

Mobile Hadoop : Retrieving Original Image From Huge Datasets in Mobile Using Big Data Analytics

Shyam Mohan J S¹

Assistant Professor,
Dept.Of CSE,

SCSVMV University,

Kanchipuram,Tamilnadu

jsshyammohan@kanchiuniv.ac.in

Dr.P.Shanmugapriya²

Associate Professor,
Dept.Of IT,

SCSVMV University,

Kanchipuram,Tamilnadu

priya_prakasam@yahoo.co.in

M.Thirunavukkarasu³

Assistant Professor,
Dept. Of CSE

SCSVMV University,

Kanchipuram, Tamilnadu

mthiru@kanchiuniv.ac.in

Abstract-Performing insights for huge data sets viz. images can be done effectively by using Big data analytics. However, gaining insights in a mobile device is a difficult task as mobile devices support only structured data. In this paper we have performed Mobile Analytics using hadoop for images spread over huge data sets and our experimental result shows how an exact image is retrieved from the unstructured data providing insight for performing other analytics successfully.

Keywords – Hadoop, Big Data

I. INTRODUCTION

Using Android for Hadoop

Google’s Android operating system was the most natural choice of a mobile platform to run Hadoop on. Android is open nature relative to other mobile platforms such as the iPhone, which is the most popular mobile application platform. It allows arbitrary applications to be installed on any number devices without any external permission . Android’s Dalvik VM implements a subset of the Apache Harmony Java implementation, which includes most of the Java classes used by Hadoop. Hadoop, an Apache project itself, depends on several Apache libraries, such as Apache log4j, Apache XML, and Apache Commons. [1] Because of this compatibility, it was possible to port Hadoop without rewriting a huge amount of code. Using Android made it possible to port Hadoop without completely overhauling its source code. Android’s debugging tool, ADB, can be used to execute shell commands, install applications, display phone logs, and push and pull files. Because it is a shell utility, it can be used in scripts which makes it acceptable platform for distributed mobile applications, whereas closed platforms such as the iPhone make the development of such systems more difficult.

II. PROBLEM STATEMENT

Generally Big Data is a Collection of huge data sets in Volume, variety and Velocity. There are variety of analytical tools available for prediction and effective decision making using Big Data Analytics. Assume that a mobile user wants to

check for his/her image from the huge data sets like social network (Facebook, Twitter, etc.) i.e, the user wants to know

how many images or videos have been uploaded in his/her account for a particular time period from their mobile phone. This is a difficult task as there are no analytical tools for performing prediction , decision making etc. for mobile phones.

III. PROPOSED SOLUTION

Having ported the DataNode and TaskTracker processes to work on Android, a Hadoop cluster was configured to run on Android phones. Running Hadoop on a cluster of phones is analogous to running Hadoop on a cluster of servers. In both cases, there is one instance of the NameNode and one instance of the JobTracker. [1] These often run on the same machine. The slave machines in the cluster each run DataNode and TaskTracker instances.

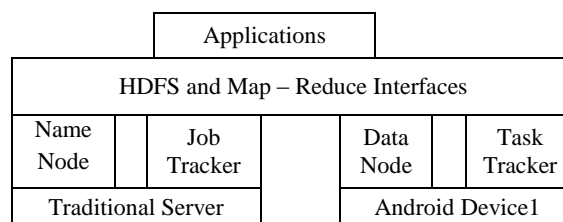


Fig 1 Hadoop on Mobile Cluster

Technique 1 :The First basic step is to Virtualizing the desired user’s Mobile phone locally which helps in retrieving the details of the particular user in case of any lost of Mobile phone.

After Virtualization is done, all the details are saved in a Data Center. Suppose if the user need to view any image he has to type the image location in the search area which is then processed to get the desired image in the mobile screen.

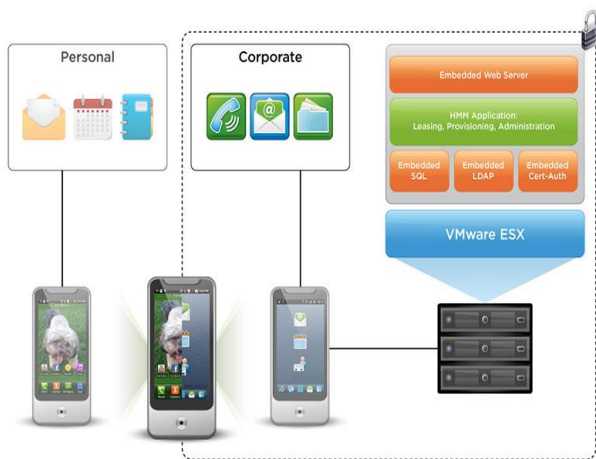


Fig 2 Sample Diagram showing Mobile Phone Virtualization.[2]

Technique 2: The second step is to perform Map Reduce Operations using Hadoop in Mobile Phone and gain insights. This is illustrated as given below:

We propose a system for mobile which captures insights from the huge data sets using Big Data Analytics. The following are some of the steps to be followed :

The Mobile user logs into any of the social networking website using their mobile phone.

The Mobile user wants to know the insights for their huge data sets say images, videos etc. in the mobile.

All the data of that particular user is stored in the Cloud say Mobile Cloud where analytics is performed.

After analytics is done i.e, from Unstructured to Structured, the data is sent in the form of snapshots to the mobile users using Cloudlet.

