Rainfall data analysis and storm prediction system

SHABARIRAM, M. E.

Published version

SHABARIRAM, M. E. (2017). Rainfall data analysis and storm prediction system. In: Computational Intelligence for Societal Development in Developing Countries (CISDIDC), Sheffield Hallam University, 17 February 2017. (Unpublished)

Rainfall data analysis and Storm Prediction System

Presented by

C.P.Shabariram M.E.,

Assistant Professor

Problem Analysis

- The main problem is that big rainfall data stored in relational database is as input.
- System is implemented on Graph search which involves multiple scan of same data.
- Finally the system is run a single server without applying any distributed technology.
- The main objective is that preprocessing technique is used to filter the unnecessary rainfall data and analyzing only the meaningfull data.

Abstract

- Rainfall fall is collected to predict the storm warning from the hydrological data. This is considered as research idea as it consumes large number of records from the distributed systems. In this work, we proposed a novel solution to manage the data based on spatial temporal characteristic and Map Reduce Framework. The Work load is classified using Support Vector Machine to initialize the Map and Reduce function. It uses the feature selection and reduction algorithm to extract feature entity attribute. Various Rainstorm concepts prediction achieved using big raw rainfall data . Three concepts are defined local, hourly and overall storms. The proposed system serves as a tool for predict rain storm from large amount of rainfall data in effective manner. This system improves the performance in terms of accuracy and efficiency.

SYSTEM REQUIREMENTS

Hardware Requirements:

Processor	:	Intel Pentium i3
RAM	:	2GB
Hard Disk	:	Minimum 2GB free space

Software Requirements:

Operating System	:	Windows 8.1
Tool	:	Hadoop
IDE	:	Eclipse
Language	:	Java

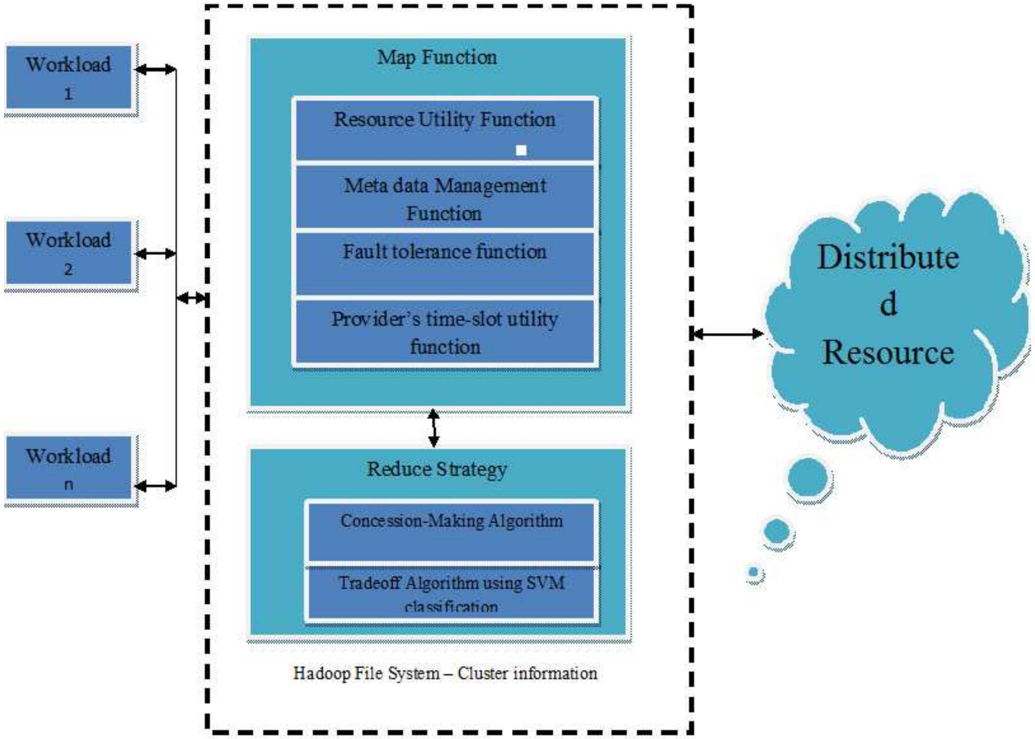
Literature Survey

Paper Tile	Author	Description	Advantages	Disadvantages
1. Using Mapreduce to Speed Up Storm Identification from Big Raw Rainfall Data	K. Jitkajornwanich, U. Gupta, R. Elmasri, L. Fegaras, and J.McEnery	Relevant storm characteristics is identified from raw rainfall data. original raw rainfall data text files instead of using the data in the relational database.	The performance of the new storm identification system is significantly improved, based on previous one	parallelization of computation in storm identification based on area and centre .
2. Simplified Data Processing on Large Clusters	J. Dean and S. Ghemawat	Implementation of Mapreduce runs on a large clusters of commodity machine	The runtime system takes care of program execution, (ie) handling errors	Network bandwidth is a scare source
3. Rainfall Depth-Duration-Frequency Curves and Their Uncertainties	A. Overeem, T. A. Buishand, and I. Holleman	effects of dependence between the maximum rainfalls for different durations on the estimation of DDF curves	It is used as Statistical literature for the estimating the maximum rainfall region	Large samples are needed to estimate this shape parameter accurately or data from several sites in a region

continued

4. Experiences on Processing Spatial Data with Mapreduce	**A. Cary, Z. Sun, V. Hristidis, and N. Rishe**	**It describes the problems of r-tree, which is used as spatial access methods**	**Computation is improved, that leads to high linear scalability.**	**It is not applied to high complex spatial problem**
5. Statistical Characteristics of Storm Inter event Time, Depth and Duration for Eastern New Mexico, Oklahoma and Texas	W. H. Asquith, M. C. Roussel, T. G. Cleveland, X. Fang, and D. B.Thompson,	The analysis is based on hourly rainfall data recorded by NWS	It helps to find the storm inter event time and duration	It is not suitable for specifying all sites in a particular location

SYSTEM ARCHITECTURE



PROPOSED SYSTEM

- The reduction in number of records allows faster querying and mining of storm data.
- The framework is compatible with the original location-specific analysis of storms
- It helps the hydrologists by helping them to analyze data easier more efficiently.

