A GRID AND CLOUD-BASED DATABASE OF PRE-COMPUTED SCENARIOS OF TSUNAMIS IN MANILA TRENCH

VAST Project: 2012-2013

Pham Thanh Giang
Institute of Information Technology
Vietnam Academy of Science Technology
ptgiang@ioit.ac.vn

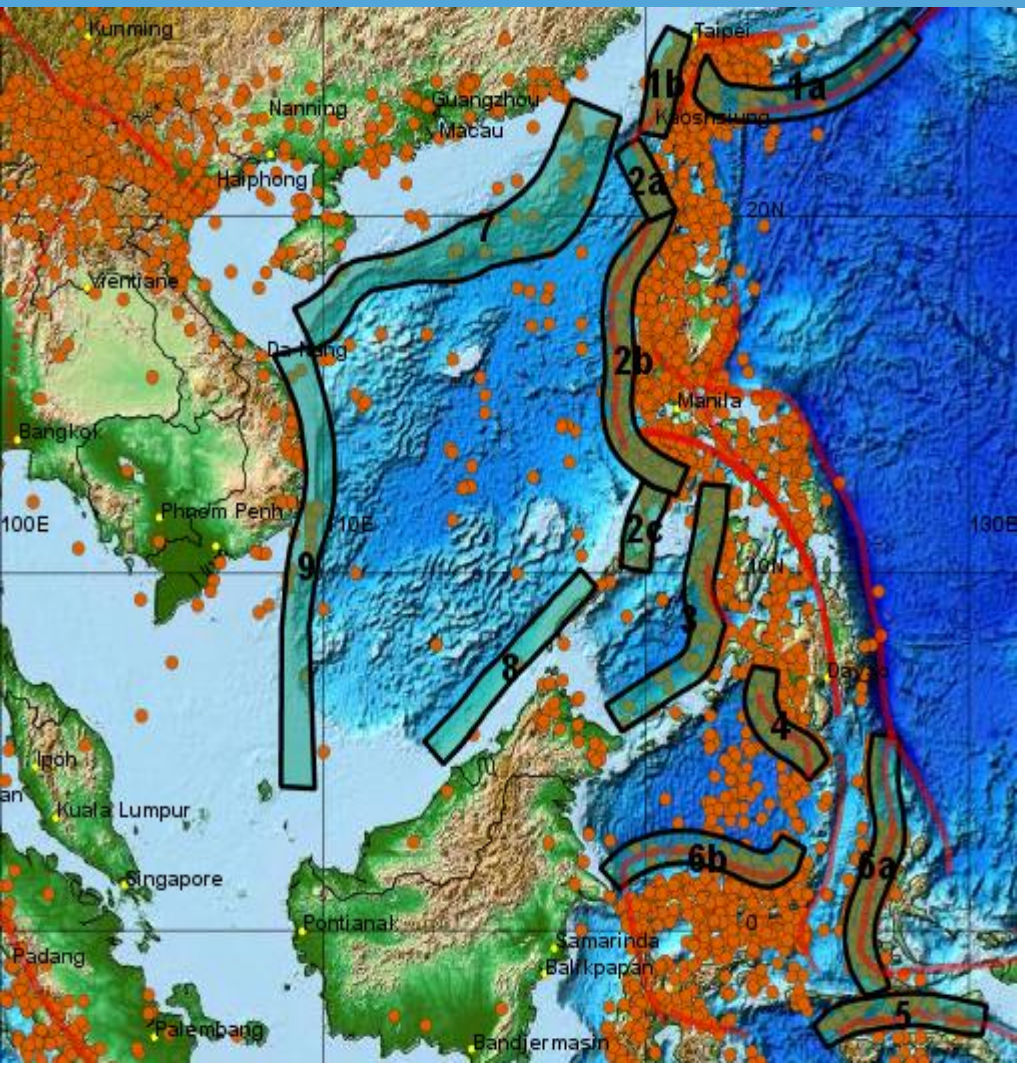
CONTENTS

- ❑ Problem
- ❑ Project members
- ❑ Proposed solution
- ❑ Challenges

PROBLEM

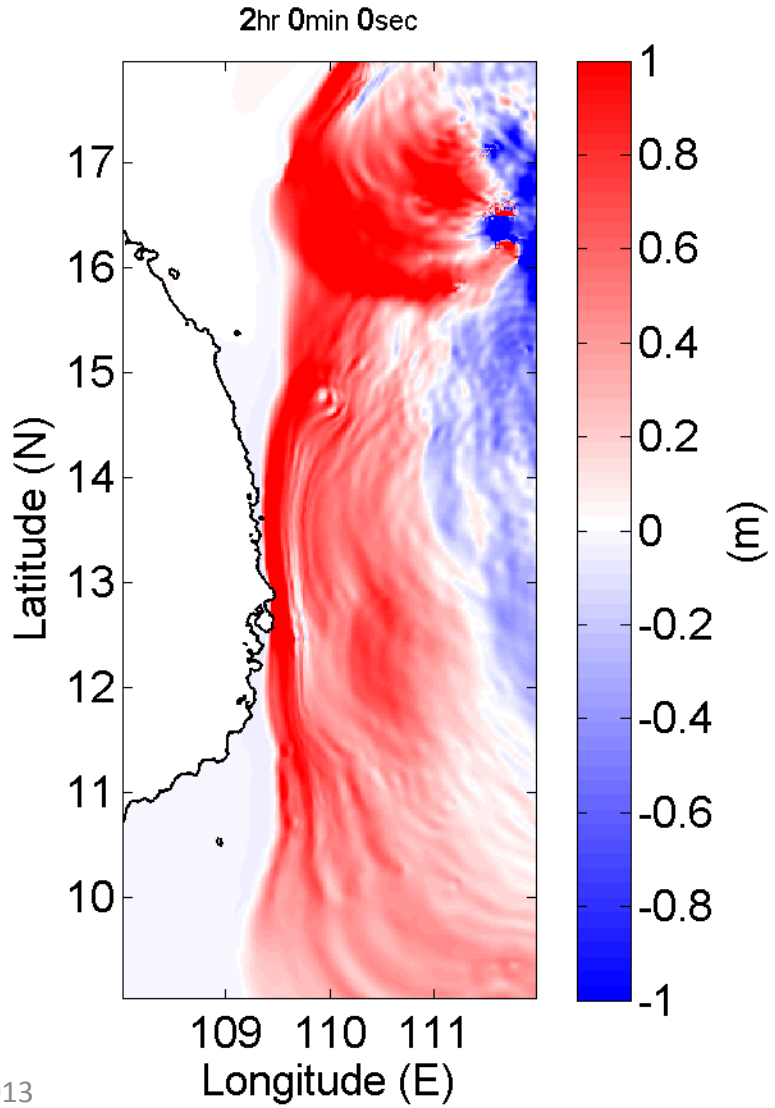
- We have long sea border (3000km)
- Vietnam East Sea has the MANILA TRENCH SOURCE ZONE
- History is said that tsunami existed.

