Silvia Delgado Olabarriaga

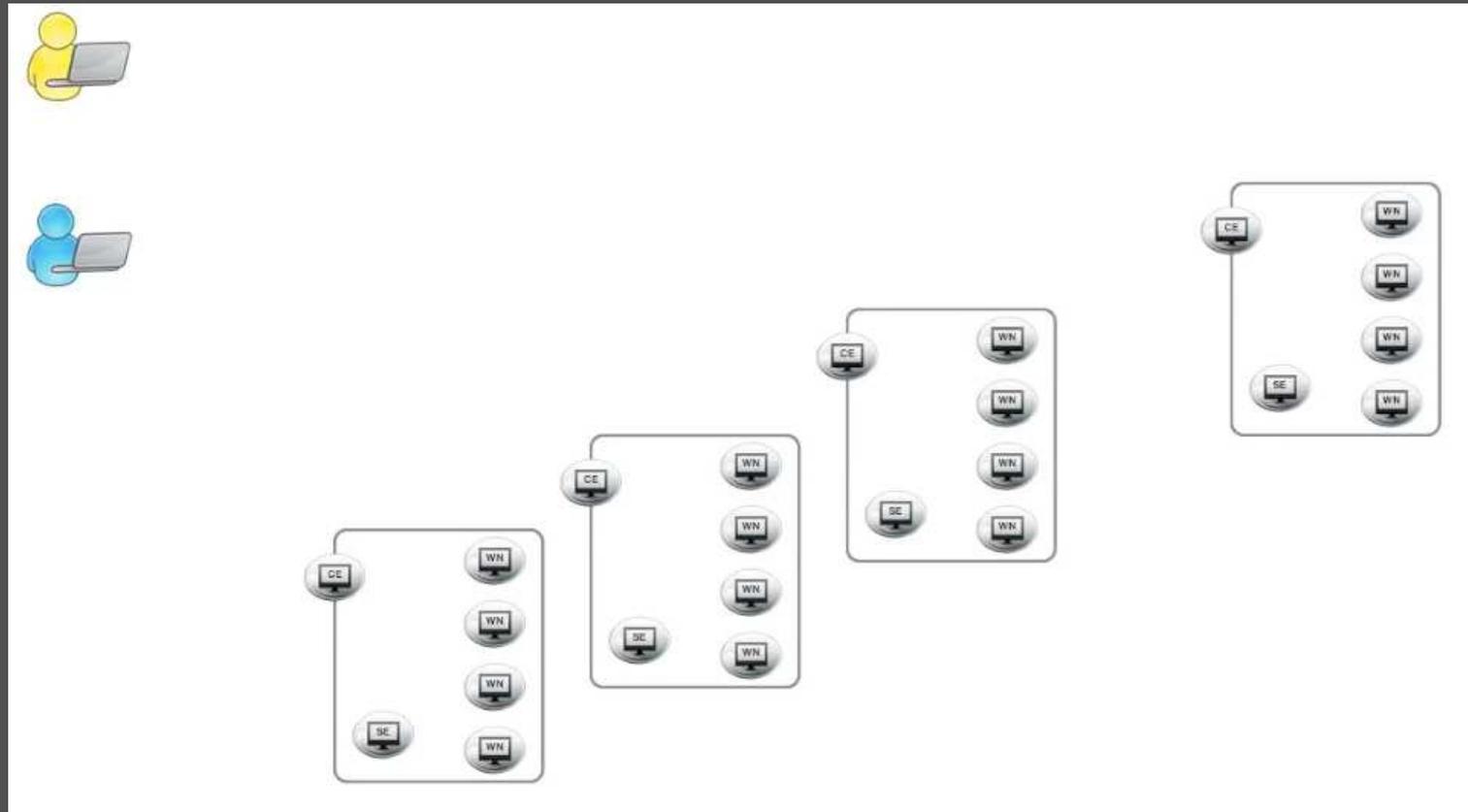# Introduction to Workflows

WORKFLOW MANAGEMEMT FOR GRID COMPUTING

Prepared by Silvia Delgado Olabarriaga

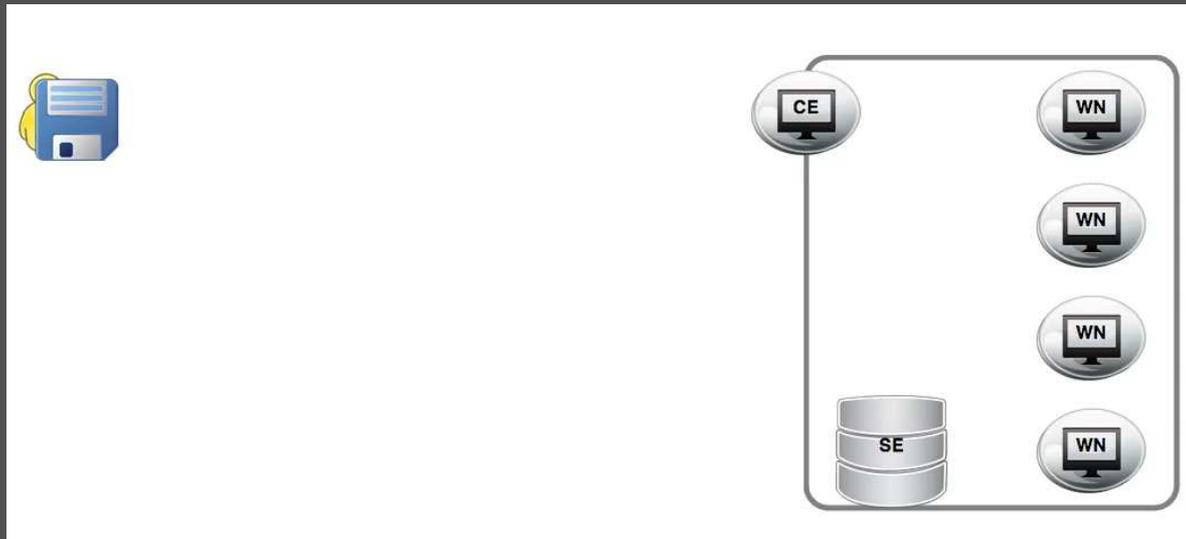Academic Medical Center of the University of Amsterdam, NL

