Typical late-afternoon in India
6-30-10

TSUNAMI PREDICTION MODEL ON CLOUD COMPUTING PLATFORM

-Keith Lee

8-25-10

Technical Notes- mapReduce

- I wrote a sequential java program that determines how many times the categories in Tsunami input files exceed a certain threshold.
- For small inputs (<500mb), the sequential program was much faster the mapReduce equivalent, but as the input size increased the mapReduce program was much faster than the sequential program.
- For example, the mapReduce program completed in 22 seconds on the University of Hyderabad cluster for a small, 4kb file while the sequential program completed in less than a second
- For a 800mb file, the mapReduce program completed in 45 seconds while the sequential program completed in 1 min. 10 seconds.
- The mapReduce program performs and scales well for larger inputs while the sequential program does not.
- This program proves that a mapReduce program can be applied to Tsunami data and increase the performance of the calculations dramatically.
- Applying the Hadoop mapReduce framework to the Tsunami Warning System can dramatically improve performance.

Data Turbine

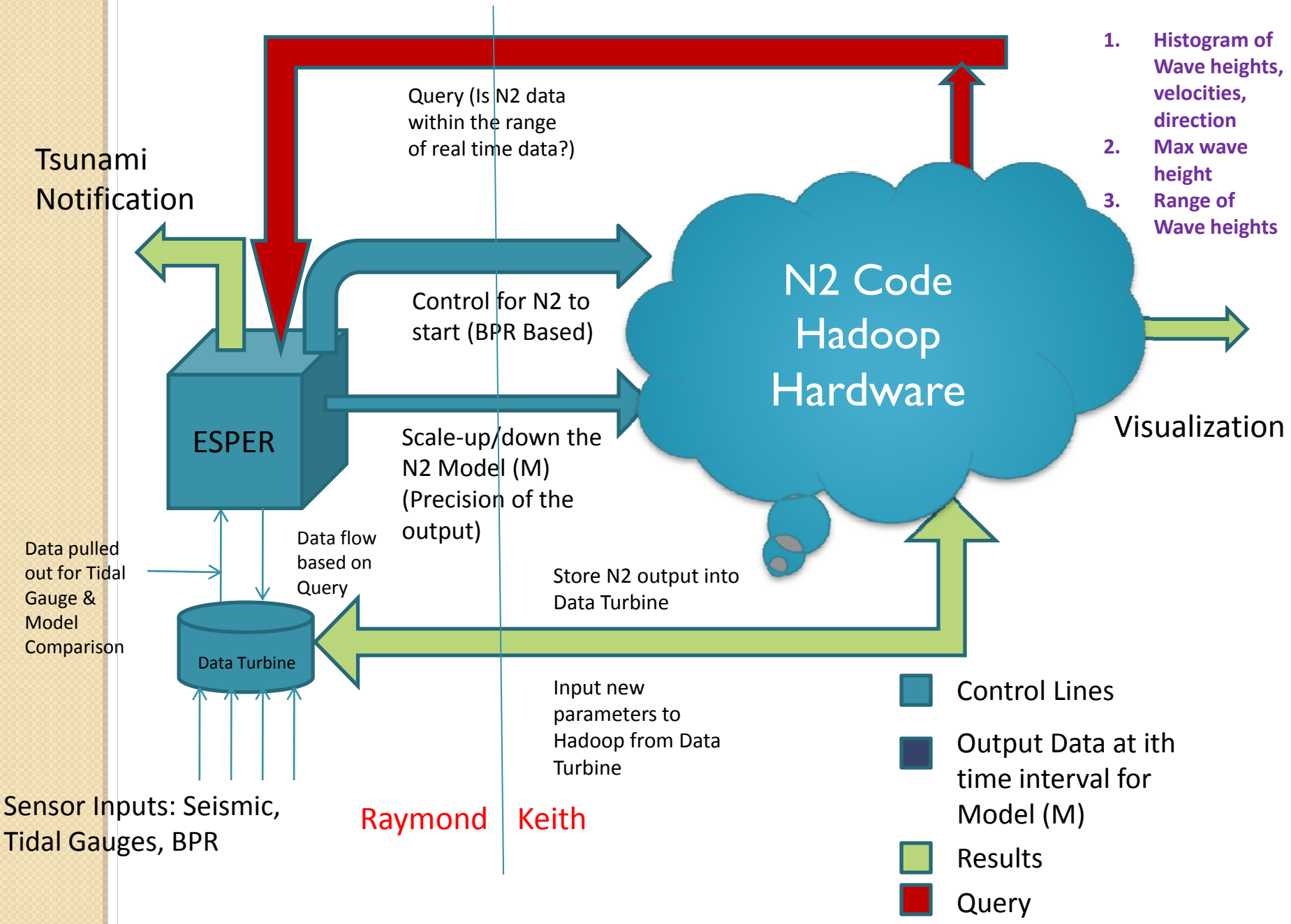
- I realized that Tunami N2 output generally consists of multiple files.
- My Data Turbine program can now sink and source all the files in a folder.
- Benefits:
- The output of a N2 run can be stored into Data Turbine and viewed by researchers, government officials and others easily and efficiently.
- Event streaming software such as ESPER can be used to analyze the output of the data.

