Seasonal Ensemble Forecasting Application on SuMegha Scientific Cloud Infrastructure

Ramesh Naidu Laveti
B. B. Prahlada Rao, Vineeth Simon, Arunachalam B

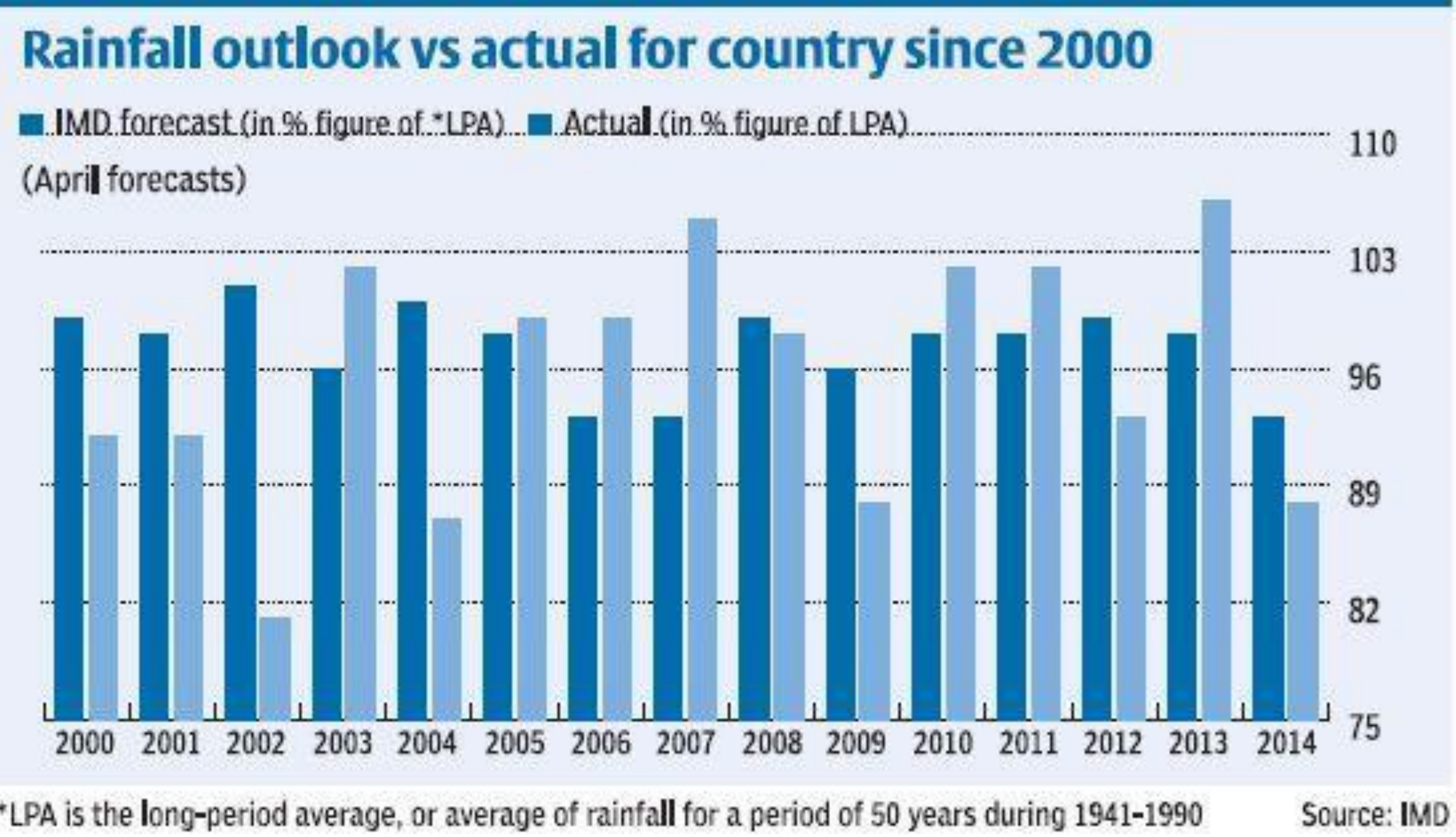
Centre for Development of Advanced Computing (C-DAC), Bangalore, India

Contact: rameshl@cdac.in

Outline

1. Introduction
2. Portable design of Seasonal Forecast Model
3. The SuMegha Cloud
4. Implementation of SFM on SuMegha
5. Discussions and Results
6. Conclusion

