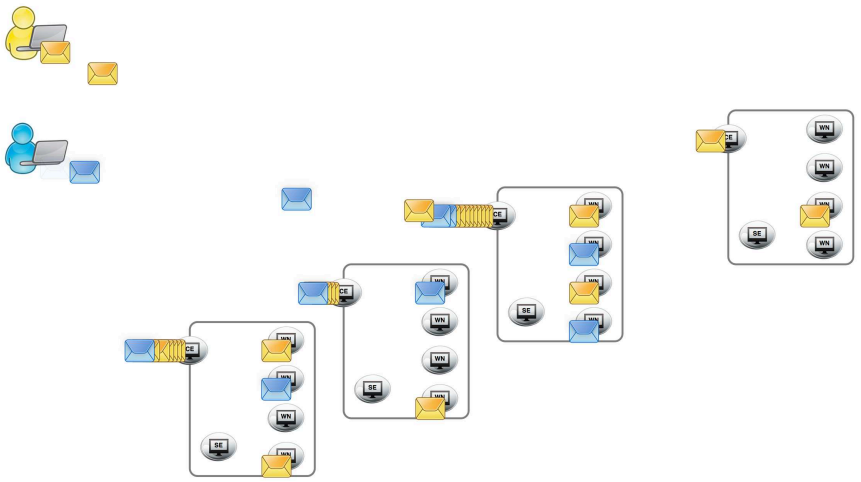
Grid Computing Overview

SURFsara – Online Grid Computing Course

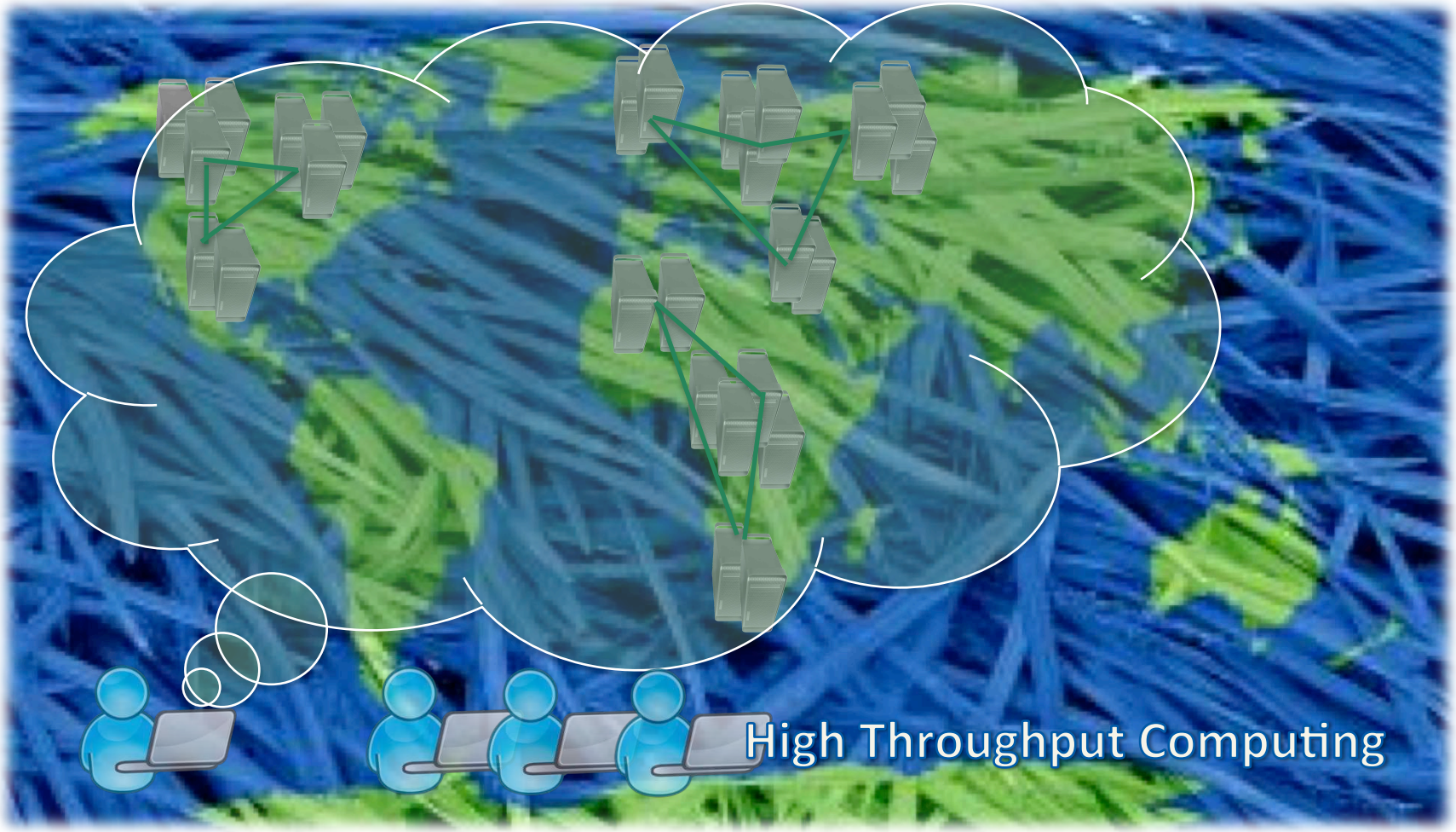
Outline

- **Grid Computing**
  - From Cluster to Grid Computing
  - Motivation
  - Beneficial attributes
- **How Grid works**
  - Building Blocks
- **To use or not to use Grid**
  - Best practices



