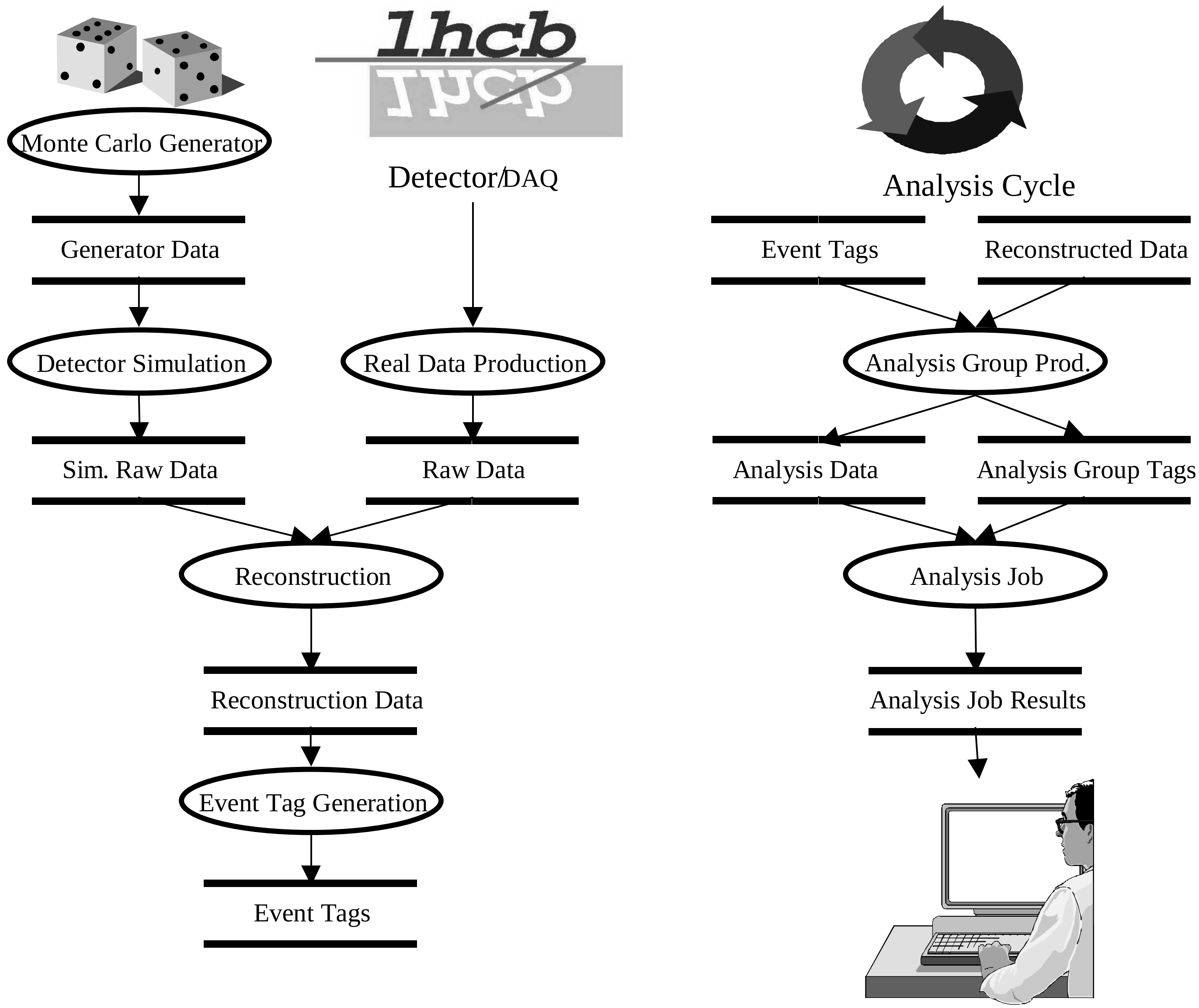
