Development of a quasi-real-time data crawling, archiving, and model workbench system with highly personalized user interface.

Eiji Ikoma
EDITORIA/
Center for Information Fusion,IIS
The University of Tokyo

My mission@IMPAC-T

a) 1-4-1 To develop a quasi real time monitoring system at Mae Waang basin, Wang Thong basin, and Sakae Krang basin

b) 2-5-1 To install test server and develop necessary tools and user interface on test server.

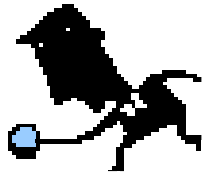
c) 3-4 To develop a system of estimating quasi-real-time risk indices for adaptation measures to water-related disasters under climate change.

Development of

a quasi-real-time data crawling, archiving, and model workbench system

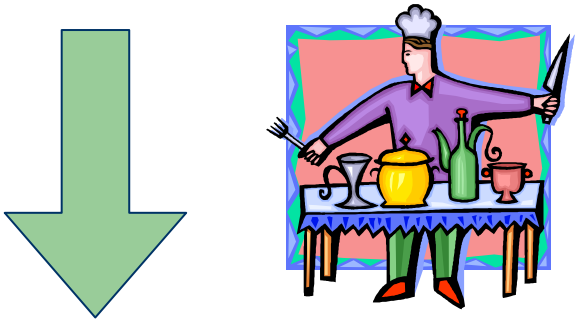
In my talk

- General framework of such system
- Key technology to realize this system @ IMPAC-T
- More important points based on my experience
- How to find research topic @ IT field



Target

- Target data = observation(telemetry) data & model output data
- Target user = Researcher (not IT person)



Final Goal = To develop “practical “ system using target data (not “prototype” !!)

How to use data “efficiently”?

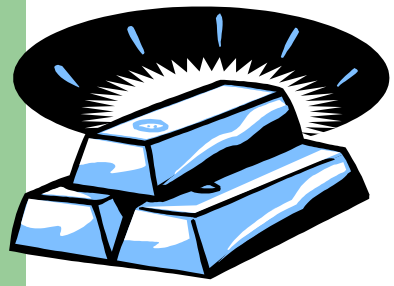
1. Efficient Data Crawling and Archiving
2. Efficient Database Management
3. Efficient User Interface

1.Data Crawling and Archiving

- Observation data is very precious

← cannot be retrieved once it is lost

- Model Output data can be generated again, however it takes much time.



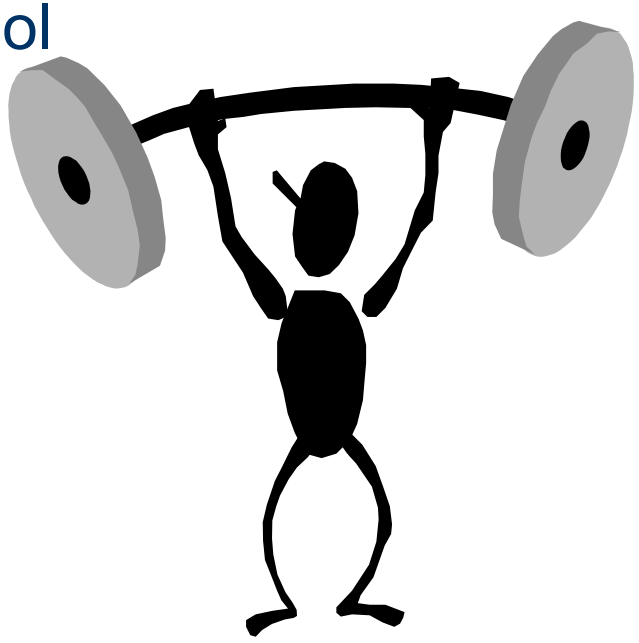
Data must not be lost!

“Strong” System

- Data Crawling
  - Checking technique for data

  =highly precious data loading tool
  - Stable network

- Data Archiving
  - Appropriate RAID system
  - Backup Framework
  - Mirror System



2. Database Management

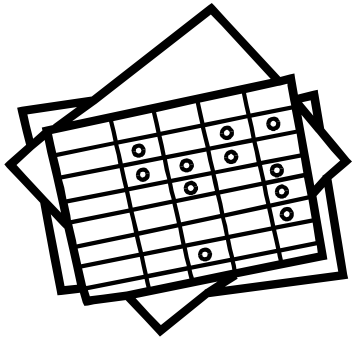
- Search data from DB at High Speed

Observation data is increasing day by day, So the technique for high speed database is quite important.