PROBLEM



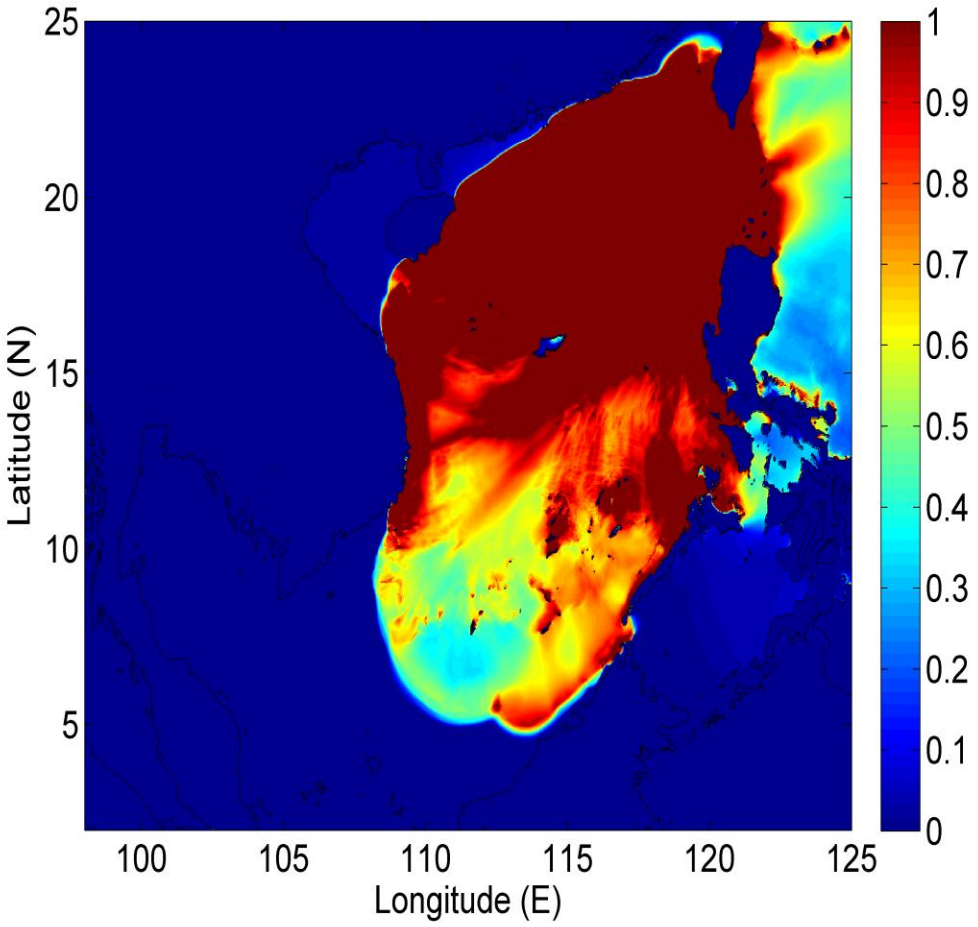
- We have long sea border (3444km)
- Vietnam East Sea has the MANILA TRENCH SOURCE ZONE
- History is said that tsunami existed.

PRE-CALCULATED TSUNAMI SCENARIOS



- 25 tsunami scenarios for Vietnam versus 100 000 ones for Japan; 7 000 for Indonesia, …
- Far not enough for warning purpose in Vietnam

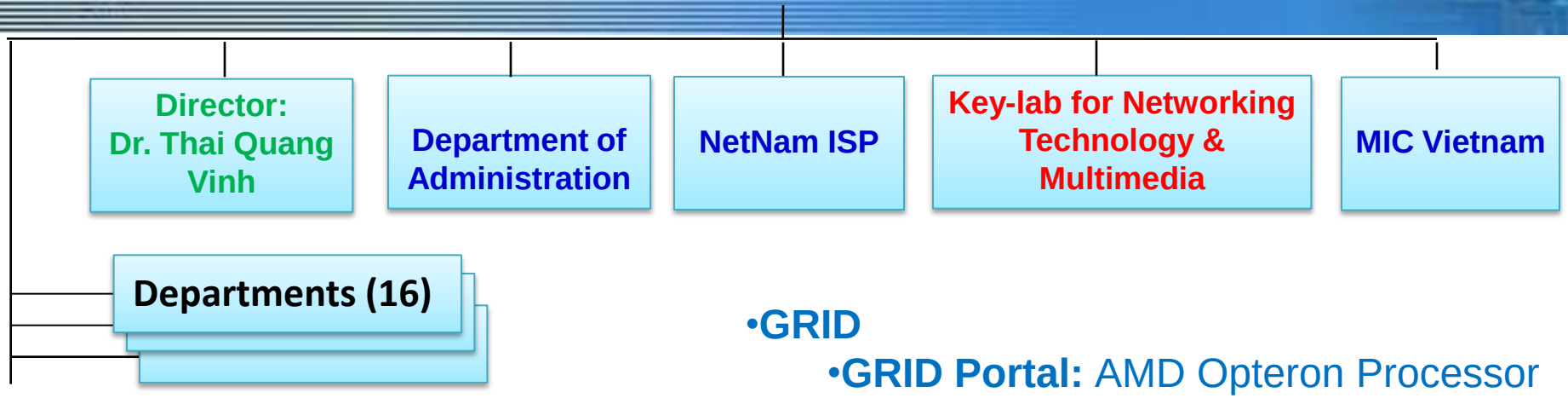
PROBLEMS OF CACULATION CAPACITY AND DATA STORAGE



- Calculating time (PC dual core 2.6GHz):
  - ~2hours/1hour-scenario
  - ~10hours/10hour scenario
- Required Storage
  - 3GB/1hour-scenario
  - 15GB/10hour scenario

PROJECT MEMBERS

INSTITUTE OF GEOPHYSICS



1. Integrated Software Systems
2. Data Management Systems
3. Software Engineering in Management
4. Databases and Programming
5. Geographical Information Systems

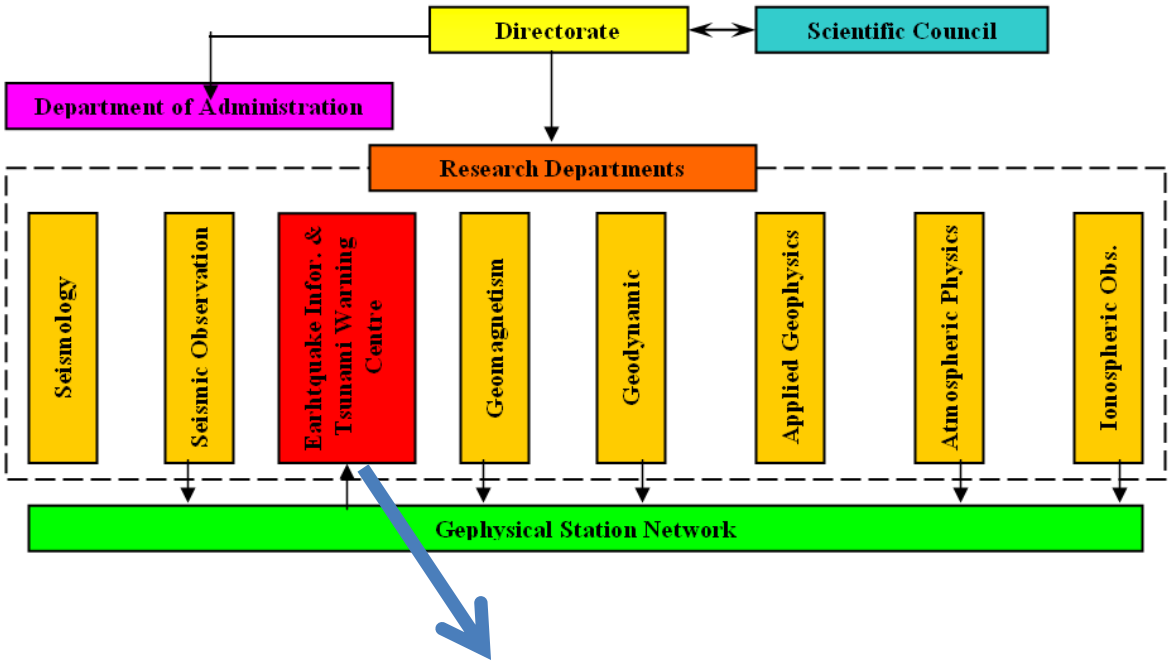
…

8. Computer Networking and IT Infrastructure
9. Telematics

…

- **GRID**
  - **GRID Portal:** AMD Opteron Processor (2.4 GHz, dual core)
  - **Globus Head Node:** AMD Opteron Processor (2.2 GHz, dual core)
  - **Linux Cluster (5nodes):** AMD Opteron Processor (2.2 GHz, dual core)
  - **Storage Server**
- **CLOUD:**
  - **4xIBM Server 12 Core, 12G RAM**
- **NETWORK:**
  - **2x40Mbps Commerce Internet & VinaREN**