Grid is...

interconnected clusters



From Cluster to Grid Computing

- http://mooc-inst.sara.cloudlet.sara.nl/mooc/grid.html

Cluster vs. Grid computing

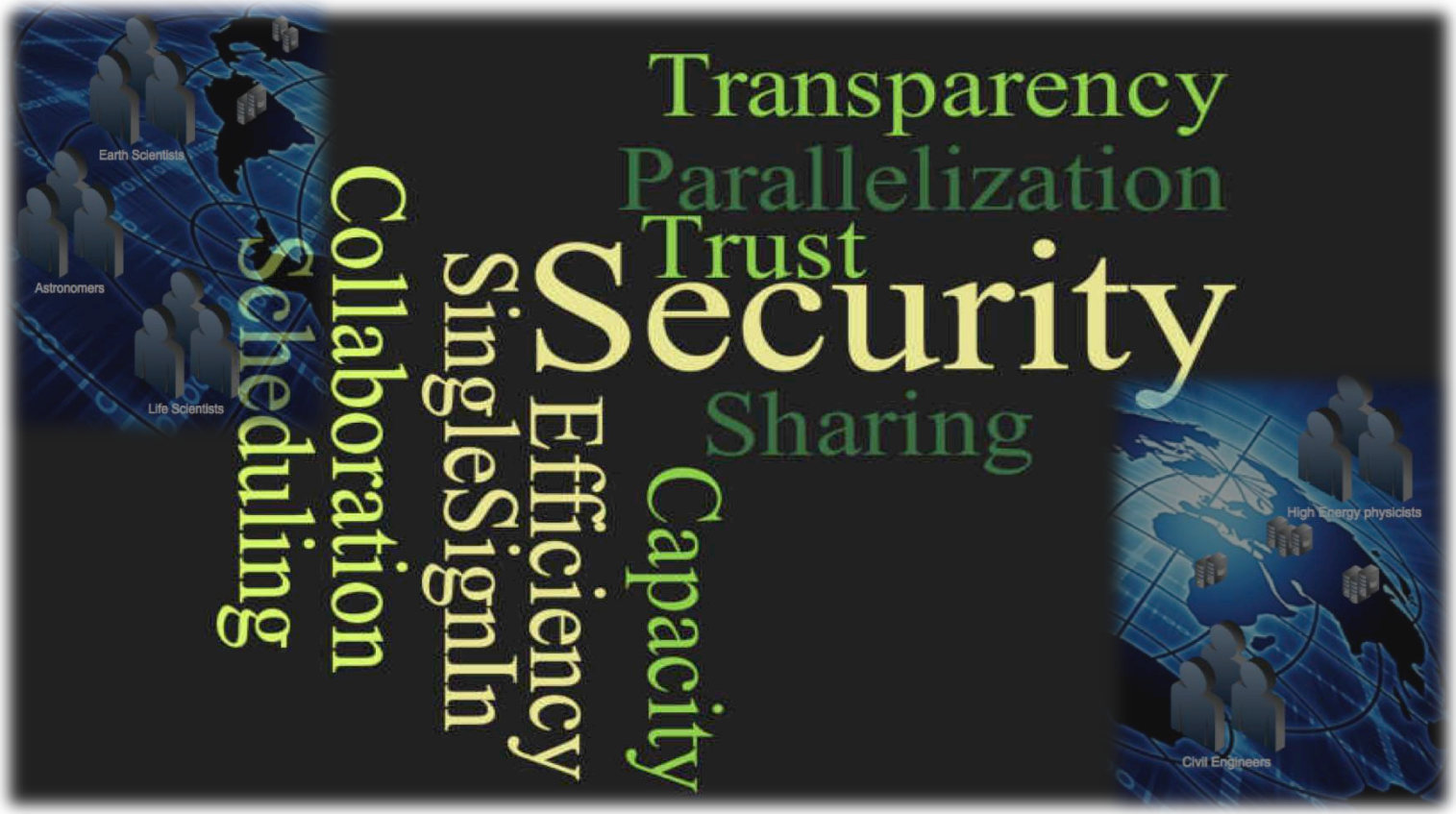
- **Cluster**
  - One site
  - Shared local storage with home account
  - Username based authentication
  - Relatively homogeneous hardware
  - Direct job submission

- **Grid**
  - Multiple sites
  - No shared storage
  - Certificate based authentication
  - Heterogeneous hardware
  - Job submission through middleware