Tunami N2 port

- INCOIS has finally sent the sample input and sample output data for a Tunami N2 scenario run.
- The Tunami N2 source code provided could not be copied and pasted and had to be manually typed into a document before it can be run.
- I will run the Tunami N2 code with the sample input and compare it with the output to make sure the Tunami N2 code works properly on the University of Hyderabad cluster.

Problems

- Installing the required software took a lot longer than expected because of the proxy blocking downloads and the lack of root access.
- Retrieving the N2 source code and sample input and output from INCOIS also took a lot longer than expected which slowed down the progress of the project.
- The initial plan of the project was to run the Tunami N2 source code on the Hadoop mapReduce framework using the F2PY (Fortran to Python) programming interface, however after installing the software and writing sample programs, I found out that Hadoop does not support F2PY, which ultimately left me without a project halfway through my PRIME internship.
- After several meetings with Dr. Agarwal, Dr. Tilak, and Prof. Narayana, I found myself designing and working on a Open Source Tsunami Warning System Architecture that will be carried on by future PRIME students.



OPEN SOURCE TSUNAMI WARNING SYSTEM ARCHITECTURE

Useful tools

- Hadoop has been setup on the aruncs account on the venus cluster.
- I have documented my Hadoop mapReduce and Data Turbine code which will help future students pick up mapReduce and Data Turbine much more quickly. My PRIME work is all located on the aruncs account on the venus cluster.
- I have fixed many of the Tunami N2 source code's compilation problems, which will speed up the porting of the N2 code onto the University of Hyderabad cluster.

Ajanta/Ellora Caves

- This weekend I went to the Ajanta/Ellora caves on a tour bus.
- The tour bus schedule was not very efficient as we spent one whole day stuck in a hotel and the next day rushing through Ajanta and Ellora.
- The caves were a marvel to look at. They were carved out of a cliff and located above a small river.
- I preferred the Ajanta Caves over the Ellora Caves because of the beautiful scenery surrounding it.

Ajanta Caves

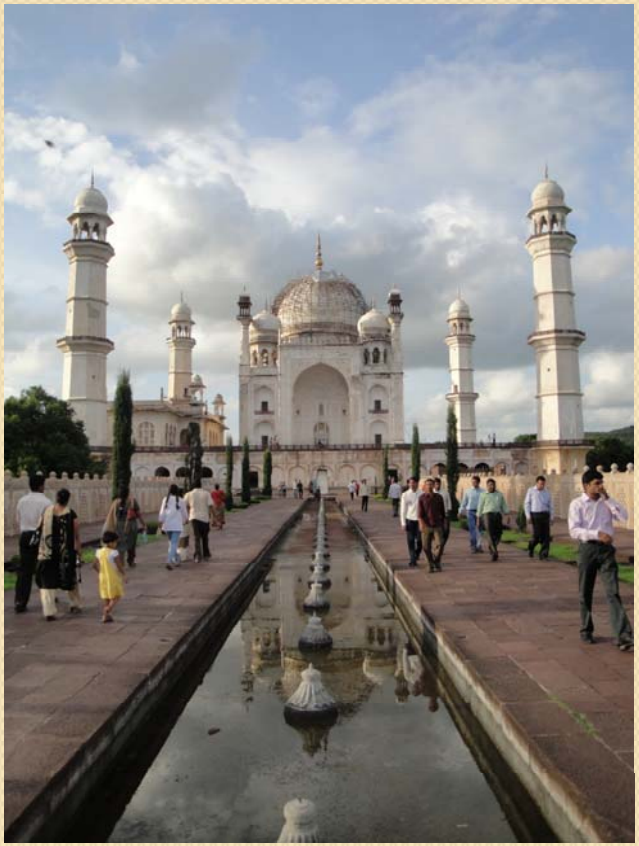
Ellora Caves

Taj Mahal

- I also visited the Taj Mahal over the weekend.
- I learned that it was built by Prince Azam Khan in 1679, son of the Mughal Empire for his mother.
- This Taj Mahal is also known as Bibi Ka Maqbara and sadly due to poor conservation had seen better days…

TAJ MAHAL

I actually visited the mini Taj Mahal over the weekend



Why I love India

- Animals are actually apart of society. Animals are not confined to being farm animals or as pets. Dogs, chickens, cats, monkeys, cows, peacocks, etc can simply live a natural life.
- The people are genuinely very nice. Some ask for money in return for their helpful services but many help because they are happy to.
- The beautiful landscapes.
- The small bumble-bee looking Auto Rickshaws.
- The food especially in the small, dirty places.
- The controlled chaos of the city.
- Riding on the back of motorcycles.
- Haggling
- Everything costs so much less.
- And of course, the smell of India.

Partying with Dr. Agarwal

Acknowledgements

- Dr. Gabrielle Wienhausen, Principal Investigator of PRIME, UCSD
- Dr. Peter Arzberger, Principal Investigator of PRAGMA, UCSD
- Dr. Arun Agarwal, Head of Department of CIS, University of Hyderabad
- Dr. Sameer Tilak, Researcher at San Diego Supercomputer Center, UCSD
- Prof. Narayana, Tsunami dynamics expert, University of Hyderabad
- Babu Rao, Graduate student at the Grid Computing Center, University of Hyderabad
- Indian National Centre for Ocean Information Sciences (INCOIS)
- Department of CIS, University of Hyderabad
- National Science Foundation, IOSE-0710726