Hadoop

Areas of expertise • Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Spark, Pyspark, HUE and having knowledge on cassandra • Programming Languages: Python, Core Java and have an idea on Scala • Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra • Tools: Eclipse, Toad, FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebool and Devops - RTC, RLM. • Scripting Languages: JSP • Platforms: Windows, UnixEducation Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Experience Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution and a rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responesbilities:

• Building and providing domain knowledge for Anti Money Laundering among team members.

• The layered architecture has Data Warehouse and Workspace layers which are used by Business Analysts.

• Actively involved in designing of star-schema model involving various Dimensions and Fact tables.

• Designed SCD2 for maintaining history of the DIM data.

• Developing Hive Queries for mapping data between different layers of architecture, and it's usage in Oozie

Workflows.

• Integration with Data Quality and Reconciliation Module.

• Regression and Integration testing of solution for any issues in integration with other modules and effectively testing the data flow from layer-to-layer.

• Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provide by BAs.

• Developing spark Jobs for various business rules.

• Learning "Machine Learning", which will be used further in the project for developing an effective model for Fraud detection for Anti Money Laundering system.

• Scheduling Jobs using Autosys tool.

• Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the multiple vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost the network performance.

Responsibilities:

• Working with Hadoop Distributed File System.

• Involved in importing data from MySQL to HDFS using SQOOP.

• Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

• Involved in Preparing Test cases Document.

• Involved in Integrating Hive and HBase to store the operational data.

• Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framwe work

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14