

Spark/Scala Developer Spark/Scala Developer Spark/Scala Developer - Subaru Piscataway, NJ ?

Extensive IT experience in Big Data technologies, Data Management/Analytics, Data visualization and java-based enterprise application using JAVA/J2EE. ? Worked in domains like e-commerce, healthcare, automotive industries etc. ? Extensive experience of Big Data Ecosystem including Hadoop2.X, HDFS, YARN, MapReduce, Spark1.4+, HIVE 2.1, Impala 1.2, Hbase 1.0+, SQOOP 1.4, Flume 1.7, Kafka 1.2+, Oozie 3.0+ and Zookeeper 3.4+ ? Experience in programming languages, namely: Java 8+, Scala 2.1+, SQL ? Experienced with real-time data processing mechanism in Big Data Ecosystem such as Apache Kafka and Spark Streaming ? Experienced in Spark Scala API, Spark Python API to transfer, process and analyse data in different formats and structures ? Experienced in writing HiveQL, and developing Hive UDFs in Java to process and analyse data ? Implemented Sqoop and Flume jobs for large sets of structured and semi-structured data migration between HDFS and/or other data storage like Hive or RDBMS ? Conducted transformation of data in formats like Avro and Parquet ? Adept at using Sqoop to migrate data between RDBMS, NoSQL and HDFS ? Knowledge of Linux/Unix Shell Commands ? Good knowledge of scheduling batch job workflow using Oozie ? Worked with RDBMS including MySQL, MsSql, Oracle and knowledge of SQL Server 2008 R2 and over and NoSQL databases including HBase, Cassandra and MongoDB ? Very Good understanding and Working Knowledge of Object-Oriented Programming (OOPS), Core Java concepts, J2EE 8, JDBC, Javascript and jQuery. ? Working knowledge of workflows and ETL batch jobs using SSIS, TSQL, Informatica and Talend ? Experience in database design using Stored Procedure, Functions, Triggers and strong experience in writing complex queries for DB2, SQL Server. ? Experience with SQL Server Reporting Services, Power BI and Tableau. ? Knowledge of Software Development Life Cycle (SDLC) methodology like Agile, Waterfall ? Familiarity with project management tools like GIT, Microsoft Team foundation server 2015+ ? Knowledge of Unit Testing with ScalaCheck, ScalaTest, JUnit and MRUnit , also used JIRA for basic issue tracking, Jenkins for continuous integration and A/B testing for certain projects. ? Excellent interpersonal and communication skills. Creative, data-oriented, problem- shooting, enthusiastic learner

Work Experience Spark/Scala Developer

Subaru - Piscataway, NJ March 2019 to Present Location: Piscataway, NJ Role: Spark/Scala Developer Subaru is Japanese multinational corporation and conglomerate primarily involved in both terrestrial and aerospace transport manufacturing. it is best known for the automobile industry.

The team was responsible to track changes, quality of work, production process as well as the safety of the products and its parts. This identifies which part of vehicle needs constant upgrades or repairs/ replacements to make high quality upgrades and help cost reduction in long run.

Responsibilities: ? Importing and exporting large amount of data using Sqoop and real time data using Flume and Kafka. ? Uploaded data to Hadoop HIVE and combined new tables with existing databases ? Created various hive external tables, staging tables and joined the tables as per the requirement. Implemented static Partitioning, Dynamic partitioning and Bucketing in Hive using internal and external table. ? Written transformations and actions on data frames, used Spark SQL on data frames to access hive tables into spark for faster processing of data. ? Developed Spark applications using Scala utilizing Data frames and Spark SQL API for faster processing of data. ? Used Talend to migrate historical data from Oracle SQL and SQL Server to HDFS and HIVE ? Extracted data from oracle SQL server and MYSQL databases to HDFS using SQOOP ? Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations and other during ingestion process itself. ? Used Scala to convert Hive/SQL queries into RDD transformations in Apache Spark. ? Implemented the workflows using Apache Oozie framework to automate tasks. Used Zookeeper to co-ordinate cluster services. ? Created visualization reports using tableau ? Designed and built a custom and generic ETL framework - Spark application using Scala; for data loading and transformations. ? Used Git for version control, JIRA for issue tracking and Jenkins for continuous integration

Environment: Hadoop2.X, Cludera CDH, HDFS, Java 8+, Python, Scala 2.1+, Spark1.4+, HIVE 2.1, Kafka 1.2+, SQOOP 1.4, Flume 1.7, Talend, Zookeeper 3.4+, Oozie 3.0+, Git, JIRA, Tableau