Modules

- Modelling the Mapreduce framework for Task Processing.
- Classification of the data to mapper phase process based on the Spatial Temporal Characteristics using SVM.

Modeling the MapReduce for Taskprocessing

- MapReduce is a programming Model that is associated with rainfall data for task processing and generating data.
- The computation takes a set input key/value pairs and produces a set of output key/value pairs.
- Map method includes

❖ Resource Utility function

❖ Metadata Management function

❖ Fault tolerance function

❖ Providers Time slot utility function

- Reduce Method includes

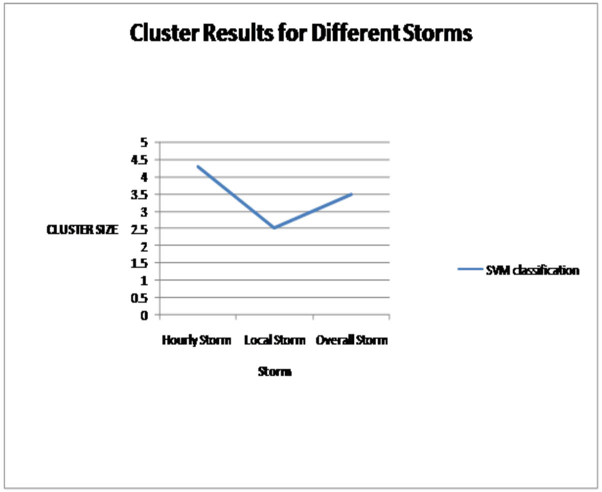
❖ Concession Making algorithm

Classification of the data to mapper phase process based on the Spatial Temporal Characteristics using SVM

- Map Process takes caries of the partitioning the spatial data of the rainfall data.
- The support vector machine is used as a data mining technique to extract informative hydrologic
- Various percentages (from 50% to 10%) of hydrologic data, including those for flood stage and rainfall data, were mined and used as informative data to characterize a flood indicated attributes.

PERFORMANCE EVALUVATION

- The evaluation is based on storms stored in different clusters.
- It describes the cluster size of the each storm class during the SVM classification with class boundaries containing the threshold limits and state values .


Cluster Results for Different Storms
CLUSTER SIZE
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Hourly Storm
Local Storm
Overall Storm
Storm
SVM classification

Hadoop Installation

localhost Hadoop Map/Reduce... HDFS:/user/rainstrom/output/...

localhost:50030/jobtracker.jsp

Quick Links

localhost Hadoop Map/Reduce Administration

State: RUNNING
Started: Thu Nov 26 22:47:30 IST 2015
Version: 0.19.1, r745977
Compiled: Fri Feb 20 00:16:34 UTC 2009 by ndaley
Identifier: 201511262247

Cluster Summary

Maps	Reduces	Total Submissions	Nodes	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node
0	0	0	1	2	2	4.00

Scheduling Information

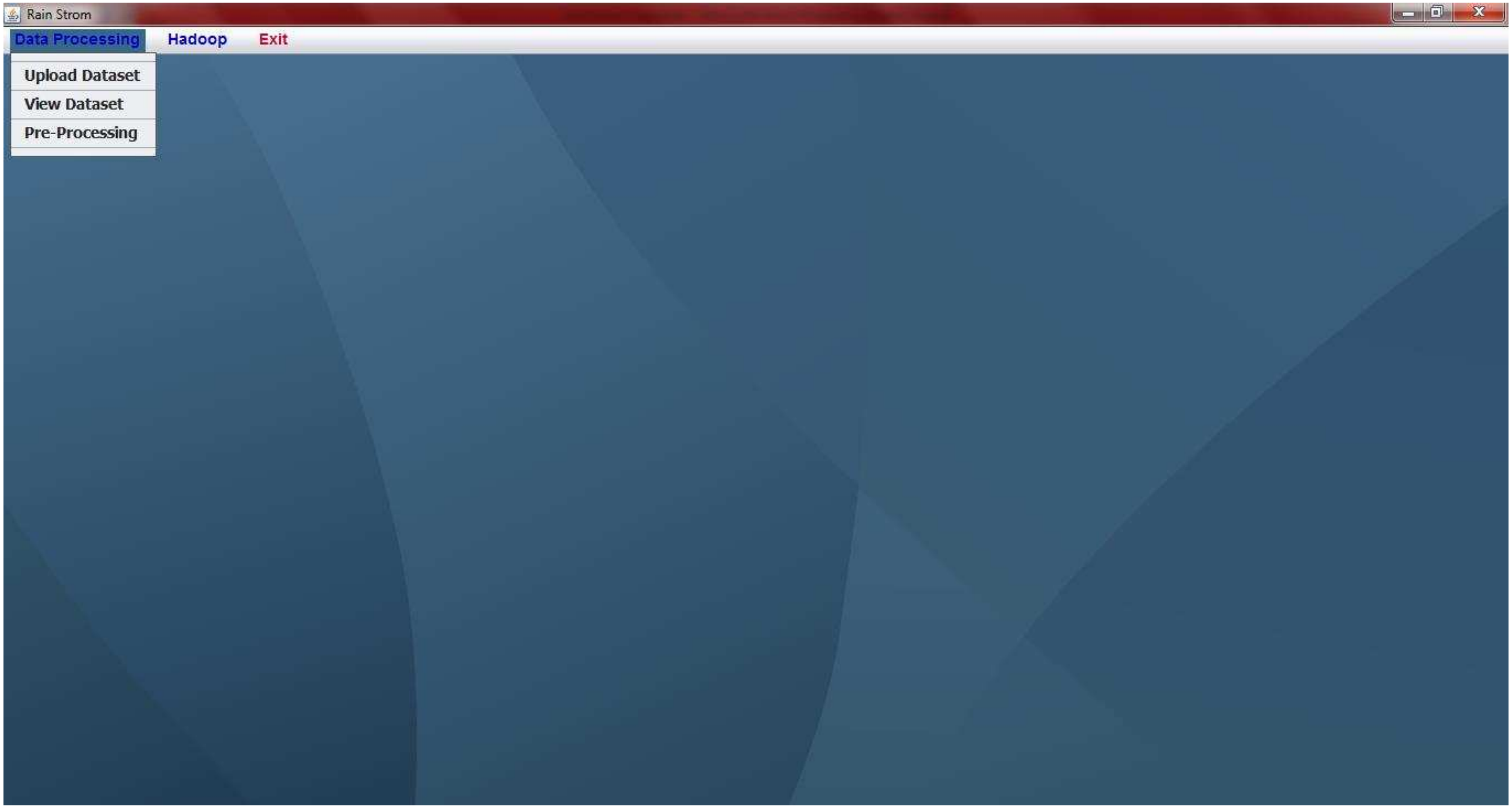
Queue Name	Scheduling Information
default	N/A

