dynamic software & engineering

science + computing

an atos company

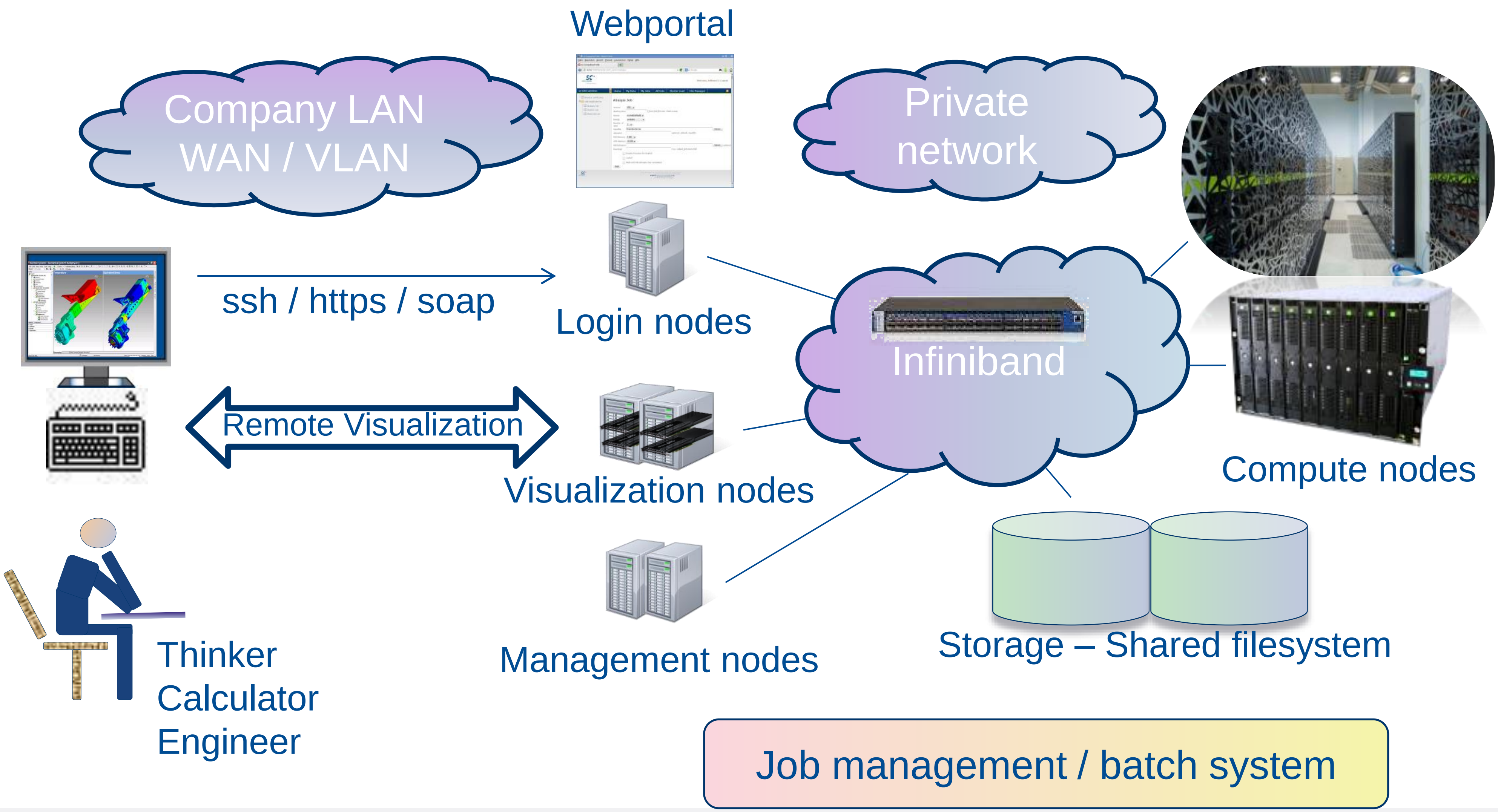
optiSLang jobs on a compute cluster

5.11.2015

Philipp Pasold
Dynardo GmbH

Dr. Ingo Seipp
science + computing ag
IT-Services and Software Solutions for Complex Computing Environments
Tübingen | München | Berlin | Düsseldorf

