From desktop to grid application

Viet Tran
Institute of Informatics
Slovakia

Information Society

Porting applications to Grid

- **Applications that are easy to port to Grid**
  - Applications that already run on Linux clusters
  - Applications that can run on distributed systems
  - Applications that can run in batch modes

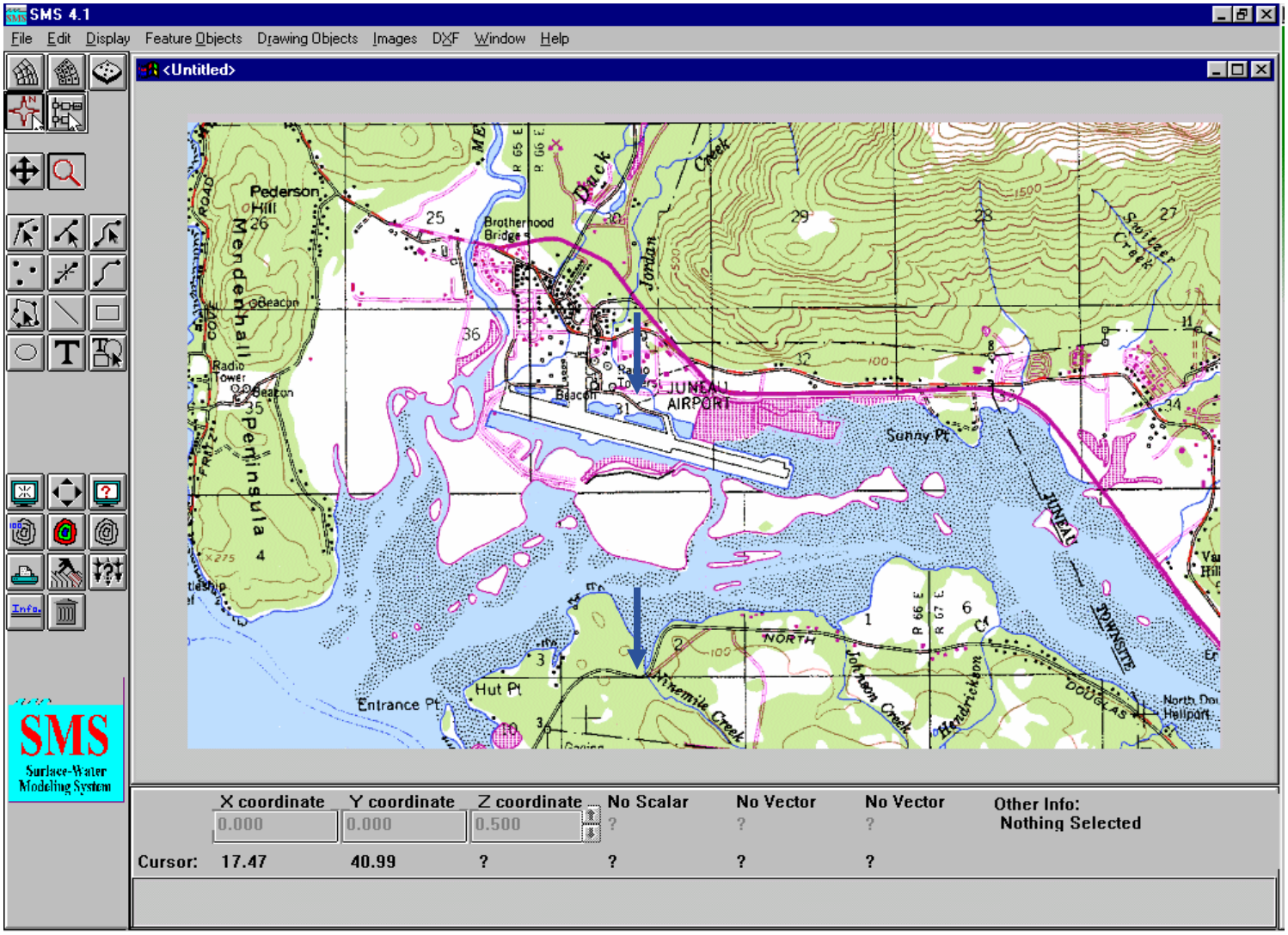
- **Applications that are difficult to port to Grid**
  - Applications that require interactive mode
  - Applications that are tightly connected with graphical user interfaces
  - Applications that use special (commercial) libraries in Windows