National Monsoon Mission



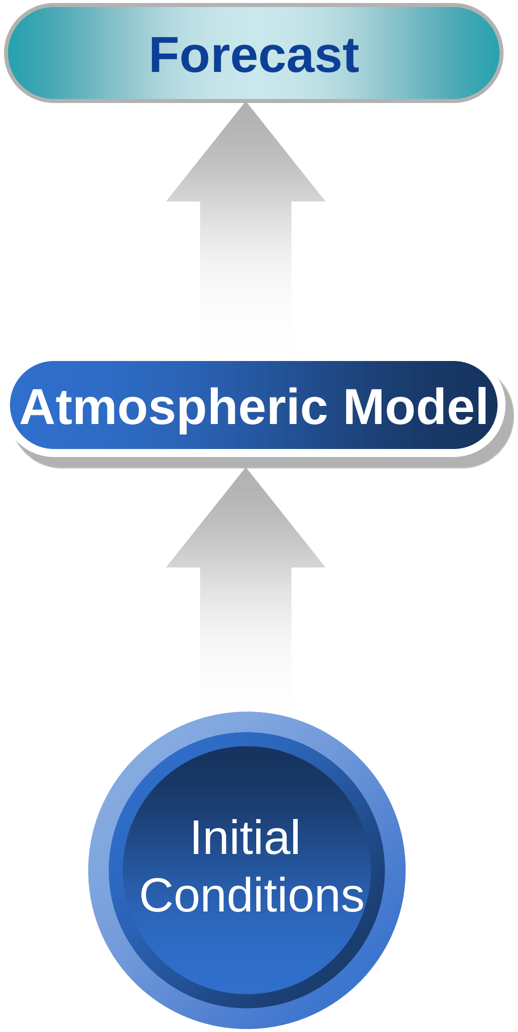
Rainfall outlook vs actual for country since 2000
IMD forecast (in % figure of *LPA)
Actual (in % figure of LPA)
(April forecasts)
110
103
96
89
82
75
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
*LPA is the long-period average, or average of rainfall for a period of 50 years during 1941-1990
Source: IMD

1. Introduction

Atmospheric model - Atmospheric models are numerical representations of various parts of the Earth's atmospheric system. Used to produce past, present and future state of the atmosphere.

Atmospheric system - The incoming and outgoing radiation, the way the air moves, the way clouds form and precipitation (rain) falls, the way the ice sheets grow or shrink, etc.

Forecast - It is an estimation of future state of the atmosphere by estimating the current state and then calculating how this evolve with time. Need to be done with high accuracy and speed – Best forecast



Ensemble forecasting

Can we accurately forecast the evolution of the atmospheric system?

Error in initial State of atmos. → **Chaotic atmospheric System** → Forecast uncertainties

Why do we need ensemble forecasting?

Butterfly effect - A small error in its present state can lead to large differences in its future state

What is ensemble forecasting?

"It consists of a number of simulations made by making small changes to the estimate of the current state which is used to initialize the simulation or making small changes to the model parameters/physics"

Parallelism in Ensemble forecasting

Ensemble forecasting on Cloud

The ensemble forecasting problem can be seen as a set of independent tasks, each task can run on a seperate clsuter or node independently

2. Seasonal Forecast Model (SFM)

Introduction

- Atmospheric General Circulation Model (AGCM) designed for seasonal prediction and climate research
- Available for research communities under research licence
- It can run as
  - Sequential
  - Shared Memory parallel (OpenMP)
  - Distributed Memory parallel (MPI)
- It can fit on HPC clusters, Grid and Cloud

Components of SFM

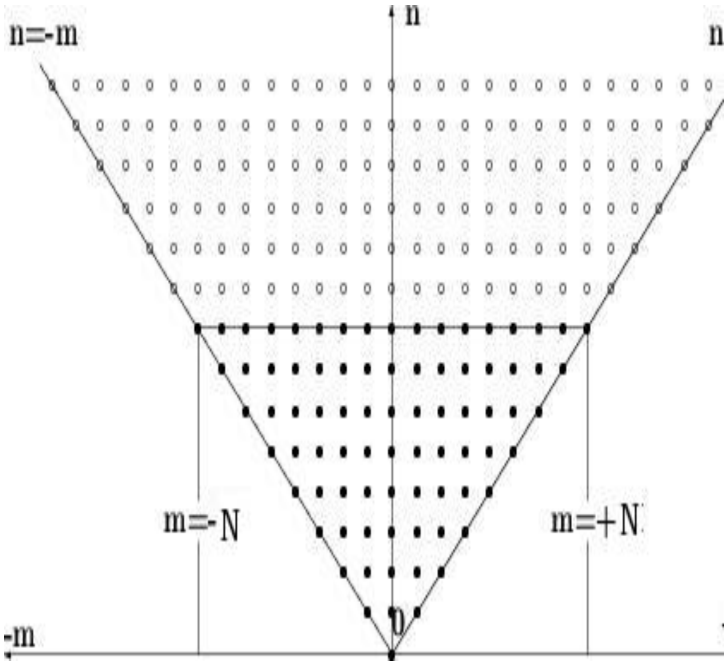
- **Libs**
  - It contains model libraries, utilities and climatological constant fields
  - It also have machine dependant and resolution independent sources
  - It is **fixed for a particular machine**, we should build it once for a particular machine
- **Model Source**
  - Contain model source codes and define model resolution and options
  - Used to create model executable
  - It is **resolution dependent**
- **Run**
  - Contain run scripts to run the model and it runs the model and stores the output
  - Allows to do different experiments for **different run lengths** (Forecast Lengths)