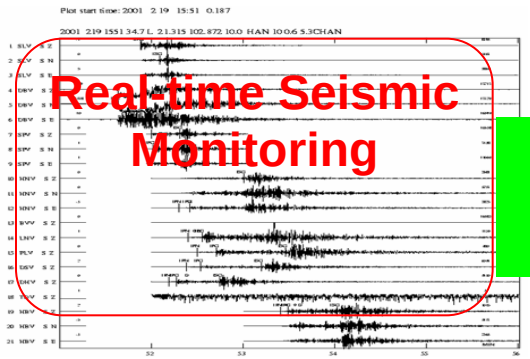
VIỆN VẬT LÝ ĐỊA CẦU
INSTITUTE OF GEOPHYSICS
Trụ sở: Nhà A8, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội - Tel: 84.4.37564380 - Fax:84.4.38364696



Centre for Earthquake Information and Tsunami Warning

- As a role coordinator between IOIT, IGP and EU GRID/CLOUD
- Have a grid node
- Educate students in Grid and Cloud technology

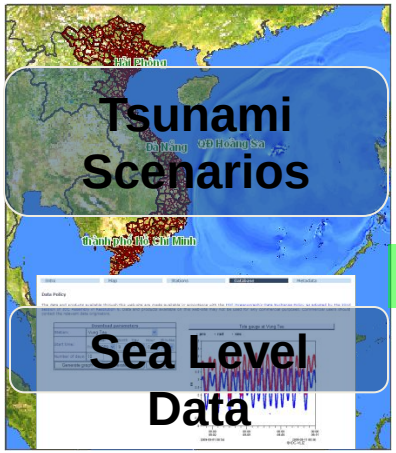
EARLY WARNING SYSTEM



Earthquake Information and Tsunami Warning Center

Rapid Determination of Hypocenter and Magnitude

Evaluation of Tsunami based on the Determined Hypocenter and Magnitude



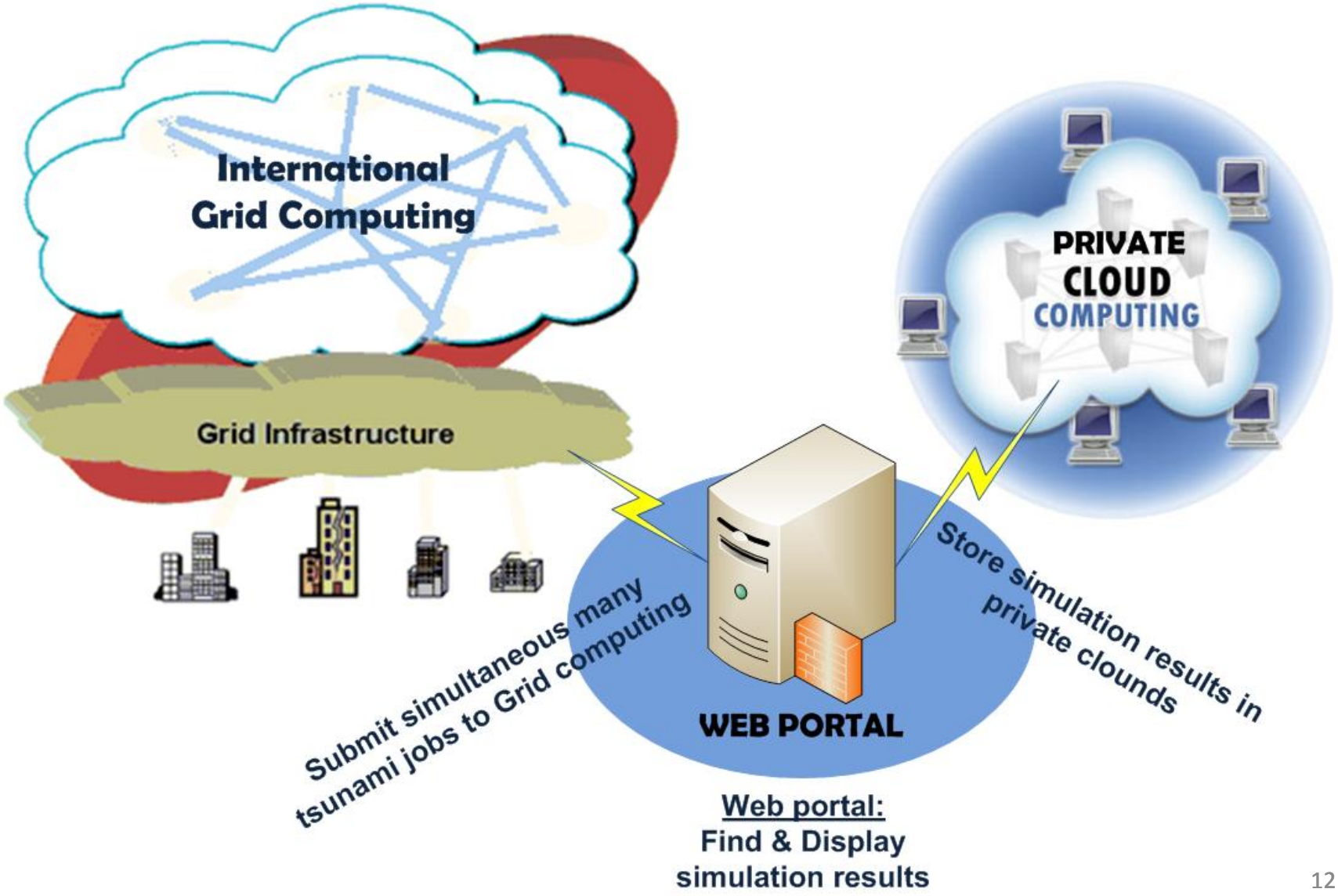
1st Warning

Re-evaluation of Warning

Subsequent Warning

Cancelation

PROPOSED SOLUTION



IDEA PROPOSED SOLUTION

- Evaluate performances of Tsunami modeling in the Grid
  - COMCOT Cornell Multi-grid Coupled Tsunami Model
  - TUNAMI-Tohoku University's Numerical Analysis Model for Investigation of Near field tsunamis
  - MOST-Method Of Splitting Tsunami
- Build Database system and private cloud
  - Stratus lab, MySQL
- Build tsunami scenarios
- Build Web portal for display and find tsunami information