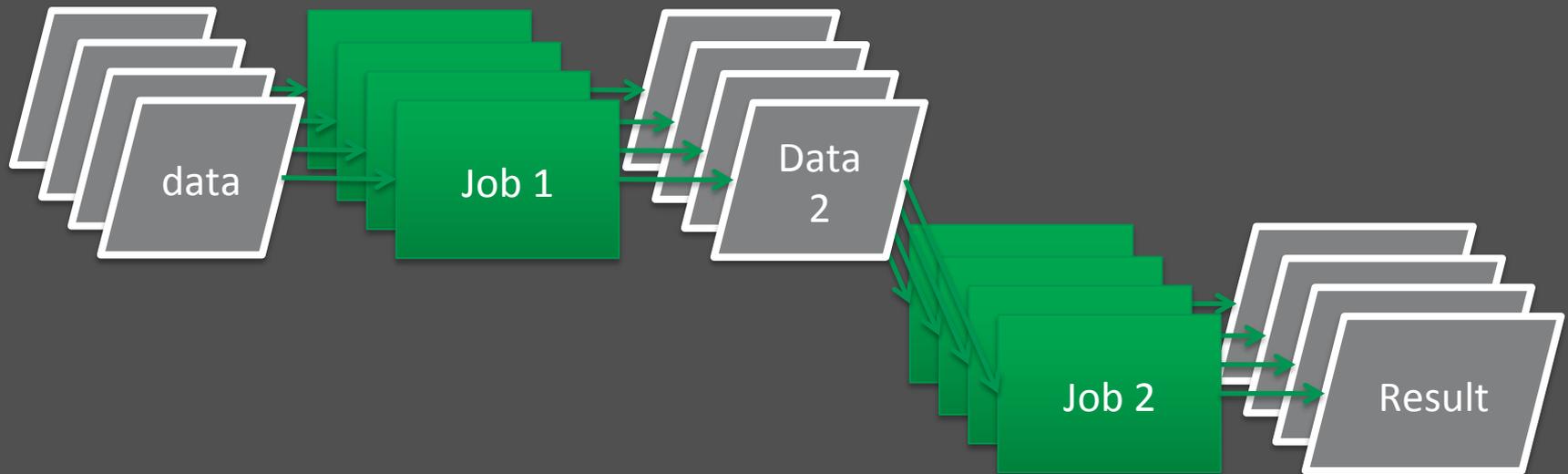# Many clusters, many jobs…

Data transport..