IT environment



Webportal

Company LAN
WAN / VLAN

Private network

ssh / https / soap

Login nodes

Infiniband

Remote Visualization

Visualization nodes

Compute nodes

Thinker
Calculator
Engineer

Management nodes

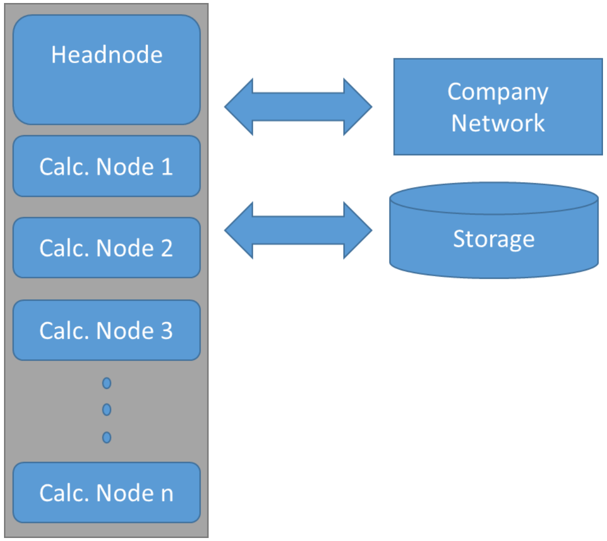
Storage – Shared filesystem

Job management / batch system

IT environment

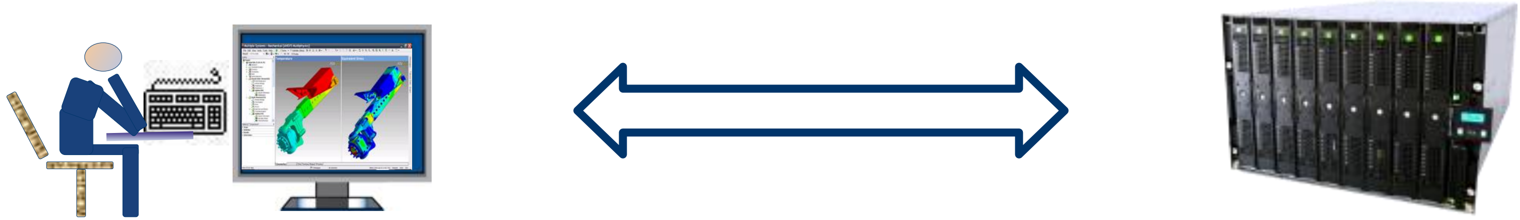
Using the cluster

- Access
  - ssh to Login-node
  - Job-Webportal (e.g. EnginFrame)
- Site policies for executing jobs on the cluster
- Submit- and Job-execution scripts
  - Customer defined jobscripts
  - scsub / flowGuide2 frameworks
- Shared file system / job execution on local disk



optiSLang batch requirements

- Most general setup
  - optiSLang node cannot run cluster jobs directly
  - ssh access to Login-node
  - No shared filesystem between optiSLang node and cluster
  - optiSLang host not configured as batch system node
  - Requirement to support customer’s batch system and job submission methods
- ssh-Access password-less with private key authentication
- Independent interface design
- Integration in optiSLang to specify job requirements

