Increase ILC Software performance using Cloud Computing

Tomasz Wojtoń

Institute of Nuclear Physics PAN, Kraków

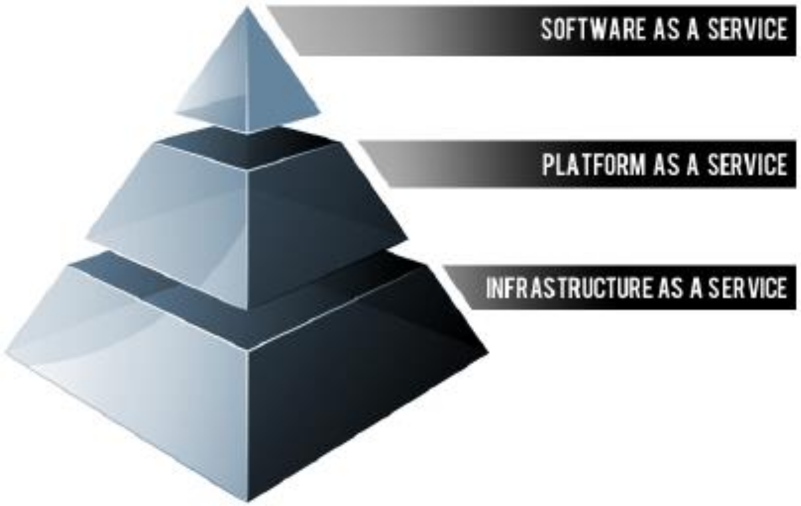
Introduction

- ILC Simulations
  - Frequently changing input parameteres
  - High statistics
- Takes a lot of time
- Fast simulation and reconstruction data to analysis

- Solution – **Cloud Computing**

Cloud Computing

- Provide computer infrastructure to end-users in friendly way **on demand**
- Elastic resource allocation
- Efficient usage of computing resources in institution
- Cloud Computing models:



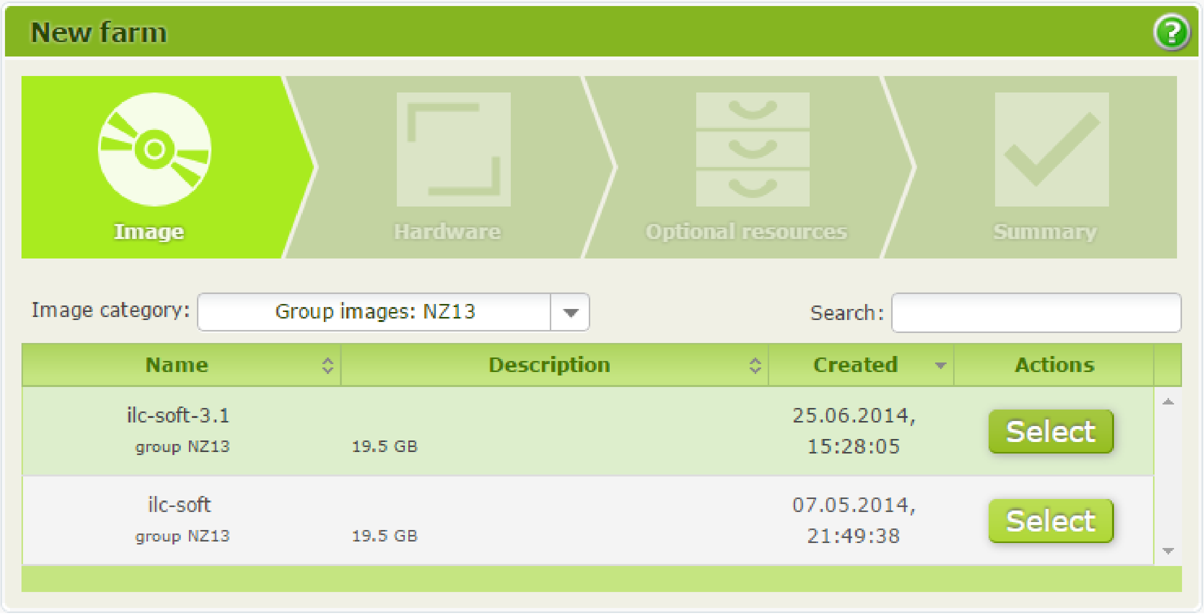
  - SaaS – applications designed for end-users, delivered over the web
  - PaaS – set of tools and services designed to make coding and deploying those applications quick and efficient
  - **IaaS** – is the hardware and software that powers it all – servers, storage, networks, operating systems