Most Important point is “Indexing” considering users’ usage

“High Speed” Database

- Efficient database schema and suitable index
  - Features of observation data
    (time sequential, spatial, correlation between elements, update frequently)
  - Considering model output data

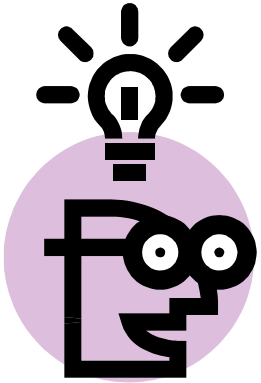


- Relational Database = Relationship with other DB

3. User Interface

- Easy-to-use for everyone

- *Searching data*
- *Downloading data*
- *Visualizing data*
- *Running Model*
- *....*

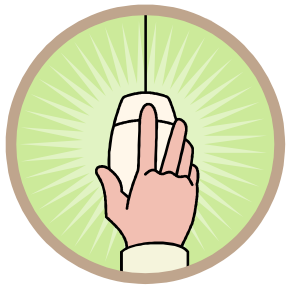


User Interface is the key of whole system

“Easy, Flexible, Efficient IF”

- Easy
  - How to search? by time, place, element….
  - How many condition?
    → Complicated condition by easy operation
- Flexible
  - More complicated condition i.e. “ Temp. > 30℃ AND Prep. < 5mm/h”
    → Flexible Interface to specify conditions
- Efficient
  - For multi users, complicated condition

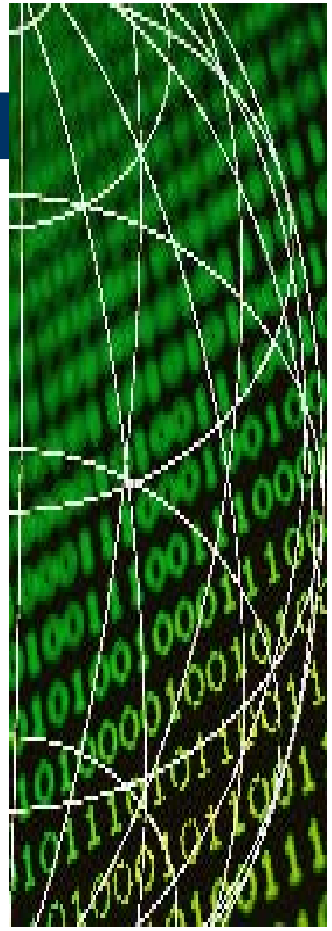
→quick response and light operation



Model Workbench

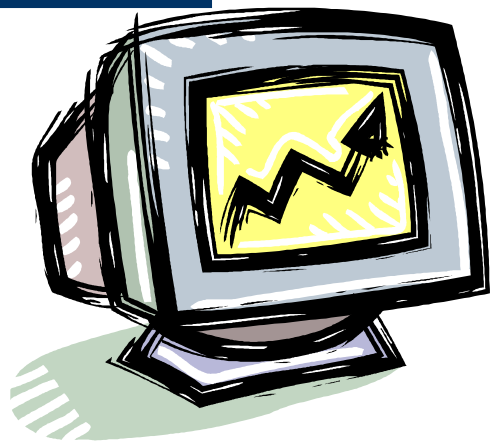
- Data Download → Data for Application
- Integration of data archive system and model execution environment is efficient for server and users

- Reduction of the amount of data transfer
- Reduction of the load of the client computers



Information Fusion of Data and Application

Advanced Interface



- Personalized Interface

← Analysis of each user log

- Understanding of user tendency

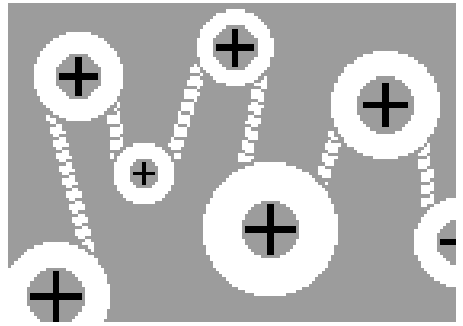
→ classification of users

+ Suitable Interface for user “group”

- Feedback = Self-growth Interface

Comfortable Interface for all users

Further more



- Non-stop system is ideal in this project. It requires much more examination for hardware and software and “administrators”.
- We need to consider the framework not only for development but also for running of this system.
- As most IT researchers know, to run system continuously is one of most difficult mission.

Running this system=Research?

- To develop and run system only is not judged as a good research @ IT field
- I'm considering which point is "unique" on this system……..??
  - Frequently updated data?
  - Comparison between model and ….
  - Analysis users' log and personalize?
- To find these point, I'd like to discuss more with Thai-side IT researchers.

Summary

- Data Collection (loading)
- Data Archiving
- Data Management
- Data for Application = Model Workbench
- Data for users = Search/DL/Vis Interface
- Running system continuously = importance of framework for this system
- Research topic for IT researchers

All System should be for users.

The best system is a system which responds to a users' request most.