

Hadoop Developer Hadoop Developer Hadoop Developer - Dell Raleigh, NC ? 7+ years of professional experience involving project development, implementation, deployment and maintenance using Java/J2EE and Big Data related technologies. ? Hadoop Developer with 5 years of working experience in designing and implementing complete end-to-end Hadoop based data analytical solutions using HDFS, MapReduce, Spark, Yarn, Kafka, PIG, HIVE, Sqoop, Storm, Flume, Oozie, Impala, HBase etc. ? Good experience in creating data ingestion pipelines, data transformations, data management, data governance and real time streaming at an enterprise level. ? Experience in importing and exporting different formats of data into HDFS, HBASE from different RDBMS databases and vice versa using Sqoop. ? Exposure to Cloudera development environment and management using Cloudera Manager. ? Experience in analyzing data using HiveQL, Pig Latin, Hbase, Mongo and custom MapReduce programs in Java. ? Experience in extending Hive and Pig core functionality by writing custom UDFs using Java. ? Developed analytical components using Spark and Spark Stream ? Background with traditional databases such as Oracle, SQL Server, MySQL. ? Good knowledge and Hands-on experience in storing, processing unstructured data using NOSQL databases like HBase and MongoDB. ? Good knowledge in distributed coordination system ZooKeeper and experience with Data Warehousing and ETL. ? Hands on experience in setting up workflow using Apache Oozie workflow engine for managing and scheduling Hadoop jobs. ? Experience in creating complex SQL Queries and SQL tuning, writing PL/SQL blocks like stored procedures, Functions, Cursors, Index, triggers and packages. ? Good knowledge of database connectivity (JDBC) for databases like Oracle, DB2, SQL Server, MySQL, NoSQL, MS Access. ? Profound experience in creating real time data streaming solutions using Apache Spark/Spark Streaming, Kafka. ? Developed analytical components using Spark and Spark Stream ? Worked on a prototype Apache Spark Streaming project, and converted our existing Java Strom Topology. ? Proficient in visualizing data using Tableau, QlikView, Microstratergy and MS Excel. ? Experience in developing ETL scripts for data acquisition and transformation using Informatica and Talend. ? Used Maven extensively for building MapReduce jar files and deployed it to Amazon Web Services (AWS) using EC2 virtual Servers in the cloud and Experience in build

scripts to do continuous integrations systems like Jenkins. ? Experienced in Java Application Development, Client/Server Applications, Internet/Intranet based applications using Core Java, J2EE patterns, Spring, Hibernate, Struts, JMS, Web Services (SOAP/REST), Oracle, SQL Server and other relational databases. ? Experience writing Shell scripts in Linux OS and integrating them with other solutions. ? Experienced in using agile methodologies including extreme programming, SCRUM and Test Driven Development (TDD). ? Experienced in creating and analyzing Software Requirement Specifications SRS and Functional Specification Document FSD. Strong knowledge of Software Development Life Cycle SDLC. ? Devoted to professionalism, highly organized, ability to work under strict deadline schedules with attention to details, possess excellent written and communication skills. Authorized to work in United States for any employer. Work Experience Hadoop Developer Dell - Austin, TX October 2018 to Present Responsibilities: ? Involved in integrating the GIT into the Puppet to ensure the integrity of applications by creating Production, Development, Test, and Release Branches. ? Responsible for loading the customer's data and event logs from Oracle database, Teradata into HDFS using Sqoop ? End-to-end performance tuning of Hadoop clusters and Hadoop MapReduce routines against very large data sets. ? Performance tuning of Hadoop clusters and Hadoop MapReduce routines ? Developed Python code to gather the data from HBase and designs the solution to implement using PySpark. ? Developed PySpark code to mimic the transformations performed in the on-premise environment. ? Analyzed the Sql scripts and designed solutions to implement using pyspark. ? created custom new columns depending up on the use case while ingesting the data into Hadoop lake using pyspark. ? Analyze Cassandra database and compare it with other open-source NoSQL databases to find which one of them better suites the current requirement. ? Integrated Cassandra as a distributed persistent metadata store to provide metadata resolution for network entities on the network. ? Implemented Spark using Scala and also used Pyspark using Python for faster testing and processing of data. ? Build servers using AWS: Importing volumes, launching EC2, RDS, creating security groups, auto-scaling, load balancers (ELBs) in the defined virtual private connection. ? Installed and configured Apache Hadoop, Hive and Pig environment on AWS. Diligently teaming with