Project Name: Supply management and logistics Business Intelligence Developer Northern Safety and industrial - Utica, NY February 2018 to October 2018 The company main business is based on construction, maintenance, agricultural, food preparation and handling, public service works,

medical, hazardous materials handling and most up-to-date safety and industrial supplies. The team was focused on migrating the data from database to database which included a large set of e-commerce data of industrial equipment. Responsibilities: ? Experience in T-SQL programming (DDL, DML) skills like creating Stored Procedures, User Defined Functions, indexes, Views, Tables ? Created aggregate, Merge Join, Sort, Execute SQL Task, Data Flow Task, and Execute Package Task etc to generate underlying data for reports and to export cleaned data from Excel Spreadsheets, Text file, MS Access and CSV files to data warehouse. ? Configure and maintain Report Manager and Report Server for SSRS. ? Created visual reports like pie charts, bar graphs of company's product sales using Power BI. Environment: Spark/Hadoop Developer Analysis of Insurance dataset using Hadoop - Utica, NY September 2017 to September 2018 WINDOWS 10, VISUAL STUDIO 2017, MS SQL Server 2016, MS Access, Microsoft Team Foundation Server (TFS). Project Name: Analysis of Insurance dataset using Hadoop Data Analyst Hyderabad, Telangana June 2015 to June 2016 This project was focused on eliminating bogus voting(multiple votes by the same person) by creating a database that will not allow the person to vote again in the election. Responsibilities ? Designed and coded certain application modules and components ? Designed the logical and physical data model, generated DDL, DML scripts ? Designed user-interface and used JavaScript to check validations. ? Wrote SQL queries, stored procedures and database triggers on the database objects ? Wrote SQL queries to extract data from archives using complex joins ? Developed various Java classes, SQL queries and procedures to retrieve and manipulate the data from backend database using JDBC ? Analysis and Reporting of data using SSRS ? Enabled reporting access to the archives for reporting tools and created documentation Environment: .Net, C#.net, IIS, ASP, JavaScript, SQL Server 2008R2, SQL-Server SSRS Project Name: Railway Stipulation System Java Developer Hyderabad, Telangana May 2014 to May 2015 The objective of Railway Stipulation system is to empower the passenger to book tickets online. It also enables them to book the tickets without getting into the queue at the comfort of their home. It also enables them to swap the tickets with other passenger without cancelling the actual ticket. Responsibilities ? Designed and coded GUI based application to facilitate the ticket booking. ?

Created an interface where the user can "swap seats" with another passenger online via request. ?
Used JDBC for database connectivity ? Designed the logical and physical data model, generated
DDL, DML scripts ? Designed user-interface and used JavaScript to check validations. ? Wrote
SQL queries, stored procedures and database triggers on the database objects Environment: Java
8+, JDBC, Tomcat 7.0, HTML, CSS, JavaScript, JSP, SQL Server 2008R2 Education Master's Skills
DATA ANALYSIS, DATABASE, SQL SERVER, SQL SERVER 2008, SQL SERVER 2008 R2, Linux,
HTML, JQuery, XML, Javascript, TECHNICAL SKILLS: Hadoop Ecosystem\ Cloud Platform\
Hadoop2.X, Spark1.4+, MapReduce, Hive2.1, \ Google Cloud Platform (Dataproc, Compute \
Impala1.2+ Sqoop1.4, Flume1.7, Kafka1.2+, \ Engine, Bucket, SQL), Amazon Web Service \
Hbase1.0+, Oozie3.0+, Zookeeper3.4+\ (EC2, S3, EMR), Databricks Cloud Community\
Programming Language\ Operating Systems\ Java 8+, C++, Scala2.1+\ Linux, Ubuntu, Mac OS,
CentOS, Windows\ Web Development\ Database\ JavaScript, jQuery, AngularJS, HTML, CSS, \
MySQL5.X, Oracle11g, PostgreSQL9.X, \ Node.js\ Netezza7.X, MongoDB3.2, HBase0.98\ IDE
Application\ Data Analysis & Visualization\ NetBeans, Eclipse, Visual Studio Code \ Python, R,
Tableau, Matplotlib, D3.js\ IntelliJ Idea, SQL Server 2008 R2+\ Scripting Language\ Machine
Learning\ UNIX Shell, HTML, XML, CSS, JSP, SQL, \ Regression, Decision Tree, Random Forest, \
Markdown\ K-Means, Neural Networks, SVM, NLP\ Environment\ Collaboration\ Agile, Scrum,
waterfall\ Git, Microsoft TFS, JIRA, Jenkins\ (5 years) Certifications/Licenses Driver's License

Name: Pamela Sellers

Email: norrisdiane@example.com

Phone: +1-652-517-8247x23210