Portability details

- **Parallelization strategy used in SFM**
  - It uses 2-D decomposition method
  - So, flexible to run any number of processors (Except a prime number)
  - Gives best performance if we choose the number of processors as $2^p \times 3^q \times 5^r$
- **Portability**
  - Can run as sequential or parallel
  - It can run on multiple platforms – CRAY, SGI, SUN, IBMSP
  - A good application to start with, on grid
- **Supports hybrid computing**
  - It can be run as hybrid – OpenMP + MPI

Resolution details

	Low resolution	High resolution
Truncation	T62	T320
Longitudes	192	972
Latitudes	94	486
Vertical levels	28	42
Resolution	200Km x 200km	40Km x 40Km



Truncation

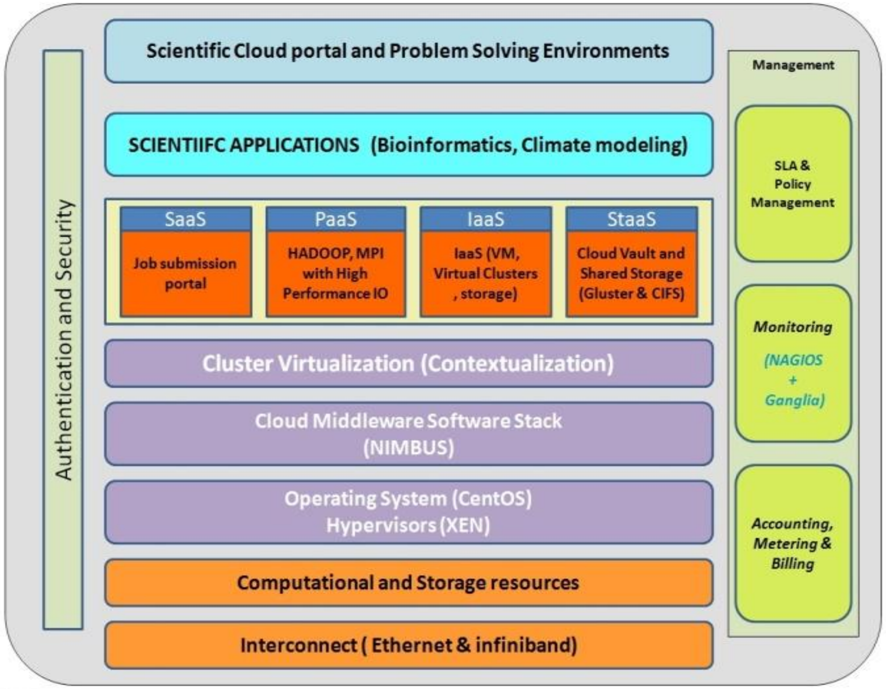
Spherical harmonic expansion truncated at wave-number 62 and 320 using triangular truncation