IV. STEPS IN MOBILE HADOOP

All the operations are done using Hadoop which is uniquely good at scheduled tasks that run at most daily, and are aimed at data that is unstructured or semi-structured, whose information can be represented in traditional databases which are abstractions of the raw data.

Create a front-end database that is derived from Hadoop but is presented in a traditional form.

The process of integrating Hadoop and mobile applications can be broken down into three segments:

- Knowing Hadoop's inherent limitations in mobile use.
- Creating frameworks for Hadoop applications built.
- Troubleshooting Hadoop in mobile applications.

Mobile applications generally should not be designed as a new Hadoop application. [3] Adapting Hadoop to meet the needs of mobile applications entails exploiting existing Hadoop applications through mobile connections.

Hadoop's value is greatest for databases that are 10 to 1,000 times larger than the typical databases used by mobile applications. Data sources outside Hadoop applications also should be examined to see if other information should be

added to improve mobile usability. This involves working backward from mobile user requirements to Hadoop's capabilities and information content.

In order to maximize the amount of information from Hadoop, users should restructure their Hadoop use as a Hive project that automates the process of converting queries from traditional SQL sources into something Hadoop can work with. The data warehouse system creates additional indexes and provides tools for real-time Hadoop access.

HBase is one of the most popular NoSQL databases. Apache Phoenix is a SQL layer on top of HBase to support the most common SQL-like operations such as CREATE TABLE, SELECT, UPSERT, DELETE, etc.

The Mobile clients (virtualized) are connected to find the Private Data Center within the premises in order to work with the data stored in HBase where the regions are the basic elements for distributing tables across the cluster.

Phoenix is implemented as a JDBC driver. A Java client using JDBC can be configured to work with Phoenix Driver and can connect to HBase using SQL-like statements.[4]

V. TESTED ENVIRONMENT AND SAMPLE RESULTS

The sample work is conducted on one Android phone and the results obtained are given below. The result is collected for the three phases for this MapReduce application.

1. Upload: Image path(HDFS)
2. Execute: MapReduce task for the above image.
3. Download : Original image from HDFS

An application was developed to test the activities of Mobile (20 Android and 10 Windows) users which allows users to browse videos and images stored on a network of phones and even sensor data which is connected in the same network for forecasting weather report which is sent to the Regional Weather Forecasting Center. All the Phones are connected to an isolated 802.11g network using a wireless router with no firmware modifications in-order to initialize the cluster, run benchmarks, and collect and post process data . NameNode and JobTracker processes are run on a desktop machine connected behind this router via Ethernet.

Quality based results are obtained by periodically executing a MapReduce job. Requests are serviced by reading results generated by the MapReduce job from HDFS. The client interface is implemented as a web application so that it can be used on mobile devices and desktop machines.

A real – time example of Mobile Hadoop is Hyrax where each machine runs one instance of NameNode and one instance of JobTracker. DataNode and TaskTracker are run on each phone in separate Android service processes within the same application.[5]

VI. TESTED ENVIRONMENT AND SAMPLE RESULTS

Android phone and the results obtained are given below. The result is collected for the three phases for this MapReduce application.

1. Upload: Image path(HDFS)
2. Execute: MapReduce task for the above image.
3. Download: Original image from HDFS

A. SAMPLE RESULTS

Hadoop job _ on Mobile 2007011529_001 run at Localhost

User: Mobile hadoop

Job Name: streamjob2223.jar

Job File: /usr/local/hadoop-datastore/hadoop-hadoop/mapred/system/job_2007011529_001/job.xml

Status: success

Started At: Mon Feb 24 18:05:14 2014

Finished At: Mon Feb 24 18:05:21 2014

Finished In: 7sec

TABLE I: FRAME FORMAT SHOWING MAP – REDUCE TASKS FOR THE ABOVE IMAGE



Kind	Type	Format	No. of Tasks	% Complete	Pending Tasks	Aborted Tasks
map		.jpg	1	100 %	0	0
reduce		.jpg	1	100 %	0	0

TABLE 2: FRAME FORMAT SHOWING MAP – REDUCE TASKS FOR THE ABOVE IMAGE

	Counter	Map	Reduce	Total
Job Counters	Launched Map Tasks	0	0	2
	Launched Reduce Tasks	0	0	1
	Local Data Tasks	0	0	2
Map-Reduce Framework	Map Input Records	12,233	0	12,233
	Map Output Records	25,566	0	25,566
	Map Input	2,13,257	0	2,13,257

	Bytes			
Map Output Bytes	1,10,114	0	1,10,114	
Reduce Input Groups	0	16,234	16,234	
Reduce Input Records	0	25,566	25,566	
Reduce Output Records	0	16,234	16,234	

ACKNOWLEDGMENTS

I thank all my Faculties, Management for giving their support in publishing this paper.

REFERENCES

- [1] “Hyrcas : Cloud Computing on Mobile Devices using MapReduce”, Eugene E. Marinalli , CMU-CS-09-164 , Sep 2009, School Of Computer Science ,Carnegie Mellon University
- [2] “http://www.vmware.com/files/images/screens_mobile/vmw-hrzn-mble-arch-dgrm-800x600.jpg”..
- [3] <http://searchsoa.techtarget.com/tip/Apache-Hadoop-applications-How-to-integrate-Hadoop-for-mobile-use>.
- [4] <http://bighadoop.wordpress.com/>
- [5] <http://ranbijay.blogspot.in/2012/05/hadoop-overview-big-data-control-though.html>