DESIGN OF COOPERATIVE VISUALIZATION ENVIRONMENT WITH INTENSIVE DATA MANAGEMENT IN PROJECT LIFECYCLE

Kenji ONO and Jorji Nonaka

High-performance Computing Team
Integrated Simulation of Living Matter Group
Computational Science Research Program
RIKEN

Ultrascale Vis Workshop 2008

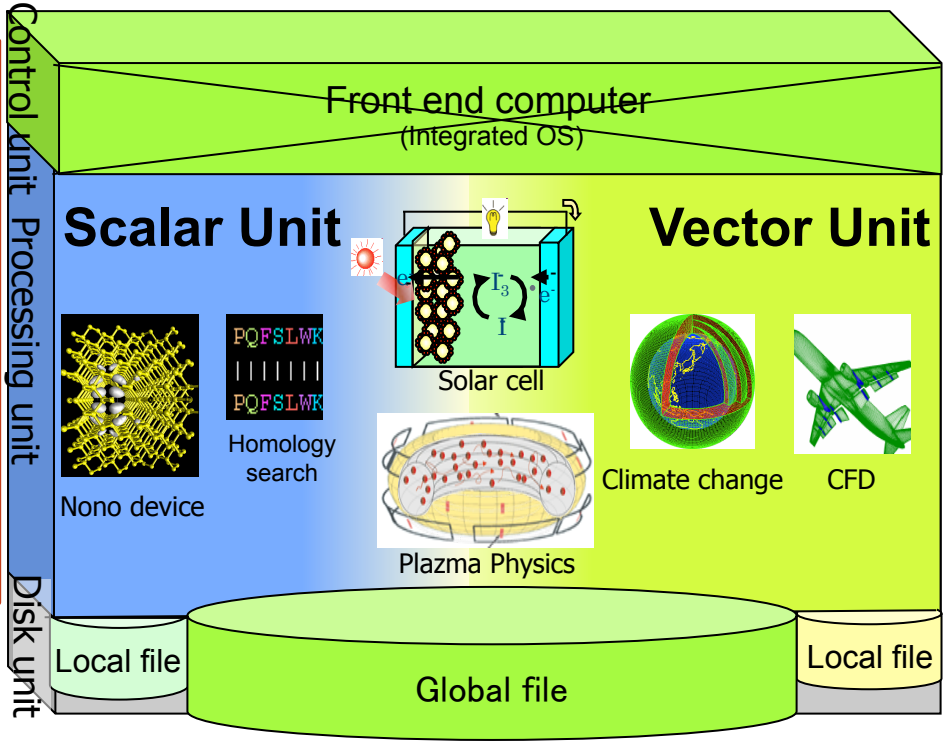
TOC

- Supercomputer development project
  - Computer Hardware development
  - Grand Challenge in Life Science
- Post-process system
  - Current status of system design

NEXT-GENERATION SUPERCOMPUTER

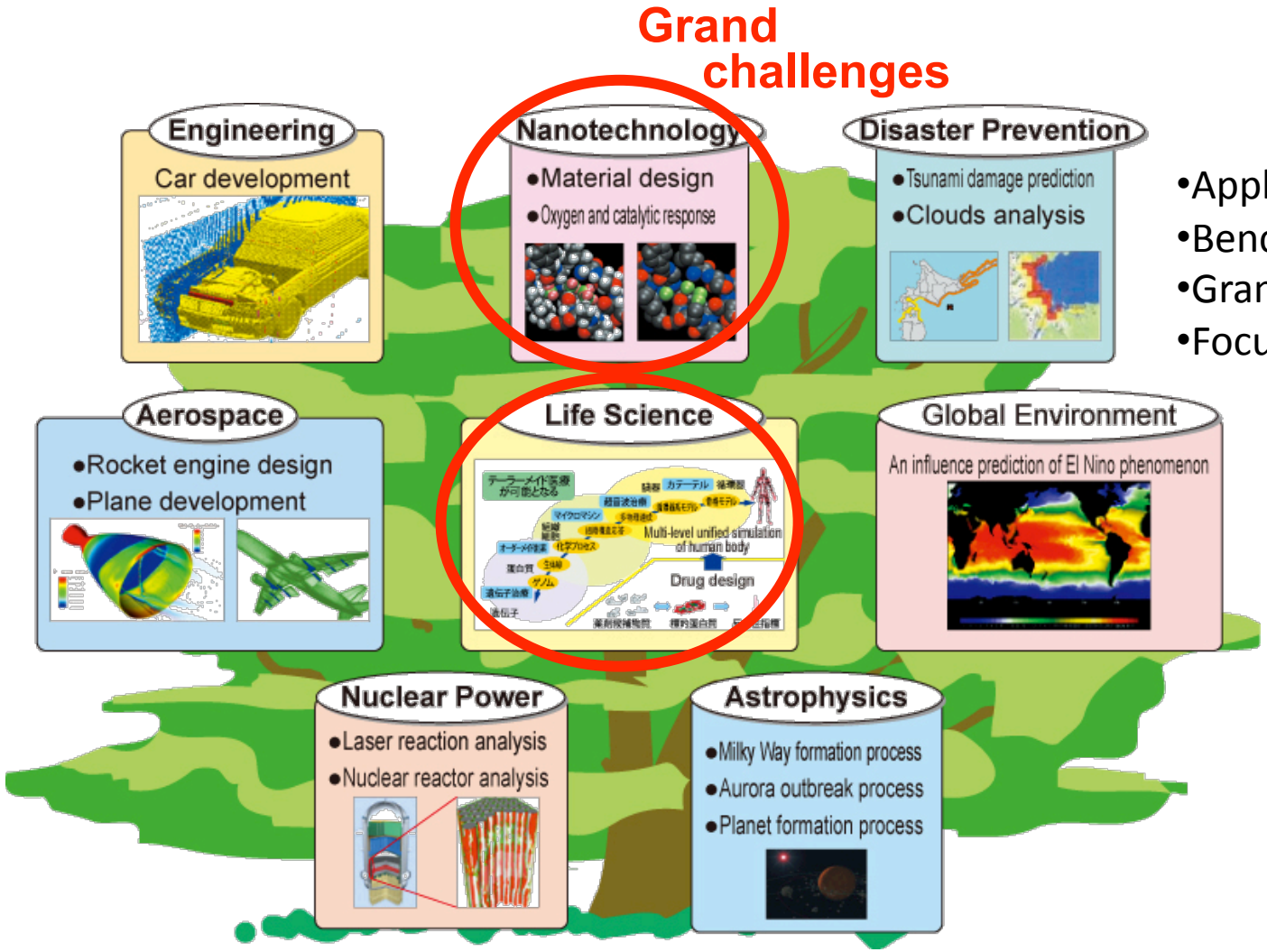
- Development
  - Under initiative of MEXT
  - Riken is responsible for dev. in collaboration with vendors
- Hybrid System composed of Scalar and Vector Units
  - LINPACK 10PFLOPS