3. The SuMegha Cloud

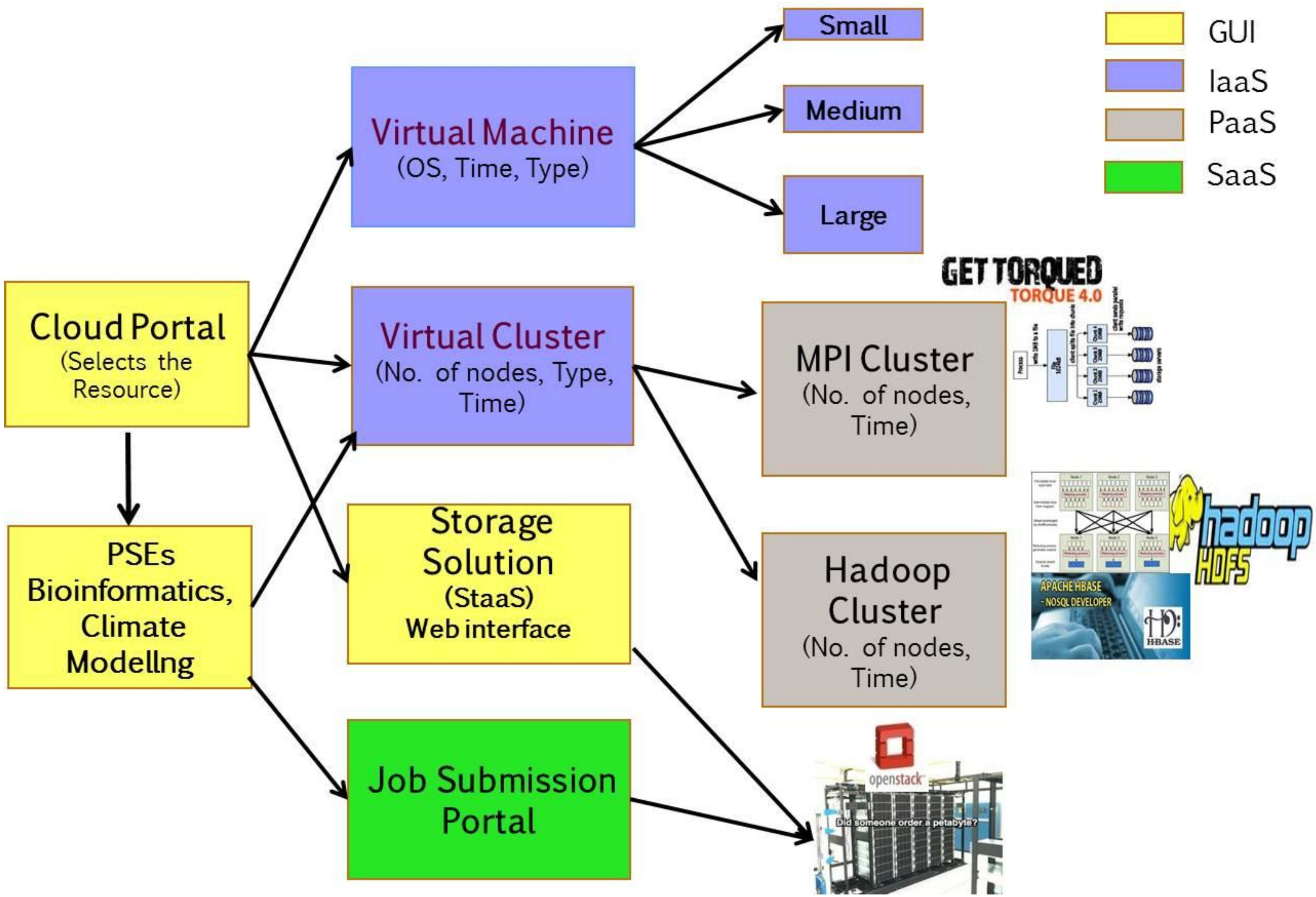
SuMegha

Scientific Cloud for on-demand access to a shared pool of HPC resources (ex: Servers, Storage, Networks, Applications) that can be easily provisioned as and when needed by the researchers/scientists.

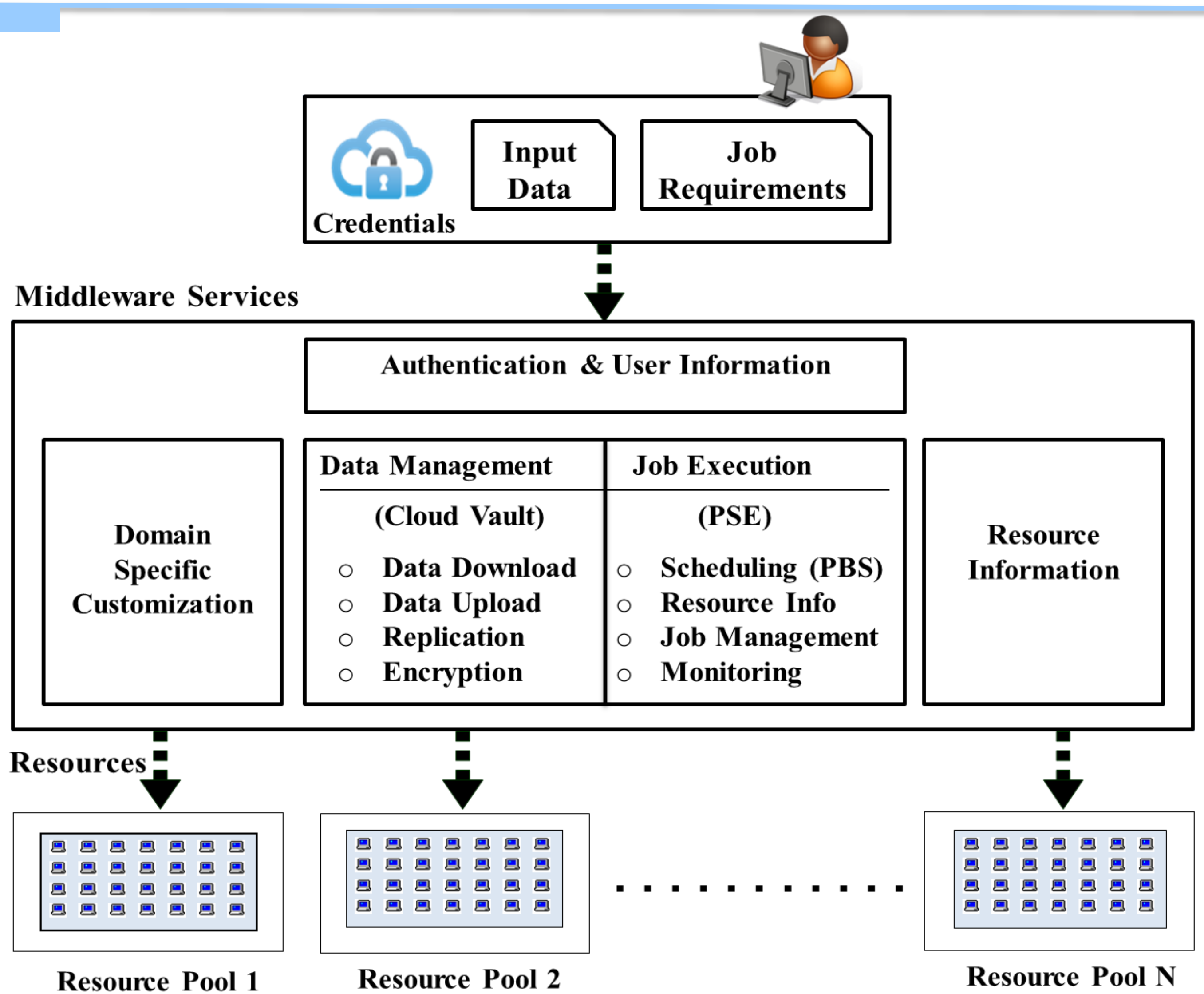
- Benefits of Scientific Cloud
- On demand access to HPC resources
- Ease of access to the available infrastructure
- Virtual ownership of resources to the users
- Ease of deployment