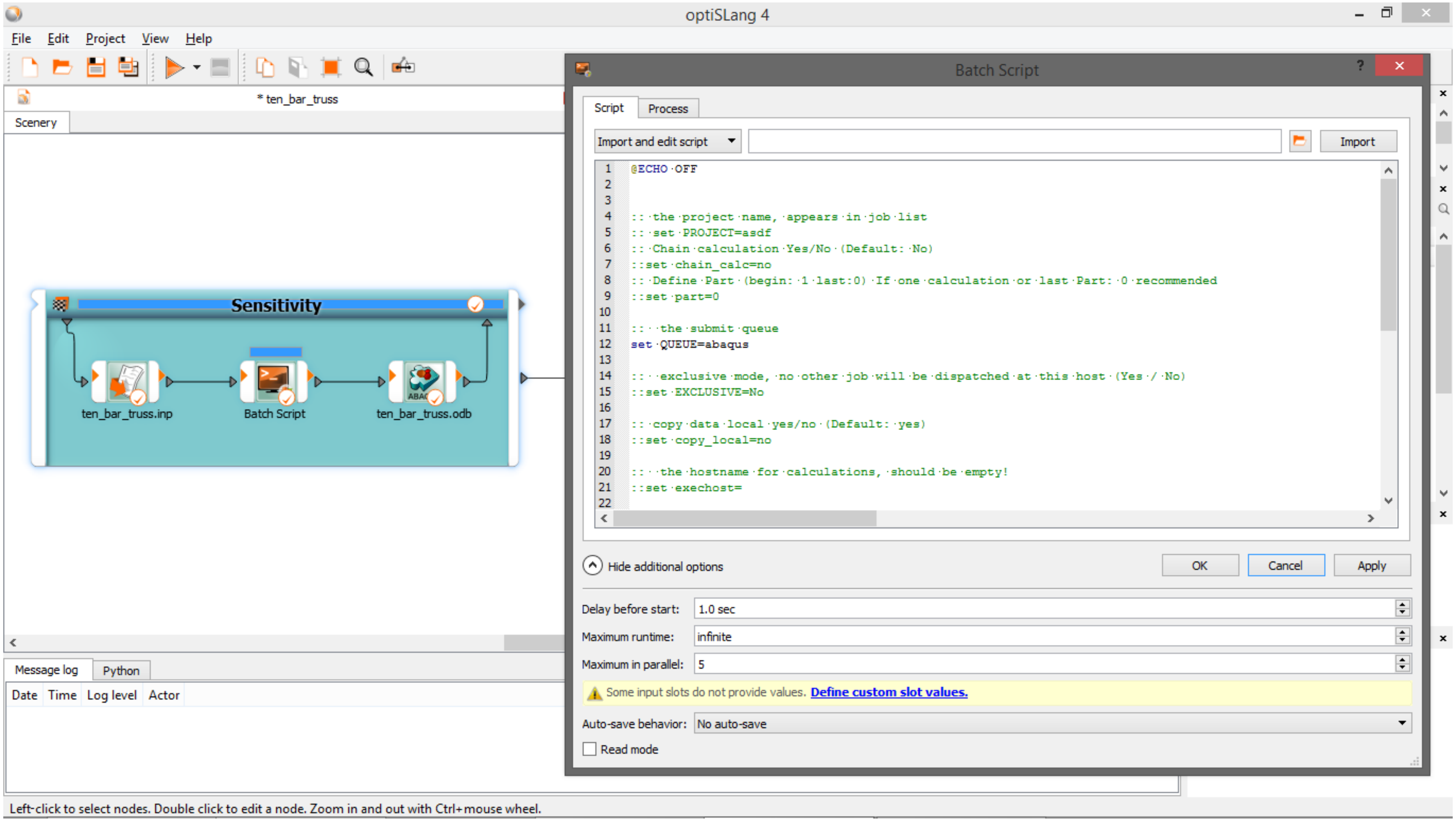


optiSLang compute cluster

- More specific scenarios
  - optiSLang host has shared filesystem with cluster
    - No need for file transfer
  - optiSLang host is batch submission host
    - Simplified batch interaction interface
    - No ssh access required
  - No job submission script available
    - Need to define job execution and create job command or script

- Alternative
  - VDI: running machines and applications in the datacenter

View from optiSLang



```
@ECHO OFF


:: the project name, appears in job list
:: set PROJECT=asdf
:: Chain calculation Yes/No (Default: No)
::set chain_calc=no
:: Define Part (begin: 1 last:0) If one calculation or last Part: 0 recommended
::set part=0

::  the submit queue
set QUEUE=abaqus

::  exclusive mode, no other job will be dispatched at this host (Yes / No)
::set EXCLUSIVE=No

:: copy data local yes/no (Default: yes)
::set copy_local=no

::  the hostname for calculations, should be empty!
::set exechost=

```

Hide additional options

Delay before start: 1.0 sec

Maximum runtime: infinite

Maximum in parallel: 5

⚠ Some input slots do not provide values. Define custom slot values.

