

# Analysis of Earthquake Occurrences Using Hadoop

Vikramaditya Pagadala, Ananthi Selvamani, Kislay Jha

Department of Computer Science

University of Bridgeport

Connecticut, United States of America

[vpagadal@my.bridgeport.edu](mailto:vpagadal@my.bridgeport.edu)

[aselvama@my.bridgeport.edu](mailto:aselvama@my.bridgeport.edu)

[kjha@my.bridgeport.edu](mailto:kjha@my.bridgeport.edu)

**Abstract:** - Hadoop has been an effective framework to store and process huge volume of data in a distributed environment. So on the earthquake data set collected from the site U.S Geological Survey for the year 1940 to 2016 this framework is used and reliable outputs are generated from it. The Earthquake data includes various parameters in it so a Map Reduce programming model available in the Hadoop is used on this data set which gives the total number of earthquake occurrences on a given criteria. The output from the Map Reduce program is processed with Hive and Spark techniques from which we can finally get the total number of earthquake occurrences in a specific area.

## 1. Introduction

The U.S Geological Survey site [2] records all the earthquake occurrences which has been occurred for a very long time. These records are a source of valuable information because by analyzing these data we can come to a prediction of an earthquake event before it occurs and what are the necessary precautions to be taken in an area when an earthquake occurs. But the real problem lies with the size of this data. This earthquake data contains various characteristics information of which the most important characteristics are explained below:

- Time – Time when the event has occurred reported in milliseconds.
- Latitude-Longitude - Typical values [-90.0, 90.0] , [-180.0 180.0]
- Depth – Depth of the event in kilometers.
- Mag – Magnitude of the event –typical values [-1.0, 10.0].
- MagType – Preferred magnitude of an event. Typical values are md, mi , ms, mw, me, ml, mb, mlg.
- Gap – The largest azimuthal gap between azimuthally adjacent stations. The smaller

the value the reliable the calculated earthquake position.

- Place- The textual description of the named geographic region near the event.
- Type –Type of Seismic event. Typical values – Earthquake, quarry etc.,

To process these information will be a tedious task as there are various readings recording for a single day. So to process this information in a easier way the hadoop framework is used which uploads these data in to HDFS through which we can process the data available in various clusters and obtain reliable result from it.

## II.Dataset

The Data for this project is taken from the site USGS where the data can be queried for a year and the data will be retrieved in the format of csv. All these data files are combined into one excel file which contains the records from the year 1940 to 2016 January. A screenshot of the data set is shown below:

File Home Insert References Formulas Data Review View Help Tell me what you want to do											
Font		Paragraph		Style		Cells		References		Tools	
Font Face Arial Calibri		Paragraph Left Center Right Justify Indent		Style General		Cells Merge & Center		References Paste Paste Special		Tools	
Font Size 10		Paragraph Alignment Left Center Right Justify		Style Bold Italic Underline		Cells Paste Paste Special		References Paste Paste Special		Tools	
Font Color		Paragraph Spacing		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells		References		Tools	
Font Weight		Paragraph		Style		Cells		References		Tools	
Font Color		Paragraph		Style		Cells		References		Tools	
Font Style		Paragraph		Style		Cells					

