International Journal of Computer Science and Engineering Research and Development (IJCSERD)

ISSN Print: 2248-9363
ISSN Online: 2248-9371

Volume 1 Issue 3, July-December (2014), pp 01-04

IJCSERD

© PRJ PUBLICATION

FILE FORMAT CONVERTER: AVI TO 3GP USING HADOOP MAP REDUCE FRAMEWORK

Dr.R.Renuga
UG HOD, CIT Sandwich Polytechnic College
Civil Aerodrome Post, Coimbatore-641014.

ABSTRACT

Digital data like images, videos contribute the major part of todays life .These data (videos) are available at various formats and in huge size. For the portable usage in many digital devices like pda, mobile phones, the video format has to be converted into a format acceptable by those devices (say 3gp in mobile phones). Conversion from one format to the other, consumes more time. So power of distributed processing is used. Hence HADOOP MAP REDUCE technique performs the conversion operation in less time and in an efficient way.

Keywords
Hadoop distributed file system, Hadoop map reduce, data nodes, name nodes, Tracker.

1. INTRODUCTION

AVI stands for Audio Video Interleave. It has become de-facto standard for storing video and audio information on PC. AVI combines audio and video into a single file in a standard container to allow simultaneous playback. Its advantage is a simple architecture, due to which AVI runs on a number of different systems like Windows, Mac, Linux, Unix and is supported by all of the most popular web-browsers. 3GP is multimedia container formats used for the delivery and playback of audio/video files over high-speed wireless networks, especially mobile cell phones. 3GP file formats were especially designed so as to decrease the bandwidth and storage requirements in order to accommodate mobile cell phones. Structurally, they are based on the ISO file format defined in ISO/IEC 14496-12-MPEG-4 Part 12. The 3GP file can store video streams in MPEG-4 Part 2 (also called H.263) and MPEG-4 Part 10 (AVC/H264) formats, and audio streams in AAC-LC, HE-AAC v1, HE-AAC v2, AMR-NB, and AMR-WB+.

For the portable usage in many digital devices like pda, mobile phones, the video format has to be converted into a format acceptable by those devices (say 3gp in mobile phones). So conversion from AVI to 3GP is necessary to view the videos on the mobile phones.

The file format conversion process in a single machine takes huge time for completion as the file sizes are in GBs. All conversion tools available today take huge time for the conversion. To overcome this problem in the proposed technique, conversion process is done using Hadoop distributed file system using the map reduce technique.

2. METHODOLOGY

Hadoop is a software platform specifically designed to process and handle vast amounts of data. It is based on the principle that moving computation to the place of data is cheaper than moving large data blocks to the place of computation. The Hadoop framework consists of the Hadoop Distributed File System (HDFS) that is designed to run on commodity hardware and MapReduce programming paradigm. HDFS is highly fault-tolerant and is designed to be deployed on low-cost commodity hardware. Hadoop is scalable, economical, efficient and reliable. Hadoop implements Map Reduce, using the HDFS. MapReduce divides applications into many small blocks of work that can be executed in parallel. HDFS creates multiple replicas of data blocks for reliability, placing them on compute nodes around the cluster. MapReduce can then process the data where it is located. HDFS has master/slave architecture.

A HDFS cluster consists of a single NameNode and a number of DataNodes. The NameNode is a master server that manages the file system namespace and regulates access to files by clients. The DataNodes manage storage attached to the nodes that they run on. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system’s clients. MapReduce is a programming paradigm that expresses a large distributed computation as a sequence of distributed operations on data sets of key/value pairs. The Hadoop MapReduce framework harnesses a cluster of machines and executes user defined MapReduce jobs across the nodes in the cluster. A MapReduce computation has a map phase and a reduce phase. The input to the computation is a data set of key/value pairs. In the map phase, the framework splits the input data set into a large number of fragments and assigns each fragment to a map task.