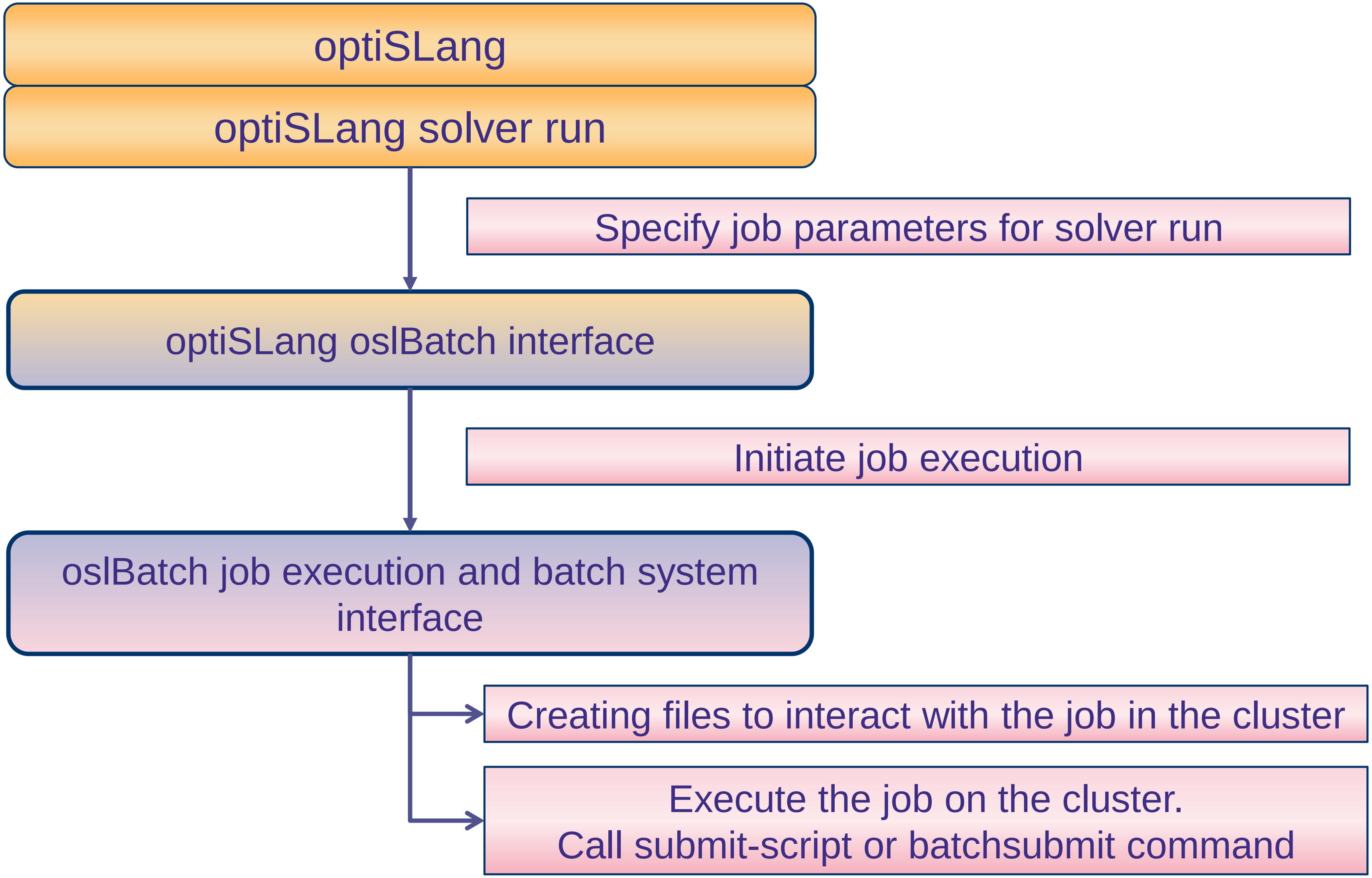
Auto-save behavior: No auto-save

□ Read mode

Left-click to select nodes. Double click to edit a node. Zoom in and out with Ctrl+mouse wheel.

A short script to set parameters and start calculations

oslBatch interface: optiSLang



optiSLang

optiSLang solver run

Specify job parameters for solver run

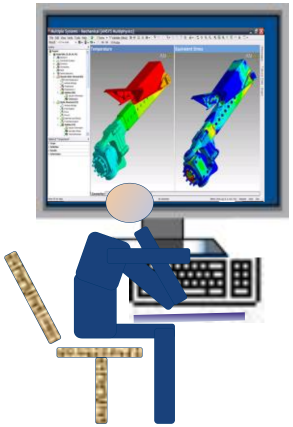
optiSLang oslBatch interface

Initiate job execution

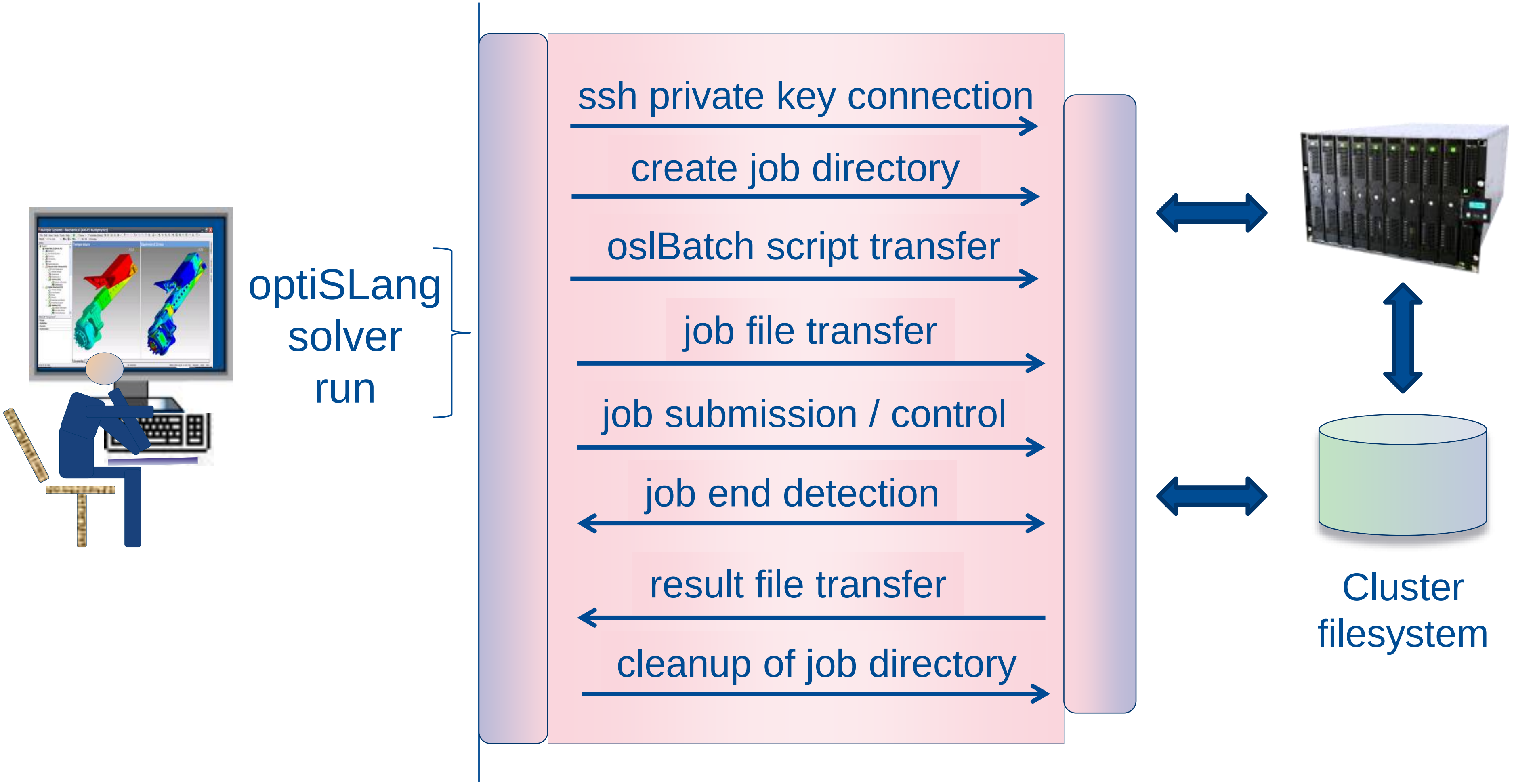
oslBatch job execution and batch system interface