	J	K	L	M	N	O	P	Q	R	S	T	U	V				
	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	magType	net	gap	dmin	net	id	updated	place	type	horizontal	depth	error	magNet	status	location	magSource	
2	ml	8	121.15	0.041	nn	nn006604	2016-09-10T04:55:02.380Z	60km NNE earthquake						automatic	nn		
3	ml	12	61.2	0.03746	0.95	ak	ak139904	2016-09-10T04:52:03.631Z	120km NE earthquake	0.3	0.2			automatic	ak		
4	ml	44	64.7999	0.054797	0.55	ak	ak139902	2016-09-10T04:49:58.953Z	76km WSW earthquake	0.5	0.4			automatic	ak		
5	ml	47	154.8	0.257037	0.63	ak	ak139901	2016-09-10T04:45:11.057Z	66km SW earthquake	1.1	0.8			automatic	ak		
6	md	5	196	0.007965	0.08	nc	nc726902	2016-09-10T04:10:31.287Z	60km WNW earthquake	1.3	4.63	0.23		2 automatic	nc		
7	md	29	104	0.07735	0.04	nc	nc726901	2016-09-10T04:10:02.187Z	56km SSE earthquake	0.21	0.89	0.25		25 automatic	nc		
8	Md	19	180	0.043323	0.4	pr	pr362601	2016-09-10T04:10:29.553Z	50km SSE earthquake	0.7	0.6	0		14 reviewed	pr		
9	ml	77	25.2	0.02944	0.72	ak	ak139970	2016-09-10T04:10:29.293Z	430km W earthquake	0.3	0.6			automatic	ak		
10	mb	85	0.062	0.18	us	us100069	2016-09-10T04:10:02.107Z	30km SE of earthquake	0.8	1.1	0.238		5 reviewed	us			
11	md	38	84	0.002897	0.04	nc	nc726900	2016-09-10T04:10:02.145Z	60km WNW earthquake	0.21	0.29	0.08		10 automatic	nc		
12	md	106	1.747	0.95	us	us100068	2016-09-10T04:10:02.003Z	70km SW earthquake	0.6	10.5	0.132		12 reviewed	us			
13	mb	6	239.2	1.32	ak	ak139970	2016-09-10T04:10:02.433Z	130km NE earthquake	1.4	0.9			automatic	ak			
14	md	10	77	0.008023	0.04	nc	nc726900	2016-09-10T04:10:02.118Z	60km NE earthquake	0.24	0.4	0.19		4 automatic	nc		
15	md	19	63	0.05517	0.23	ci	ci176930	2016-09-10T04:10:29.171Z	50km SW of earthquake	0.46	0.99	0.252		22 automatic	ci		
16	md	13	264	0.3764	0.2	ci	ci176930	2016-09-10T04:10:02.536Z	34km SE of earthquake	4.54	6.64	0.085		7 automatic	ci		
17	ml	15	151	0.02891	0.24	ci	ci176930	2016-09-10T04:10:29.611Z	70km SW of earthquake	0.59	0.65	0.149		17 automatic	ci		
18	ml	10	122	0.1225	0.14	ci	ci176930	2016-09-10T04:10:02.360Z	50km WNW earthquake	0.54	1.92	0.09		11 automatic	ci		
19	Md	3	124	0.1299	0.02	pr	pr362601	2016-09-10T04:10:29.069Z	150km S of earthquake	0.6	7.3	0		3 reviewed	pr		
20	ml	26	50.4	0.77	ak	ak139970	2016-09-10T04:10:02.780Z	107km W earthquake	0.2	0.2				automatic	ak		
21	md	11	108	0.007365	0.02	nc	nc726902	2016-09-10T04:10:02.094Z	40km WNW earthquake	0.3	0.44	0.01		4 automatic	nc		
22	ml				0.6	ak	ak139971	2016-09-10T04:10:02.558Z	58km WSW earthquake	0.4	0.7			automatic	ak		
23	ml	13	105	0.1175	0.14	ci	ci176930	2016-09-10T04:10:02.145Z	150km SW earthquake	0.26	0.46	0.089		10 automatic	ci		
24	ml				0.75	ak	ak139970	2016-09-10T04:10:02.579Z	430km W earthquake	0.3	0.4			automatic	ak		
25	md	7	96	0.5977	0.16	nc	nc726901	2016-09-10T04:10:02.294Z	40km SWW earthquake	1.02	13.13	0.2		8 automatic	nc		
26	ml				0.58	ak	ak139971	2016-09-10T04:10:02.229Z	76km W of earthquake	0.4	0.7			automatic	ak		
27	ml				0.72	ak	ak139970	2016-09-10T04:10:02.959Z	70km SWW earthquake	0.4	0.3			automatic	ak		

### III. Data Characteristics

The following are the data characteristics of the earthquake dataset:

#### Time :

Data Type - Long Integer

Description - Time when the event occurred. Times are reported in *milliseconds* since the epoch ( 1970-01-01T00:00:00.000Z), and do not include [leap seconds](#). In certain output formats, the date is formatted for readability.