Filter (Jobid, Priority, User, Name)
Example: 'user:smith 3200' will filter by 'smith' only in the user field and '3200' in all fields

Running Jobs

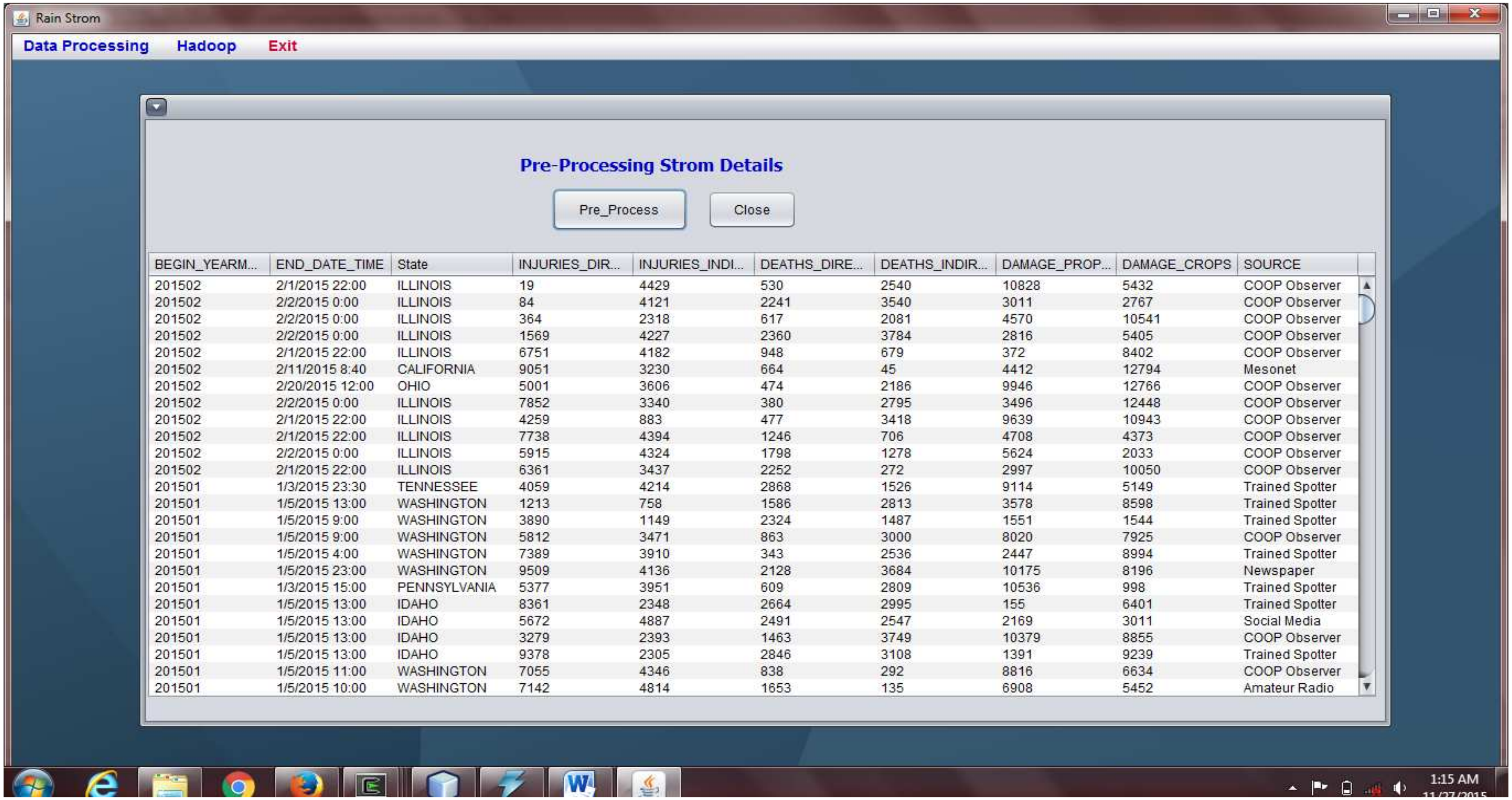
none

MainPage



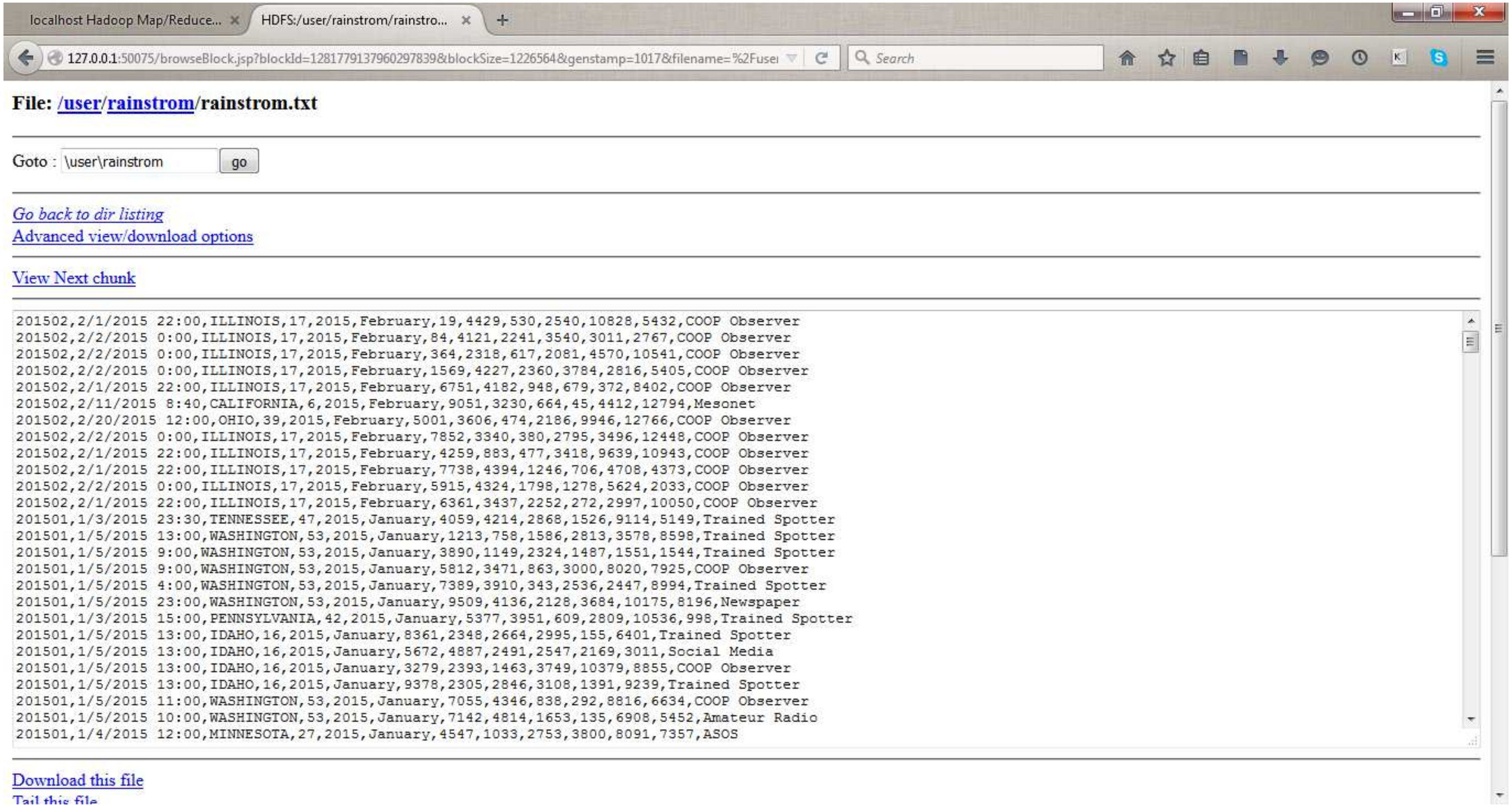
Rain Strom
Data Processing
Hadoop
Exit
Upload Dataset
View Dataset
Pre-Processing