Creating files to interact with the job in the cluster

Execute the job on the cluster.
Call submit-script or batchsubmit command



oslBatch interface: node to cluster



optiSLang
solver
run

ssh private key connection

create job directory

oslBatch script transfer

job file transfer

job submission / control

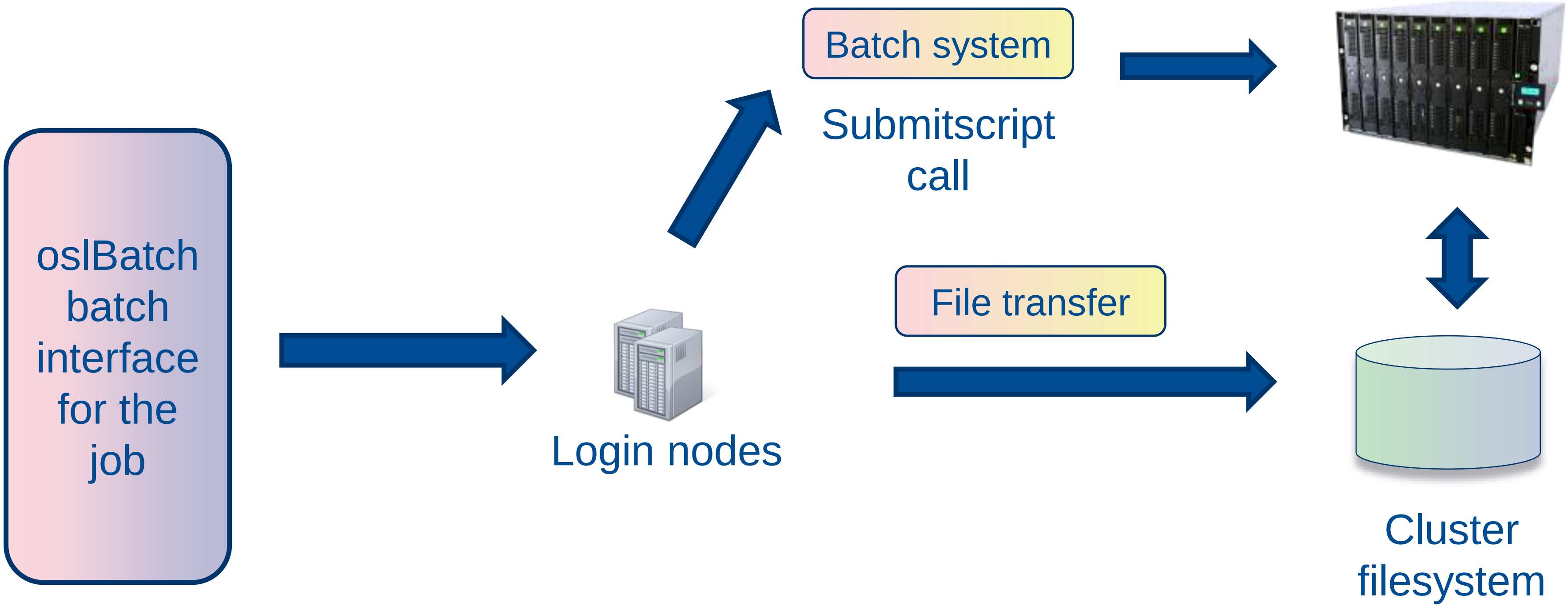
job end detection

result file transfer

cleanup of job directory

Cluster
filesystem

oslBatch interface: cluster side



oslBatch batch interface for the job

Login nodes

Batch system

Submitscript call

File transfer

Cluster filesystem

oslBatch options

- Configuring Batch System
  - Define Memory
  - Number of Cores
  - Select Queue
  - Host Exclusive
  - Use a local directory for job execution on the execution host
  - Specify Calculation Node
  - Design chained Calculations