#### Latitude :

Data Type – Decimal

Typical Values[-90.0, 90.0]

Description - Decimal degrees latitude. Negative values for southern latitudes

#### Longitude :

Data Type – Decimal

Typical Values - [-180.0, 180.0]

Description - Decimal degrees longitude. Negative values for western longitudes.

#### Depth :

Data Type – Decimal

Typical Values - [0, 1000]

Description - Depth of the event in kilometers.

#### Mag:

Data Type – Decimal

Typical Values - [-1.0, 10.0]

Description - The magnitude for the event. Learn more about magnitudes

#### MagType:

Data Type – String

Typical Values - “Md”, “Ml”, “Ms”, “Mw”, “Me”, “Mi”, “Mb”, “MLg”

Description - The method or algorithm used to calculate the preferred magnitude for the event. Learn more about magnitude types.

#### Nst:

Data Type – Integer

Description - The total number of seismic stations used to determine earthquake location.

#### Gap:

Data Type – Decimal

Typical Values - [0.0, 180.0]

Description - The largest azimuthal gap between azimuthally adjacent stations (in degrees). In general, the smaller this number, the more reliable is the calculated horizontal position of the earthquake.

#### Dmin:

Data Type – Decimal

Typical Values - [0.4, 7.1]

Description - Horizontal distance from the epicenter to the nearest station (in degrees). 1 degree is approximately 111.2 kilometers. In general, the smaller this number, the more reliable is the calculated depth of the earthquake.

#### Rms:

Data Type – Decimal

Typical Values - [0.13,1.39]

Description - The root-mean-square (RMS) travel time residual, in sec, using all weights. This parameter provides a measure of the fit of the observed arrival times to the predicted arrival times for this location. Smaller numbers reflect a better fit of the data. The value is dependent on the accuracy of the velocity model used to compute the earthquake location, the quality weights assigned to the arrival time data, and the procedure used to locate the earthquake.

#### Net:

Data Type – String

Typical Value - sak, at, ci, hv, ld, mb, nc, nm, nn, pr, pt, se, us, uu, uw

Description - The ID of a data contributor. Identifies the network considered to be the preferred source of information for this event.

#### Place:

Data Type – String

Description - Textual description of named geographic region near to the event. This may be a city name, or a Flinn-Engdahl Region name.

#### Type:

Data Type – String

Typical Values - “earthquake”, “quarry”  
Description - Type of seismic event.

**HorizontalError:**

Data Type – Decimal

Typical Values - [0, 100]

Description - Uncertainty of reported location of the event in kilometers.

**DepthError:**

Data Type – Decimal

Typical Values - [0, 100]

Description - Uncertainty of reported depth of the event in kilometers.

**MagError:**

Data Type – Decimal

Typical Values - [0, 100]

Description - Uncertainty of reported magnitude of the event.

**MagNst:**

Data Type – Integer

Description - The total number of seismic stations used to calculate the magnitude for this earthquake.

**Status:**

Data Type – String

Typical Values - “automatic”, “reviewed”, “deleted”

Description - Indicates whether the event has been reviewed by a human.

## IV. MapReduce

The research paper is about Map Reduce design[1] and its implementation in processing a large amount of datasets. In this model there are two phase (1) Map phase which creates an intermediate value based on the key/value pair and (2) phase is Reduce phase which uses these intermediate key and merges all its values. The map reduce programs are parallelized and to there are executed on a large group of machines called clusters.