Scalar Unit
Data base, homology search
Similarity with PC
- high speed low power CPU
- New strong network for enormous parallelism



Vector Unit
Vector processor based on the Earth Simulator
- New generation low-power vector processor with optical network

EXPECTED MAJOR APPLICATIONS



- Applications
- Benchmark
- Grand challenge
- Focusing on life science

ISSUES TO BE CONSIDERED

- Large-scale nature of data
  - computational space, time-varying, multivariate
  - Expected data size is order of 1PB (Peak)
- Data Tsunami
  - How to do vis., data processing, analysis?
  - Depend on each researcher
  - Various scenarios
- Complex hardware
  - Various configurations

CURRENT STATUS AND GOAL

- System design
  - To develop user-friendly post-processing system
  - Derive useful information from Large-scale dataset
  - Easy to use
  - Enhance productivity

  - To assist scientific knowledge and understanding physical phenomena

- Operating this post-process system to assist research

ISSUES ON POST-PROCESS OF LARGE-SCALE DATASET

- **Size of dataset**
  - Space, Temporal data, multivariate
- **Distribution of dataset**
  - Distributed parallel, GRID
- **Complexity of HW system**
  - Heterogeneous environment, File system, Network

- High cost of data copy, movement
  - MMU/HDD capacity
- can not move any large-scale data
  - need appropriate tools to access large-scale data

- Many cumbersome procedures
  - File handling, preparation of process
  - need an environment to focus on “THINKING”

