

Hadoop / Spark Developer Hadoop / Spark Developer Hadoop Developer - Esvee Technologies Inc
Chantilly, VA Authorized to work in the US for any employer Work Experience Hadoop / Spark
Developer Esvee Technologies Inc - Chantilly, VA September 2015 to Present Description: Worked
with financial and production dataset. Our main aim was to help finance department in finding any
patterns in the stock datasets, which get influence with many natural and political news. We did
collect real-time data from different sources using flume and then data was analyzed using Spark
Scala. Responsibilities: Design efficient Spark code using Scala and Spark SQL, which can be
forward engineered by our code generation developers. Did spark streaming and micro-batch
processing using Scala as a programming language. Using Hive Script in Spark for data cleaning
and transformation purpose. Consult with and code developers on how to improve their algorithms
for generating high performing Spark code. Consult with product management on the features that
are applicable and valuable for a product which designs and submits Spark applications. Spark
Streaming, for Fast failure and straggler recovery, Load balancing, unification of streaming, batch
and interactive workloads and advanced analytics. Used Apache Phoenix to get access to HBase
database and process using SQL commands. Created Hive customized UDF for cleaning the data
and storing it in Hive warehouse from HBase. Manage Hadoop operations with multi-node HDFS
cluster using Cloudera Manager. Created Oozie for fetching out data on the periodic basis and in
periodic timely fashion. Installing and configuring Hadoop ecosystem like Pig, Hive. Handle
importing of data from various data sources; perform transformations using Hive, Map Reduce, load
data into HDFS and extract the data from My SQL into HDFS using Sqoop. Extensively used Pig
for data cleansing. Create partitioned tables in Hive. Use Hive to analyze the partitioned and
bucketed data and compute various metrics for reporting. Install and configure Pig and also write Pig
Latin scripts. Environment: - Hadoop 1.2.1, MapReduce, Sqoop 1.4.4, Hive 0.10.0, Pig 0.11.1,
Hbase 0.94.11, Scala, Zookeeper 3.4.3, Mac EL Caption Database Administrator Indore, Madhya
Pradesh December 2014 to August 2015 Description: - Implemented an MY-SQL database for a
pathology lab client. We made a central database which helps in linking data between testing
department and finance. We developed some stored procedures for generating reports on periodic

basis. Responsibilities: Lead and participated in designing, coding, validating, and implementing database developments in My-SQL. Backing up and recovering data from data warehouse. Validating schema and create necessary data objects. Writing stored procedures, triggers with development team for application. Manipulate, clean data as per requirement of the application. Educate coworker with the structure of the database and their standard practice. Providing information on request basis by management for BI purpose. Environment: - My-SQL, SQL-Workbench, and OS windows XP Education Bachelor of Science in Bioinformatics in Bioinformatics Soft vision Institute of Biotechnology - Indore, Madhya Pradesh Skills Apache (2 years), APACHE HADOOP HDFS (2 years), APACHE HADOOP MAPREDUCE (2 years), database (2 years), SQL (2 years) Additional Information Over 3 years of professional experience in IT industry with hands-on experience in Developing, Implementing and maintenance of various applications using Java technologies. Around 2+ years of experience with Hadoop Ecosystem including MapReduce, Spark, HDFS, SQOOP, HIVE, Avro-tools, Pig, HBase and YARN Around 1 year of professional experience in Database Administration and Data Analysis Hands on experience with Amazon Web Services infrastructure (EC2, RDS, EMR, etc). Experience in using and implementing Machine Learning Algorithms Experience in using PL/SQL to write Stored Procedures, Functions and Triggers. Experience in Collecting log data from various sources and integrating into HDFS using Flume and experience in developing custom UDFs for Pig and Hive. Experience working with Oozie workflow engine to schedule time-based jobs to perform multiple actions. Good understanding of Data analysis, Data engineering, and Data Science Ability to plan, prioritize and work under pressure with an excellent analytical, leadership and support skills. Experience in working with agile methodologies and suggesting process improvements in agile. Used JIRA for an issue and project tracking. Loaded data from flat data files into SQL Server 2008 database tables using bulk insert, and table export/import; loaded data from flat files and excel sheet into Oracle database using SQL*Loader. Provide Consulting to Customers in identifying Big Data use cases and guiding them to implementation RDBMS experience includes Oracle, PostgreSQL, MySQL and programming using PL/SQL, SQL. Load and transform large sets of structured, semi

structured and unstructured data. Technical Skills: Data Analytics SAS 9.3, Tableau, Apache Hadoop 2.6.3, Spark, HDFS, Hbase, MapReduce, R, Sqoop, Flume, Avro-tools, Pig, Hbase, oozie Languages/Web Technologies Java, C++, Python(package:- TensorFlow, Anaconda, Numpy), C, .Net, HTML, XML, Java Script, JSON, JSP, CUDA, Scala, R (packages: - Bioconductor, random forest, kernlab, nnet, and caret), Angular JS Web/Application Servers Apache Tomcat, Oracle VM Database My-SQL, SQL server 2005, HIVE, Impala, GitHub (code Management) Developer tools Eclipse, R-Studio Operating System Linux, Windows, Mac

Name: Carolyn Mills

Email: lawsonnicole@example.com

Phone: (913)739-0513