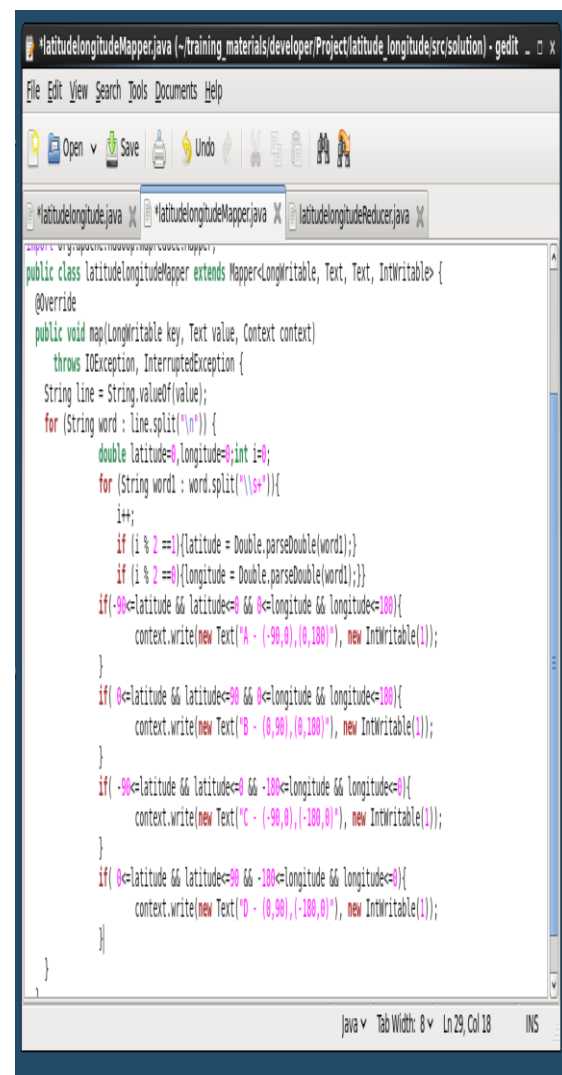
The Map Reduce program is highly scalable and can be executed on a cluster of machines which can handle huge volume of data. For instance when huge amount of data has to be processed on a cluster of machines in case of any failure of the machines in the cluster then the data processing has to be handled properly. This involves some tedious task so to overcome this we have a abstracted technique which is Map Reduce which creates some intermediate key/values in the map phase and in the reduce phase these intermediate key values are merged to get a result. So in this research paper the MapReduce technique is described in detail which involves the implementation of it in large clusters, the working of

its programming model, performance measurements of this technique in various workloads.

The map reduce function first splits the input file into M pieces of size 64mb. The library then starts copies of these splits in many machines. Out of these copies one of them acts as a master which assigns task to other worker machines. The worker process this split and generates key value pairs which are written in the memory. The partitioning function plays a major role in writing this intermediate result to the memory and notify the master regarding this. The master in turn notifies the reducer worker about this location and the reducer uses sort functionality to process this intermediate result and creates a final output. The output result is then appended to the output file. Once the map reduce is done the master notifies the user program about the output file.

The Mapper, Reducer, Driver code for the latitude longitude word count program is shown below:

**Mapper Code:**



```
public class latitudeLongitudeMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    @Override
    public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException {
        String line = String.valueOf(value);
        for (String word : line.split("\n")) {
            double latitude=0, longitude=0; int i=0;
            for (String word1 : word.split("\s+")){
                i++;
                if (i % 2 == 1){latitude = Double.parseDouble(word1);}
                if (i % 2 == 0){longitude = Double.parseDouble(word1);}
            }
            if (-90<latitude && latitude<= 90 && 0<longitude && longitude<=180){
                context.write(new Text("A - (-90,0),(0,180)"), new IntWritable(1));
            }
            if (0<latitude && latitude<=90 && 0<longitude && longitude<=180){
                context.write(new Text("B - (0,90),(0,180)"), new IntWritable(1));
            }
            if (-90<latitude && latitude<=0 && -180<longitude && longitude<=0){
                context.write(new Text("C - (-90,0),(-180,0)"), new IntWritable(1));
            }
            if (0<latitude && latitude<=90 && -180<longitude && longitude<=0){
                context.write(new Text("D - (0,90),(-180,0)"), new IntWritable(1));
            }
        }
    }
}
```