COMCOT Cornell Multi-grid Coupled Tsunami Model

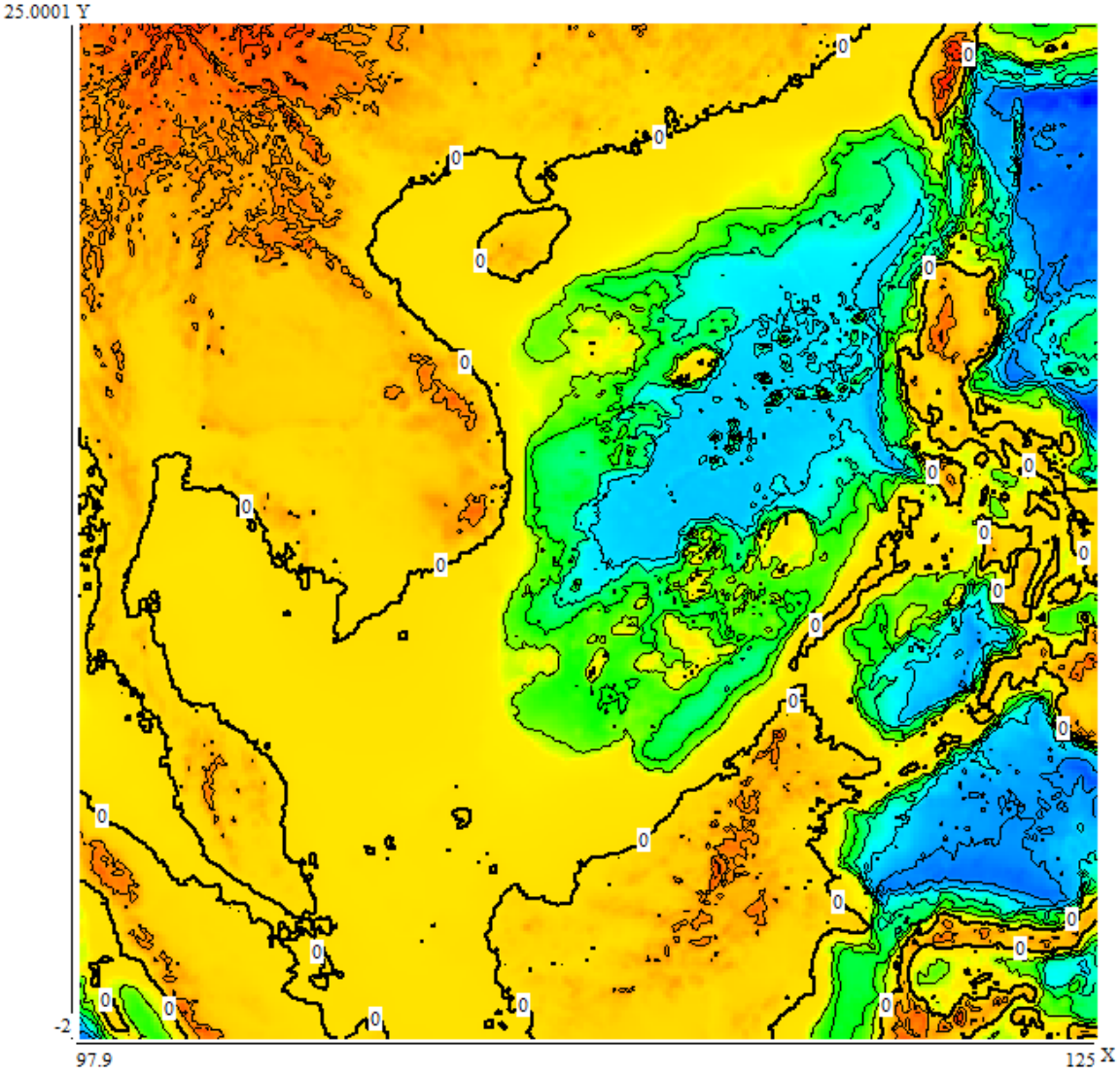
- Evaluate performances of Tsunami modeling in the Grid
  - COMCOT Cornell Multi-grid Coupled Tsunami Model
- Deploy to GRID through DIRAC (Distributed Infrastructure with Remote Agent Control)
- Implement & Process Input/Output Data

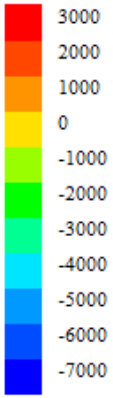
PRE-CALCULATE TSUNAMI MODEL IN THE GRID

- Input parameters
  - General Parameters fo Simulation
  - Parameters for Fault Model
  - Parameters for Wave Maker
  - Parameters for Submarine LS/Transient Motion
  - Configurations for all grids
- Terrain data: (East Vietnam Sea.xyz, DaNang.xyz,…):
  - X: latitude
  - Y: longitude
  - Z: height
  - Data size: tens to hundreds MBs.

```
98.000000   25.000000   2255
98.033333   25.000000   1954
98.066667   25.000000   1624
```

East Vietnam Sea.xyz





X:98 – 125
Y:-2 – 25
Resolution: 2′
Size: 20MB

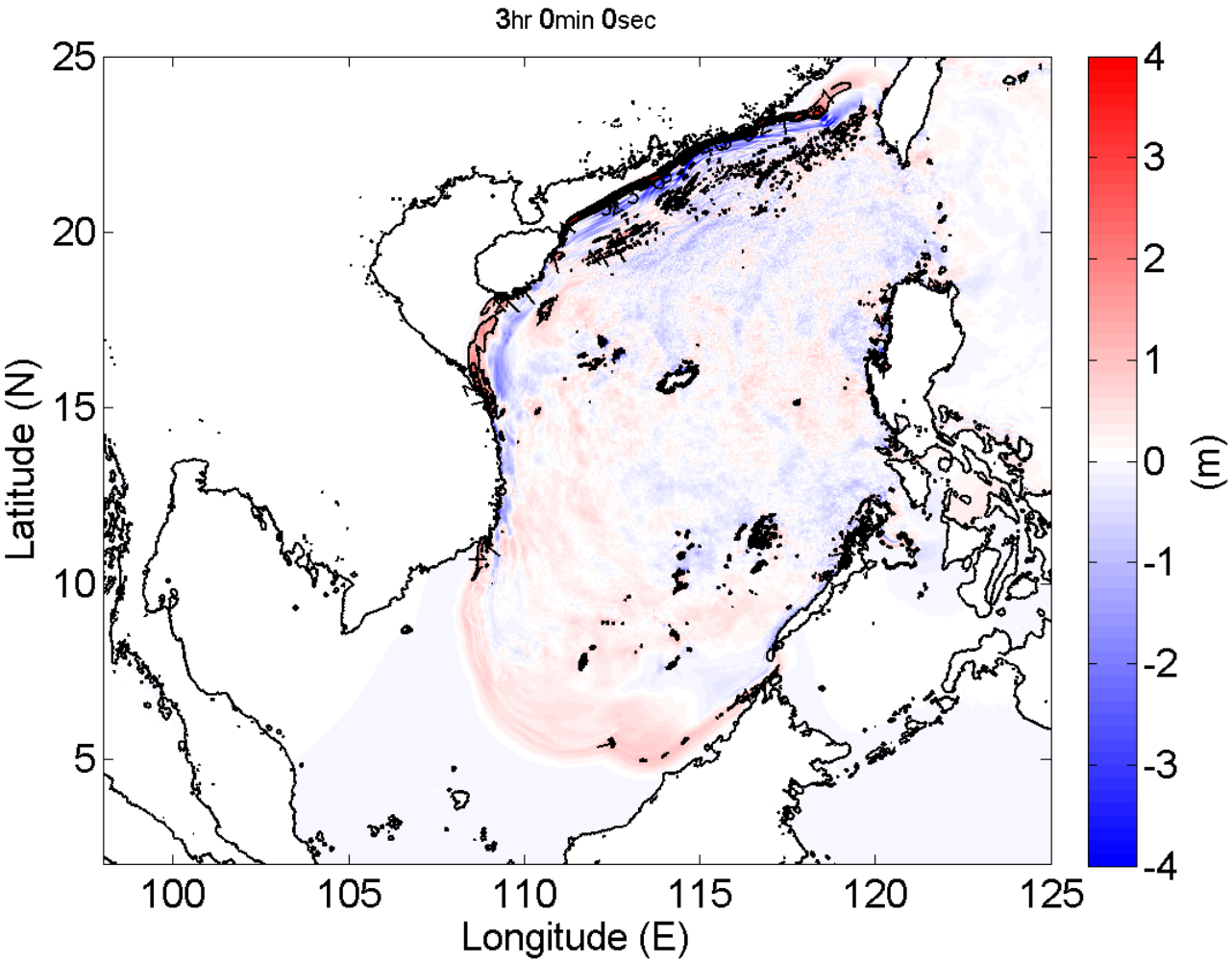
PRE-CALCULATE TSUNAMI MODEL IN THE GRID

- Results:
  - z_xx_yyyyyy.dat: Wave height of region xx at time yyyyyy.
  - zmax_xx_yyyyyyhrs.dat: Wave height max (zmax) of region xx at yyyyyy hrs.
  - zmin_layerxx_yyyyyyhrs.dat Wave height min (zmix) of region xx at yyyyyy hrs hrs.
  - zmax_layerxx.dat: Wave height max (zmax) of region xx for the whole simulation time
- Size of file depends on size of Terrain data
- Number of file depends on Total and interval time
- Total output data ~10sGB.