CC1 – Private Cloud Solution

- Complete solution for Cloud Computing developed and deployed in IFJ PAN
- Use by many scientists in several experiments
- Convenient and flexible access to computing power
- Web access interface
- Dedicated feature for research institutes and universities – **Farms (Virtual Clusters)**

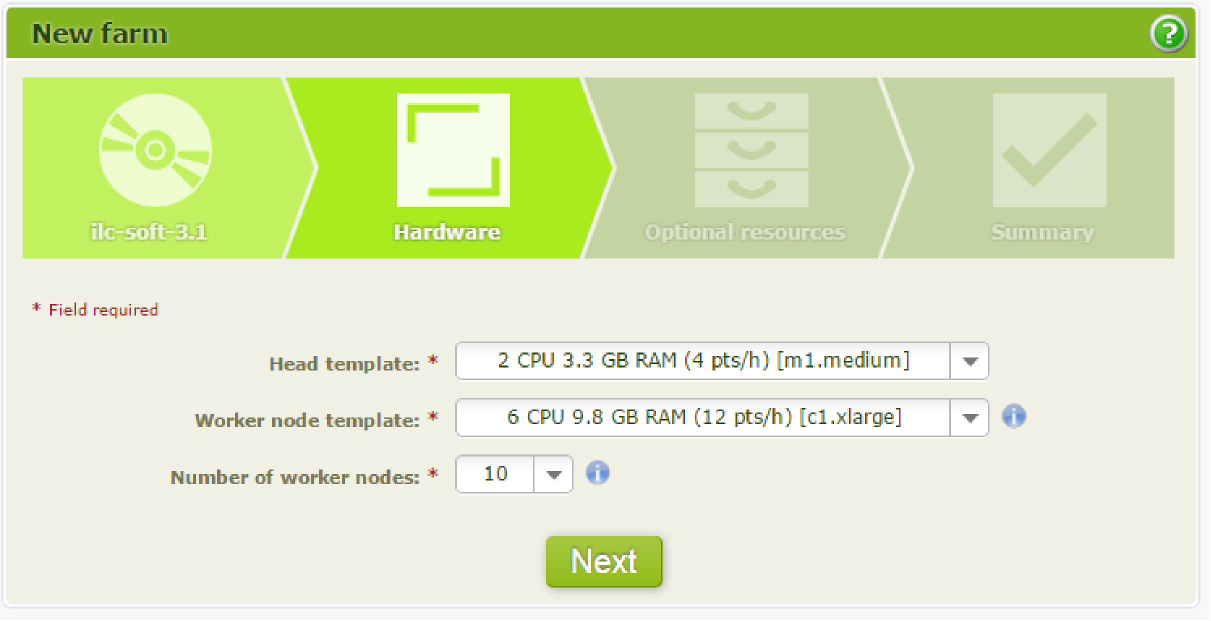
CC1 Virtual Clusters – in few easy steps