From desktop to Grid

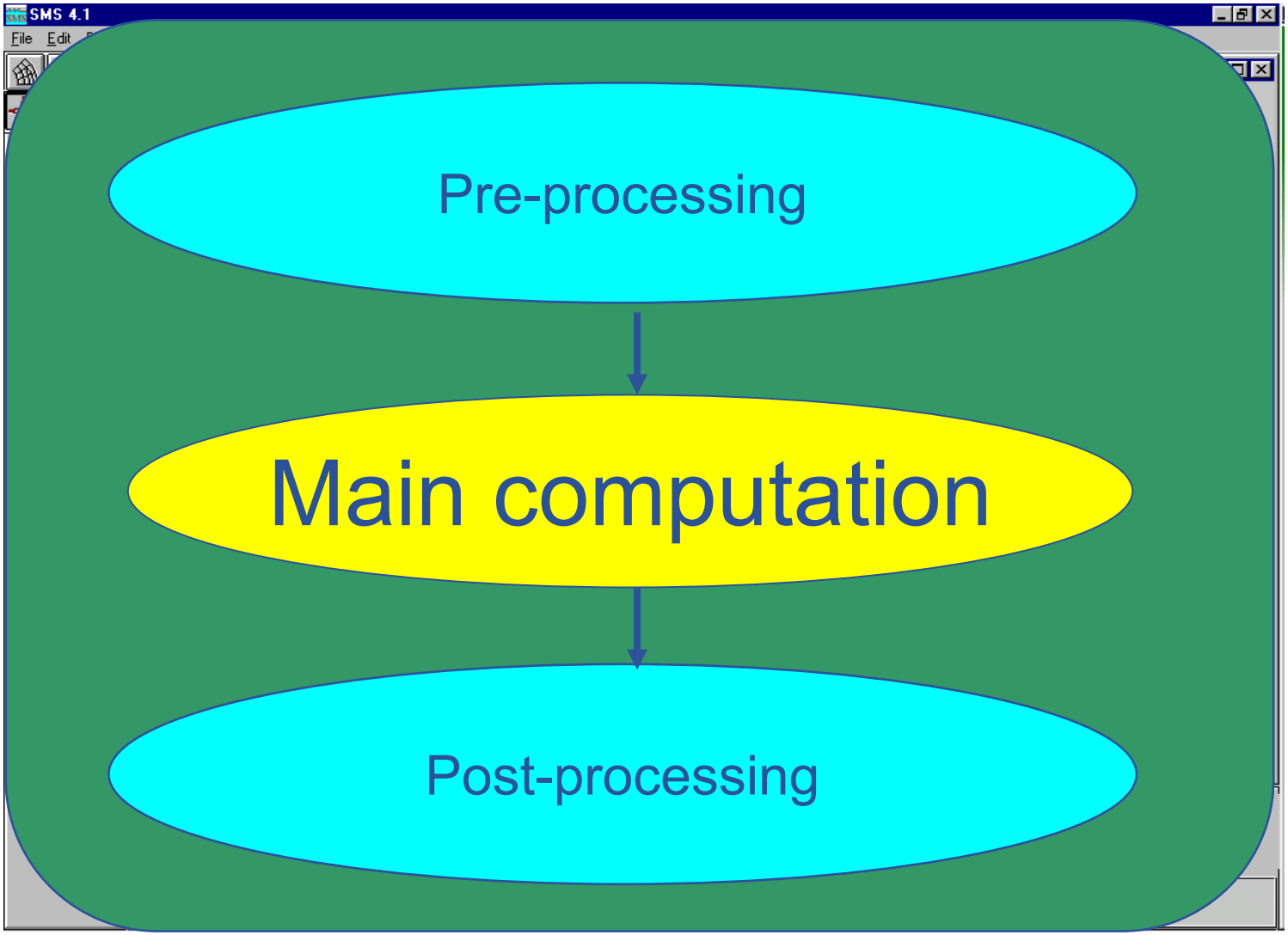
- Separating computations from user interfaces
- Remote processing
- Parallelization
- Using portal
- Using collaborative services