SuMegha Cloud Services



User view of SuMegha Cloud



4. Implementation of SFM on SuMegha

SFM-T62 — Prototype experiments

SFM-T320 — Scalability experiments

SFM-T320 — Ensemble experiments

- Implemented on 5 virtual clusters
- Low resolution (T62) & high resolution (T320) configurations
- Compiled using “gcc-v4.x” and “mpiifort”
- Linked with Intel MPI library

	SFM-T62	SFM-T320
C compiler	gcc-v4.x or later	gcc-v4.x or later
FORTRAN compiler	mpiifort	mpiifort
MPI Library	Intel MPI	Intel MPI
Disk space	1 GB *	27 GB *

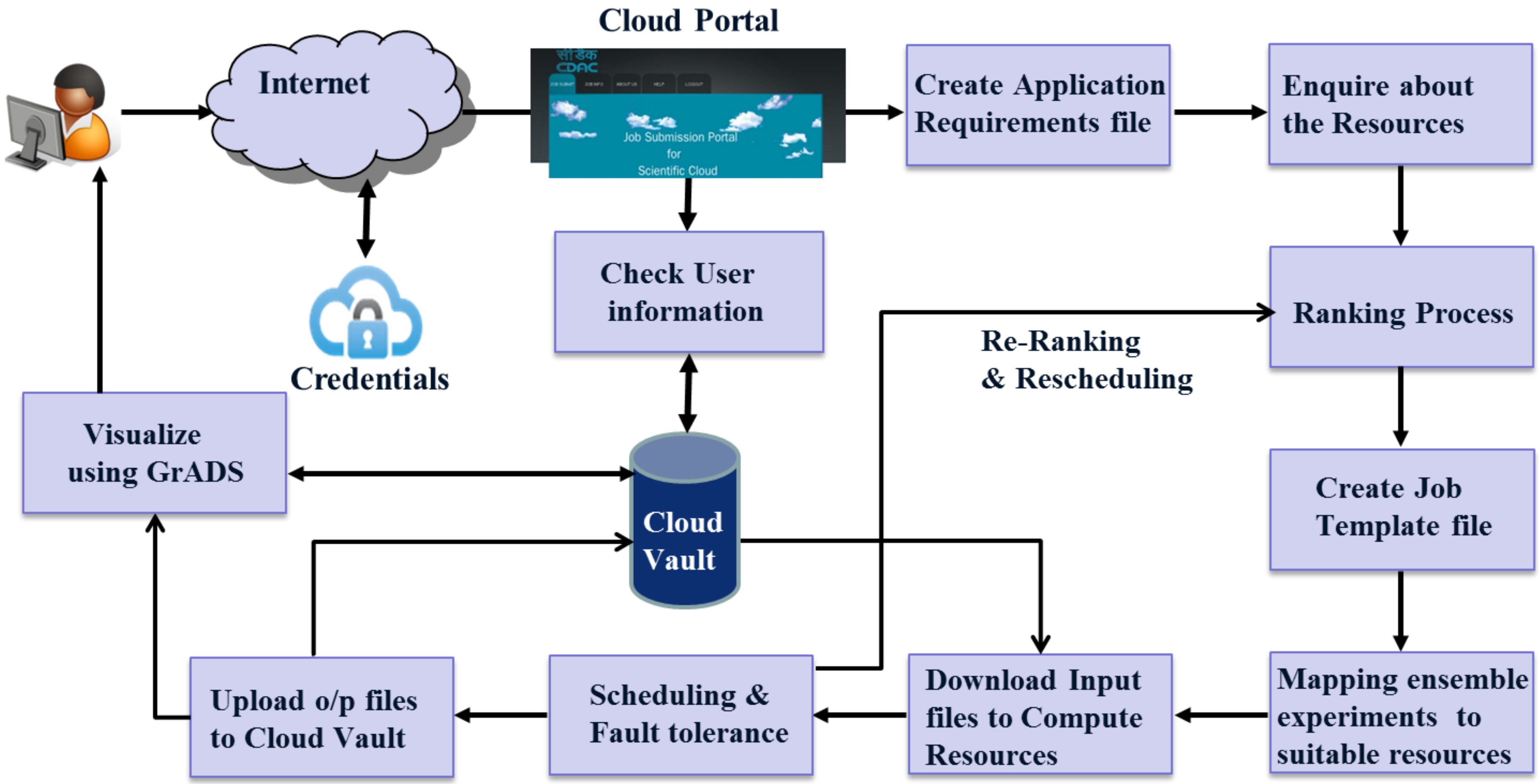
* Disk space is for a seasonal run (JJAS) of an year per member