DataPreprocessing

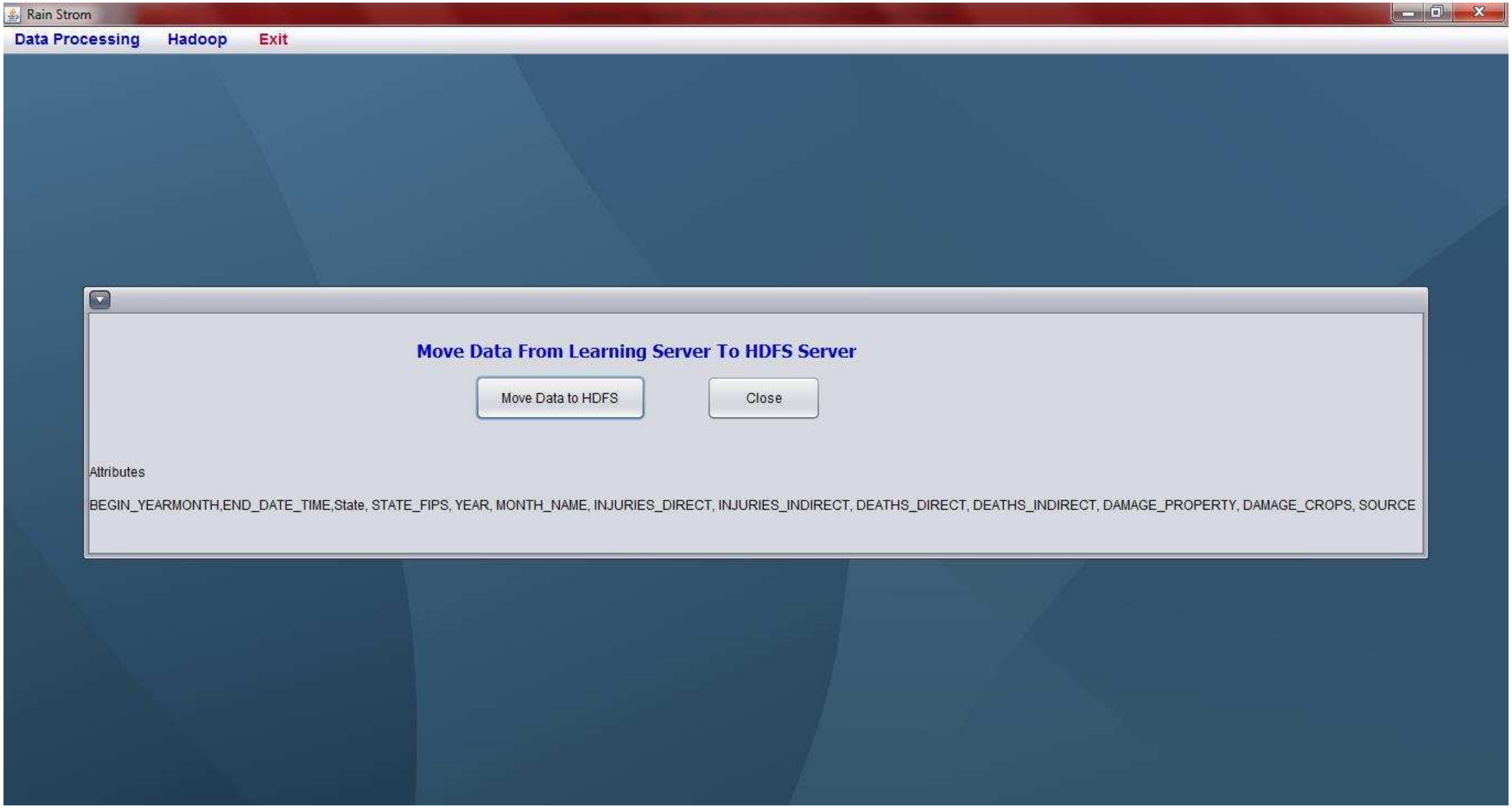


BEGIN_YEARM...	END_DATE_TIME	State	INJURIES_DIR...	INJURIES_INDI...	DEATHS_DIRE...	DEATHS_INDIR...	DAMAGE_PROP...	DAMAGE_CROPS	SOURCE
201502	2/1/2015 22:00	ILLINOIS	19	4429	530	2540	10828	5432	COOP Observer
201502	2/2/2015 0:00	ILLINOIS	84	4121	2241	3540	3011	2767	COOP Observer
201502	2/2/2015 0:00	ILLINOIS	364	2318	617	2081	4570	10541	COOP Observer
201502	2/2/2015 0:00	ILLINOIS	1569	4227	2360	3784	2816	5405	COOP Observer
201502	2/1/2015 22:00	ILLINOIS	6751	4182	948	679	372	8402	COOP Observer
201502	2/11/2015 8:40	CALIFORNIA	9051	3230	664	45	4412	12794	Mesonet
201502	2/20/2015 12:00	OHIO	5001	3606	474	2186	9946	12766	COOP Observer
201502	2/2/2015 0:00	ILLINOIS	7852	3340	380	2795	3496	12448	COOP Observer
201502	2/1/2015 22:00	ILLINOIS	4259	883	477	3418	9639	10943	COOP Observer
201502	2/1/2015 22:00	ILLINOIS	7738	4394	1246	706	4708	4373	COOP Observer
201502	2/2/2015 0:00	ILLINOIS	5915	4324	1798	1278	5624	2033	COOP Observer
201502	2/1/2015 22:00	ILLINOIS	6361	3437	2252	272	2997	10050	COOP Observer
201501	1/3/2015 23:30	TENNESSEE	4059	4214	2868	1526	9114	5149	Trained Spotter
201501	1/5/2015 13:00	WASHINGTON	1213	758	1586	2813	3578	8598	Trained Spotter
201501	1/5/2015 9:00	WASHINGTON	3890	1149	2324	1487	1551	1544	Trained Spotter
201501	1/5/2015 9:00	WASHINGTON	5812	3471	863	3000	8020	7925	COOP Observer
201501	1/5/2015 4:00	WASHINGTON	7389	3910	343	2536	2447	8994	Trained Spotter
201501	1/5/2015 23:00	WASHINGTON	9509	4136	2128	3684	10175	8196	Newspaper
201501	1/3/2015 15:00	PENNSYLVANIA	5377	3951	609	2809	10536	998	Trained Spotter
201501	1/5/2015 13:00	IDAHO	8361	2348	2664	2995	155	6401	Trained Spotter
201501	1/5/2015 13:00	IDAHO	5672	4887	2491	2547	2169	3011	Social Media
201501	1/5/2015 13:00	IDAHO	3279	2393	1463	3749	10379	8855	COOP Observer
201501	1/5/2015 13:00	IDAHO	9378	2305	2846	3108	1391	9239	Trained Spotter
201501	1/5/2015 11:00	WASHINGTON	7055	4346	838	292	8816	6634	COOP Observer
201501	1/5/2015 10:00	WASHINGTON	7142	4814	1653	135	6908	5452	Amateur Radio

