

# **Globus toolkit**

## **Introduction to file replication systems**

22 Nov 2002

Andrei.Chevel@pnpi.spb.ru

# Our real requirements

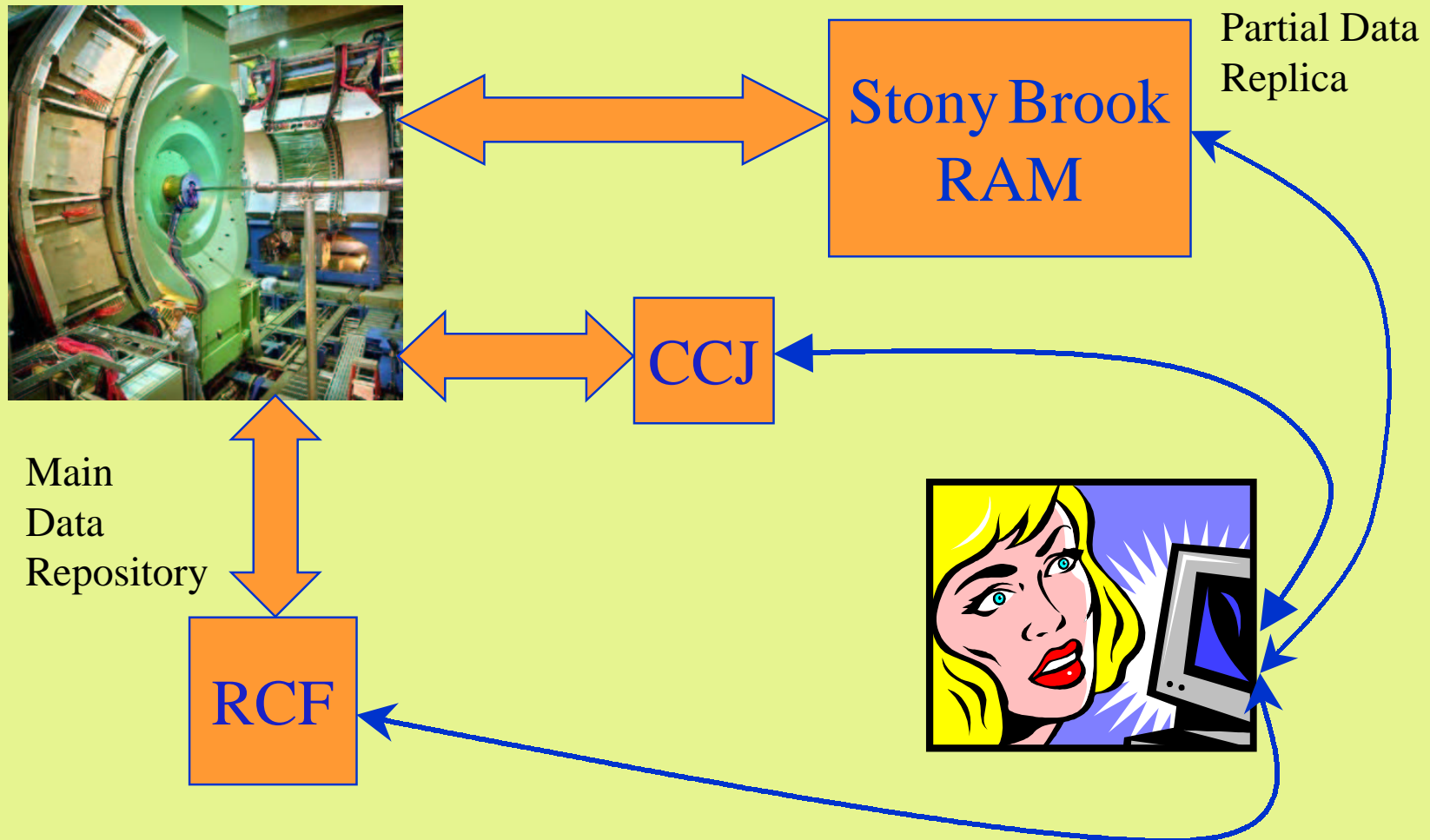
## Ü Job submission

Ü Globus utilities to perform the commands on trusted hosts (in context of GSI).

## Ü Data replication

Ü Grid Data Mirroring Package on top of Globus Replica Catalog  
(<http://www.globus.org/datagrid/replica-catalog.html>)

# General scheme: jobs are going where data are and less loaded clusters



22 Nov 2002

Andrei.Chevel@pnpi.spb.ru

# Replica management in HEP

- Ü Replica is (in our context) the byte by byte copy in general on another host of original file or directory.
  - Ü There might be several replicas in different locations (in one host or in different hosts).
  - Ü It is required to keep the replica catalog which could keep all information about replicas and original files.
  - Ü There are several available systems.

# Replication details

## Ü Files

- Ü A set of identical files (replicas) at multiple location is logical filename (LFN). Standard name contains lfn://suny-ncg//Data/data02/nanodst;
- Ü Each physical file is assigned a physical filename for example pfn://host/file

# Several terms

- Ü **REPLICA** - is a **COPY** of a file in **DIFFERENT PLACE** from **MAIN LOCATION** (**SECONDARY COPY**).
- Ü **MAIN REPLICA** (**Master COPY, MAIN COPY**) - replica in **MAIN LOCATION**.
- Ü **A DIFFERENT PLACE** - might mean different directory, different media or different host.
- Ü **VIRTUAL ORGANIZATION (VO)** - name of any group: one person, small team, detector group, University team, collaboration.

## Terms (continuation)

- Ü LOGICAL FILE NAME - means line in format:  
lfn:/VO/NAME\_OF\_FILE\_ABSTRACTIO  
N
- Ü NAME\_OF\_FILE\_ABSTRACTION -  
name which mean some abstraction of  
the file. Abstraction is the same in spite  
of location.
- Ü PHYSICAL FILE NAME - means line  
in format pfn:/HOST/PATH

## Terms (continuation)

- Ü CONSISTENCY - means REPLICA CATALOG has proper up to date state
- Ü REPLICA CATALOG - means the table (in any form) where LFNs and PFN are located.
- Ü REPLICATION SYSTEM - software (system) which could copy the files over WAN and keep consistent REPLICA CATALOG