## Reducer Code:

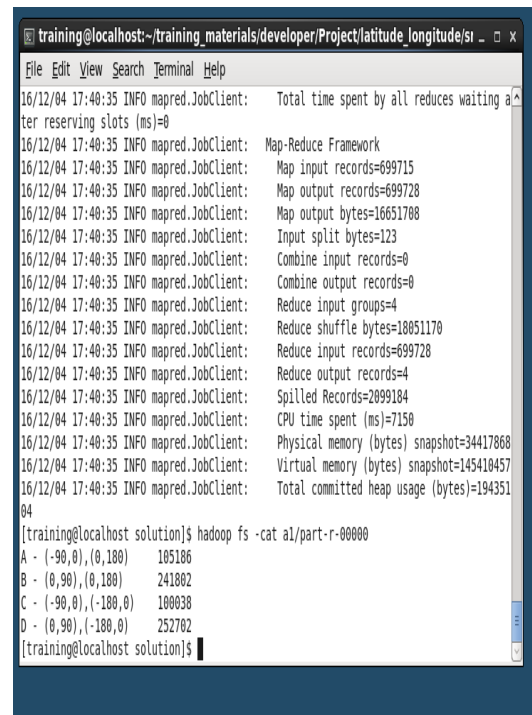


```
File Edit View Search Tools Documents Help
*latitudelongitude.java *latitudelongitudeMapper.java *latitudelongitudeReducer.java

import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.FloatWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

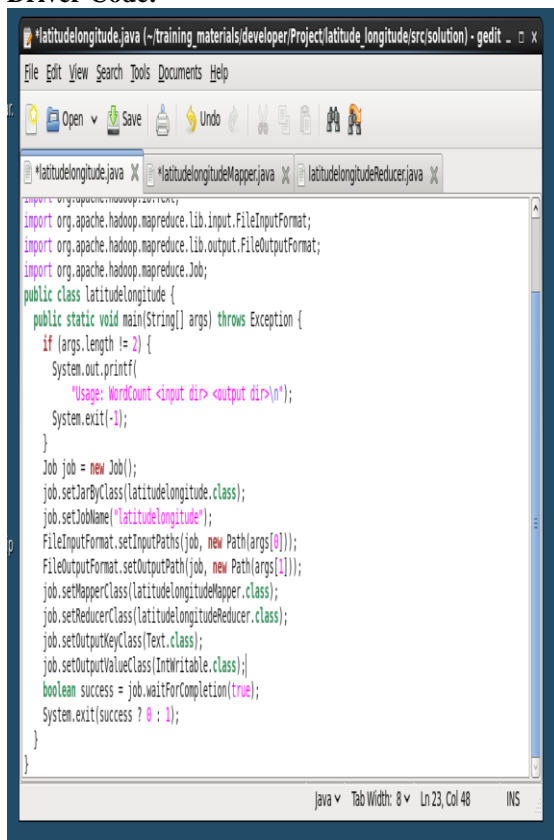
public class latitudelongitudeReducer extends Reducer<Text, IntWritable> {
    @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context context)
        throws IOException, InterruptedException {
        int wordCount = 0;
        for (IntWritable value : values) {
            wordCount += value.get();
        }
        context.write(key, new IntWritable(wordCount));
    }
}
```

The output when the map reduce program runs will be as follows:



```
training@localhost:~/training_materials/developer/Project/latitude_longitude/sr
File Edit View Search Terminal Help
16/12/04 17:40:35 INFO mapped.JobClient: Total time spent by all reduces waiting a
ter reserving slots (ms)=0
16/12/04 17:40:35 INFO mapped.JobClient: Map-Reduce Framework
16/12/04 17:40:35 INFO mapped.JobClient: Map input records=699715
16/12/04 17:40:35 INFO mapped.JobClient: Map output records=699728
16/12/04 17:40:35 INFO mapped.JobClient: Map output bytes=16651788
16/12/04 17:40:35 INFO mapped.JobClient: Input split bytes=123
16/12/04 17:40:35 INFO mapped.JobClient: Combine input records=0
16/12/04 17:40:35 INFO mapped.JobClient: Combine output records=0
16/12/04 17:40:35 INFO mapped.JobClient: Reduce input groups=4
16/12/04 17:40:35 INFO mapped.JobClient: Reduce shuffle bytes=18051170
16/12/04 17:40:35 INFO mapped.JobClient: Reduce input records=699728
16/12/04 17:40:35 INFO mapped.JobClient: Reduce output records=4
16/12/04 17:40:35 INFO mapped.JobClient: Spilled Records=2099184
16/12/04 17:40:35 INFO mapped.JobClient: CPU time spent (ms)=7150
16/12/04 17:40:35 INFO mapped.JobClient: Physical memory (bytes) snapshot=34417868
16/12/04 17:40:35 INFO mapped.JobClient: Virtual memory (bytes) snapshot=145410457
16/12/04 17:40:35 INFO mapped.JobClient: Total committed heap usage (bytes)=194351
04
[training@localhost solution]$ hadoop fs -cat a1/part-r-00000
A - (-90,0),(0,180) 105186
B - (0,90),(0,180) 241802
C - (-90,0),(-180,0) 100038
D - (0,90),(-180,0) 252782
[training@localhost solution]$
```