Hardware Details

- All are LINUX based resources
- Interconnect - Infiniband

Resource Pool	Processor	Speed	Mem.	CPUS/ Node	Total CPUs
VC 1	Intel Xeon	3.16 GHz	16 GB	8	256
VC 2	Intel Xeon	3.16 GHz	16 GB	8	256
VC 3	Intel Xeon	3.16 GHz	16 GB	8	256
VC 4	Intel Xeon	2.95 GHz	64 GB	16	256
VC 5	AMD Opteron	2.50 GHz	64 GB	16	256

Framework of SFM on SuMegha



(a) Prototype experiments with SFM-T62

- Run is divided into sequential & Parallel
- Experiments on Physical & Virtual recourses separately
- Same user control parameters in all the runs
- Similar experiment on five resource pools
- Performance variations observed

User control parameters

Variable	Value	Description
MACHINE	Linux	Machine type (sgi/ibmsp/sun/dec/hp/cray/linux)
MARCH	mpi	Machine functionality (single/thread/mpi/hybrid)
MODEL	gsm	Name of the model (gsm/rsm)
DEFINE	gsm6228/gsm32048	model resolutions
DIR	gsm	Model executable directory
NCPUS	1/8/16/32	Number of Nodes
NPES	8/64/128/256	Number of processing elements
F77	mpiifort	Model compiler (mpiifort - Intel MPI library)

(a) Prototype experiments with SFM-T62…

Performance Metrics (Before tuning & After tunign)

	Physical Cluster	VC1	VC2	VC3	VC4
Processor speed	2.93 GHz	3.16 GHz	3.16 GHz	3.16 GHz	2.5 GHz
Processor family	Intel Xeon	Intel Xeon	Intel Xeon	Intel Xeon	AMD Opteron
Total run time (%T)	74m 46s	75m 46m	191m 37s	191m 20s	273m 38s
%T w.r.t Physical Resources	**100%**	**101.3%**	**256.3%**	**255.9%**	**365.9%**
Total run time (Using Framework)	74m 46s	75m 46m	81m 37s	77m 20s	92m 38s
%T w.r.t (Using Framework)	**100%**	**101.3%**	**116.3%**	**104.1%**	**124.3%**

Observations

- Performance is always more when we use framework
- Variations in performance are due to various reasons like small variations in CPU speed, Wall time spent in queue, MPI libraries, the differences in bandwidth, errors during the execution

(b) Reliability experiments with SFM-T320

- Run is divided into sequential & Parallel
- **3-Day forecast** experiments with SFM-T320
- Same user control parameters as T62 configuration except DEFINE, NCPUS and NPES variables
- Similar experiment on clusters
- Studied scalability – scaling up to 256 processes
- Studies reliability of the resources

SFM T320 Scalability

Number of Cores	Total Execution Time
64	1hr 17min
128	43min (~80% gain)
256	31 min (~40% gain)

SFM T320 Reliability

Type	Failure Rate
Without Framework	24%
With Framework	8%

(c) Ensemble experiments with SFM-T320

- Five virtual pools of resources have been chosen.
- Each pool/VC can run an experiment with one ensemble member.
- Proposed framework is used.
  - PSE for job submission and management
  - Storage – Cloud Vault
  - Data visualization (Grid Analysis and Display System – GrADS is integrated into PSE.
- Source code is not modified, modified run scripts to integrate with the ensemble framework.

(c) Ensemble experiments with SFM-T320

Benefits

- We can run the model with several ensemble members simultaneously
- It saves lot of wall clock time
  - **One seasonal run** with one ensemble member needs around **80 hours** of wall clock time if I use 64 processors (2 x Quad core Xeon @ 3.16 GHz) as a single job, for **100 such experiments** we need **8000 CPU hours** (approx. **1 year**).
  - We could complete these 5 experiments in **1 month** using the above framework on 5 virtual pools of resources of SuMegha.
- Cloud Vault allowed us to keep the replicas of the output at different sites.
- **Failure rates have been decreased from 24% to 8%.**

(c) Ensemble experiments with SFM-T320…

Requirements from the Middleware

Cloud Middleware should have the following features.

- It should provide a mechanism to address the issues such as **non-uniform memory sizes** that are available on the virtual clusters of the cloud.
- It should be able to **identify the failed jobs as early as possible.**
- It should **hide the virtualization layer** completely from the application.
- It should **seamlessly transfer the huge output data files to the user** from cloud during the experiment which will avoid the accumulation of huge data on the compute clusters.
- It should allow **automatic migration of failed jobs** to other reliable resources.
- It should provide **dynamic scaling of resources** without user's intervention.