the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability ? Installed and configured various components of Hadoop ecosystem and maintained their integrity ? Designed, configured and managed the backup and disaster recovery for HDFS data. ? Migrating physical Linux/Windows servers to cloud (AWS) and testing. ? Installing, Upgrading and Managing Hadoop Cluster on Cloudera distribution. ? Commissioned Data Nodes when data grew and decommissioned when the hardware degraded ? Development, Acceptance, Integration, and Production AWS Endpoints. ? Configured various property files like core-site.xml, hdfs-site.xml, mapred-site.xml based upon the job requirement ? Monitored multiple Hadoop clusters environments using Ganglia and Nagios. Monitored workload, job performance and capacity planning ? Develop Hive scripts for end user / analyst requirements to perform ad hoc analysis. ? Plan, deploy, monitor, and maintain Amazon AWS cloud infrastructure consisting of multiple EC2 nodes and VMware Vm's as required in the environment. ? Expertise in AWS data migration between different database platforms like SQL Server to Amazon Aurora using RDS tool. ? Installing, Upgrading and Managing Hadoop Cluster on Cloudera distribution ? Installed and configured Hadoop HDFS, MapReduce, Pig, Hive, and Sqoop. ? Wrote Pig Scripts to generate MapReduce jobs and performed ETL procedures on the data in HDFS. ? Exported analyzed data to HDFS using Sqoop for generating reports. ? Scheduling all hadoop/hive/sqoop/Hbase jobs using Oozie. Environment: Hadoop, Map Reduce, Shell Scripting, spark, Pig, Hive, Cloudera Manager, AWS, CDH 5.4.3, Pyspark, HDFS, Yarn, Hue, Sentry, Oozie, Zoo keeper, Impala, Solr, Kerberos, cluster health, Puppet, Ganglia, Nagios, Flume, Sqoop, storm, Kafka, KMS Python/Java Consultant Walmart - Bentonville, AR September 2016 to October 2018 Responsibilities: ? Implemented user interface guidelines and standards throughout the development and maintenance of the website using the HTML, CSS, JavaScript and jQuery. ? Developed GUI using Django and Python for dynamically displaying the test block documentation and other features with Python code for a web browser. ? Implemented AJAX for dynamic functionality of a webpages for front end applications. ? Developed and tested many features for dashboard, created using Bootstrap, CSS, and JavaScript. ? Developed front end using Angular.js, React.JS, Node.js, bootstrap.js, backbone.js, JavaScript,

where back end is java with REST webservice. ? Involved in the application development using Spring Core, MVC modules and Java web based technologies: such as, Servlets, JSP, Java Web Service (REST/SOAP based), WSDL. ? Utilized standard Python modules such as csv, robot parser, iterators and pickle for development. ? Developed views and templates with Python and Django's view controller and templating language to create user-friendly website interface. ? Worked on Python OpenStack APIs and used NumPy for Numerical analysis. ? Developed Wrapper in Python for instantiating multi-threaded application. ? Managed datasets using Panda data frames and MySQL, queried MYSQL database queries from python using Python-MYSQL connector and MySQL dB package to retrieve information. ? Used Ajax and jQuery for transmitting JSON data objects between frontend and controllers. ? Used Angular MVC and two-way data binding. Worked on automation scripts using Selenium in JAVA. ? Developed entire frontend and backend modules using on Django including Tastypie Web Framework using Git. ? Designed, coded and tested key modules of the project using java oops concepts. ? Install KAFKA on Hadoop cluster and configure producer and consumer coding part in java to establish connection from twitter source to HDFS with popular hash tags. ? Developed Splunk infrastructure and related solution for application toolsets. ? Helped team to on-board data, create various knowledge objects, Install and maintain the Splunk Apps. ? Creating Application on Splunk to analyse the data. ? Manage Splunk configurations files like input, props, transforms and lookups. ? Configured Maven for Java automation projects and developed Maven Project Object Model (POM). ? Setup automated cron jobs to upload data into database, generate graphs, bar charts, upload these chartsto wiki and backup the database. ? Software Development in Linux Environment, utilized XShell to build, deploy Java applications ? Developed Founctional Package with Java, Erlang and Python ? Added several options to the application to choose particular algorithm for data and address generation. ? Maintained the versions using GIT and sending the release notes for each release. ? Supported the issues seen in the tool across all the teams for several projects

Environment: Python 3.2, Django, HTML5/CSS, MS SQL Server 2013, MySQL, JavaScript, Eclipse, Linux, Shell Scripting, PyCharm, Urllib, jQuery, GitHub, Angular.JS, Jira. Sr. Hadoop Developer CoxAuto - New York, NY August