Separating computation from user interface

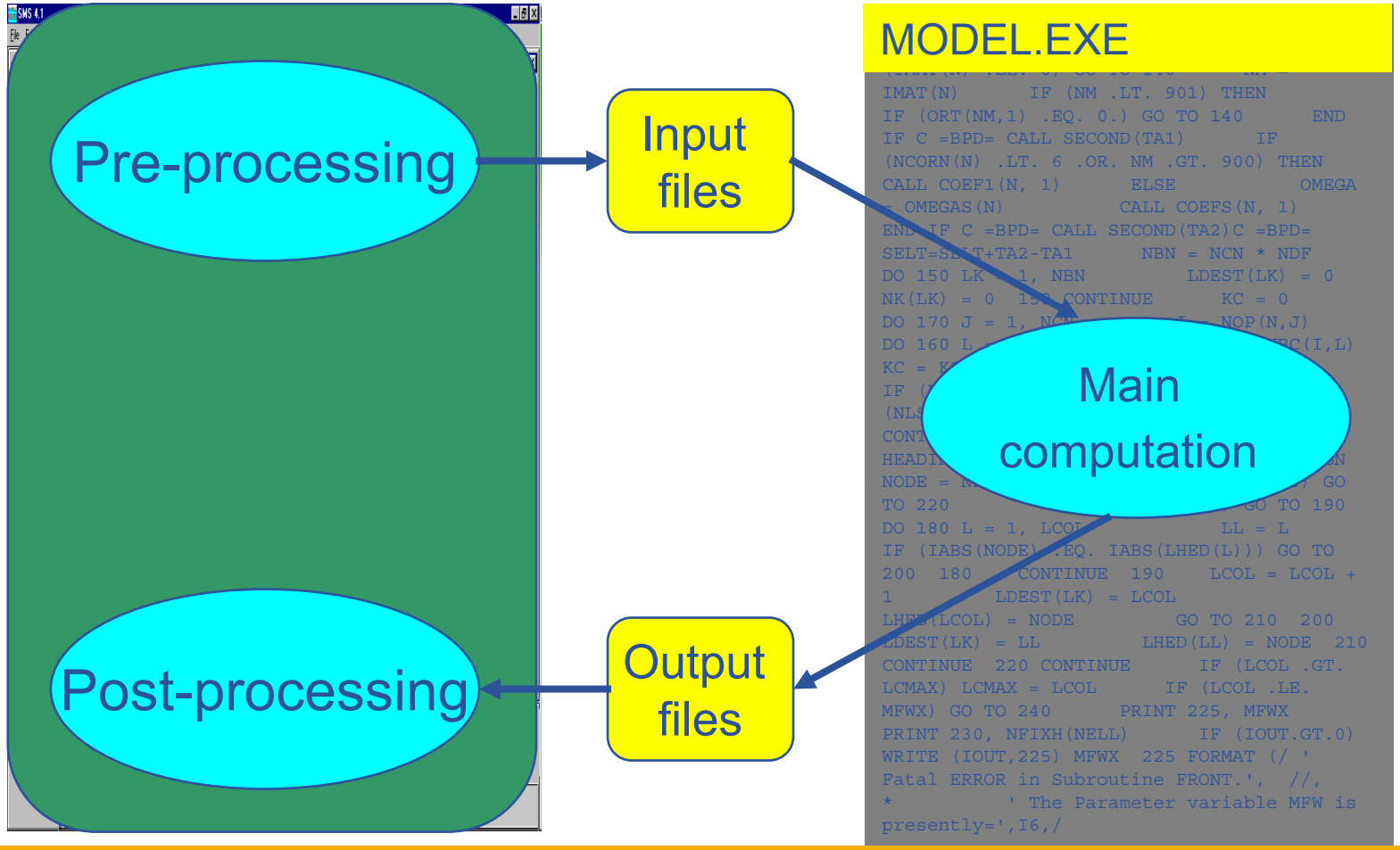
Desktop application



Desktop application

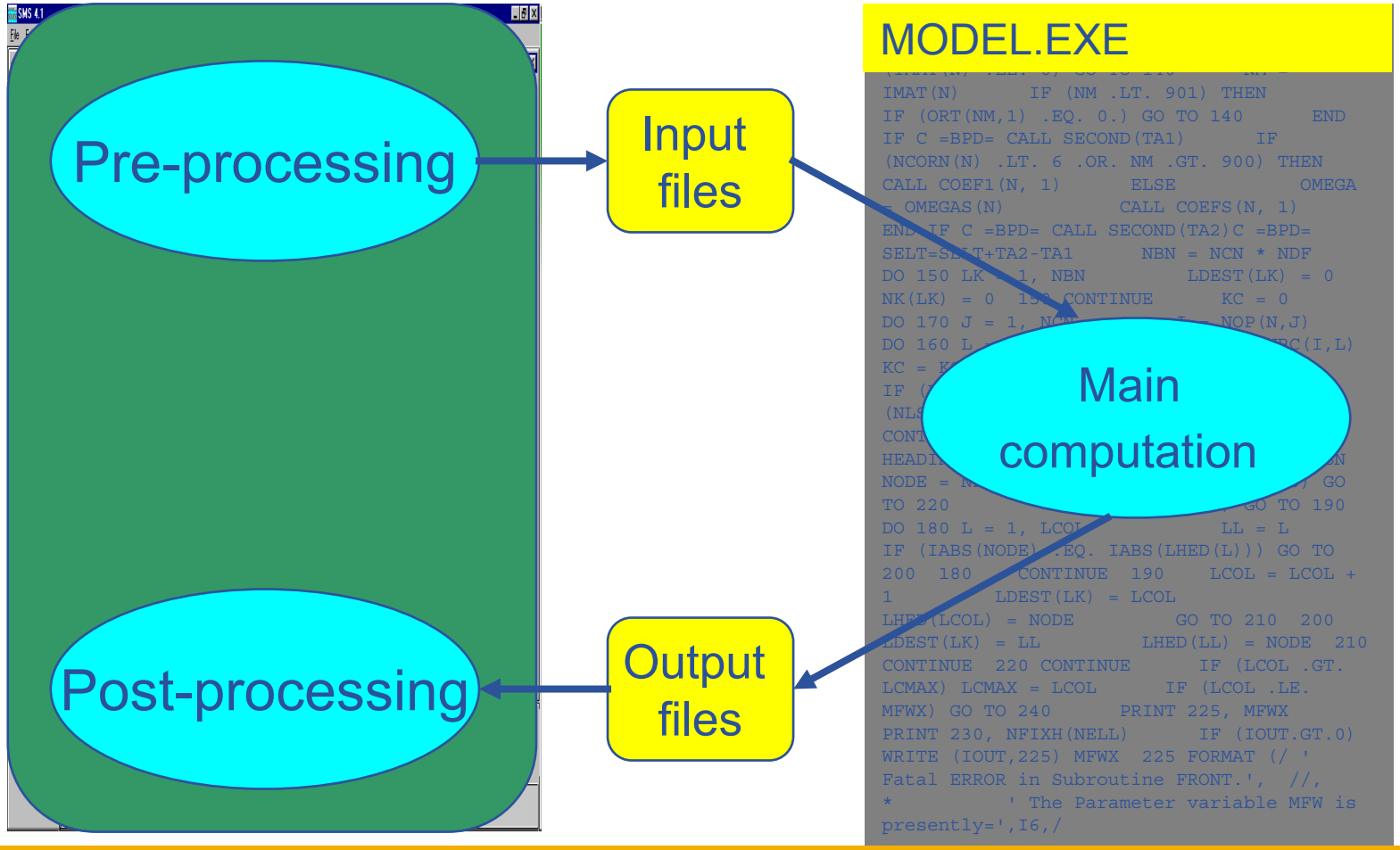


Separating computation

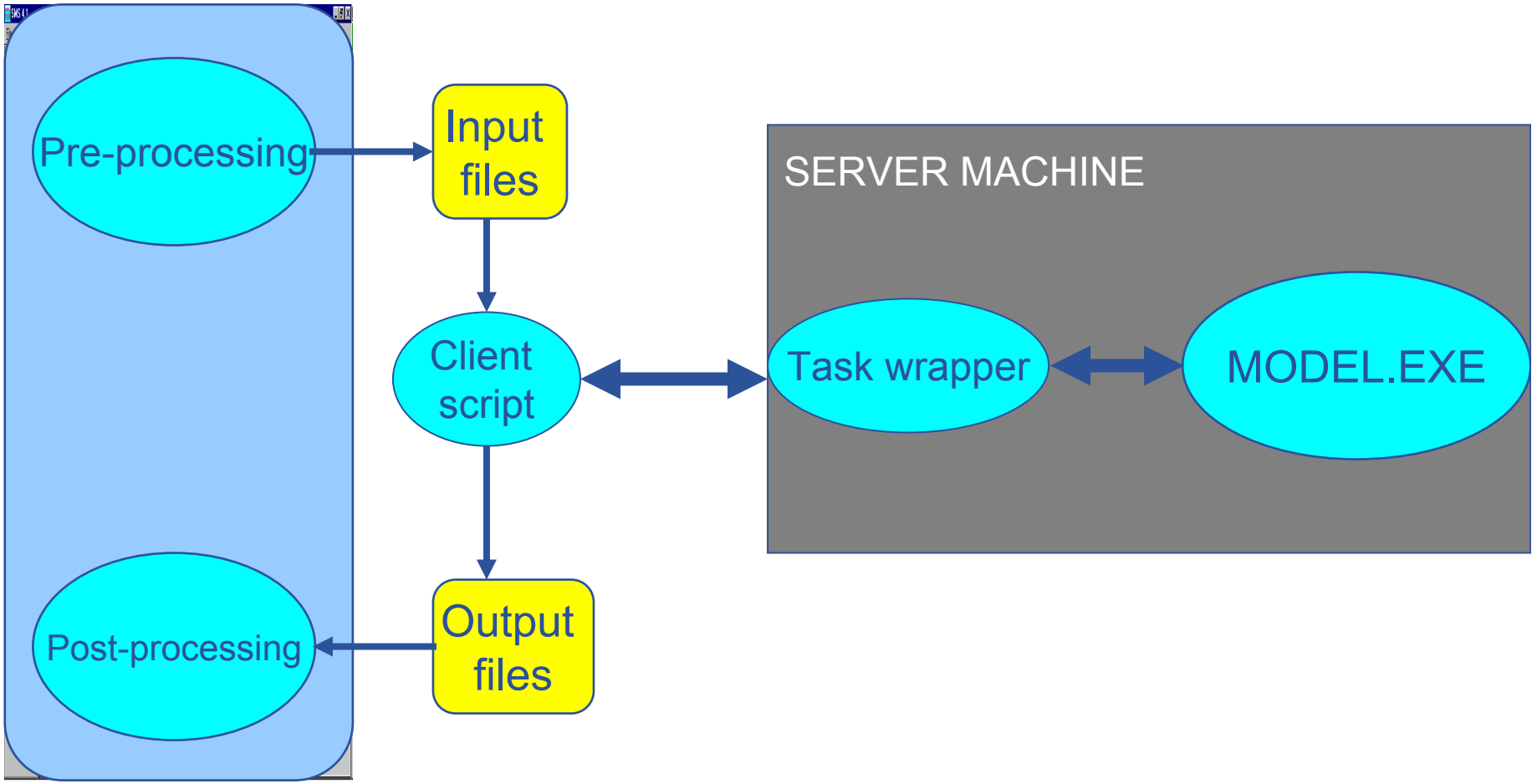


Remote processing

Desktop application



Remote processing



Example of task wrapper

```
globus-url-copy gsiftp://storage.sav.sk/code/model
                file://$PWD/model

globus-url-copy gsiftp://storage.sav.sk/data/input-data
                file://$PWD/input-data

./model input-data output-data

globus-url-copy file://$PWD/output-data
                gsiftp://storage.sav.sk/data/output-data
```

Parallelization

Parallelization

- **Parallelization on single cluster**
  - MPI program
  - No grid-related instructions, just cluster computing

- **Parallelization on multiple clusters**
  - MPI-G2 program
  - Only applicable for very coarsely-grained computation because of high latencies of inter-cluster communications
  - Modified mpirun script for starting process on multiple clusters

Parallelization

- **Distributed computing**
  - Multiple executions of the same program with different input data
  - Very suitable for grid computing (linear speedup, practically no limit on number of CPUs)