## Driver Code:

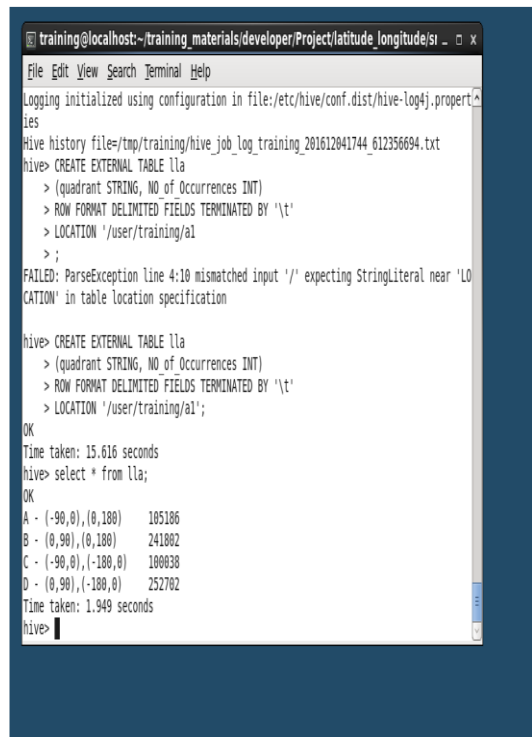


```
File Edit View Search Tools Documents Help
*latitudelongitude.java *latitudelongitudeMapper.java *latitudelongitudeReducer.java

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Reducer;

public class latitudelongitude {
    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.out.printf(
                "Usage: WordCount <input dir> <output dir>\n");
            System.exit(-1);
        }
        Job job = new Job();
        job.setJarByClass(latitudelongitude.class);
        job.setJobName("latitudelongitude");
        FileInputFormat.setInputPaths(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.setMapperClass(latitudelongitudeMapper.class);
        job.setReducerClass(latitudelongitudeReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        boolean success = job.waitForCompletion(true);
        System.exit(success ? 0 : 1);
    }
}
```

The Hive implementation of the same code is shown below:



```
training@localhost:~/training_materials/developer/Project/latitude_longitude/sr
File Edit View Search Terminal Help
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
Hive history file=/tmp/training/hive_job_log_training_201612041744_612356694.txt
hive> CREATE EXTERNAL TABLE lla
> (quadrant STRING, NO of Occurrences INT)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
> LOCATION '/user/training/al'
> ;
FAILED: ParseException line 4:10 mismatched input '/' expecting StringLiteral near 'LO
CATION' in table location specification

hive> CREATE EXTERNAL TABLE lla
> (quadrant STRING, NO of Occurrences INT)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
> LOCATION '/user/training/al';
OK
Time taken: 15.616 seconds
hive> select * from lla;
OK
A - (-90,0),(0,180) 105186
B - (0,90),(0,180) 241802
C - (-90,0),(-180,0) 100038
D - (0,90),(-180,0) 252782
Time taken: 1.949 seconds
hive>
```

