Techlist 1
I524

Kumar Satyam : S17-IR-2031

January 25, 2017

1. Google Cloud storage

This is the cloud storage offered by Google. It is unified object storage. To have high availability and performance among different regions in the geo-redundant storage offering. If you want high availability and redundancy with a single region one can go for Regional storage. Nearline and Coldline are the different archival storage techniques. Nearline storage offering is for the archived data which the user access less than once a month . Coldline storage is the storage which is used for the data which is touched less than once a year.

All the data in Google Cloud storage belongs inside a project. A project will contains different buckets. Each bucket has different objects. We need to make sure that the name of the bucket is unique across all Google cloud name space . And the name of the objects should unique in a bucket.

2. EclipseLink

It is an open source persistence Services project from Eclipse foundation. It is a framework which provide developers to interact with data services including database and web services, Object XML mapping etc. This is the project which was developed out of Oracle's Toplink product. The main difference is EclipseLink does not have some key enterprise feature. Eclipselink support a number of persistence standard model like JPA, JAXB, JCA and Service Data Object. Like Toplink, the ORM (Object relational model) is the technique to convert incompatible type system in Object Oriented programming language. It is a framework for storing java object into relational database.

3. Torch

It is a open source machine learning library, a scientific computing framework. It implements LuaJIT programming language and implements C/CUDA. It implements N-dimensional array. It does routines of indexing, slicing, transposing etc. It has in interface to C language via scripting language LuaJIT. It supports different artificial intelligence models like neural network and energy based models. It is compatible with GPU. The core package of is torch. It provides a flexible N dimensional array which supports basic routings. It has been used to build hardware implementation for data flows like those found in neural networks,

4. Caffe

It is a deep learning framework made with three terms namely expression, speed and modularitywww-caffe. Using Expressive architecture, switching between CPU and GPU by setting a single flag to train on a GPU machine then deploy to commodity cluster or mobile devices.Here the concept of configuration file will comes without hard coding the values . Switching between CPU and GPU can be done by setting a flag to train on a GPU machine then deploy to commodity clusters or mobile devices.

It can process over 60 million images per day with a single NVIIA k40 GPU It is being used bu academic research projects, startup prototypes, and even large-scale industrial applications in vision, speech, and multimedia.

5. Parquet

Apache parquet is the column Oriented data store for Apache Hadoop ecosystem and available in any data processing framework like Hadoop, spark etc, data model or programming language. It stores data such that the values in each column are physically stored in contiguous memory locations. As it has the columnar storage, it provides efficient data compression and encoding schemes which saves storage space as the queries that fetch specific column values need not read the entire row data and thus improving performance.