- 1. Select Virtual Machine image



CC1 Virtual Clusters – in few easy steps

- 2. Choose hardware (CPU, memory amount) and number of nodes in cluster

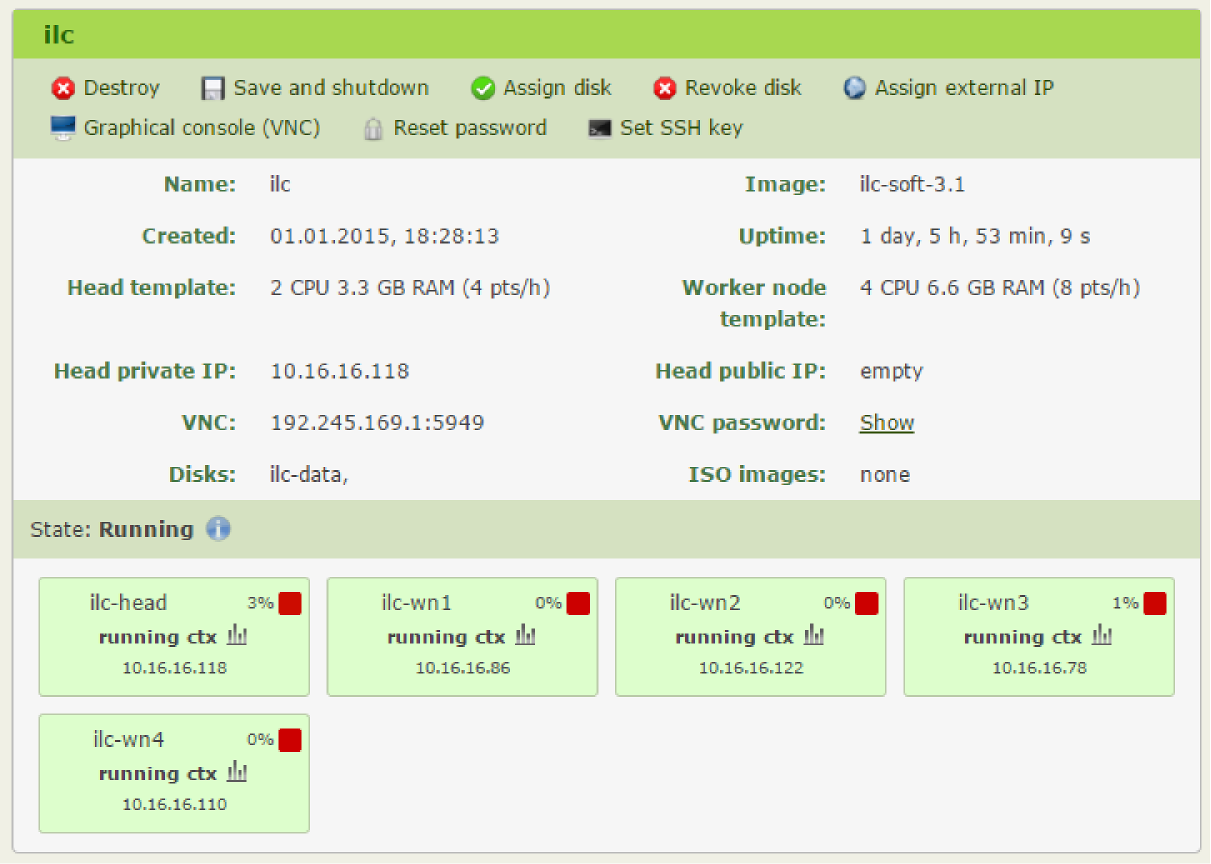


CC1 Virtual Clusters – „Farms”

- 3. Assign IP or attach external storage

CC1 Virtual Clusters – „Farms”