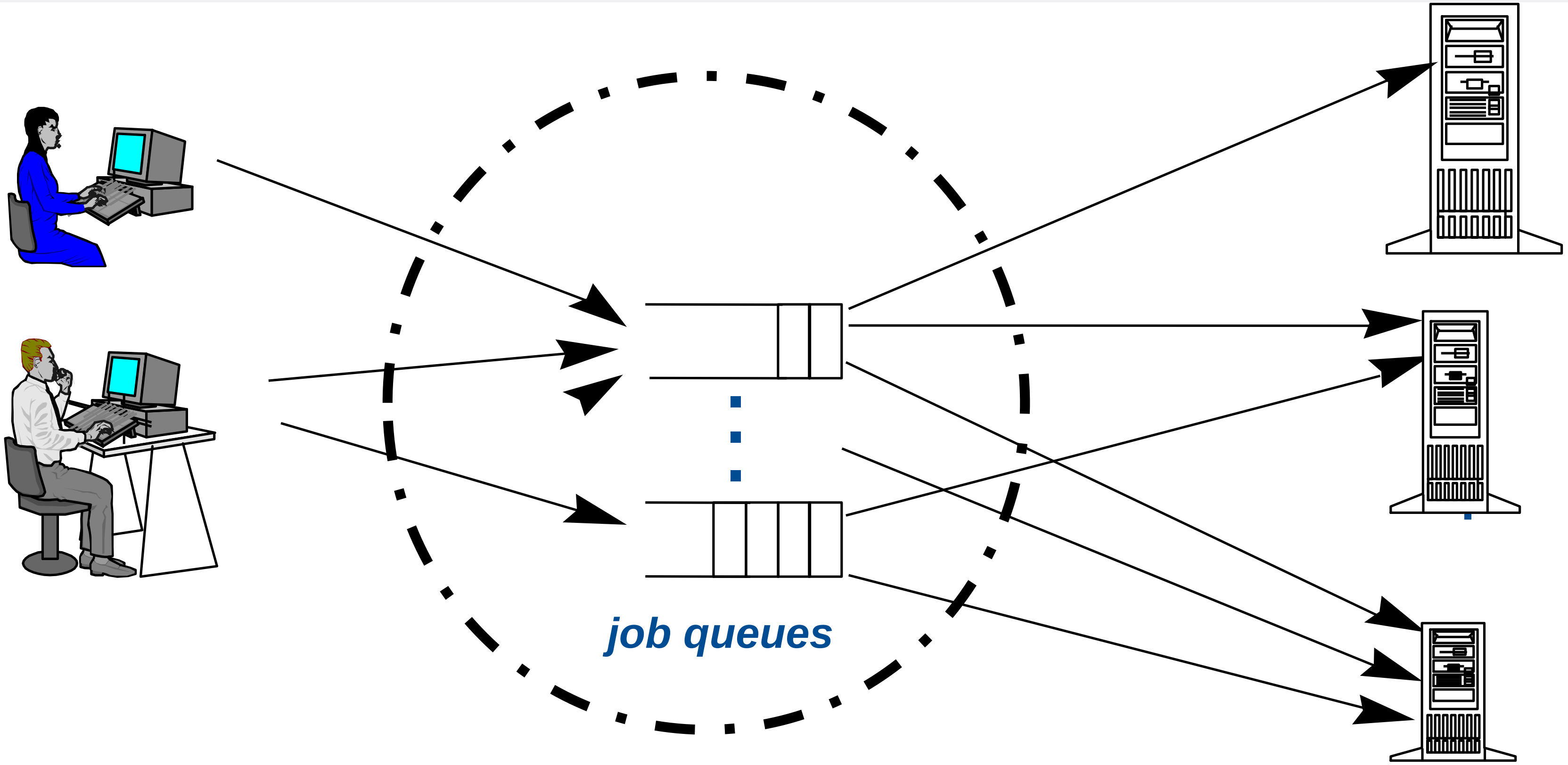
- Configuring Job
  - Project Name
  - Specific Solver Options (e.g. MPI Options, Version, License)