## SPARK:

The same work can be run in a SPARK framework which is an application program interface based on a data structure called as resilient distributed dataset(RDD). A spark object has to be created first and in this the map reduce code has be written which will provide the result as key value pairs

```

training@localhost:~
File Edit View Search Terminal Help
Welcome to

      _ _ _ _ _
     / /   \ \
    / /     \ \
   / /       \ \
  / /         \ \
 / /           \ \
/_/             \_\

version 1.3.0

Using Python version 2.7.8 (default, Aug 27 2015 05:23:36)
SparkContext available as sc, HiveContext available as sqlCtx.

In [1]: sc
Out[1]: <pySparkContext object at 0x7f4fec70e1d0>

In [2]: mydata = sc.textFile(
...:     "file:/home/training/training_materials/Project/type_data.txt")
File <ipython-input-2-7543d612f1a0>, line 2
...:     "file:/home/training/training_materials/Project/type_data.txt")
SyntaxError: invalid syntax

In [3]: mydata = sc.textFile(
...:     "file:/home/training/training_materials/Project/type_data.txt")
...:     .flatMap(lambda line: line.split())
...:     .map(lambda word: (word, 1))
...:     .reduceByKey(lambda v1, v2: v1+v2)
16/11/27 13:01:25 INFO storage.MemoryStore: ensureFreeSpace(280171) called with
currentMem=0, maxMem=280248975
16/11/27 13:01:25 INFO storage.MemoryStore: Block broadcast_0 stored as values i
n memory (estimated size 273.6 KB, free 267.0 MB)
16/11/27 13:01:25 INFO storage.MemoryStore: ensureFreeSpace(21284) called with c
urrentMem=280171, maxMem=280248975
16/11/27 13:01:25 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as b

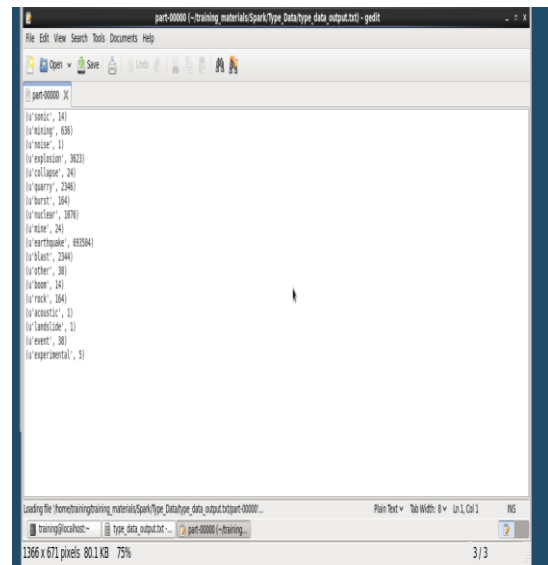
```

```

training@localhost:~
File Edit View Search Terminal Help
n memory (estimated size 273.6 KB, free 267.0 MB)
16/11/27 13:01:25 INFO storage.MemoryStore: ensureFreeSpace(21284) called with c
urrentMem=280171, maxMem=280248975
16/11/27 13:01:25 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as b
ytes in memory (estimated size 28.7 KB, free 267.0 MB)
16/11/27 13:01:25 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in mem
ory on localhost:48036 (size: 28.7 KB, free: 267.2 MB)
16/11/27 13:01:25 INFO storage.BlockManagerMaster: Updated info of block broadca
st_0_piece0
16/11/27 13:01:25 INFO spark.SparkContext: Created broadcast_0 from textFile at
NativeMethodAccessorImpl.java:2
16/11/27 13:01:25 INFO mapred.FileInputFormat: Total input paths to process : 1

In [4]: mydata.saveAsTextFile("file:/home/training/training_materials/Project/ty
pe_data_output.txt")
16/11/27 13:02:29 INFO Configuration.deprecation: mapred.tip.id is deprecated. I
nstead, use mapreduce.task.id
16/11/27 13:02:29 INFO Configuration.deprecation: mapred.task.id is deprecated.
Instead, use mapreduce.task.attempt.id
16/11/27 13:02:29 INFO Configuration.deprecation: mapred.task.is.map is deprecate
d. Instead, use mapreduce.task.ismap
16/11/27 13:02:29 INFO Configuration.deprecation: mapred.task.partition is depre
cated. Instead, use mapreduce.task.partition
16/11/27 13:02:29 INFO Configuration.deprecation: mapred.job.id is deprecated. I
nstead, use mapreduce.job.id
16/11/27 13:02:29 INFO output.FileOutputCommitter: File Output Committer Algorit
hm version is 1
16/11/27 13:02:29 INFO spark.SparkContext: Starting job: saveAsTextFile at Nativ
eMethodAccessorImpl.java:2
16/11/27 13:02:29 INFO scheduler.DAGScheduler: Registering RDD 4 (reduceByKey at
<ipython-input-3-4b5f94a8b6b4>:5)
16/11/27 13:02:29 INFO scheduler.DAGScheduler: Got job 0 (saveAsTextFile at Nati
veMethodAccessorImpl.java:2) with 1 output partitions (allowLocal=false)

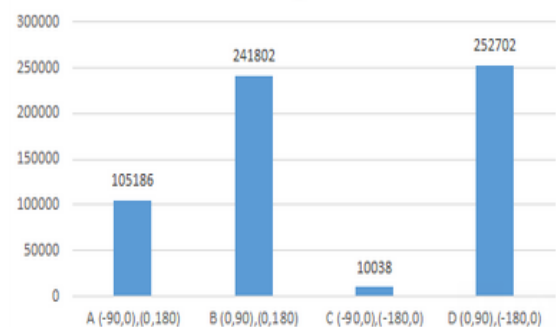
```