- 4. Create and run



ILC Software & Virtual Clusters

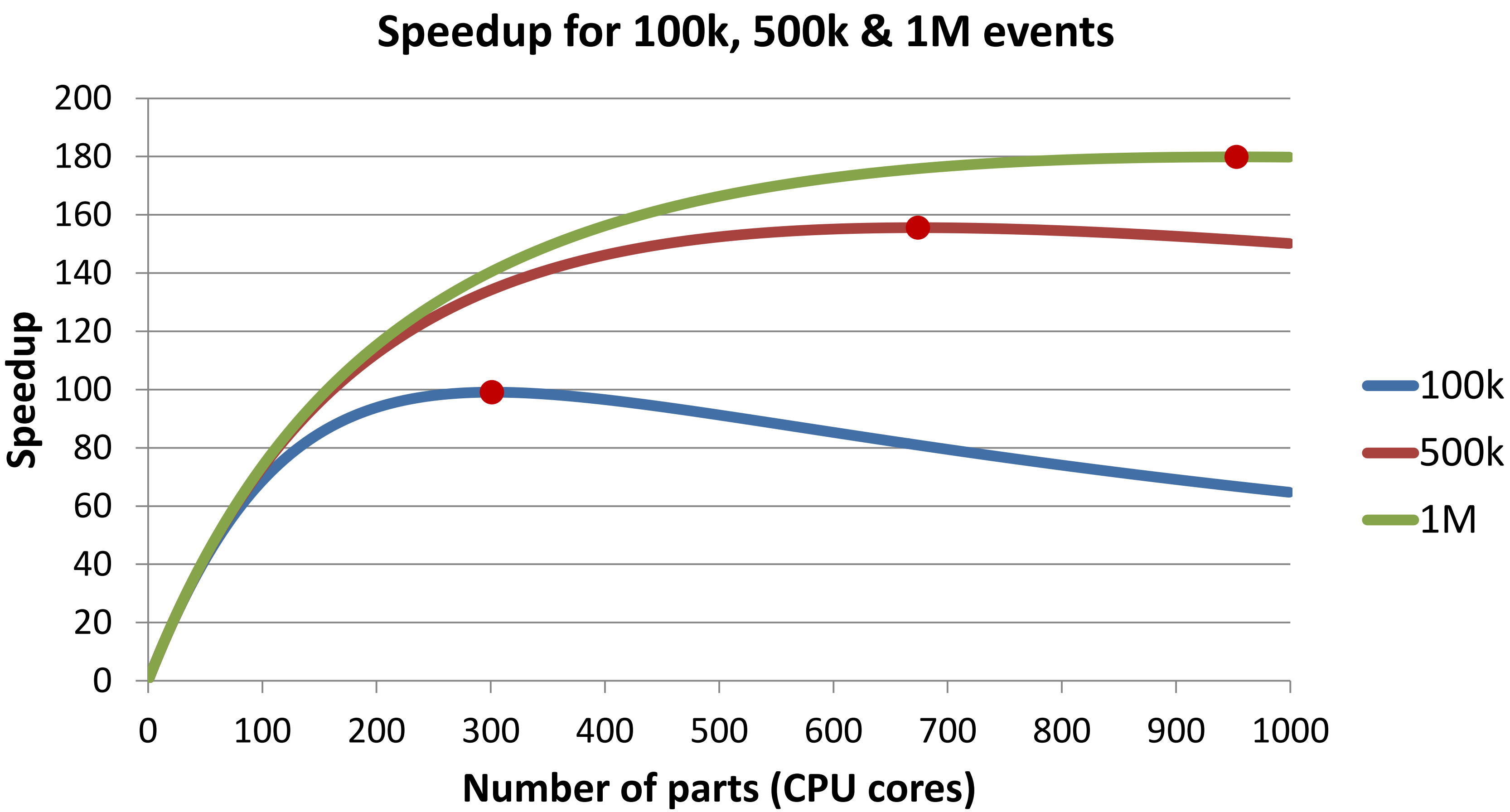
- **Example –** Generate 1000 events (500 GeV, $e^+e^- \rightarrow e^+e^-X$), simulate and reconstruct data – more than **10 hours**
- The same case using farm of nodes with total 25 CPU cores takes **~1 hour**