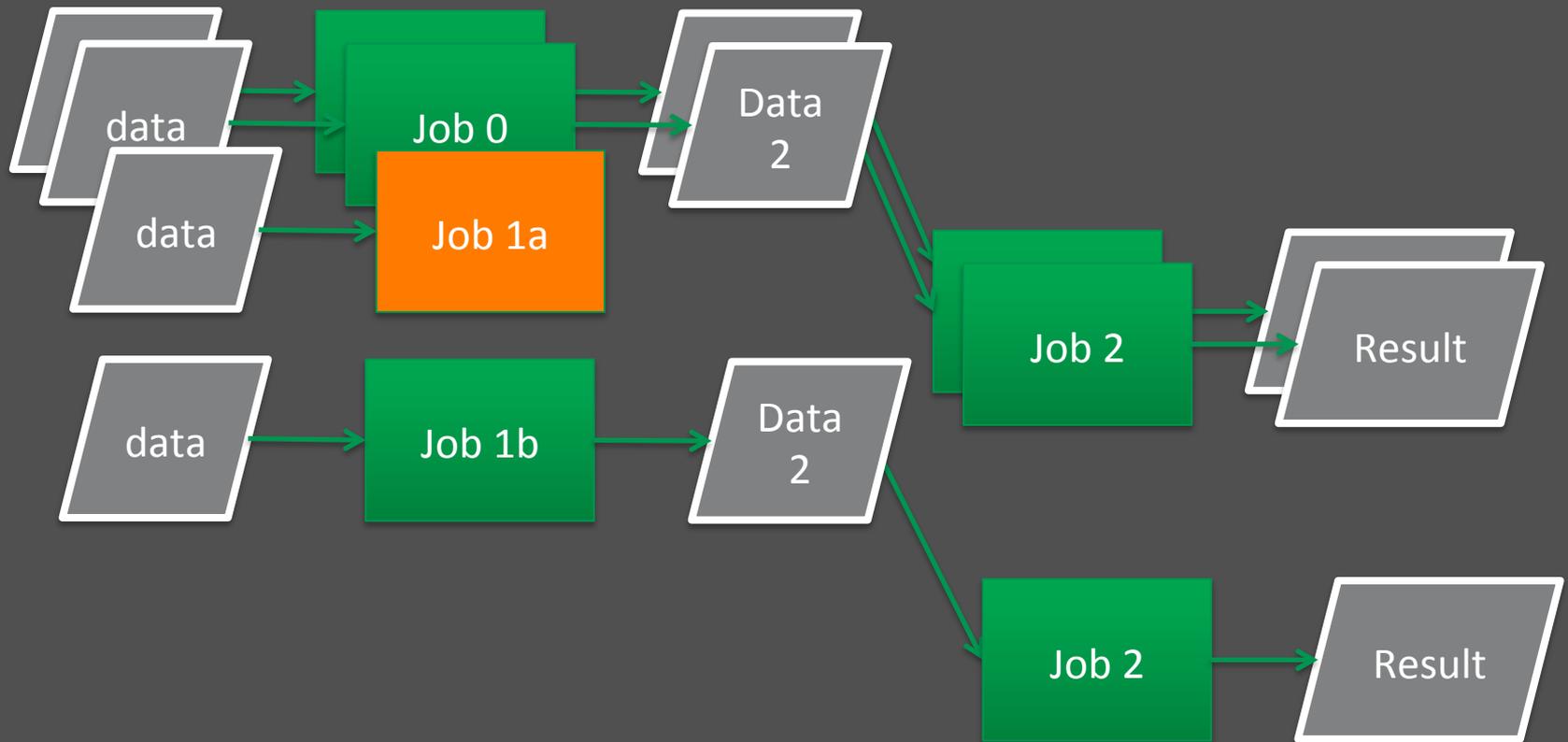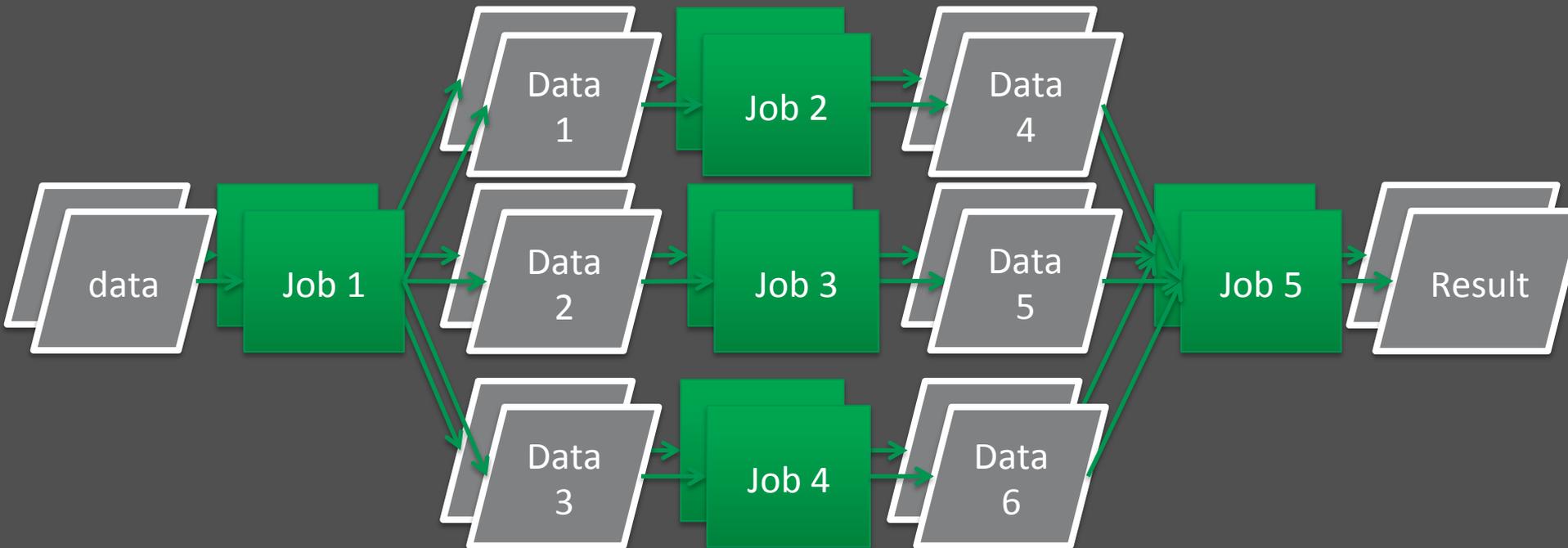
Jobs and data don't come alone..