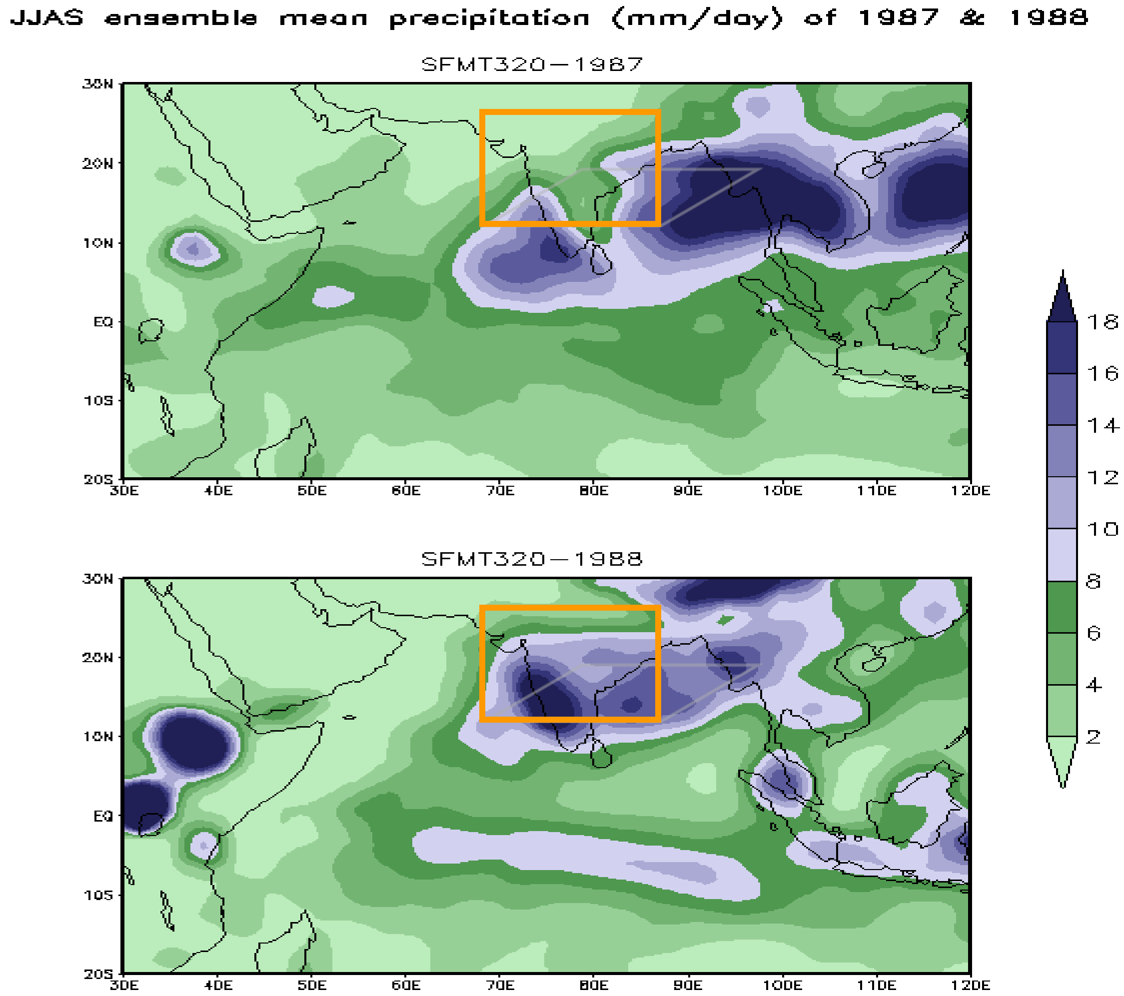
Few Results

Top panel

Ensemble mean rainfall of the Indian summer monsoon season of 1987

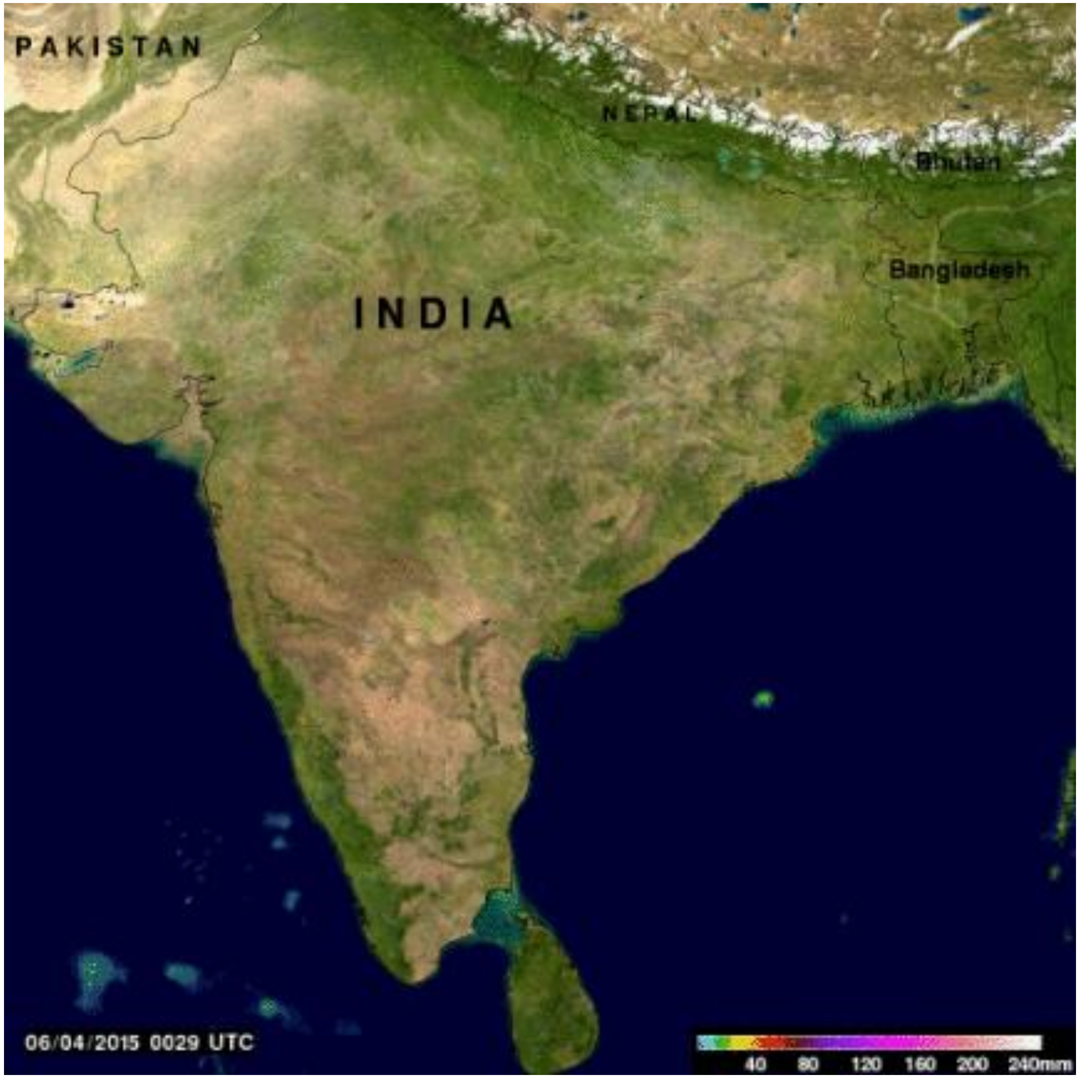
Bottom panel

Ensemble mean rainfall of the Indian summer monsoon season of 1988



Excess monsoon rainfall occurred in 1988, drought occurred in 1987. SFM is capable of simulating these extremes.

Few Results



PAKISTAN
NEPAL
Bhutan
Bangladesh
INDIA
06/04/2015 0029 UTC
40
80
120
160
200
240mm

Conclusions

- Ensemble forecasting is a suitable application from climate modeling domain which can harness the power of cloud computing paradigm.
- Perfomrance variations were observed on different resources even though they are homogenous, and fine-tuned resources.
- A framework is developed which provides a foundation on which to build a reliable cloud environment for huge climate applications which are time sensitive and/or critical.
- Provides a platform to climate researchers to counduct experiments with ease and comfort.
- Future work: Enhance the framework to deal with complexity and instability of the cloud infrastructure to conduct the experiments with more comfort and reliability, and planning to extend the framework for other climate applications.

Thanks to

- SuMegha Cloud Operations and Administration Team, C-DAC, Bangalore
- National PARAM Supercomputing Facility (NPSF) Team, C-DAC, Pune