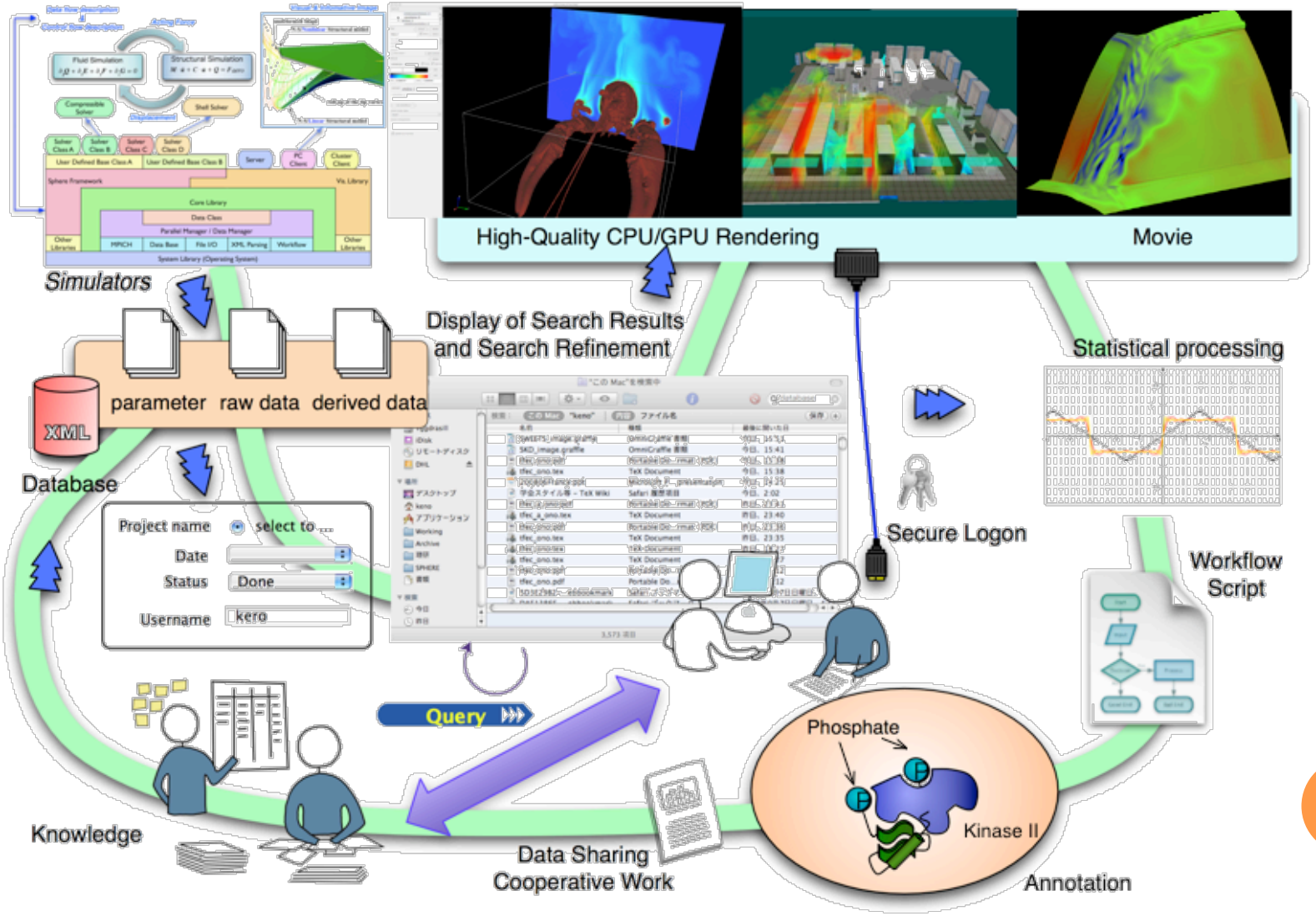
ORIENTATION OF POST-PROCESS

- do note move data
  - beyond that…
    - Sharing data, collaboration
    - Remote, virtual organization

- Data intensive service
  - Data access, visualization, analysis, processing
  - Sharing data, results, knowledge, resource
    - Data repository for group
    - Data browse, search, break down

  - Comprehension of phenomena
  - Sharing information, knowledge

WHAT WE WANT TO DO



SWEETS
SCIENTIFIC KNOWLEDGE DISCOVERY TOOLS

- Script based loosely coupling
  - script glues modules in sub-systems
  - scalability, flexibility and sustainability
- Sub-system
  - Visualization
  - Data / project management
  - Automation of a routine task using a workflow
  - Data sharing
  - ...
  - module works independently, but provide full capability when working together
- Efficiently extract useful information from simulated results

Tools to obtain knowledge, then to magnify value of simulations

REQUIRED TASKS FOR RESEARCH

Thinking

Cumbersome

Visualization
Data analysis

Task

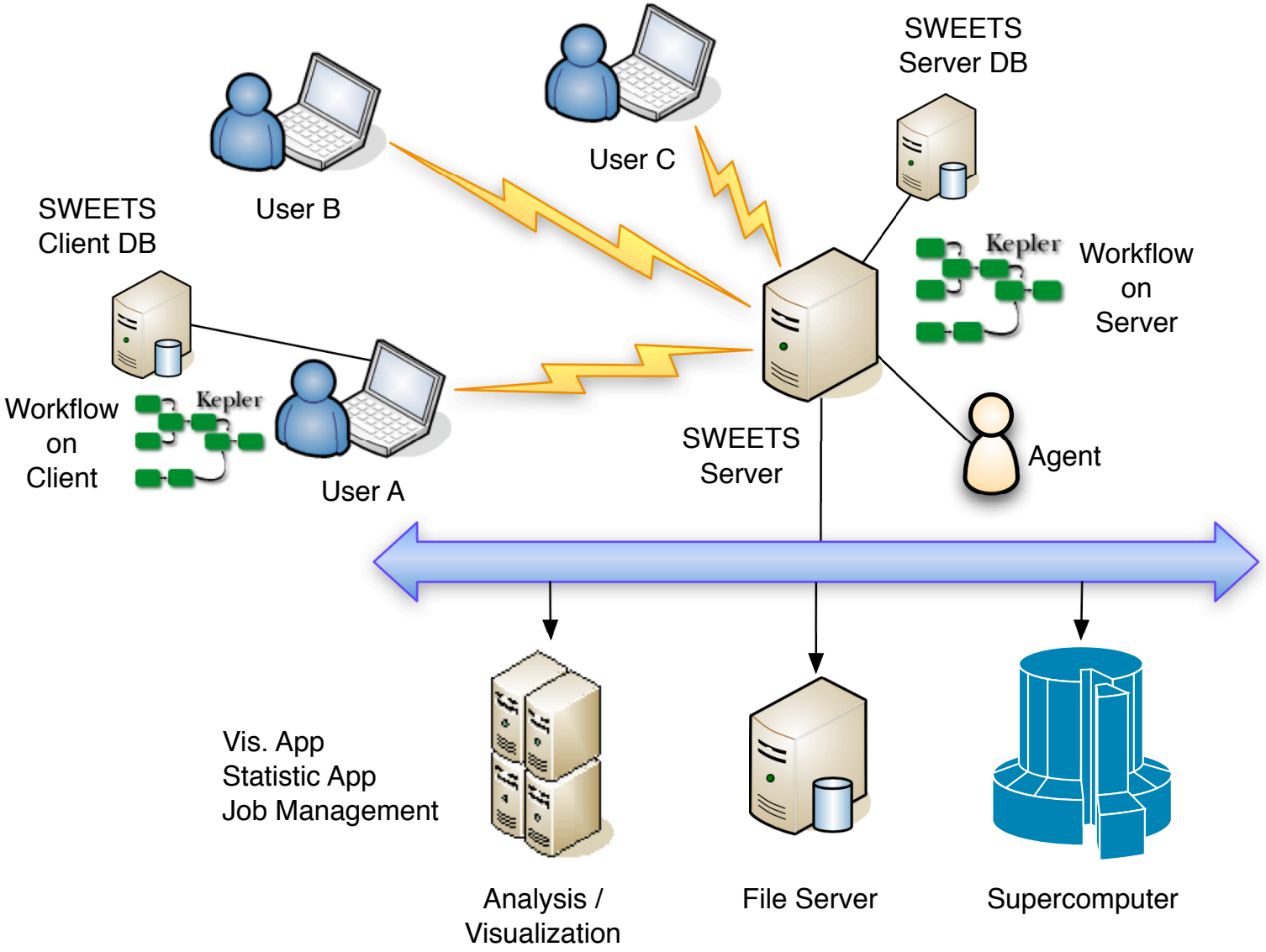
Visualization
Data/project mgmt
Automation
Data sharing
Analysis