ILC Software & Virtual Clusters

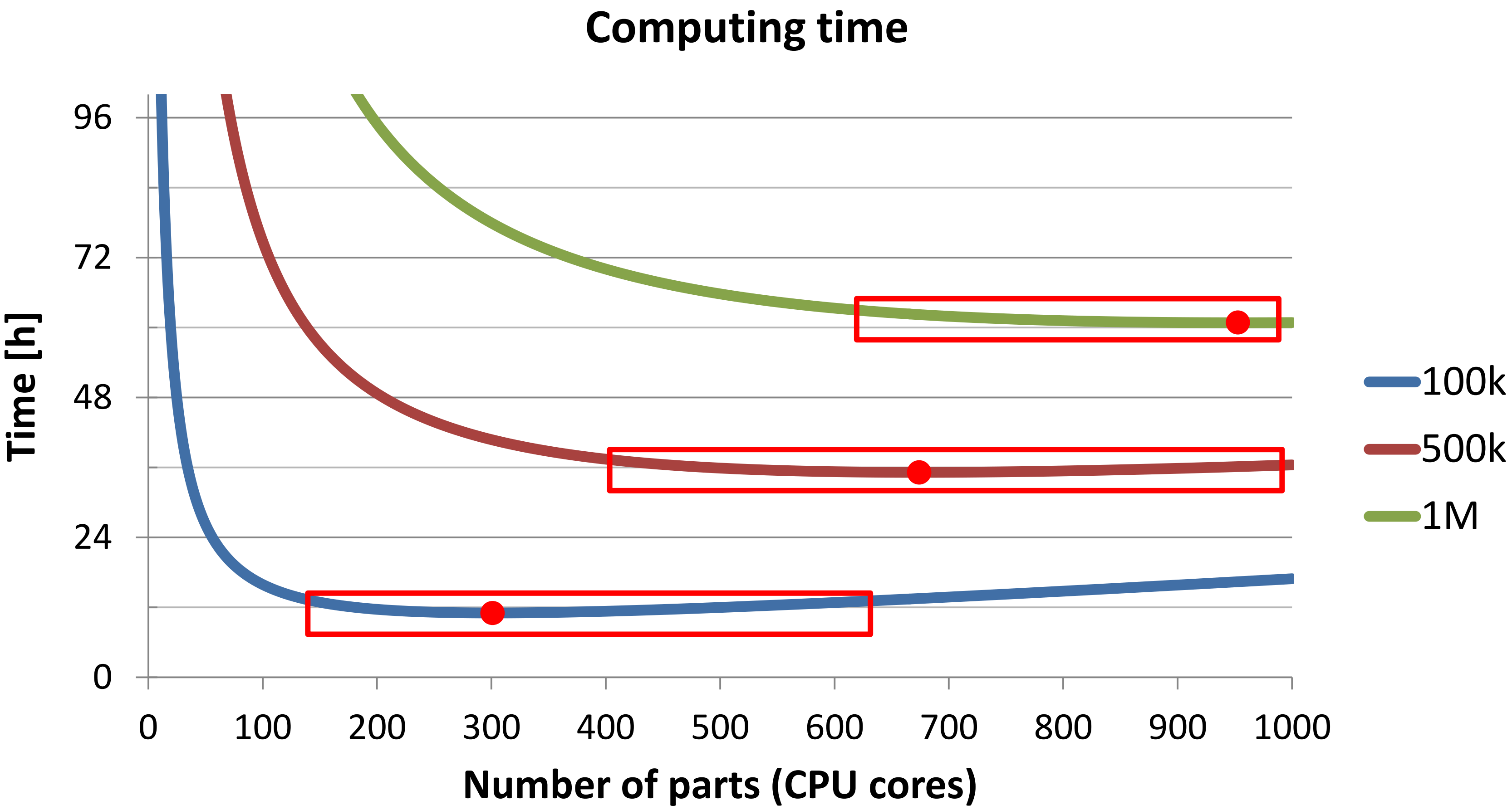
- ILC Software automation
  - Splitting job to smaller parts
  - Automatic tasks processing
  - Merge parts
- Preparing:
  - Farm with attached storage for data
  - Events generator (pythia)
  - Mokka and Marlin steering files
  - Fill up the starting script
- Just `./start`

```
EVENTS=100000
PARTS=240
GAMMA_FILE="/data/res.HEPEvt"
GAMMA_DIR_PARTS="/data/events-parts"
STORAGE_DEVICE="/dev/vdb1"
STORAGE_PATH="/data"
```

Performance analysis – speedup



Performance analysis – computing times



Summary

- More CPU cores is **not** always better
- Performance measurements are **important**
- It could be used as a benchmark to select how many events could be processed at specific time with defined number of CPU cores.
- Cloud computing is **good candidate** for ILC Computing backend