Batch system features overview

Batch system functions

- Job distribution to cluster nodes
- Load balancing between cluster nodes
- Resource management for jobs
  - Memory requirements of jobs
  - Application license availability
  - Dependency on specific hard-, software
  - …
- Policies for job execution
  - Fairshare scheduling (user- and usergroup priorities)
  - Service levels for special jobs (e.g. throughput)
  - Advance reservation of resources
  - Topology awareness (racks, IB switches)
  - Cleanup after job finish
  - Reporting and accounting

Batch system job scheduling

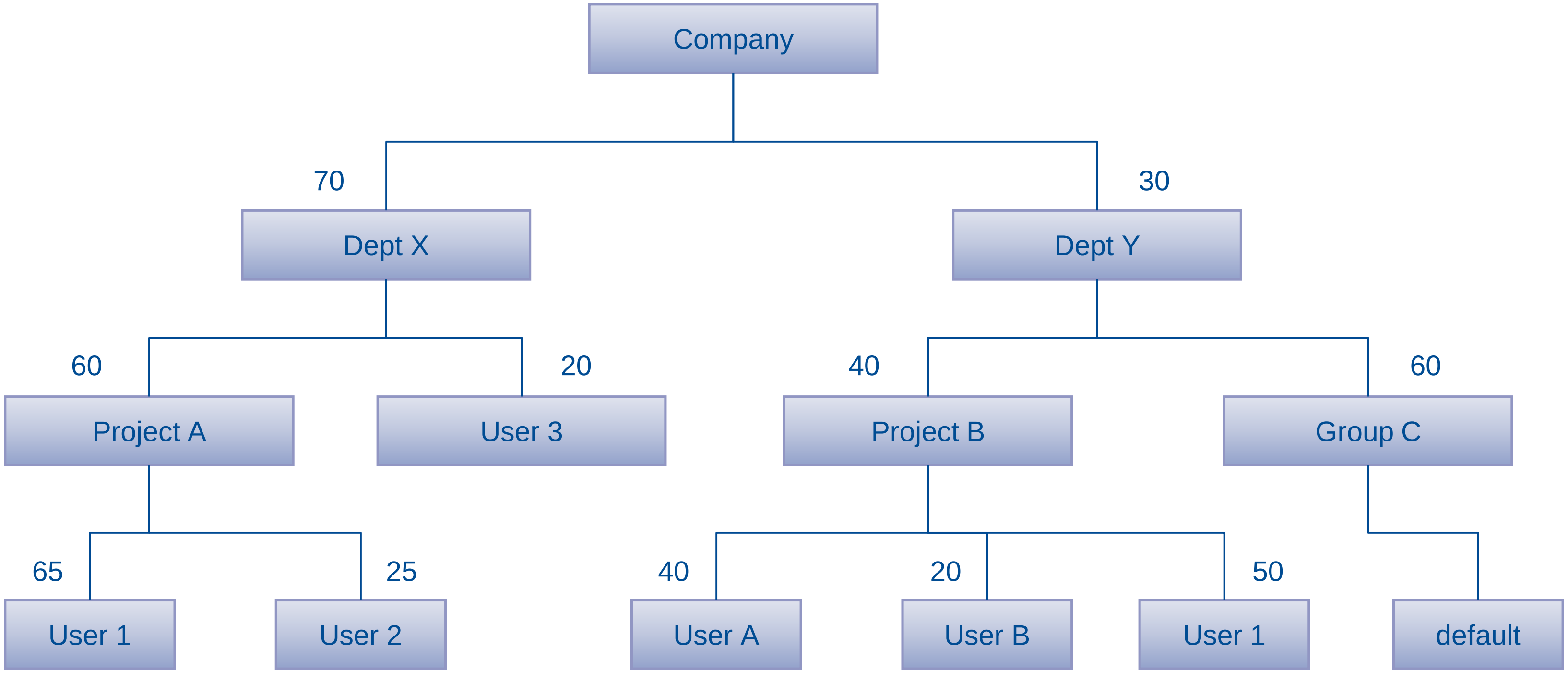


Jobs ⟶ *Information of current resource availibility* ⟶ *Scheduling decision*

Batch master

Compute nodes

Example: Hierarchical Fairshare Tree



Batch system job control

Example: Memory management with LSF

- Reserve memory on execution nodes for a job
  - Reserved memory taken into account for scheduling of the next jobs' memory requirements
- Limit maximum memory usage of a job
  - Job may/will be killed when the limit is reached
- Host-based threshold
  - Define memory threshold on a host basis
  - When free memory becomes less than threshold LSF suspends jobs automatically. Jobs will be resumed automatically when free memory is higher than a second threshold