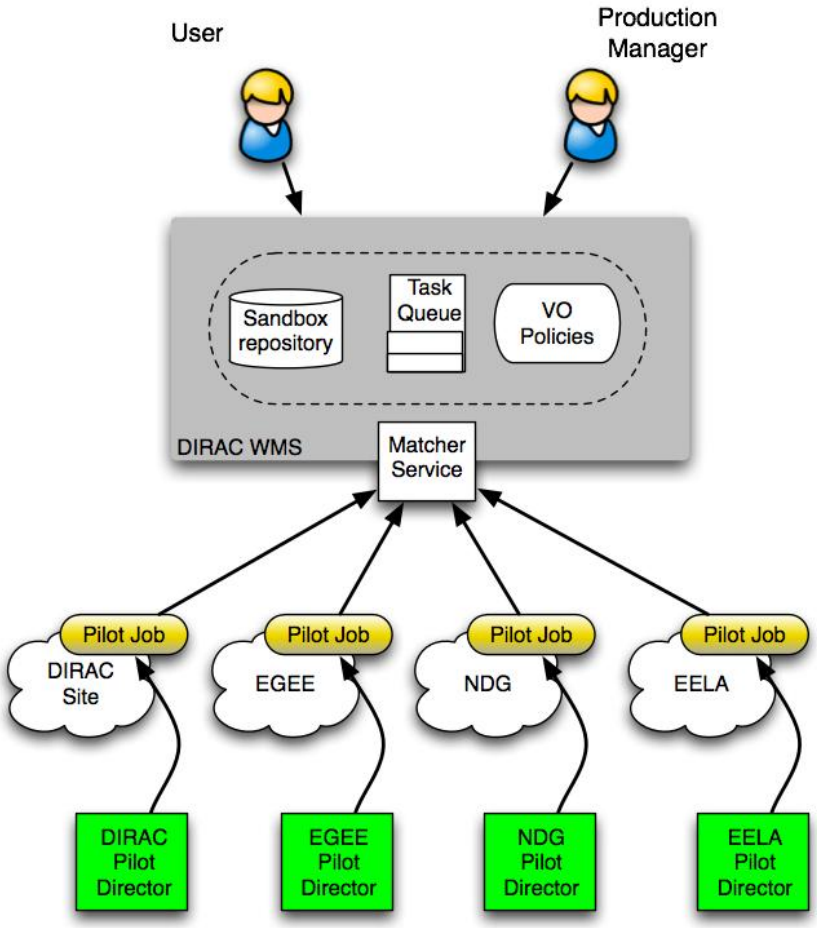
PRE-CALCULATE TSUNAMI MODEL IN THE GRID

- Results:
  - z_xx_yyyyyy.dat: Wave height of region xx at time yyyyyy.
  - zmax_xx_yyyyyyhrs.dat: Wave height max (zmax) of region xx at yyyyyy hrs.
  - zmin_layerxx_yyyyyyhrs.dat Wave height min (zmix) of region xx at yyyyyy hrs hrs.
  - zmax_layerxx.dat: Wave height max (zmax) of region xx for the whole simulation time
- Size of file depends on size of Terrain data
- Number of file depends on Total and interval time
- Total output data ~10sGB.

PRE-CALCULATE TSUNAMI MODEL IN THE GRID



DIRAC (Distributed Infrastructure with Remote Agent Control)



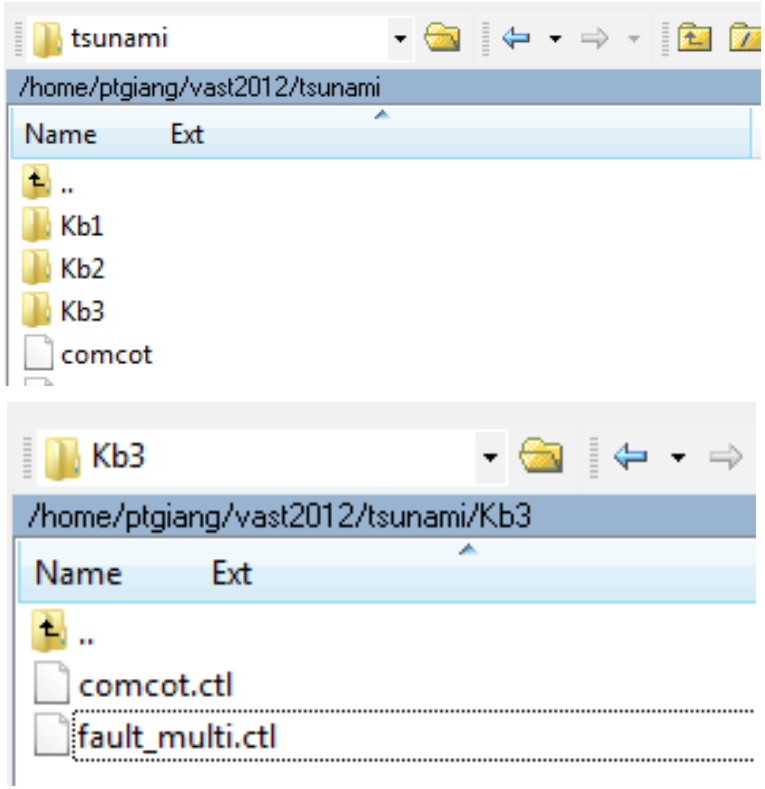
- DIRAC developers has lots of experience with HEP applications but less with other domains
- Large user communities (Virtual Organizations)

DIRAC (Distributed Infrastructure with Remote Agent Control)

- Python is the main development language
  - `from DIRAC.Interfaces.API.Dirac`
  - `import Diracfrom DIRAC.Interfaces.API.Job import Job`
- Submitting job
  - `jobID = dirac.submit(j)`
- Job Monitoring
  - `print dirac.status(jobid)`
- Job Output
  - `print dirac.getOutput(jobid)`

