Data Pipeline to load data from different sources like Web, RDBMS, NoSQL to Apache Kafka or Spark cluster. Worked with xml's extracting tag information using xpaths and Scala XML libraries from compressed blob data types. Involved in creating Data Lake by extracting customer's Big Data from various data sources into Hadoop HDFS. This included data from Excel, Flat Files, Oracle, SQL Server, MongoDB, Cassandra, HBase, Teradata, Netezza and also log data from servers Define data governance rules and administrating the rights depending on job profile of users. Developed Pig and Hive UDF's to implement business logic for processing the data as per requirements and developed Pig UDFs in Java and used UDFs from PiggyBank for sorting and preparing the data. Developed Spark scripts by using Scala IDEas per the business requirement. Configured and optimized the Cassandra cluster and developed real-time java based application to work along with the Cassandra database. Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS. Developed Spark jobs using Scala on top of Yarn/MRv2 for interactive and Batch Analysis and involved in querying data using SparkSQL on top of Spark

engine for faster data sets processing and worked on implementing Spark Framework, a Java based Web Framework. Created Hive tables, loaded data and wrote Hive queries that helped market analysts spot emerging trends by comparing fresh data with EDW reference tables and historical metrics. Developed Spark code by using Scala and Spark-SQL for faster processing and testing and performed complex HiveQL queries on Hive tables. Creating Nagios, Grafana and Graphite dashboard for infrastructure monitoring. Environment: Spark, AWS, EC2, EMR, Hive, SQL Workbench, Genie Logs, Kibana, Sqoop, Spark SQL, Spark Streaming, Scala, Python, Hadoop (Cloudera Stack), Hue, Spark, Netezza, Kafka, HBase, HDFS, Hive, Pig, Sqoop, Oracle, ETL, AWS S3, AWS EMR, GIT, Grafana. SR. JAVA DEVELOPER Hi-Tech Solutions, Inc - Chicago, IL, US January 2010 to June 2012 RESPONSIBILITIES: Involved in the design and development phases of Agile Software Development and analyzed current Mainframe system and designed new GUI screens. Used Sqoop to ingest from DBMS and Python to ingest logs from client data centers. Develop Python and bash scripts for automation and implemented Map Reduce jobs using Java API and Python using Spark Imported data from RDBMS systems like MySQL into HDFS using Sqoop and developed Sqoop jobs to perform incremental imports into Hive tables. Demonstrated experience in managing the collection of geospatial data and understanding of data systems managed policies concerning the compilation of information and coordination of data through the GIS program; coordinating and overseeing the implementation of policies Involved in loading and transforming of large sets of structured and semi structured data and created Data Pipelines as per the business requirements and scheduled it using Oozie Coordinators. Developed the application using 3 Tier Architecture i.e. Presentation, Business and Data Integration layers in accordance with the customer/client standards. Played a vital role in Scala framework for web based applications and used File net for Content Management and for streamlining Business Processes. Created Responsive Layouts for multiple devices and platforms using foundation framework and implemented printable chart report using HTML, CSS and jQuery. Applied JavaScript for client side form validation and worked on UNIX, LINUX to move the project into production environment. Created Managed Beans for handling JSF pages and include logic for processing of the data on the

page and Created simple user interface for application's configuration system using MVC design patterns and swing framework. Used Object/Relational mapping tool Hibernate to achieve object to database table persistency. Worked with Core Java to develop automated solutions to include web interfaces using HTML, CSS, JavaScript and Web services. Developed web GUI involving HTML, Java Script under MVC architecture and creation of WebLogic domains and setup Admin & Managed servers for JAVA/J2EE applications on Non Production and Production environments. Implemented code according to coding standards and Created AngularJS Controller Which Isolate scopes perform operations and used GWT, GUICE, JavaScript, and Angular JS for client side implementation and extensively used Core Java such as Exceptions, and Collections. Configured the Web sphere application server to connect with DB2, Oracle and SQL Server in the back end by creating JDBC data source and configured MQ Series with IBM RAD and WAS to create new connection factories and queues. Extensively worked on TOAD for interacting with data base, developing the stored procedures and promoting SQL changes to QA and Production Environments.