Batch system job control

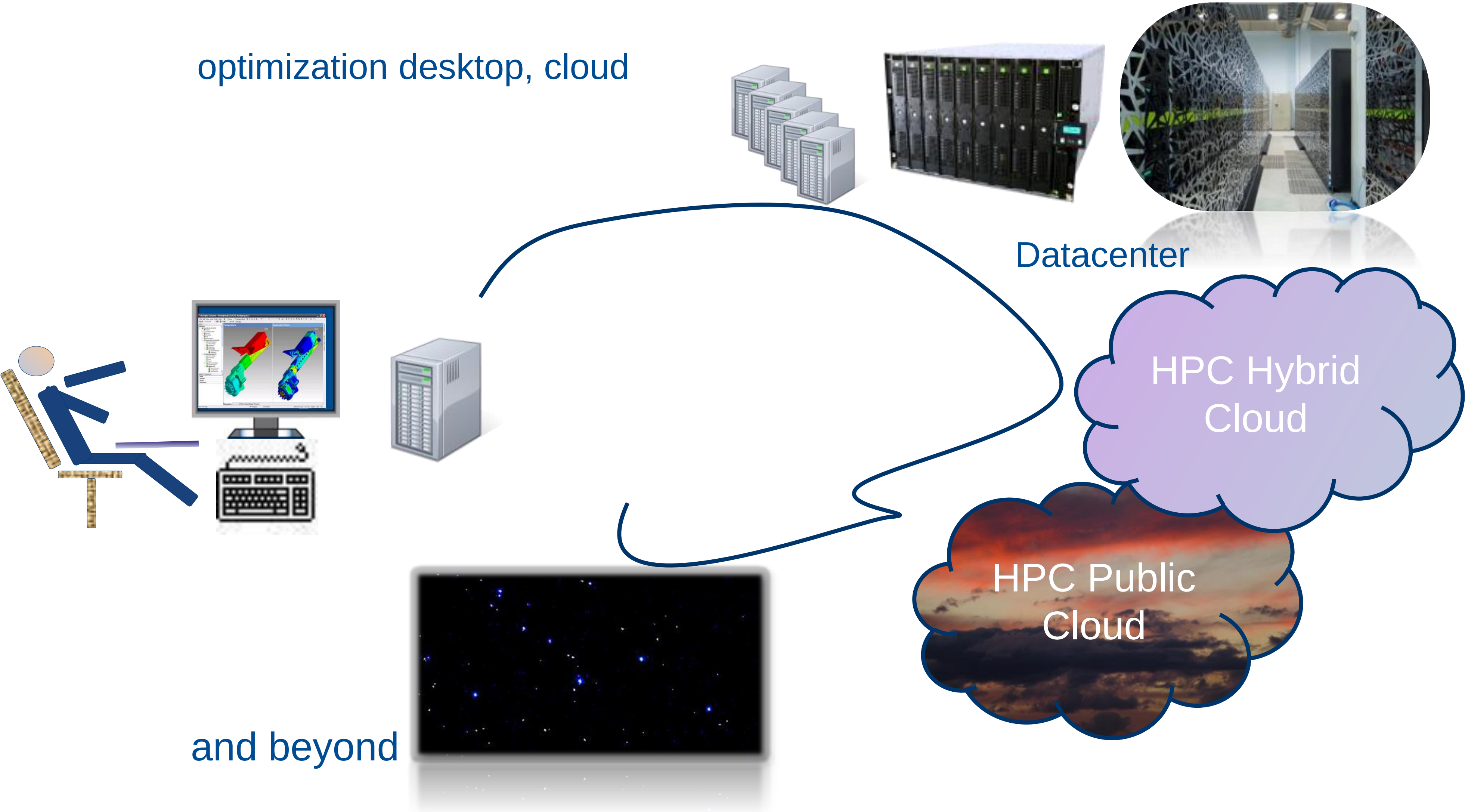
Example: Memory limit control on Linux with LSF

- cgroups (Linux)
  - Hard limit enforcement by “control group” for job
- Linux OS memlimit handling
  - Memlimit enforced per process when OS thinks it might be necessary - might be too late
- Memlimit enforced by LSF
  - Job’s memory controlled by LSF, memory limit enforced over all processes of a job (on a host- or job-basis)
- Smart memlimit enforcement by LSF
  - Define threshold for overall mem- and swap-usage (e.g. 90%, 10%)
  - When threshold is reached LSF kills jobs which overruns

optiSlang batch benefits

- Modular Structure
  - Not restricted to one batch system (e.g. Platform LSF)
  - Complete integration in any company environment
  - Use existing jobscripts
  - No installation of other tools required (except ssh-key)
  - Start any solver with command line availability
- Usage
  - Define all job parameters in optiSLang
  - Save time by one click submission
  - Secure data transfer

optiSLang batch



optimization desktop, cloud

Datacenter

HPC Hybrid Cloud

HPC Public Cloud

and beyond

Thank you for your attention.

i.seipp@science-computing.de
science + computing ag
www.science-computing.de

Tel.: +49 7071 9457-219

e-mail: info@science-computing.de

philipp.pasold@dynardo.de
Dynardo GmbH
www.dynardo.de

Tel.: +49 3643 9008-45

e-mail: sales@dynardo.de