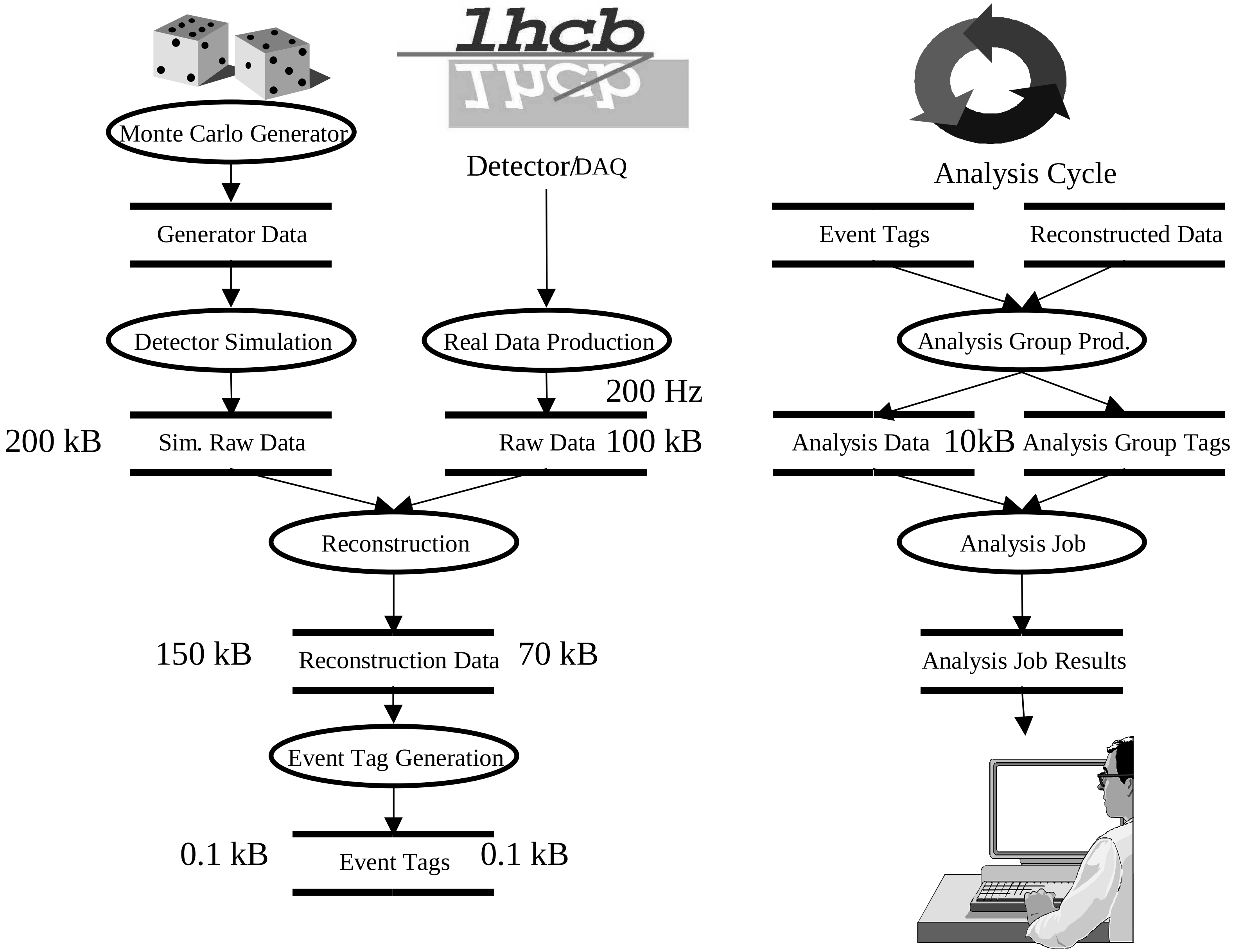
LHCb datasets and processing stages



