International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 8, Issue 12 (October 2013), PP. 19-22

Big Data Using Hadoop Database Using Python Language to Implement Real Time Applications

E.Punarselvam[1], G.Sumathi[2], R.Parthasarathy[3], M.suresh[4]
[1]*Assistant Professor Department of Information Technology Muthayammal Engineering College Rasipuram – 637 048, Tamilnadu, India.*
[2]*Assistant Professor Department of computer science and engineering Muthayammal Engineering College* Rasipuram – 637 048, Tamilnadu, India
[3,4]*II Year ME [CSE] Muthayammal Engineering College, Rasipuram – 637 048, Tamilnadu, India.*

Abstract:- In previous day lot of databases are use to store and reterive the data but it is complex when the data size reach at (petabytes).in recent days we use Hadoop database for big data. The Hadoop database is use to store datasets based on java language. In our paper talk about using python coding to store datasets . Hadoop databases for real time applications and projects. Python is user friendly environment it works all platform. But in java coding is large and it's not easy to debug and execute. Python is very easy to execute the code line by line.

Keywords:- Big data, Hadoop database, Python, Map reduce algorithm, geo datasets.

I. INTRODUCTION

Big data usually includes data sets with sizes ranging from a few dozen terabytes to many petabytes of data in a single data set. big data concept includes in big science, science and research, government. International development.

WHERE THE HADOOP DATABASE USED:

Hadoop database used in real time projects, organisations and finance etc. Image shack(image hosting website),ISI(information science institute),Amzon.com, Foursquare(social networking)Twitter.

a) BASIC PLATFORM OF HADOOP

In Hadoop database store the datasets usig java program. But it is complement to write the coding for datasets in real time projects sometimes it is difficult to execte the datasets.datasets are execute with mapreduce algorthim and it failed sometimes.so we replace the language java to python.How map reduce algorithm work in big data: real time **Example: word count**

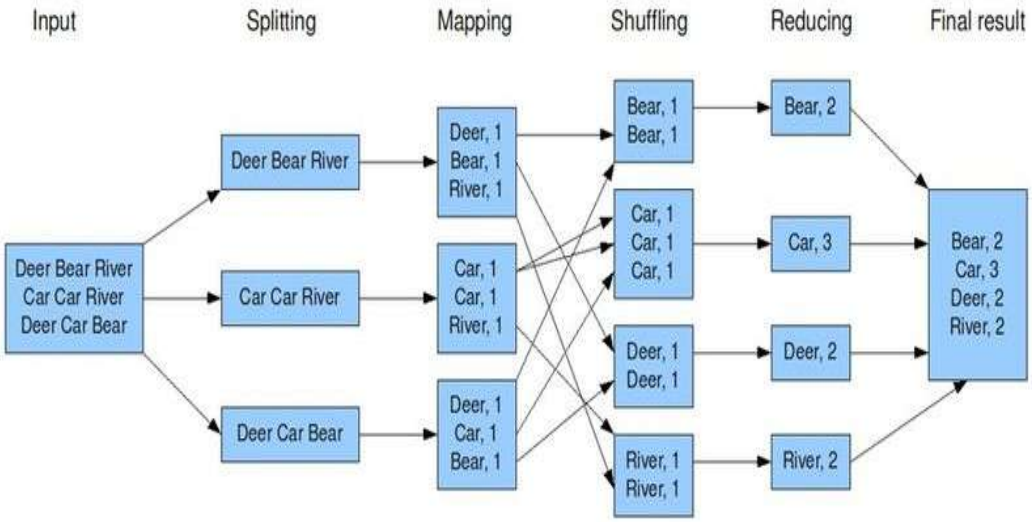

Figure 1 : Example for Map Reduce Algorithm

II. WHAT IS PYTHON

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. The Python interpreter and the extensive standard library are freely available in source or binary form for all major platforms.

Why the python is suitable for Hadoop database:

Step 1: Compare to java language the python is code efficient programming language.

Step 2: In python check the errors line by line.

Step 3: After correcting the errors it goes to next step for producing the applications

Some basic examples for python:

Now i am shown to write **Welcome to Anna University** program in python:

Syntax:

```
Print ”welcome to anna university”
```

In java hello program is written us:

```
Public class hello world
{
 public static void main (string[] args)
{
system.out.println(“welcome to anna university”);
}
}
```

Compare to these two codings which one is efficient and now we get some idea to create dataset in python language.

What is the major difference between python and java:

SI.No	PYTHON	JAVA
1	Python programs are also take much less time to develop the applications	Java programs take much more time to develop the applications
2	Python programs are typically 3-5 times shorter than equivalent Java program	Java programs take more time for execution when compare to Python language
3	This difference can be attributed to Python's built-in high-level data types and its dynamic typing	Java program is static where as in python is dynamic typing
4	**Example:** Python programmer wastes no time declaring the types of arguments or variable	**Example:** java programmer wastes time declaring the types of arguments or variable
5	Python's run time must work harder than Java's.	Java run time less work harder than python.
6	Python coding not a Case sensitive.	Java coding is a Case sensitive.
7.	There is no semicolon in the end of the statement	There is semicolon in the end of the statement for a suitable statements.
8.	Open and close brackets not necessary.	Open and close brackets is necessary for a suitable statements.