Used Apache Maven for project management and building the application and CVS was used for project management and version management. Creating and updating existing build scripts using Ant for deployment Tested and implemented/deployed application on WAS server and used Rational Clear Case for Version Control. Environment: Hadoop, Map Reduce, Yarn, Hive, Pig, HBase, Sqoop, Spark, Scala, MapR, Core Java, R Language, SQL, Python, Eclipse, Linux, Unix, HDFS, Map Reduce, Impala, Cloudera, SQOOP, Kafka, Apache Cassandra, Oozie, Impala, Zookeeper, MySQL, Eclipse, PL/SQL JAVA DEVELOPER APLOMB Technologies - Princeton, NJ, US May 2007 to December 2009 RESPONSIBILITIES: Involved in the configuration of Spring Framework and Hibernate mapping tool and monitoring WebLogic/JBoss Server health and security.

Creation of Connection Pools, Data Sources in WebLogic console and implemented Hibernate for Database Transactions on DB2. Implemented CI, CD using Jenkins for continuous development and delivery. Involved in configuring hibernate to access database and retrieve data from the database and written Web Services (JAX-WS) for external system via SOAP/HTTP call. Used Log4j framework to log/track application and involved in developing SQL queries, stored procedures,

and functions. Developed a new CR screen from the existing screen for the LTL loads (Low Truck Load) using JSF. Used spring framework configuration files to manage objects and to achieve dependency injection. Implemented cross cutting concerns like logging and monitoring mechanism using Spring AOP. Implemented SOA architecture with web services using SOAP, WSDL, UDDI and XML and made screen changes to the existing screen for the LTL (Low Truck Load) Accessories using Struts. Developed desktop interface using Java Swing for maintaining and tracking products. Used JAX-WS to access the external web services, get the xml response and convert it back to java objects. Developed the application using Eclipse IDE and worked under Agile Environment and worked with Web admin and the admin team to configure the application on development, training, test and stress environments (Web logic server). Executed and coordinated the installation for the project and worked on web-based reporting system with HTML, JavaScript and JSP. Build PL\SQL functions, stored procedures, views and configured Oracle Database with JNDI data source with connection pooling enabled. Developed the Training and Appraisal modules using Java, JSP, Servlets and JavaScript. Used Hibernate based persistence classes at data access tier and adopted J2EE design patterns like Service Locator, Session Facade and Singleton. Worked on Spring Core layer, Spring ORM, Spring AOP in developing the application components. Modified web pages using JSP and Used Struts Validation Framework for form input validation. Created the WSDL and used Apache Axis 2.0 for publishing the WSDL and creating PDF files for storing the data required for module. Used custom components using JSTL tags and Tag libraries implementing struts and used Web Logic server for deploying the war files and used Toad for the DB2 database changes. ENVIRONMENT: Java, J2EE, JSF, Hibernate, Struts, Spring, Swing/JFC, JSP, HTML, XML, Web Logic, iText, DB2, Eclipse IDE, SOAP, Maven, JSTL, TOAD, DB2, JDK, Web Logic Server, WSDL, JAX-WS, Apache Axis. Education Bachelor's Skills Apache (9 years), APACHE HADOOP HDFS (8 years), APACHE HADOOP SQOOP (8 years), Java (10+ years), Oracle (9 years), Javascript Additional Information Technical Skills: Bigdata Ecosystem HDFS and Map Reduce, Pig, Hive, Impala, YARN, HUE, Oozie, Zookeeper, Solr, Apache Spark, Apache STORM, Apache Kafka, Sqoop, Flume, Flink, Elasticsearch NoSQL

Databases HBase, Cassandra, and MongoDB Hadoop Distributions Cloudera, Hortonworks
Programming languages Java, C, SCALA, Pig Latin, HiveQL PySpark. Scripting Languages Shell
Scripting Databases MySQL, oracle, Teradata, DB2 Build Tools Maven, Ant, sbt Reporting Tool
Tableau Version control Tools SVN, Git, GitHub Cloud AWS, Azure, S3, EC2, EMR App/Web
servers WebSphere, WebLogic, JBoss and Tomcat Web Design Tools HTML, AJAX, JavaScript,
JQuery, CSS and JSON. Operating Systems WINDOWS 10/8/Vista/ XP Development IDEs
NetBeans, Eclipse IDE, Python(IDLE) Packages Microsoft Office, putty, MS Visual Studio.

Name: Kevin Garcia

Email: druiz@example.net

Phone: 900-911-3584x69302