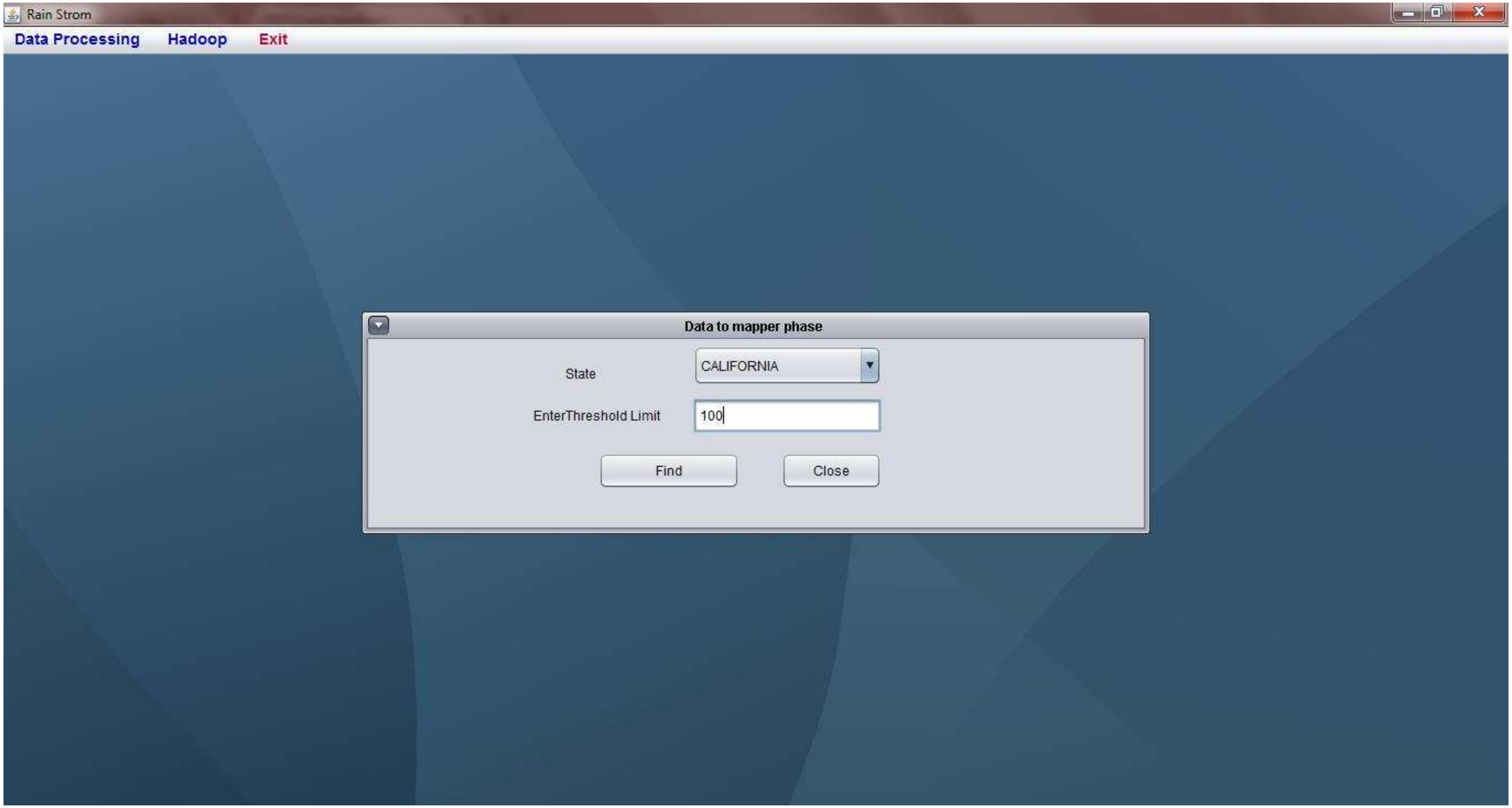
Analysing of Preprocessed Data as TextFiles