Jobs fail…

Jobs can split and merge data...
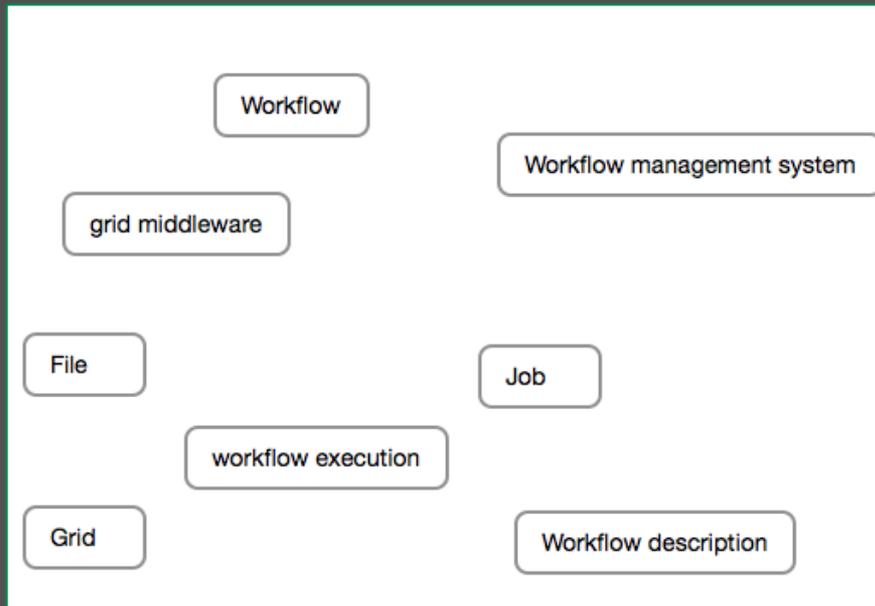
Many times...

In practice...

HELP!