  - Typical algorithms: Monte-Carlo, searching
  - Typical applications: High energy physics, bioinformatics, SETI
  - Usually does not require modification of the executable binary, only add pre- and post-processing (dividing input domain and collecting output data)

Using portal

What is portal

- Web-based User interface
- Act as gateway between users and services/resource
- Provides personalization, single sign-on, content aggregation from different sources

Advantages of portal

- **Uniform user interface**
  - Same interface for all systems
- **Universally accessible**
  - Everywhere with internet connection
- **Low requirements on clients**
  - Web browser and network connection
- **Additional security layer**
  - Additional authentication and authorization in portal
  - User can perform only predefined actions on portal

Why portal

- **Accessible from everywhere**
  - Like electricity, just connect to Internet and have access to all resources
- **“Meeting point” for collaboration**
  - People in virtual organizations need collaboration, exchanging resources, data, …
- **Hide implementation details**
  - Independent from concrete grid middleware and testbed
- **Additional security**
  - Security is one of primary concerns of grid computing

Available portal technologies

- **Portlet (JSR 168)**
  - An autonomous entity in portal
  - Has its own content (a small “window” in working area)
  - Has its own behavior (independent from portal)
  - Is configurable (has its own configuration and status)
  - Is manageable (can be maximized, minimized, can be inserted or removed from portal view area)
  - In other word, portlet is a small “application” in portal
- **Generic portal framework uPortal, Jetspeed**
- **Grid portal framework Gridsphere, OGCE**

Using collaborative service

Collaborative services

- **There are many computational resources, many applications, many data files in grid**
  - Need coordination and management
- **Some useful services**
  - Resource broker: finding suitable computation resource for a given task
  - Replica catalog: finding best replica of a data file
  - Workflow: executing jobs with data dependence
  - Metadata catalog: managing data according descriptions
  - Information index: information about grid

Summary

- Separating computations from user interfaces
- Remote processing
- Parallelization
- Using portal
- Using collaborative services

Thank you for your attention