Implementation Input Data

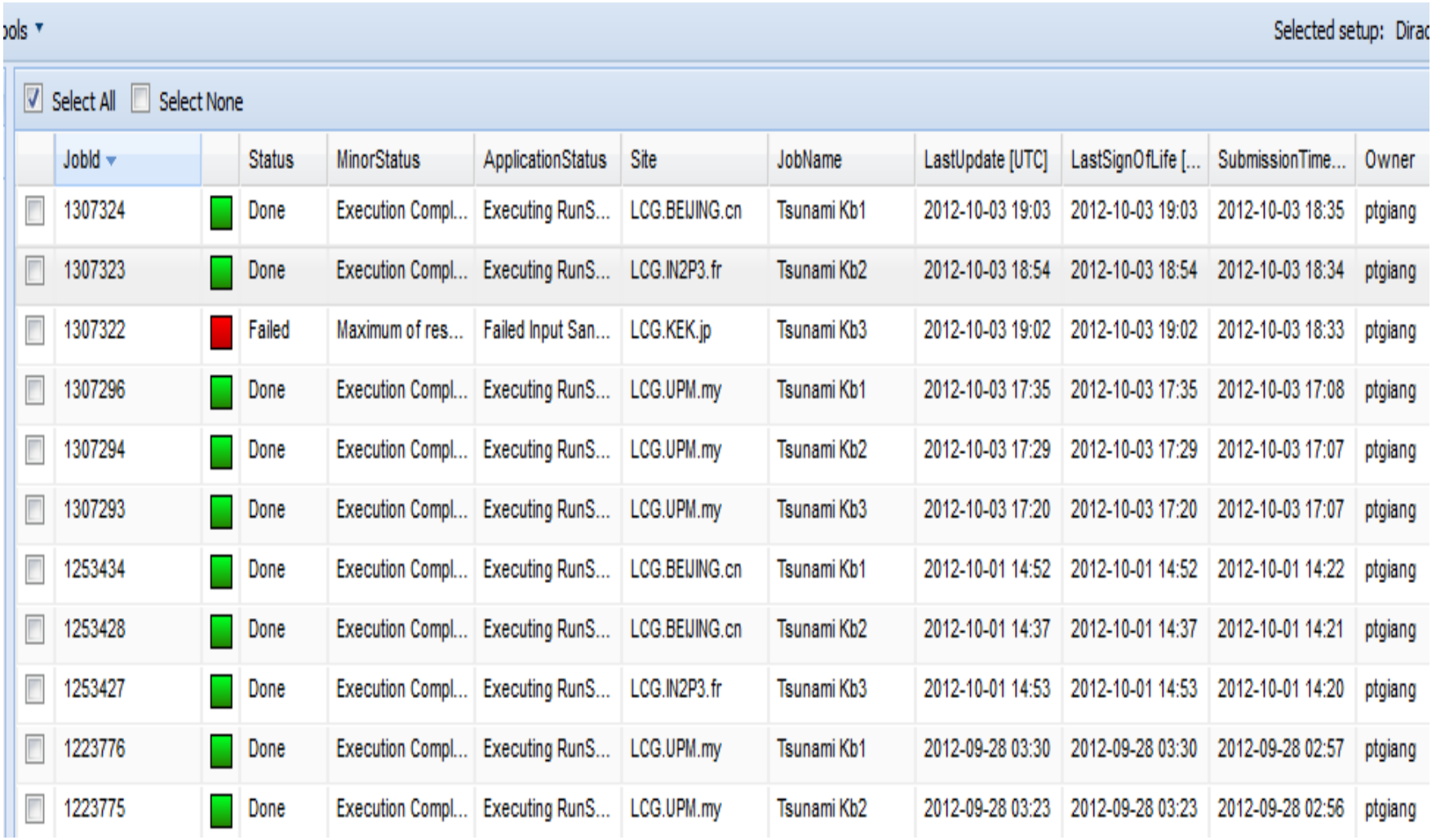
- Each simulation parameters in on folder
- A script generates each simulation to a DIRAC job
- Terrain data was pre-store in Grid to reduce upload time:
  - `dirac-dms-add-file LFN:/esr/user/p/ptgiang/tsunami/EVSea.xyz EVSea.xyz DIRAC-USER`
  - …



Submit job

```
InputDataPath = 'LFN:/esr/user/p/ptgiang/tsunami‘
j = Job()
j.setExecutable('tsunami.sh', kichban)
j.setName('Tsunami ' + kichban)
j.setInputSandbox(‘tsunami.sh‘, ‘comcot.ctl',
‘fault_multi.ctl', ‘comcot‘,
   InputDataPath + '/EVSea.xyz‘, InputDataPath +
'/Middle.xyz', InputDataPath + '/DaNang.xyz’)

j.setOutputData (['tsunami_' + kichban + '.tar',
'tsunami_' + kichban + '.tar.md5sum'])

dirac = Dirac()
jobID = dirac.submit(j)
```

Submit job



ools ▾ Selected setup: Dirac

Select All Select None

JobId	Status	MinorStatus	ApplicationStatus	Site	JobName	LastUpdate [UTC]	LastSignOfLife [...	SubmissionTime...	Owner
1307324	Done	Execution Compl...	Executing RunS...	LCG.BEIJING.cn	Tsunami Kb1	2012-10-03 19:03	2012-10-03 19:03	2012-10-03 18:35	ptgiang
1307323	Done	Execution Compl...	Executing RunS...	LCG.IN2P3.fr	Tsunami Kb2	2012-10-03 18:54	2012-10-03 18:54	2012-10-03 18:34	ptgiang
1307322	Failed	Maximum of res...	Failed Input San...	LCG.KEK.jp	Tsunami Kb3	2012-10-03 19:02	2012-10-03 19:02	2012-10-03 18:33	ptgiang
1307296	Done	Execution Compl...	Executing RunS...	LCG.UPM.my	Tsunami Kb1	2012-10-03 17:35	2012-10-03 17:35	2012-10-03 17:08	ptgiang
1307294	Done	Execution Compl...	Executing RunS...	LCG.UPM.my	Tsunami Kb2	2012-10-03 17:29	2012-10-03 17:29	2012-10-03 17:07	ptgiang
1307293	Done	Execution Compl...	Executing RunS...	LCG.UPM.my	Tsunami Kb3	2012-10-03 17:20	2012-10-03 17:20	2012-10-03 17:07	ptgiang
1253434	Done	Execution Compl...	Executing RunS...	LCG.BEIJING.cn	Tsunami Kb1	2012-10-01 14:52	2012-10-01 14:52	2012-10-01 14:22	ptgiang
1253428	Done	Execution Compl...	Executing RunS...	LCG.BEIJING.cn	Tsunami Kb2	2012-10-01 14:37	2012-10-01 14:37	2012-10-01 14:21	ptgiang
1253427	Done	Execution Compl...	Executing RunS...	LCG.IN2P3.fr	Tsunami Kb3	2012-10-01 14:53	2012-10-01 14:53	2012-10-01 14:20	ptgiang
1223776	Done	Execution Compl...	Executing RunS...	LCG.UPM.my	Tsunami Kb1	2012-09-28 03:30	2012-09-28 03:30	2012-09-28 02:57	ptgiang
1223775	Done	Execution Compl...	Executing RunS...	LCG.UPM.my	Tsunami Kb2	2012-09-28 03:23	2012-09-28 03:23	2012-09-28 02:56	ptgiang

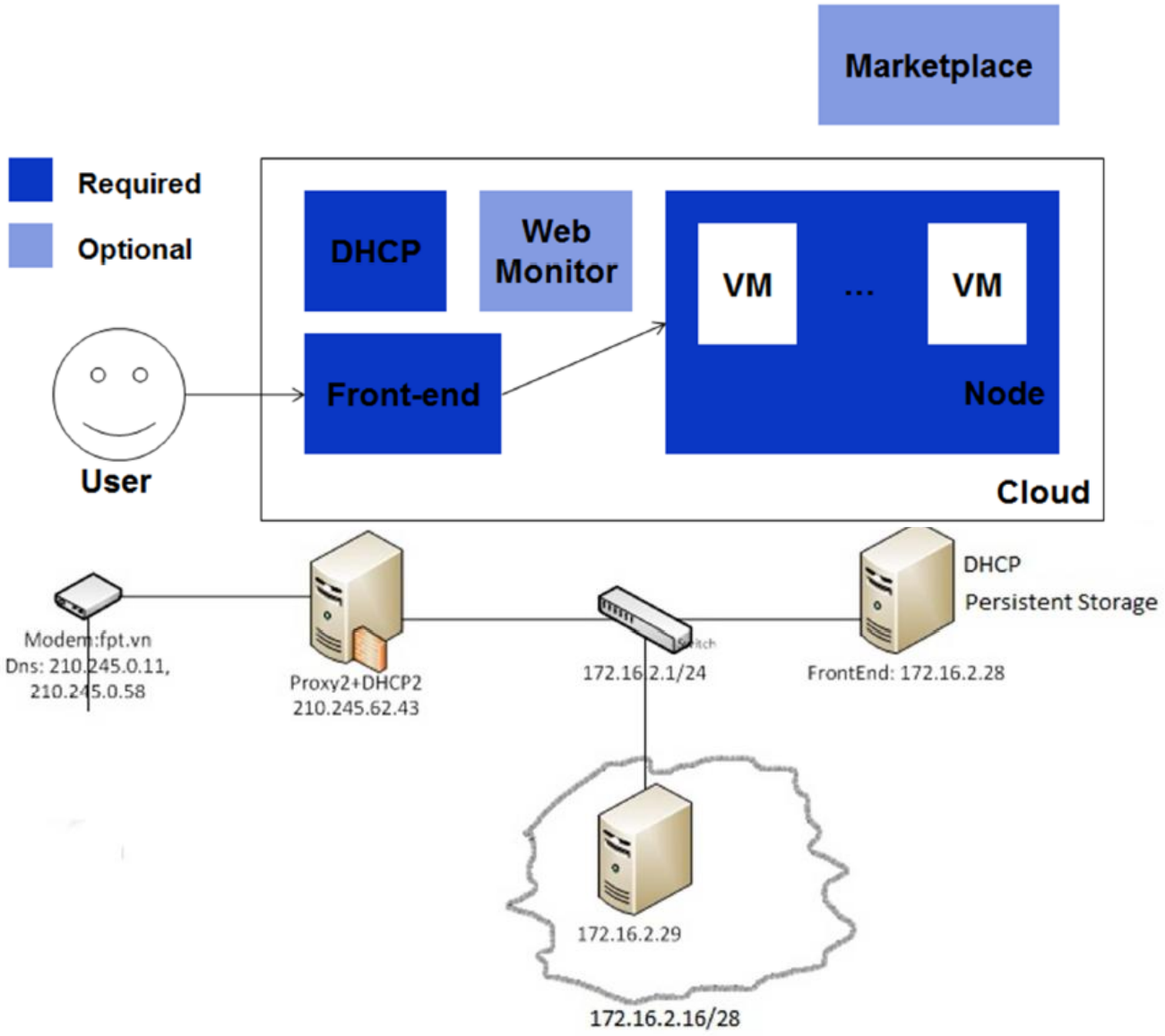
Get results

```
fJobIDname = "JobID.txt"
fJID = open(fJobIDname )

for jobid in fJID:
   print dirac.status(jobid)
   print dirac.getJobOutputData(jobid)

fJID.close()
```