How the python implement in various platform projects:

In recent days python is used in various real time projects.in that projects there is not easy to create the data set using C, C++,java languages.but in python it is very easy to written the codes.

List some various platform use in python language:

1. ACCESSIBILITY:

Python in the blind audio tactile mapping system,The Blind Audio Tactile Mapping System (BATS) seeks to provide access to maps for the blind and visually impaired.

2. DATABASES

Forecast watch.com it is useful for meteorologist to find weather reports around 800 cities

3. DOCUMENTATION DEVELOPMENT

Honeywell reduce the document cost using python

4. NETWORK DEVELOPMENT

Super League Uses Python and PostgreSQL for Rapid Development

The Devil Framework: A Python-based Distributed System For Technology Integration

5.PRODUCT DEVELOPMENT

Acqutek Uses Python to Control CD/DVD Packaging Hardware

GravityZoo: Bringing Your Desktop Applications to The Internet As A Service

b) GEODATA:

The geodatabase is the common data storage and management framework for ArcGIS. It combines "geo" (spatial data) with "database" (data repository) to create a central data repository for spatial data storage and management. It can be leveraged indesktop, server, or mobile environments and allows you to store GIS data in a central location for easy access and management.The geodatabase offers you the ability to **Store** a rich collection of spatial data in a centralized location.Apply **sophisticated rules and relationships** to the data.

- Define **advanced geospatial relational models** (e.g., topologies, networks).
- **Maintain integrity** of spatial data with a consistent, accurate database.
- Work within a **multiuser access and editing** environment.
- **Integrate spatial data** with other IT databases.
- Easily **scale** your storage solution.
- Support **custom features** and behaviour.
- **Leverage** your spatial data to its full potential.

General example to create dataset:

Create the dataset. (java)

```
dataset_id =
        H5Dcreate_wrap (file_id, "/dset",
                        H5.J2C (HDF5CDataTypes.JH5T_STD_I32BE),
                        dataspace_id, HDF5Constants.H5P_DEFAULT);
```

create dataset in python:

```
create dataset=f.create dataset("mydataset", (100,), dtype='i')
```

create dataset in java language geodata: public class Dataset implements MachineImageSupport

```
{
   private SmartDataCenter provider;
        Dataset(@Nonnull SmartDataCenter sdc)
        {
         provider = sdc;
         }
   @Override
   public void addImageShare(@Nonnull String providerImageId, @Nonnull String accountNumber) throws
CloudException, InternalException
{
throw new OperationNotSupportedException("Image sharing is not supported");
}
public void addPublicShare(@Nonnull String providerImageId) throws CloudException, InternalException {
throw new OperationNotSupportedException("Image sharing is not supported");
   }
   @Override
   public @Nonnull String bundleVirtualMachine(@Nonnull String virtualMachineId, @Nonnull MachineIm
ageFormat format, @Nonnull String bucket, @Nonnull String name) throws CloudException, InternalException

{
throw new OperationNotSupportedException("Image bundling is not supported");
        }
```

```
Create dataset in python for geodataset:
def crunch_data():
        cloud.bucket.get('dataset.txt')

        with open('dataset.txt', 'r+') as dataset:
          modify_data(dataset)

        cloud.bucket.put('dataset.txt')

    cloud.call(crunch_data)

import cloud

    def crunch_data():
      with open('/bucket/dataset.txt', 'r+') as dataset:
        modify_data(dataset)

    cloud.call(crunch_data)
```

III. CONCLUSION

In this paper show about various difference between java and python language used in big data. Some useful examples shown for future work how the python language is use and easy code writing and debugging concept. This paper described a systematic flow of survey on the big data processing in the context of python language. We respectively discussed the key issues, including Python, java and map reducing architecture, popular parallel processing framework, major applications and optimization of MapReduce. Big Data is not a new concept but very challenging. It calls for scalable storage index and a distributed approach to retrieve required results near real-time. It is a fundamental fact that data is too big to process conventionally. Nevertheless, big data will be complex and exist continuously during all big challenges, which are the big opportunities for us. In the future, significant challenges need to be tackled by industry and academia. It is an urgent need that computer scholars and social sciences scholars make close cooperation, in order to guarantee the long-term success of big data.

REFERENCES

[1]. Global Research Data Infrastructures: Towards a 10-year vision for global research data infrastructures. Final Roadmap, March 2012. [online] http://www.grdi2020.eu/Repository/FileScaricati/6bdc07fbb21d-4b90-81d4-d909fdb96b87.pdf

[2]. Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. October 2010. [online] Available at http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

[3]. Birney E. 2012, The making of ENCODE: Lessons for big-data projects, Nature, vol.489, pp.49-51.

[4]. Hadoop: The Definitive Guide, Third Edition by Tom White (O'Reilly - 2012)

[5]. http://source.mozillaopennews.org/en-US/code/dataset-python/

[6]. http://refcardz.dzone.com/refcardz/data-mining-discovering-and#refcard-download-social-buttons-display