Service
(sub-system)

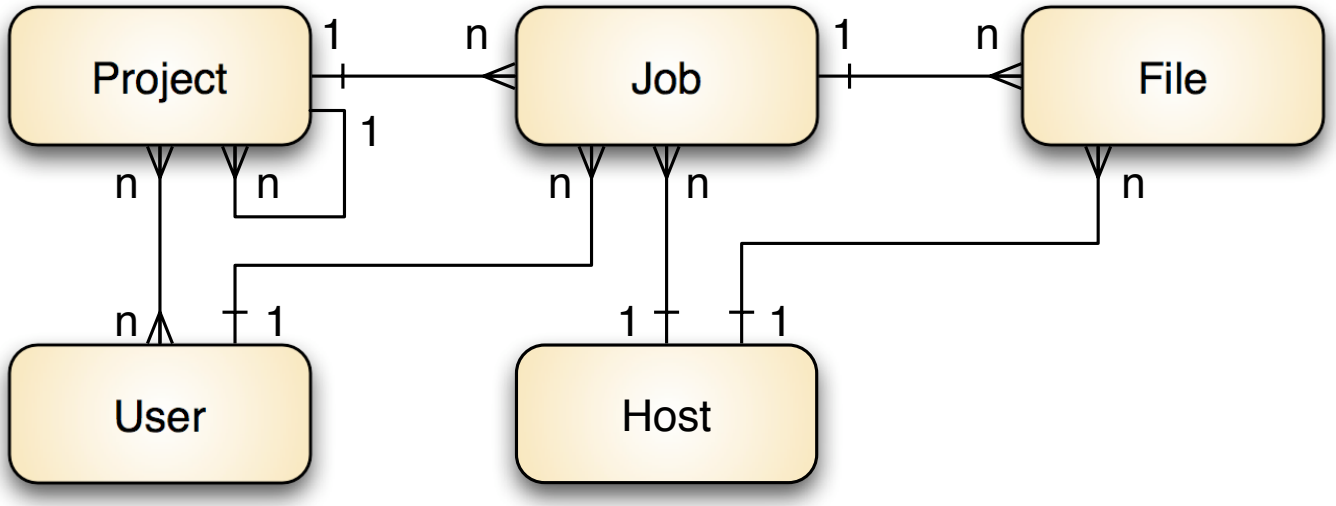
File access
Parameter edit
Data copy/move
Submit a batch job
Program launch
…