2014 to September 2016 Responsibilities:

- ? Loading Defined, designed and developed Java applications, specially using Hadoop [Map/Reduce] by leveraging frameworks such as Cascading and Hive.
- ? Developed workflow using Oozie for running Map Reduce jobs and Hive Queries.
- ? Worked on loading log data directly into HDFS using Flume.
- ? Worked on Cloudera to analyze data present on top of HDFS
- ? Responsible for managing data from multiple sources.
- ? Load data from various data sources into HDFS using Flume.
- ? Involved in requirement and design phase to implement Streaming Lambda Architecture to use real time streaming using Spark and Kafka.
- ? Created and maintained Technical documentation for launching Cloudera Hadoop Clusters and for executing Hive queries and Pig Scripts
- ? Worked on SPARK engine creating batch jobs with incremental load through STORM, KAFKA, SPLUNK, FLUME, HDFS/S3, KINESIS, Sockets, AWS etc.,
- ? Writing scala classes to interact with the database.
- ? This plugin allows Hadoop MapReduce programs, HBase, Pig and Hive to work unmodified and access files directly.
- ? Involved in analyzing existing architecture on premise datacenters and designed to migrate applications from onprem to AWS Public Cloud.
- ? Designed and implemented MapReduce-based large-scale parallel relation-learning system
- ? Successfully loaded files to Hive and HDFS from Mongo DB Solar.
- ? Familiarity with a NoSQL database such as MongoDB Solar.
- ? Successfully loaded files to Hive and HDFS from Mongo DB Solar.
- ? Extracted files from MySQL through Sqoop and placed in HDFS and processed.
- ? Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.
- ? Developed Pig Latin scripts to extract data from the web server output files to load into HDFS.
- ? Built reusable Hive UDF libraries for business requirements which enabled users to use these UDF's in Hive Querying.
- ? Worked on debugging, performance tuning of Hive & Pig Jobs.
- ? Created Hbase tables to store various data formats of PII data coming from different portfolios.
- ? Developed Pig Scripts, Pig UDFs and Hive Scripts, Hive UDFs to load data files into Hadoop
- ? Worked closely on parallel computing with Spark team to explore RDD in Datastax Cassandra.
- ? Generating scala and java classes from the respective APIs so that they can be incorporated in the overall application.
- ? Writing entities in Scala and Java along with named queries to interact with database
- ? Writing user console page in lift along with the

snippets in Scala. The product is responsible to give access to the user to all their credentials and privileges within the system Writing scala test cases to test scala written code. Environment: Python 3, Django, Hadoop, HDFS, Map Reduce, Shell Scripting, spark, solr, Pig, Hive, HBase, Sqoop, Flume, Oozie, Zoo keeper, cluster health, monitoring security, Redhat Linux, impala, Cloudera ManagerGolang, bitbucket, pdb, AWS, Jira, Jenkins, dockers, Pyspark, Rest, Virtual Machine, Ajax, jQuery, JavaScript, LINUX. Software Developer Cognizant - Greenville, SC July 2012 to August 2014 Responsibilities: ? Design, develop, test, deploy and maintain the website. ? Interaction with client to understand expectations and requirements. ? Designed and developed the UI of the website using HTML, AJAX, CSS and JavaScript. ? Developed entire frontend and backend modules using Python on Django Web Framework. ? Performed testing using Django's Test Module. ? Assisting junior Perl developers in the development of scripts ? Developed a fully automated continuous integration system using GIT, Gerrit, Jenkins, MySQL and custom tools developed in Python and Bash. ? Managed, developed, and designed a dashboard control panel for customers and Administrators using Django, Oracle DB, PostgreSQL. ? Extensively used python modules such as requests, urllib, urllib2 for web crawling. ? Implemented configuration changes for data models. ? Used Pandas datamining library for statistics Analysis & NumPY for Numerical analysis. ? Managed large datasets using Panda data frames and MySQL. ? Exported/Imported data between different data sources using SQL Server Management Studio. ? Maintained program libraries, users' manuals and technical documentation. Responsible for debugging and troubleshooting the web application. ? Successfully migrated all the data to the database while the site was in production. ? Developed GUI using webapp2 for dynamically displaying the test block documentation and other features of python code using a web browser. Environment: Python 2.7, Django 1.8, CSS, HTML, JavaScript, JQuery, webbapp2, AJAX, MYSQL, Linux, Heroku, GIT, urllib, urllib2, Oracle DB, PostgreSQL, and VMWare API.

Name: Taylor McLaughlin

Email: petersonrichard@example.net

Phone: 001-538-644-5294x1750