The framework also distributes many map tasks across the cluster of nodes on which it operates (Figure). Each map task consumes key/value (K,V) pairs from its assigned fragment and produces a set of intermediate key/value (K',V') pairs. The framework sorts the intermediate data set by key and each reduce task consumes the fragment of (K',V'*) tuples assigned to it. For each such tuple it invokes a user-defined reduce function that transmutes the tuple into an output key/value pair (K,V). The Hadoop MapReduce framework has a master/slave architecture (Figure 3.9). It has a single master server or jobtracker and several slave servers or tasktrackers, one per node in the cluster. The jobtracker is the point of interaction between users and the framework. Users submit map/reduce jobs to the jobtracker, which puts them in a queue of pending jobs and executes them on a first-come/first-served basis. The jobtracker manages the assignment of map and reduce tasks to the tasktrackers. The tasktrackers execute tasks upon instruction from the jobtracker and also handle data motion between the map and reduce phases.

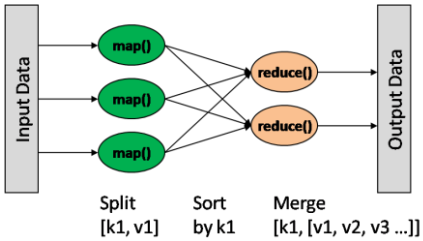

Figure 1 Hadoop Map-Reduce Framework

3. FILE SPLITTING

"Moving the computation is cheaper than moving data", so instead of moving the entire data for conversion, the proposed technique move the process of conversion towards the data.Fig 1.1 shows the architecture of the proposed system. Once a hadoop file system has setup the input AVI file that has to be converted is given to the Map phase. As the Map operation is parallelized the input file is first split to several pieces called FileSplits.

AVI is a derivative of the Resource Interchange File Format (RIFF), which divides a file's data into blocks, or chunks. Each chunk is identified by a FourCC tag. An AVI file takes the form of a single chunk in a RIFF formatted file, which is then subdivided into two mandatory chunks and one optional chunk. The first sub-chunk is identified by the hdrl tag. This sub-chunk is the file header and contains metadata about the video, such as its width, height and frame rate. The second sub-chunk is identified by the movi tag. This chunk contains the actual audio/visual data that make up the AVI movie. The third optional sub-chunk is identified by the idx1 tag which indexes the offsets of the data chunks within the file.

The splitting is done based on various parameters of the input file like file size, configuration etc. Then for each File split a <key,value> pair is generated. Based on this the filespilts are distributed across the data nodes, which are provided with the AVI to 3GP conversion tool.

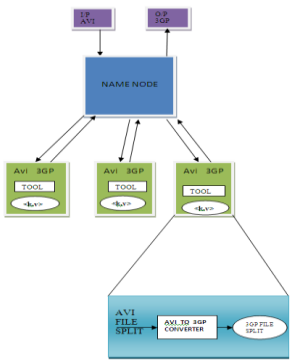

Figure 2 File format conversion using Hadoop

4. CONVERSION

By the end of the map phase the input file is splitted into various chunks and appropriate key-value pair will be available. At this stage the file format is converted to a desire format using the video format conversion tool and stored in the computing nodes itself, now these files forms the input to the next phase of the process.

5. REDUCE PHASE

When a reduce task starts, its input is scattered in many files across all the nodes where map tasks and the conversion tasks ran. These files are then merge sorted so that the key-value pairs for a given key are contiguous. As Sequential Output Format is used , the output key and value classes must also be specified.

6. CONCLUSION

The proposed system uses the hadoop file system-map reduce technique efficiently to convert the given file format (avi to 3gp) in optimized fashion.

REFERENCES

[1] Hadoop 0.20 Documentation

[2] http://hadoop.apache.org/mapreduce/

[3] IEEE paper on ”A Design of Grid Supported Services for Mobile Learning System” by M. Norazizi Sham Mohd Sayuti,Universiti Sains Islam Malaysia (USIM)

[4] “Supplementary Video S1.avi.”

[5] “Information technology. UPnP Device Architecture.”

[6] “IEEE Standard for Advanced Audio and Video Coding.”

[7] M. Camras, “Experiments with electron scanning for magnetic recording and playback of video,” IEEE Transactions on Audio, vol. AU-11, no. 3, pp. 93–96, May 1963.

[8] Koehler, Stephan. *Video Streams in a Computing Grid*. 2010.