Low level
procedure

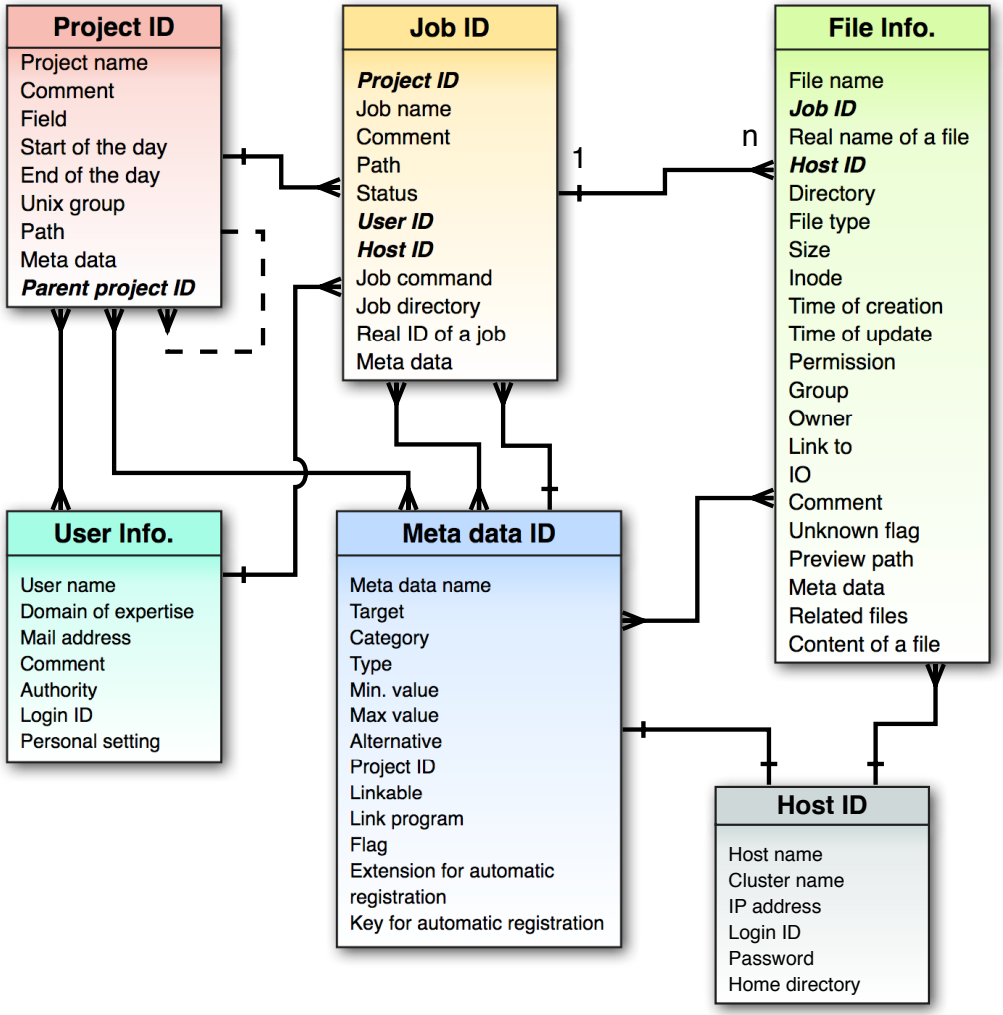
AN EXAMPLE OF HARDWARE CONFIGURATION



A CONCEPT MODEL OF DATA ENTITY



DATA ENTITY OF SWEETS SYSTEM



PREREQUISITE OF VIS. SUB-SYSTEM

Various scenarios for each researcher's approach

- **Remote or local visualization**
  - Provides unified environment by a common client
- **Real-time or post visualization**
  - Basically, file based visualization
- **Interactive or batch visualization**
- **Software or hardware rendering**

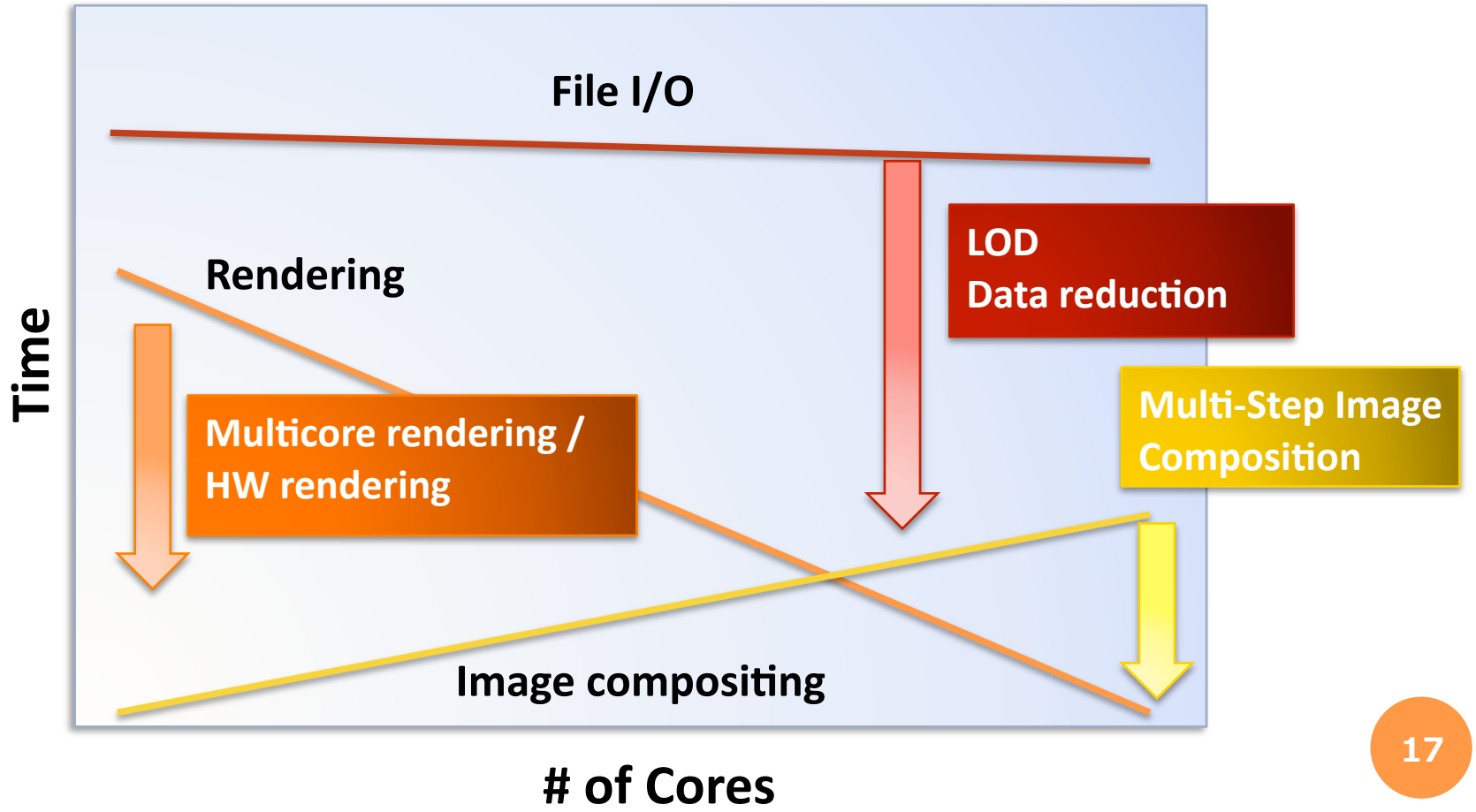
- **Large-scale data handling**
- **Parallel rendering**
- **Platform**
  - Linux, Windows, Max OSX, PC cluster, Supercomputer