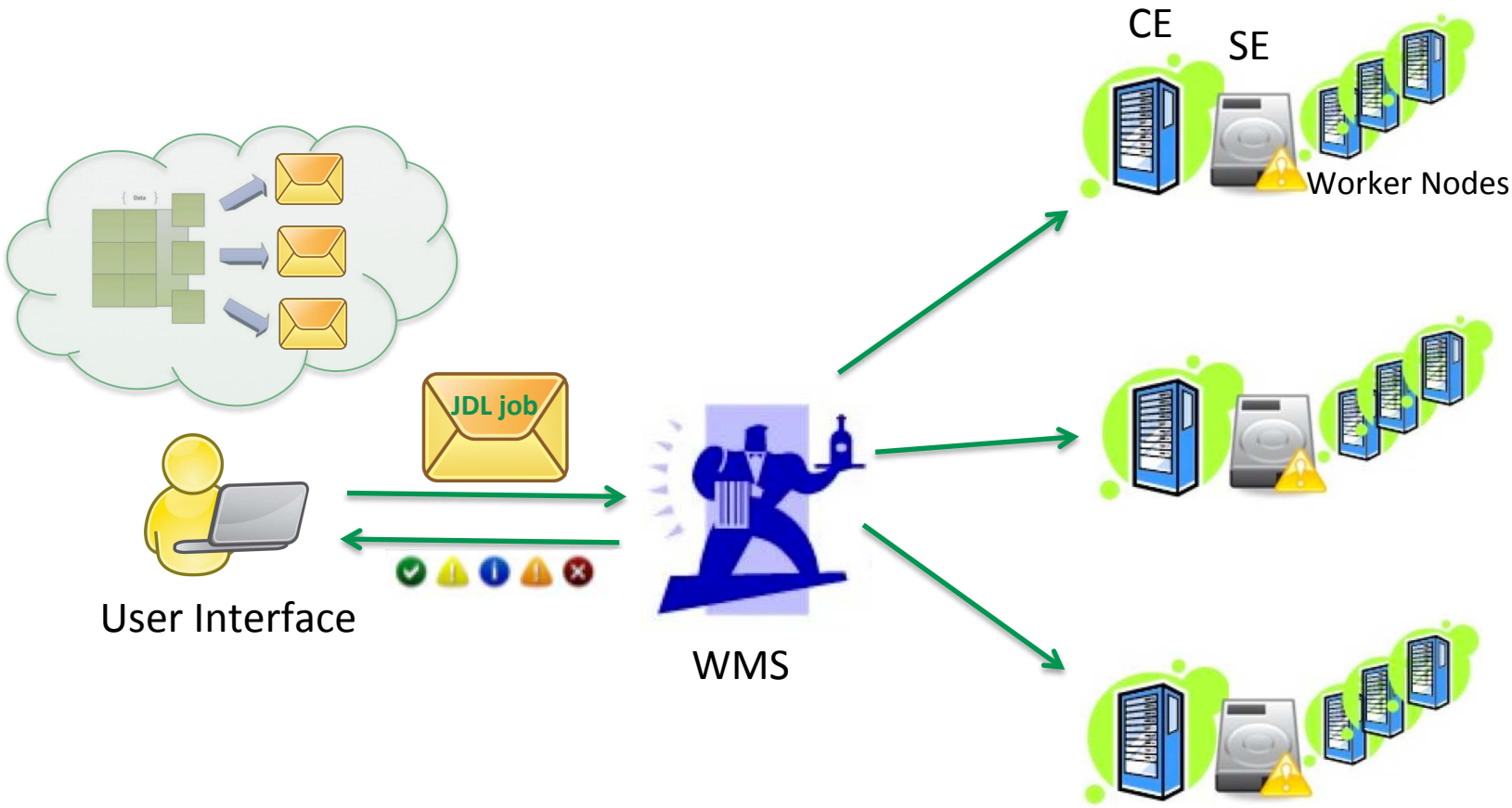
Motivation… is science

- Geographically disperse resources
- Enormous Compute and Storage capacity
- Security and Authorization
- Collaboration among virtual communities
- Large Hadron Collider (LHC) experiments
  - 150 sites
  - 250000 CPU cores
  - 160 PB disk and 90 PB tape

Beneficial attributes



Building Blocks



Animation- Grid WMS

- http://mooc-inst.sara.cloudlet.sara.nl/mooc/wms.html

To use or not to use Grid

- **Embarrissingly parallel problem**
  - Easily partitionable in independent parallel tasks (jobs)
  - Simultaneous independent calculations
- **Submitting hundreds, thousands jobs simultaneously.**
  - Large scale computational problems
  - Large-scale data processing
  - Parallel processing
  - Scheduling and bookkeeping overhead
  - A single grid job is slow

Best practices

Users have to do a lot of things themselves…

- **Prepare the application**
  - What program? What data? Other requirements?
  - Don't rely on hard-coded paths. Program portably.
- **Submit a bunch of jobs**
  - Track status/Retrieve output
  - Deal with failures
- **Error handling is not easy**
  - Embed logging steps in your scripts for debugging
  - Use Pilot jobs
- **Adapt your implemantation design considering**
  - memory, wallclock times, available scratch space
- **Start small…scale up gradually; learn**