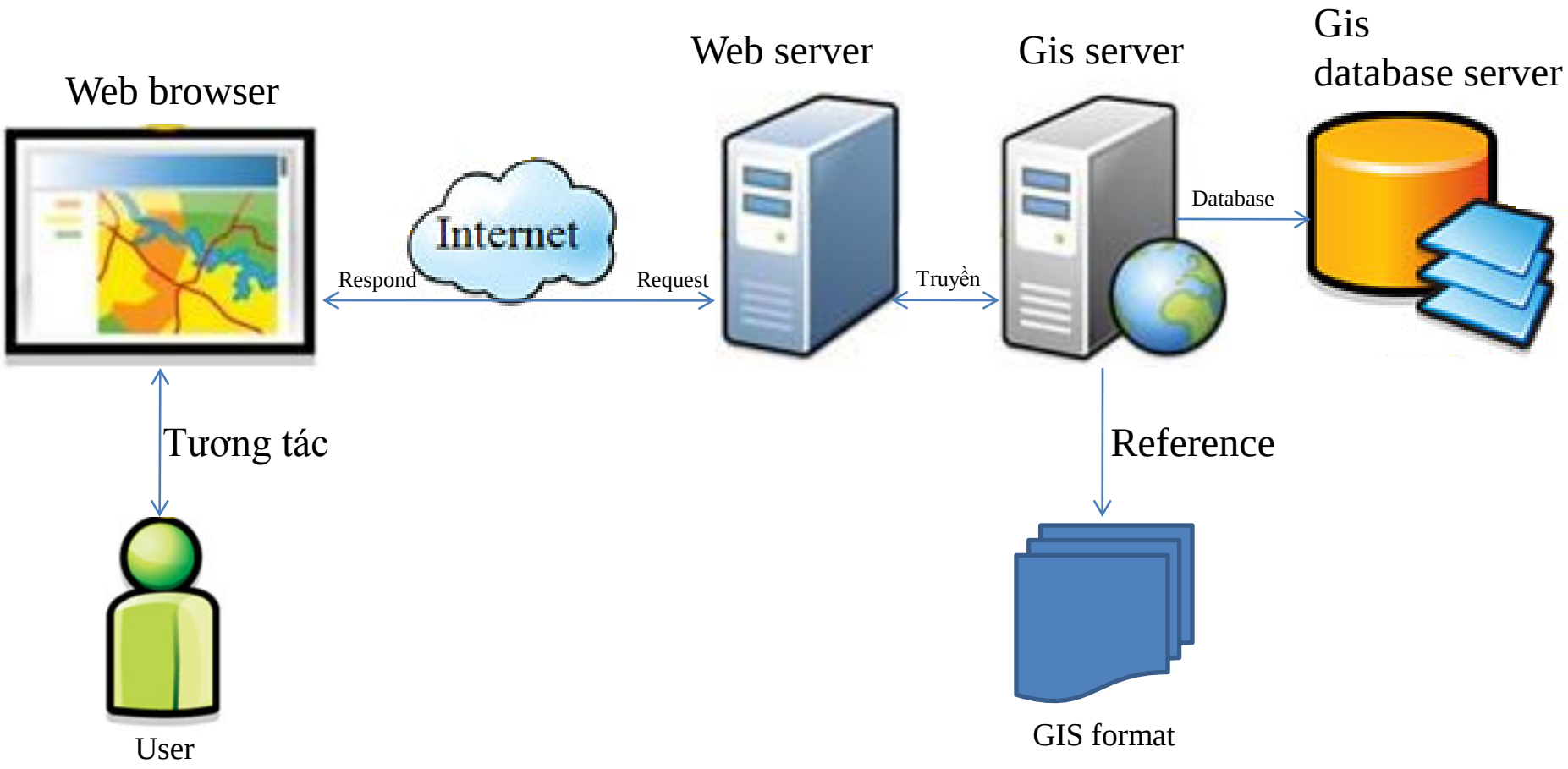
Build Database system and private cloud

- Build Database system and private cloud
  - Stratus lab, MySQL
- 2xIBM Server
  - Processor: Intel xeon(R) CPU X5650 @ 2.67GHz x12
  - Disk: 500GB
  - Memory: 12GB

Build Database system and private cloud

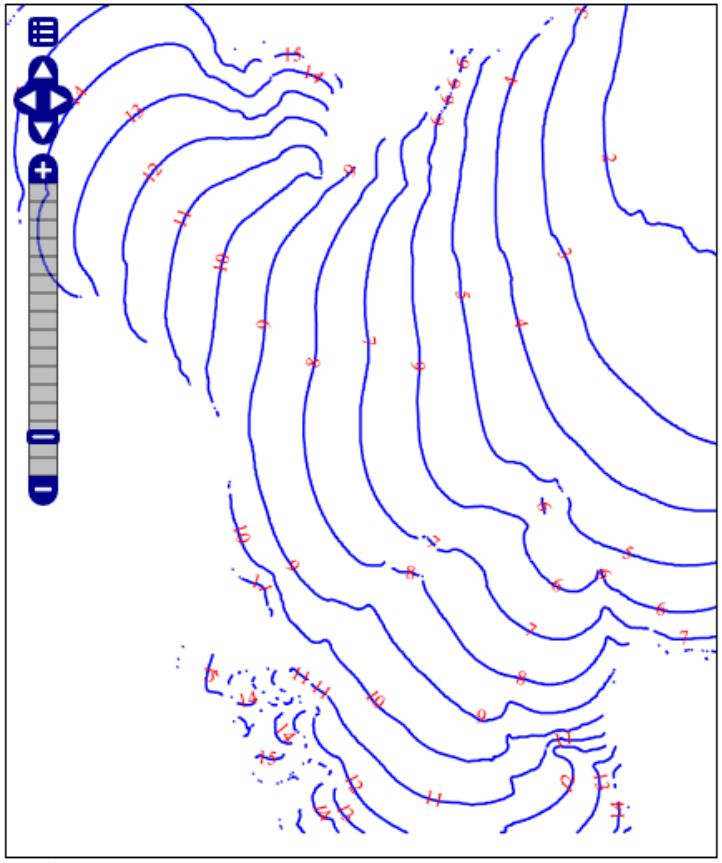


Web portal


Web browser
Web server
Gis server
Gis database server
Internet
Respond
Request
Truyền
Database
Tương tác
Reference
User
GIS format