2-WAY SCENARIO

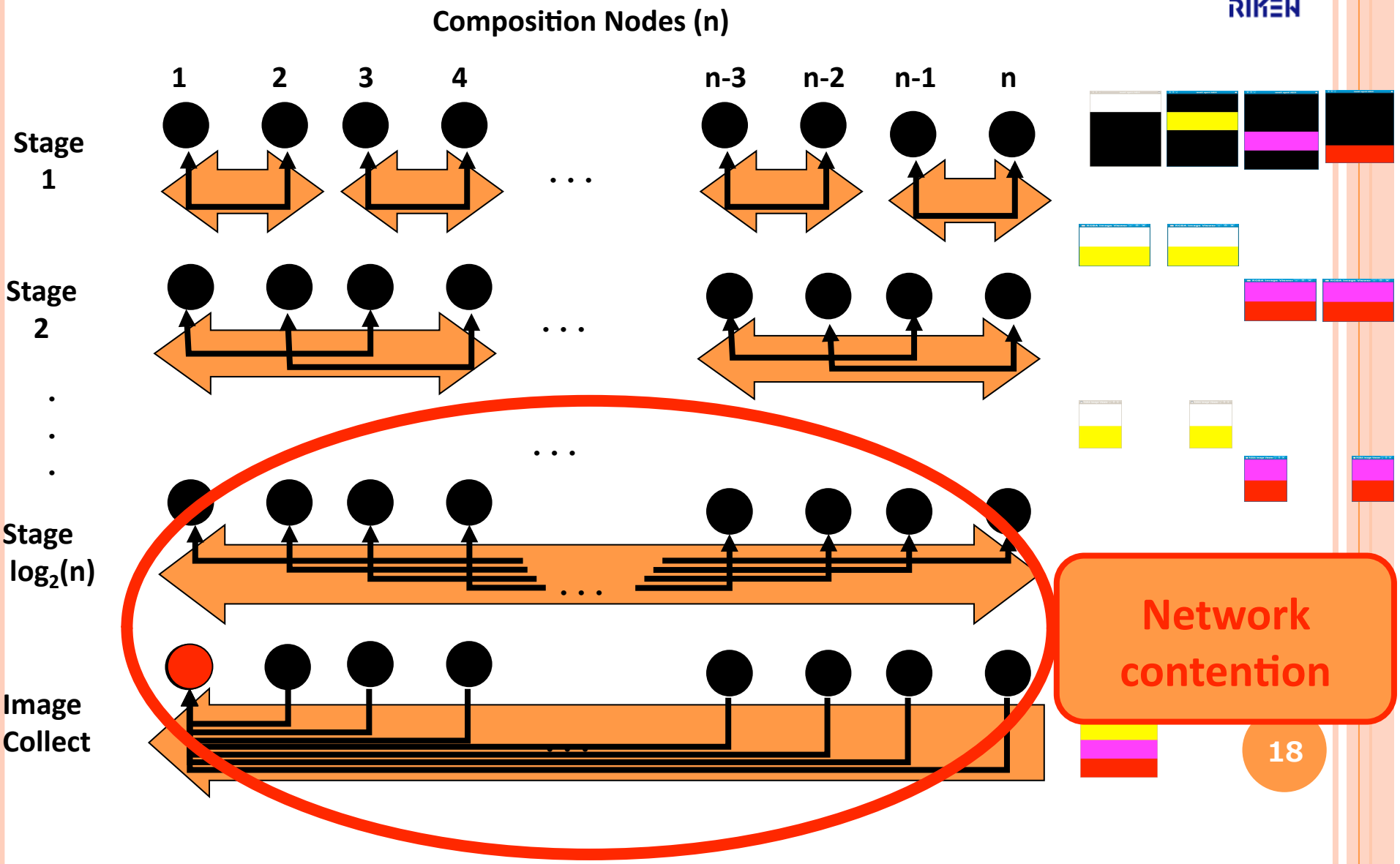
- 1st step : Visualization on supercomputer
  - Sever client system
    - with basic visualization function
    - Data reduction, ROI
  - Operation is only batch job by its policy

- 2nd step : Visualization on CPU/GPU cluster
  - Capability of interactive visualization by GPU
  - Relatively small MMU (Data reduction is necessary)
  - Reduce risk of development
  - Reduce utility time of supercomputer
  - can use existing software, COTS