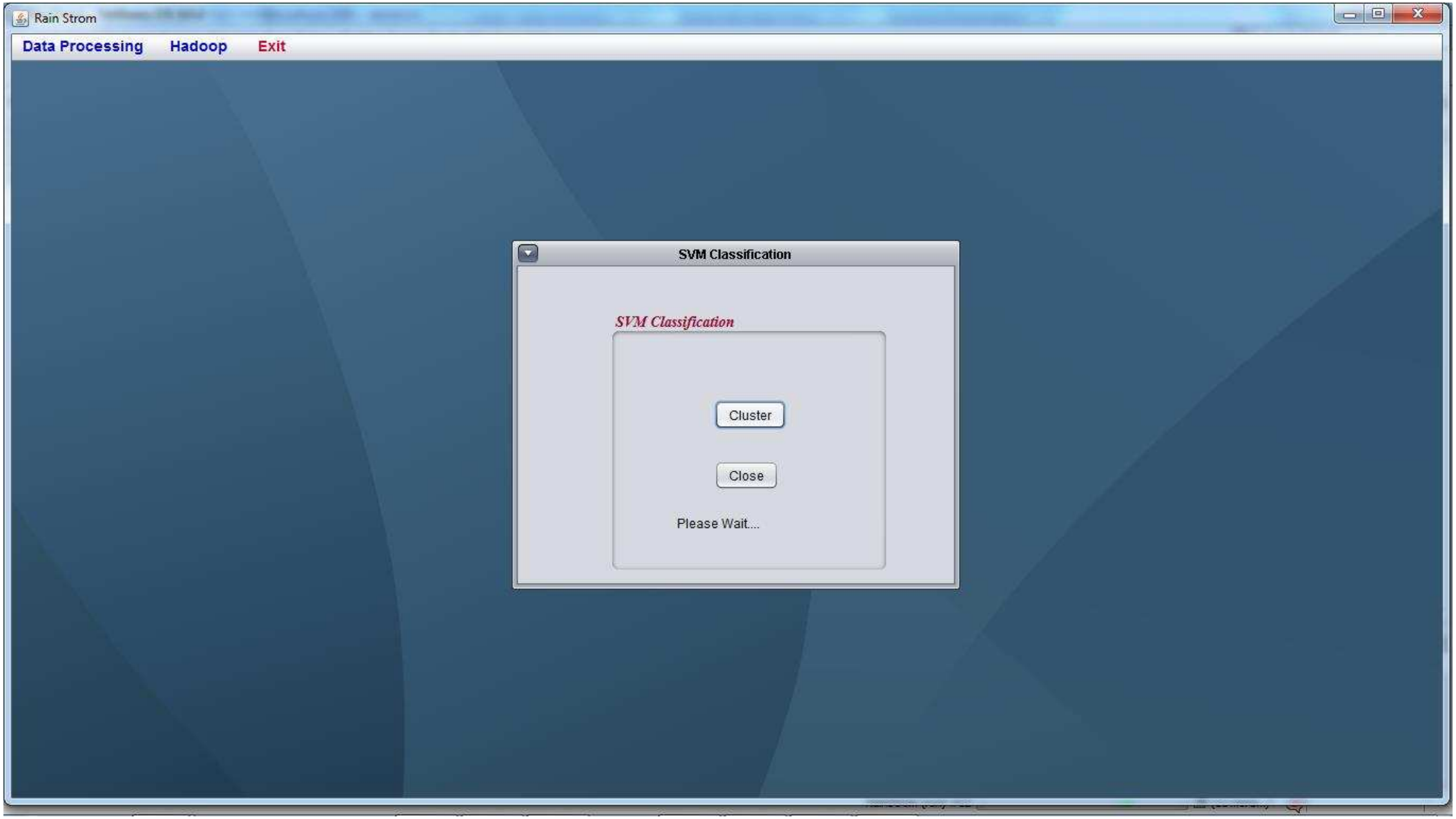
Moving data to HDFS server



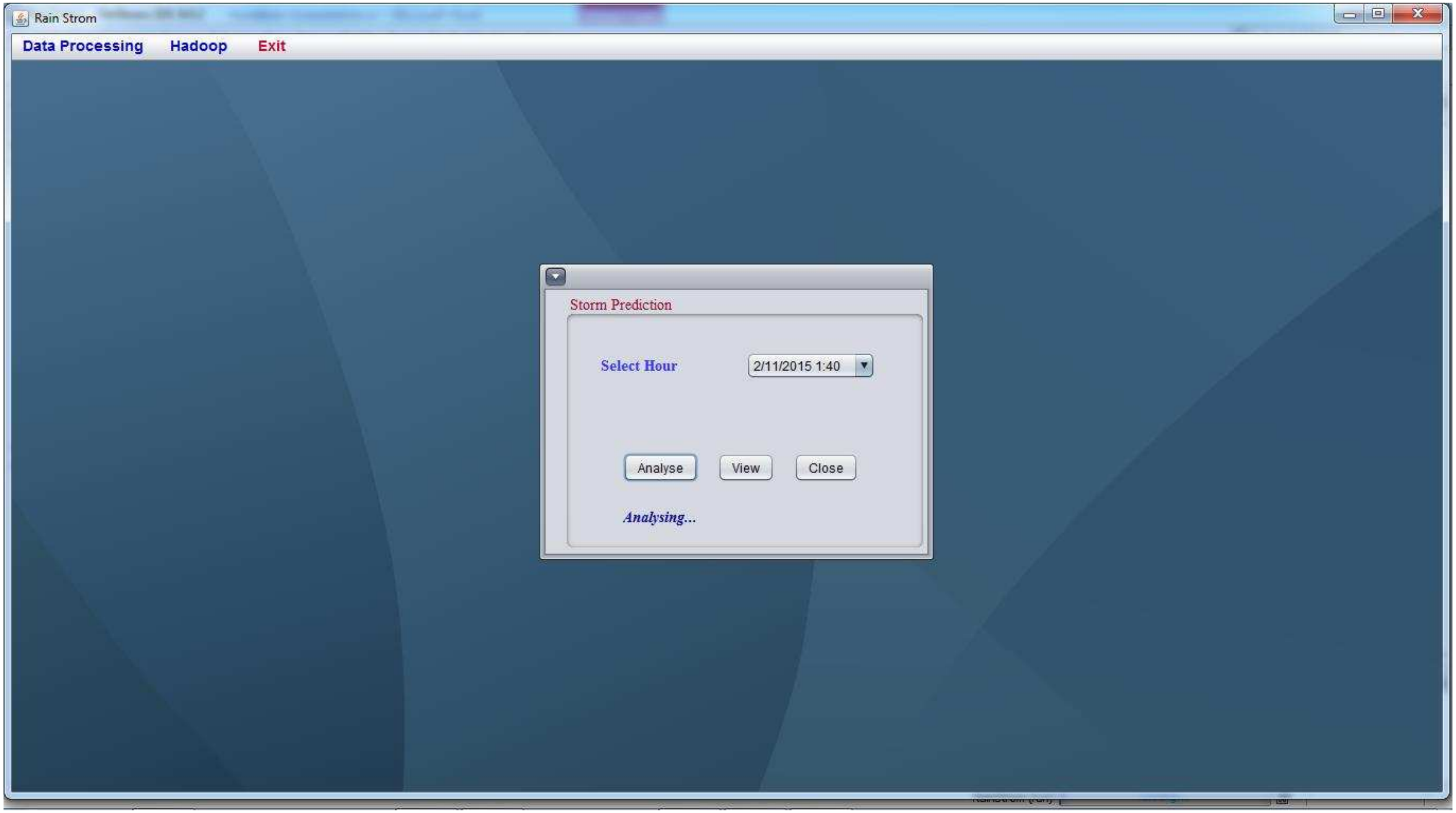
LOCAL STORMS



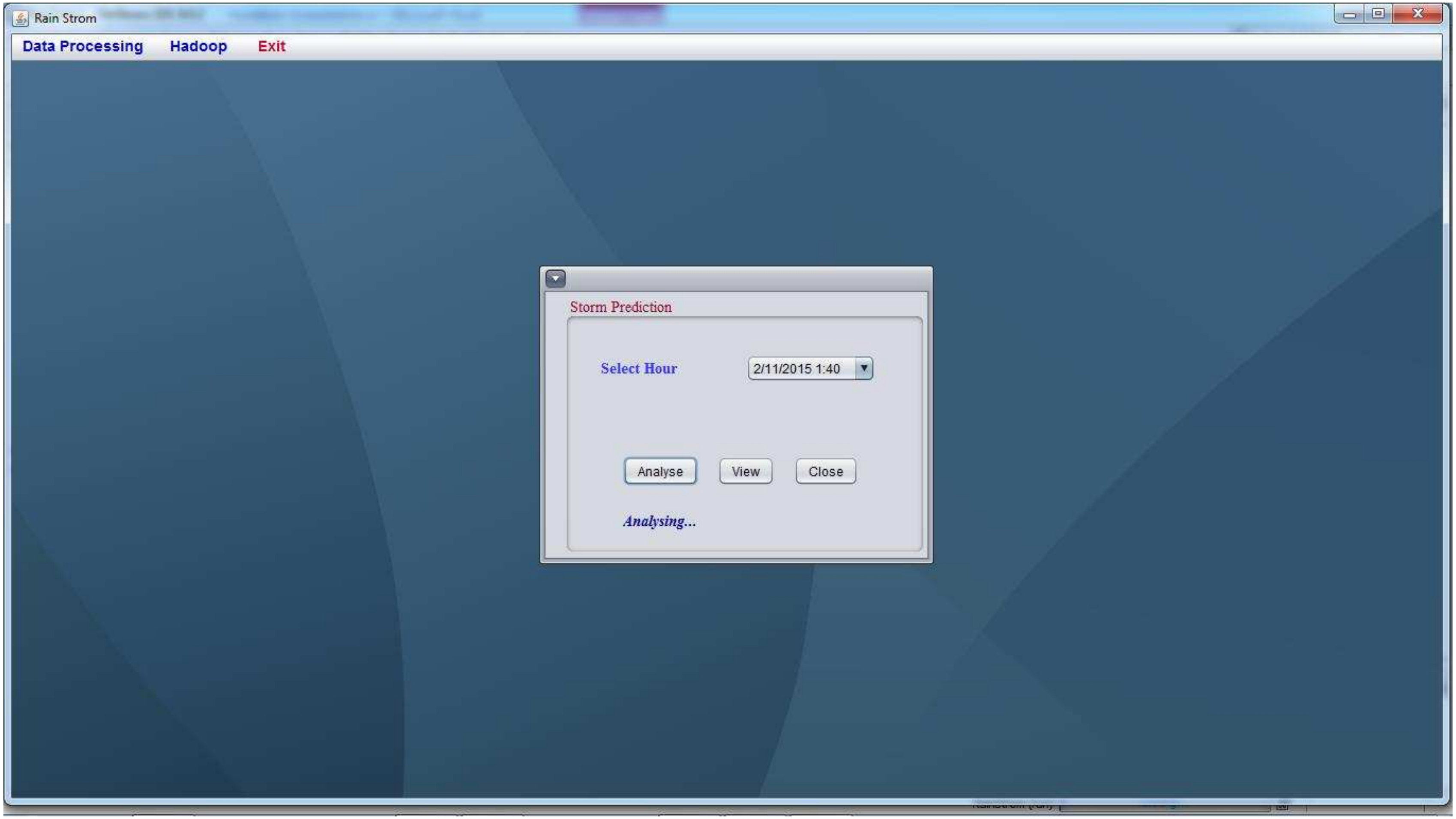
CLUSTER CLASSIFICATION USING SVM



HOURLY STORMS



OVERALL STORMS



ANALYSING OF OVERALL STORMS

After Analysing Records

BEGIN_Y...	BEGIN_D...	BEGIN_TI...	END_YEA...	END_DAY	END_TIME	EPISODE...	EVENT_ID	STATE	STATE_FI...	YEAR	MONTH_...	EVENT_T...	CZ_TYPE	CZ_FIPS	CZ_NAME	WFO	BEGIN_D...
201502	1	200	201502	1	2200	1417343	560121	ILLINO...	17	2015	Febru...	Winter ...	Z	47	Cass	ILX	2/1/201...
201502	1	200	201502	1	2200	1417343	560111	ILLINO...	17	2015	Febru...	Heavy ...	Z	36	Fulton	ILX	2/1/201...
201502	1	200	201502	1	2200	1417343	560125	ILLINO...	17	2015	Febru...	Winter ...	Z	41	Mason	ILX	2/1/201...
201502	1	200	201502	1	2200	1417343	560126	ILLINO...	17	2015	Febru...	Winter ...	Z	48	Menard	ILX	2/1/201...
201502	1	200	201502	1	2200	1417343	560128	ILLINO...	17	2015	Febru...	Winter ...	Z	40	Schuyl...	ILX	2/1/201...
201502	1	200	201502	1	2300	1417343	560117	ILLINO...	17	2015	Febru...	Heavy ...	Z	29	Peoria	ILX	2/1/201...