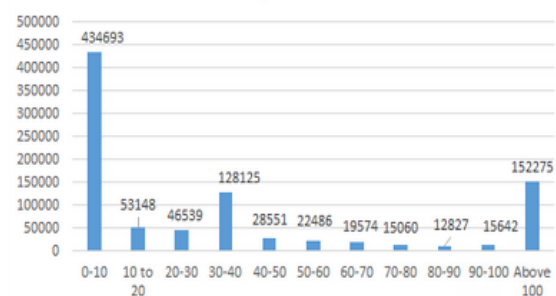
## V. Results and Evaluation

The above mentioned Map Reduce Program can be run on various characteristics of the Earthquake dataset and the result retrieved from it is presented below:

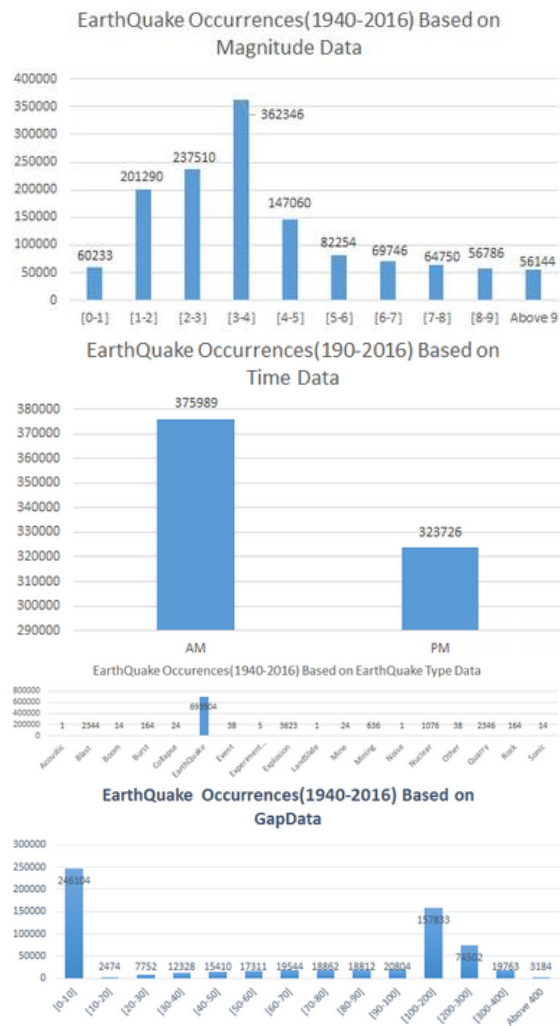
Earthquake Occurrences(1940-2016) based on Latitude Longitude Data



Earthquake Occurrences(1940-2016) Based on Depth Data







## VI. Conclusion:

As observed from the output of the above map reduce code we conclude that hadoop framework can be used to store and process data of huge volume. The output from the map reduce code will be key value pairs which are used to find some pattern in the output and with this pattern we predict the occurrences of earthquake in a given place.

## References:

1. *Jeffrey Dean and Sanjay Ghemawat*  
MapReduce : Simplified Data  
Processing on Large Clusters
2. *U.S Geological Survey* <https://www.usgs.gov/>