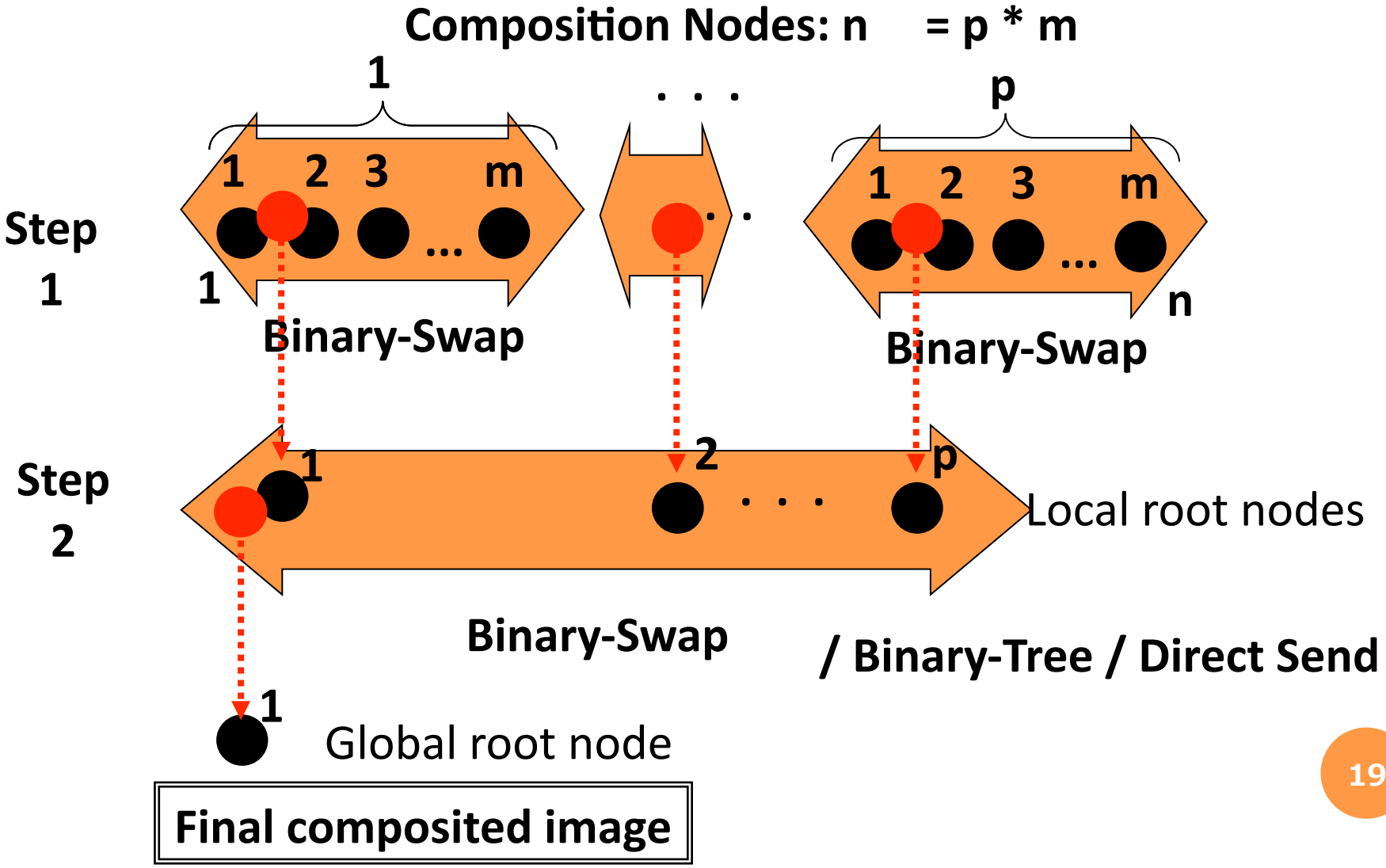
STRATEGY: INTERACTIVITY AND SCALABILITY



BINARY-SWAP IMAGE COMPOSITION

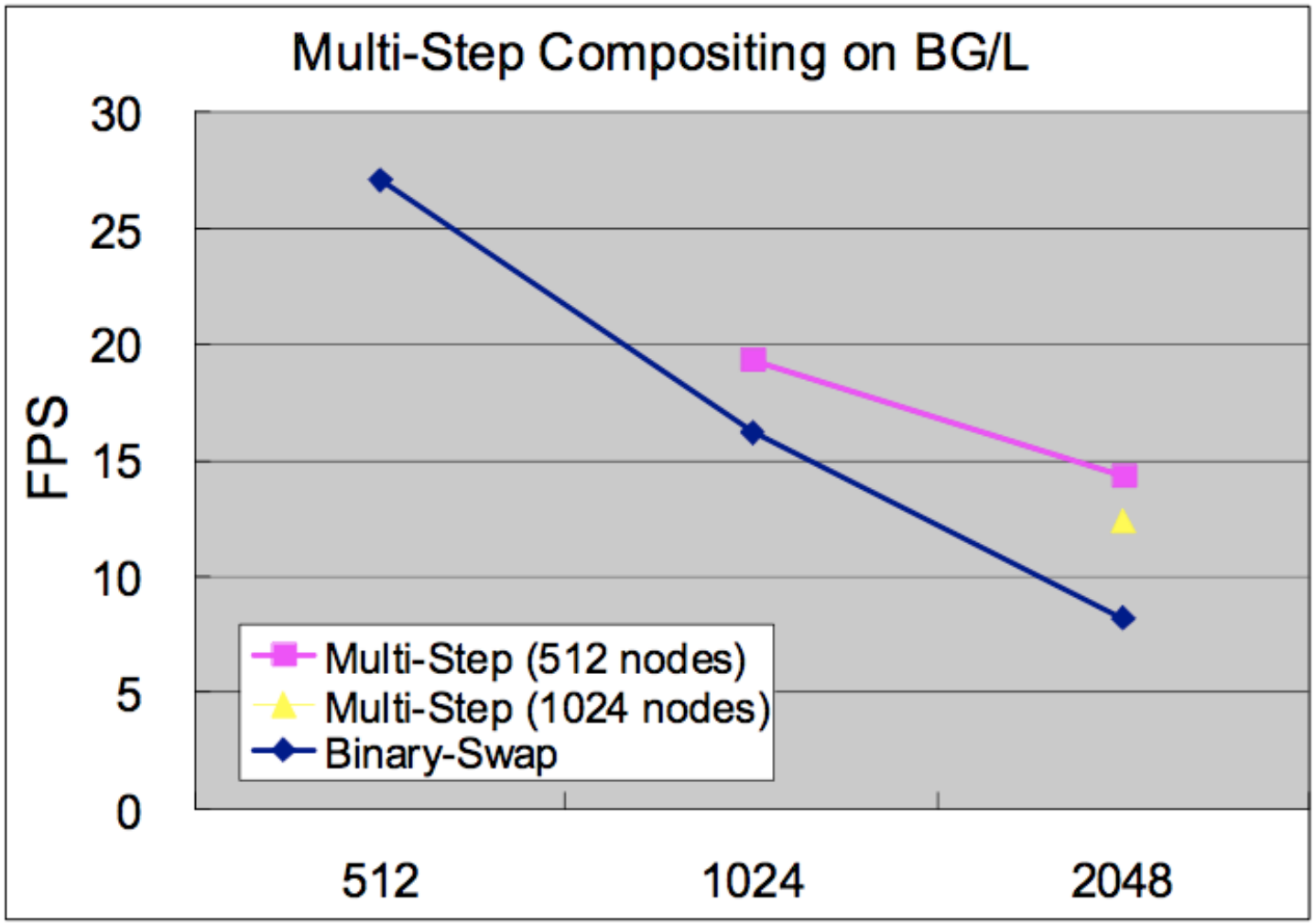


Multi-Step Image Composition



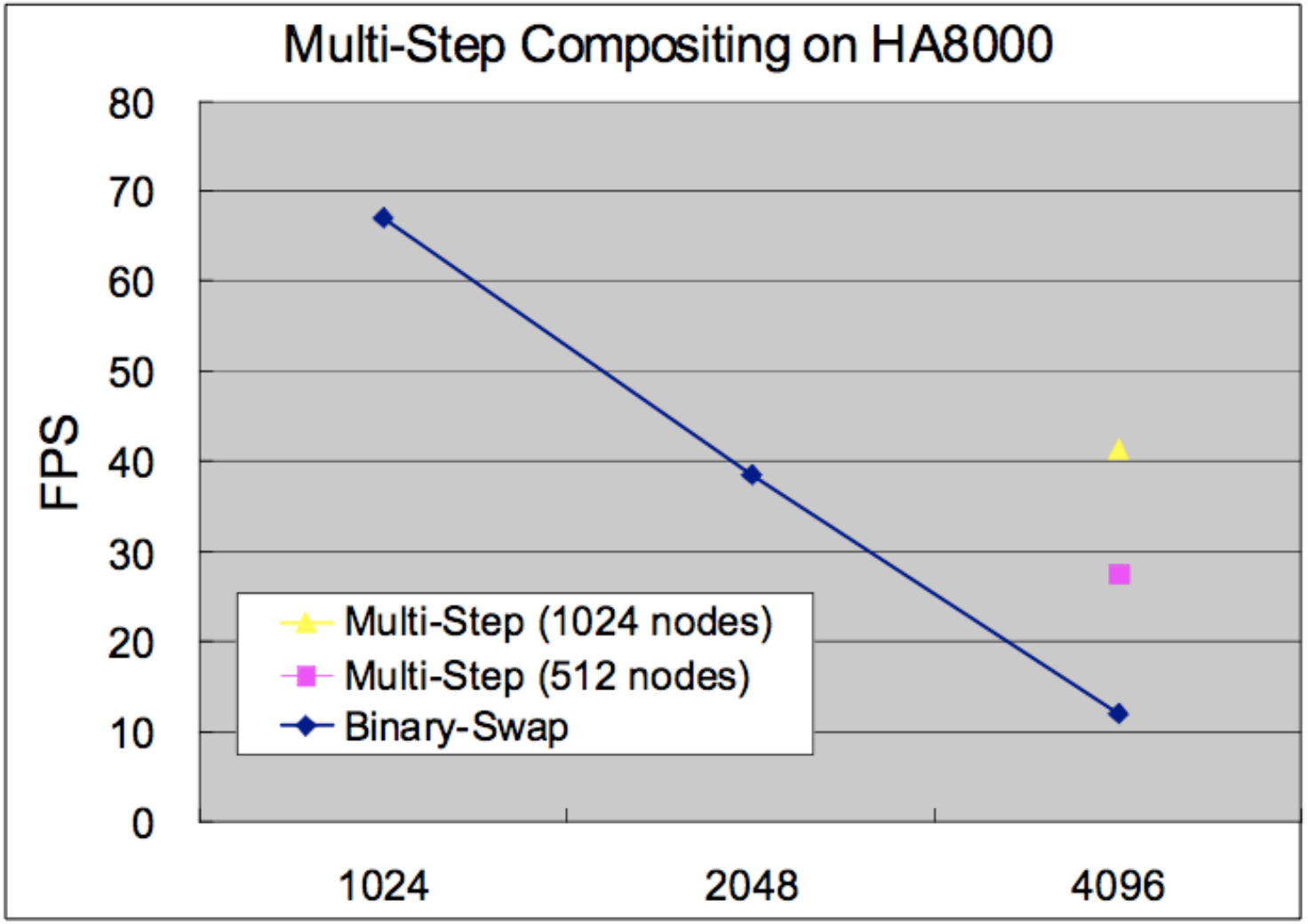
MSIC BG/L (RIKEN)

BlueGene / L Dual core x 1024 nodes



MSIC - T2K (Univ. of Tokyo)

AMD Opteron Quad Core x 4 Sockets



OTHER SUB-SYSTEMS

- Workflow
  - Kepler
- Data base
  - RDB ? XML feature
- Scripting
  - python ... plays glue
- Analysis
  - R
  - user program
- Other useful existing tools...

CONCLUDING REMARKS

- Design and current development status
  - to assist discovery and understanding
  - Database centered structure
  - sub-systems take part in the system
    - Visualization
    - Workflow
    - Other programs
  - resource mgmt and access control provide data sharing
  - reuse of existing useful software
  - script base
- Supercomputer will be fully operational on 2011
- Combining existing software, enhancing capability