Web portal

Wave Transmistion time

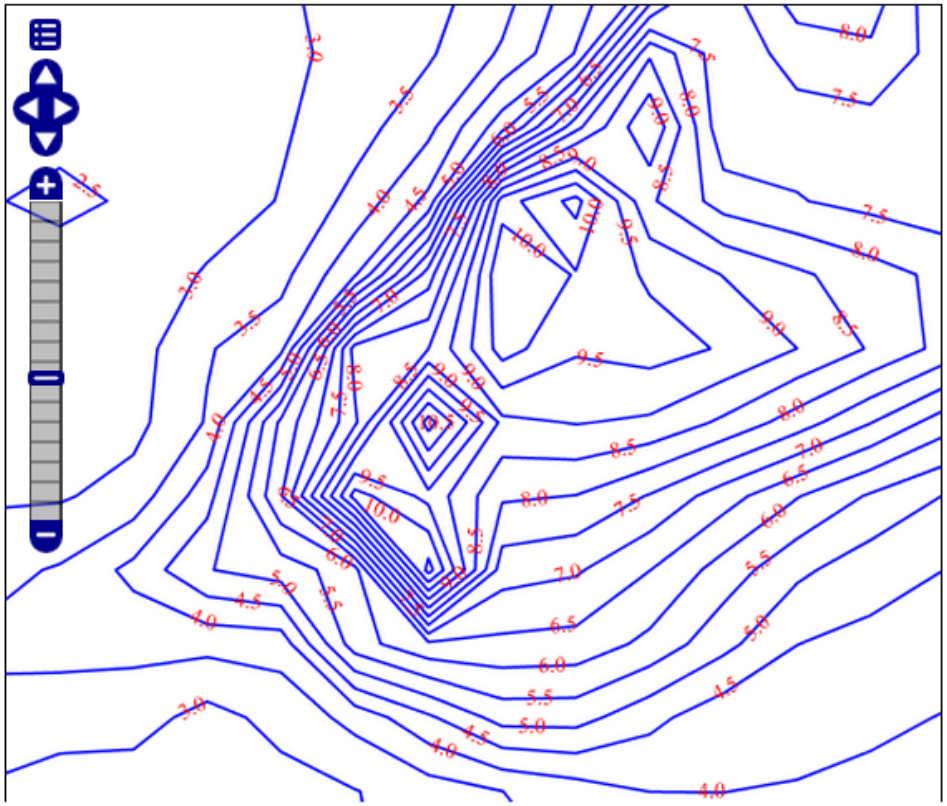


Scale = 1 : 10M 107.97055, 1.19613

CTTIMESPREAD

FID	ID	CONTOUR
ctTimeSpread.695	695	8

Wave Height

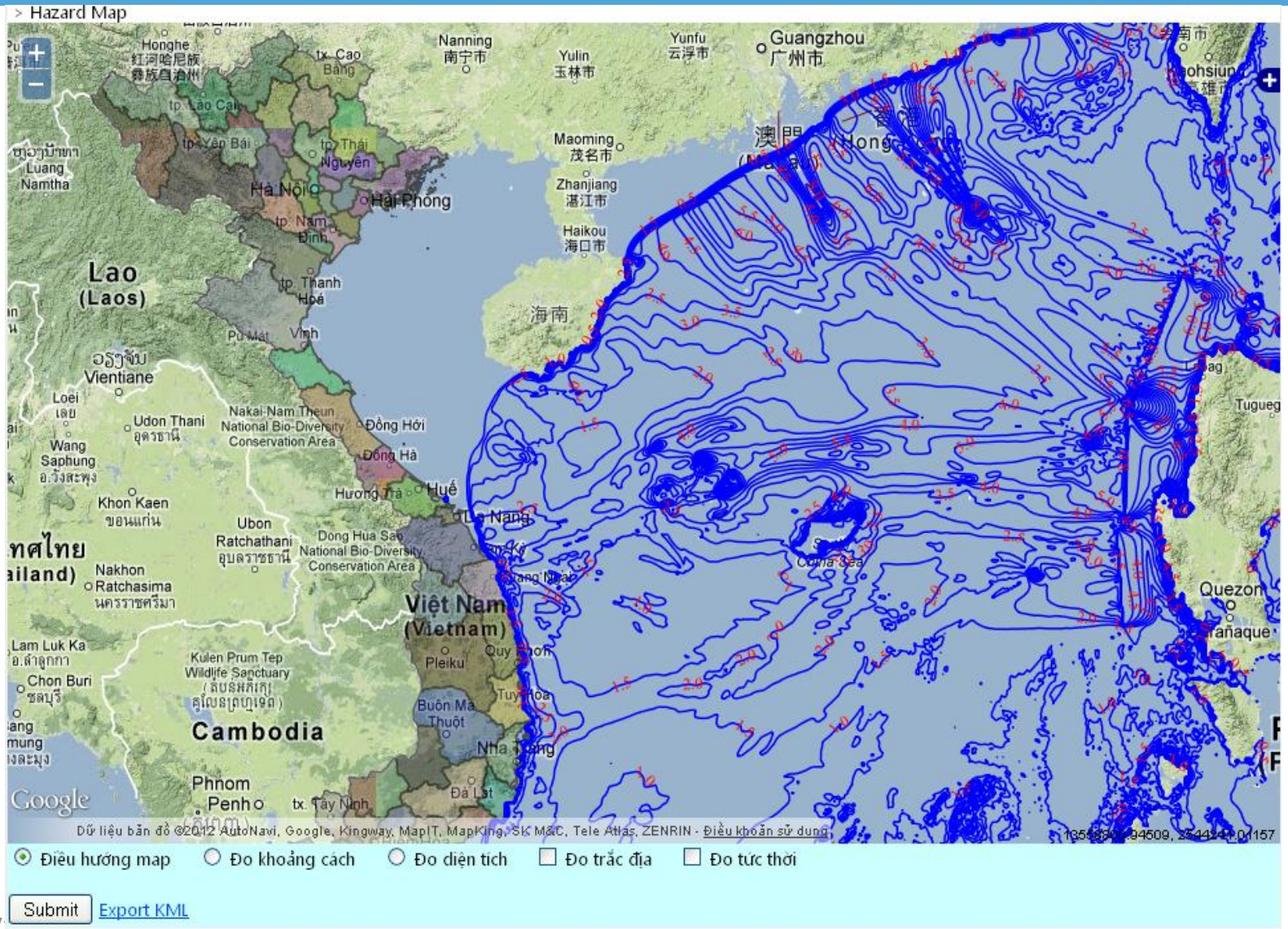


Scale = 1 : 327K 112.78708, 16.33970

ZMAX_LAYER01

FID	ID	HEIGHT
zmax_layer01.2683	2682	9.5

Web portal


> Hazard Map
Hà Nội
Hải Phòng
Lao
(Laos)
Việt Nam
(Vietnam)
Cambodia
Vientiane
Phnom
Penh
Huế
Đà Lạt
Pleiku
Nha Trang
Đồng Hới
Đồng Hà
Hương Trà
Udon Thani
Khon Kaen
Nakhon
Ratchasima
Ubon
Ratchathani
Nakai-Nam Theun
National Bio-Diversity
Conservation Area
Dong Hua Sao
National Bio-Diversity
Conservation Area
Kulen Prum Tep
Wildlife Sanctuary
Guangzhou
广州市
Nanning
南宁市
Maoming
茂名市
Zhanjiang
湛江市
Haikou
海口市
海南
Google
Điều hướng map
Đo khoảng cách
Đo diện tích
Đo trắc địa
Đo tức thời
Submit
Export KML

CHALLENGES

- Don’t have a high-computing center for calculating enough tsunami scenarios
- A current cluster computing nodes/workers are not enough for both calculating and storing large tsunami scenarios
- Network infrastructure is limited for high-computing
- Only few faculties teach cluster, cloud and grid Computing. So only few master and understand them
- Don’t have enough fund for:
  - Investment hardware (computers)
  - Permanent human resources that manage the Grid, Cloud