Goals of this lecture

❑ **Introduce concepts of workflow management systems for grid computing**

❑ **Briefly present some examples**

❑ **Provide pointers to further study**

Keywords of today

Workflow

Workflow management system

grid middleware

File

Job

workflow execution

Grid

Workflow description

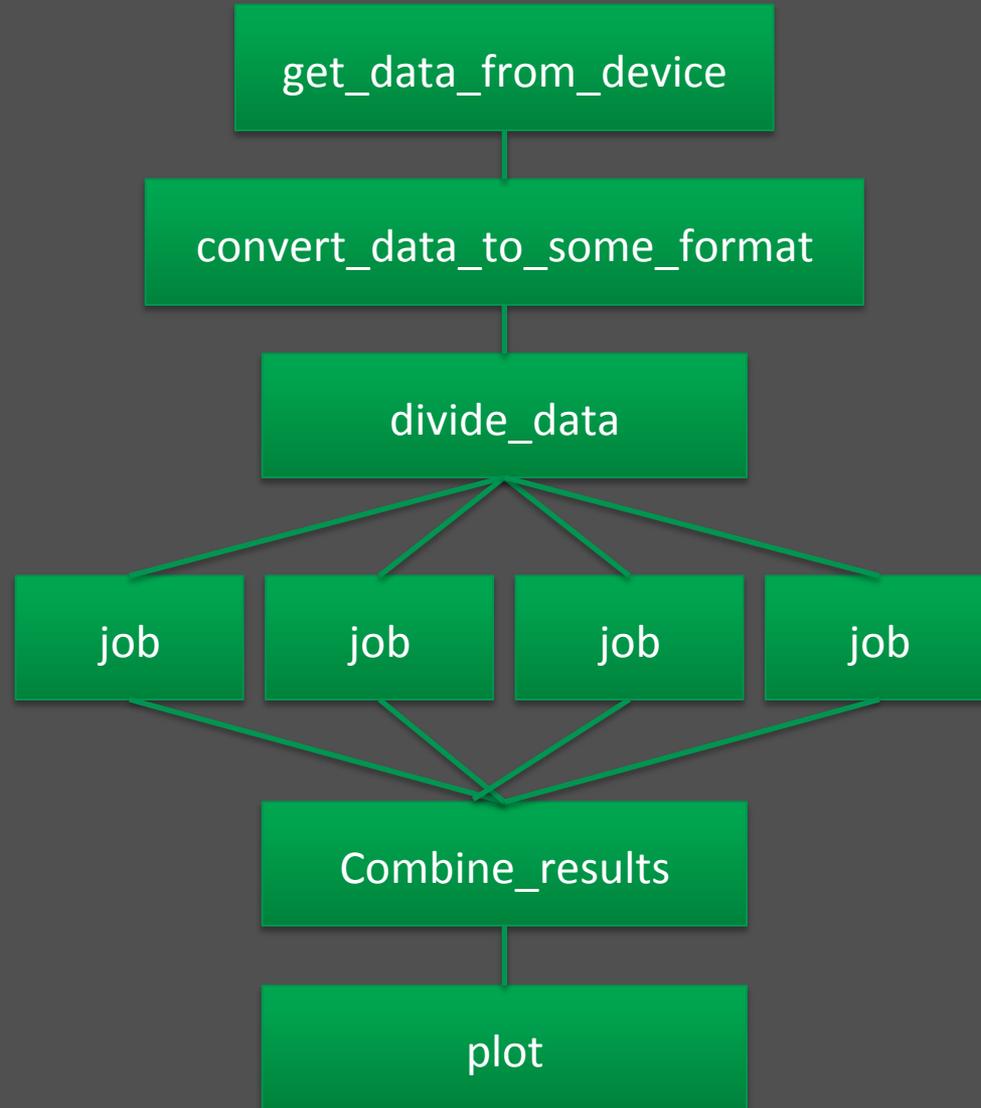# Workflow is...

- ❑ **a sequence of connected steps;**

- ❑ **a drawing of a sequence of operations;**

- ❑ **designed to achieve processing intents of some sort, such as physical transformation, service provision, or information processing.**

Example

```
get_data_from_device
        |
convert_data_to_some_format
        |
    divide_data
    /   |   |   \
 job  job  job  job
    \   |   |   /
   Combine_results
        |
       plot
```
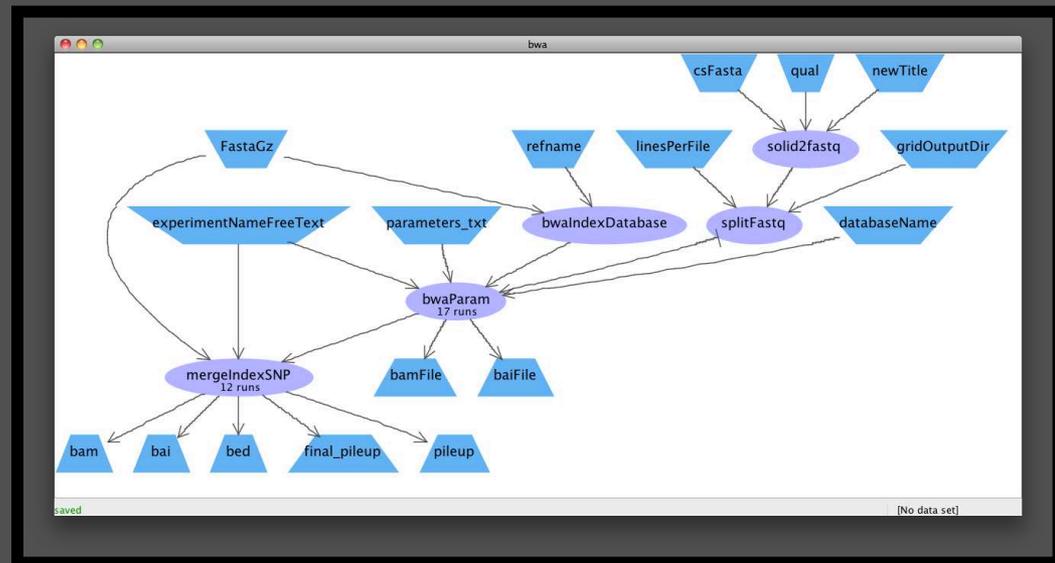
Why workflows?