201502	1	200	201502	1	2200	1417343	560112	ILLINO...	17	2015	Febru...	Heavy ...	Z	27	Knox	ILX	2/1/201...

CONCLUTION

- The design and implementation a storm classification mechanism using SVM classification
- The data classification is carried in the map reduce paradigm using Hadoop framework.
- As the dataset is available in large scale and hence to improve the performance of the cluster scalability, it has been utilized and classify the rainfall data into cluster using the mapper and reduce functions.

FUTURE WORK

- The challenge to proposed system is to guarantee the quality of discovered relevance features in rainfall dataset for describing storm prediction large scale terms and data patterns.
- Most popular classification methods have adopted term-based approaches suffered from the problems of feature evolution.
- It discovers rainfall conditions as higher level features and deploys them over low-level features.

References

- K. Jitkajornwanich, R. Elmasri, C. Li, and J. McEnery, “Extracting Storm-Centric Characteristics from Raw Rainfall Data for Storm Analysis and Mining,” Proceedings of the 1st ACM SIGSPATIALInternational Workshop on Analytics for Big Geospatial Data (ACM SIGSPATIAL BIGSPATIAL’12), 2012, pp. 91-99.

- K. Jitkajornwanich, U. Gupta, R. Elmasri, L. Fegaras, and J.McEnery, “Using MapReduce to Speed Up Storm Identification from Big Raw Rainfall Data,” Proceedings of the 4th International Conference on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING’13), 2013, pp. 49-55.

- J. Dean and S. Ghemawat, “MapReduce: Simplified Data Processing on Large Clusters,” Proceedings of the 6th Symposium on Operating Systems Design and Implementation (OSDI’04), 2004

Continued..

- A. Overeem, T. A. Buishand, and I. Holleman, “Rainfall Depth-Duration-Frequency Curves and Their Uncertainties,” Journal of Hydrology, vol. 348, 2008, pp. 124-134.

- W. H. Asquith, M. C. Roussel, T. G. Cleveland, X. Fang, and D. B.Thompson, “Statistical Characteristics of Storm Interevent Time,Depth, and Duration for Eastern New Mexico, Oklahoma, and Texas,” Professional Paper 1725. U.S. Geological Survey, 2006.

- W. H. Asquith, “Depth-Duration Frequency of Precipitation for Texas,” Water-Resources Investigations Report 98-4044. U.S.Geological Survey (USGS), 1998.

THANK YOU