References

1. M. Leutbecher et al, *"Ensemble Forecasting",* Journal of Computational Physics, Vol. 227, pp. 3515-3539, 2008.
2. John M. Lewis, *"Roots of Ensemble Forecasting"*, Monthly Weather Review, Vol. 133, pp 1865-1885, 2005.
3. V. Fernadez etal, *"Benefits and requirements of grid computing for climate applications. An example with the community atmospheric model"*, Journal of Environmental and Modelling software, Vol. 269, pp 1057-1069, 2011.
4. Kanamitsu et al.,*"NCEP dynamical seasonal forecast system",* BAMS, 83, 1019-1037, 2000.
5. Sumegha – CDAC Scientifi Cloud, *http://www.sumegha.in/cloud-services/*. (Accessed on 22nd September, 2015).
6. J. G. Sela, *"Weather forecasting on parallel architectures",* Journal of parallel computing, Vol. 21, Issue 10, pp 1639-1654, ISSN 0167-8191, 10.1016/0167-8191(95)00039-1, October 1995.
7. Natioanl Knowledge Network – Connecting Knowledge Institutions, *http://www.nkn.in/*. (Accessed on 22nd September, 2015).
8. Raj, A.,Kaur, et al, *"Enhancement of Hadoop Clusters with Virtualization Using the Capacity Scheduler"*, Third International Conference on Services in Emerging Markets(ICSEM), December 2012.
9. Grid Analysis and Display System – An interactive visualization tool for Weather and Climate,*http://iges.org/grads* (Accessed on 22nd September, 2015).

Thank You