❑ **Manage complex sequences of operations**

❑ **Separate**

- Description of the process to ████████████ ████mer)
- Execution steps need███ ██████ (by the execution system)
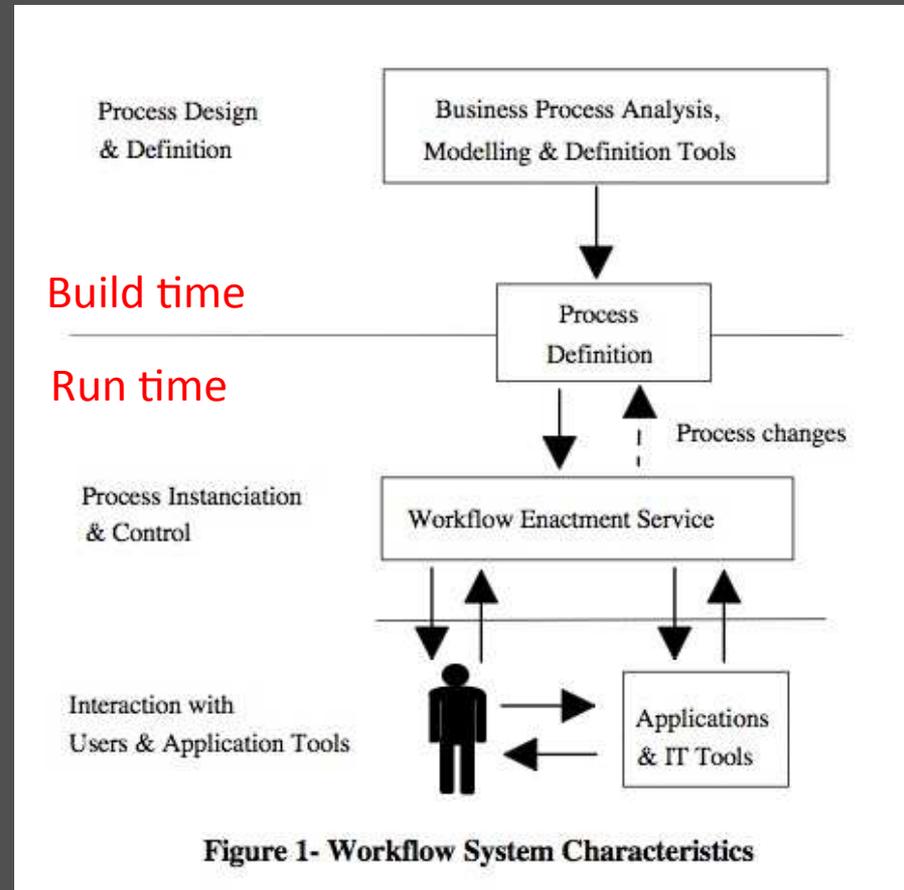
❑ **Adva████**

- Pro███ documentation
- Automation of process execution

For business and science

Workflow Reference Model

❑ **Workflow Management Coallition (WfMC)**

http://www.wfmc.org/reference-model.html

Figure 1- Workflow System Characteristics

Types of workflows

❑ **Process**

- Order a new computer in your organization

❑ **Scientific**

- Align the DNA sequences of 100 patients with the Human Genome, then discover the genetic variations that may lead to a disease

- Find an interesting event in the sun between March and June 2013

Grid

- Run a program (job) with 1000 combinations of three parameters