Hadoop

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume.

Other Language: Scala, Core-Java, SQL, PLSQL, Sell Scripting ETL Tools: Informatica Power Center8.x/9.6, Talend 5.6

Tools: Eclipse, Intellij Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL, Oracle.10/11gEducation Details

 M.C.A Pune, MAHARASHTRA, IN Pune University

Hodoop Developer

Hodoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure behavior. The customers spending ranges from online shopping, medical expenses in hospitals, cash transactions, and dwebit card usage etc. the behavior allows the bank to create an analytical report and based on which the bank used to display the product offers on the customer portal which was built using java. The portal allows the customers to login and see their transactions which they make on a day to day basis .the analytics also help the customers plan their budgets through the budget watch and my financial forecast applications embedded into the portal. The portal used hadoop framework to analyes the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done using the hadoop framework as big data analytics system.

Role & Responsibilities:

Import data from legacy system to hadoop using Sqoop, flume.

Implement the business logic to analyses the data

Per-process data using spark.

Create hive script and loading data into hive.

Sourcing various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the wweb application and extract meaningful information out of it.the solution was based on the open source s/w hadoop. The data will be stored in hadoop file system and processed using Map/Reduce jobs. Which in trun includes getting the raw html data from the micro wwebsites, process the html to obtain product and user information, extract various reports out of the vistor tracking information and export the information for further processing

Role & Responsibilities:

Move all crawl data flat files generated from various micro sites to HDFS for further processing.

Sqoop implementation for interaction with database

Write Map Reduce scripts to process the data file.

Create hive tables to store the processed data in tabular formats.

Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main objective of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream application are like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to which the performance of OLTP system goes slows down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLINICAL, EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional layer are created which are used for reporting application.

Role & Responsibilities:

To understand & analyze the requirement documents and resolve the queries.

To design Informatica mappings by using various basic transformations like Filter, Router, Source qualifier, Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and bug fixing.

Create workflow/batches and set the session dependencies.

Implemented Change Data Capture using mapping parameters, SCD and SK generation.

Developed Mapplet, reusable transformations to populate the data into data warehouse.

Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

Involved in Unit Case Testing (UTC)

Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041