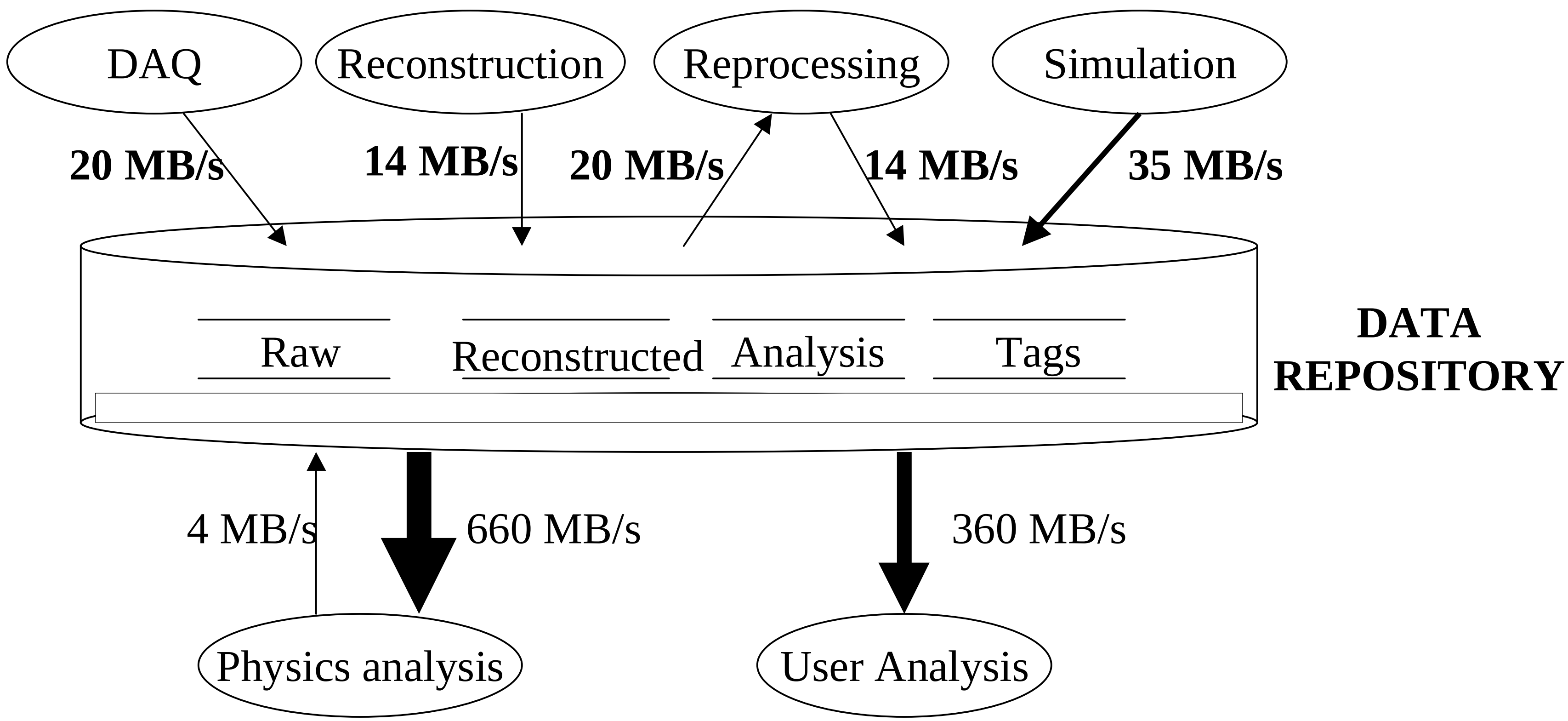
Data Production **Data analysis**

LHCb datasets and processing stages



Data Production **Data analysis**

Data Rates


DAQ
Reconstruction
Reprocessing
Simulation
20 MB/s
14 MB/s
20 MB/s
14 MB/s
35 MB/s
Raw
Reconstructed
Analysis
Tags
DATA REPOSITORY
4 MB/s
660 MB/s
360 MB/s
Physics analysis
User Analysis

Data Volumes

Assume run for 10^7 secs each year

Data type	**Rate /sec**	**Volume /year**
a) Raw data	20 MB/s	200 TB
b) Interesting physics data	1 MB/s	10 TB
c) Simulated data (bx10)	10 MB/s	100 TB
d) Reconstructed Raw	14 MB/s	140 TB
e) Reconstructed Sim	7 MB/s	70 TB
f) Analysis data	4 MB/s	40 TB
TOTAL		**560 TB**

CPU Resources

Assumption is that the cpu power / processor in 2004 will be ~4000 Mips

CPU Resources at the experiment

data production and triggering	1,400x1000 Mips		350 nodes
reconstruction	800x1000 Mips		200 nodes
		Total	**550 nodes**

Event simulation and reprocessing (4 month duty cycle)

monte carlo production	1,400x1000 Mips		350 nodes
reprocessing, event tag creation	800x1000 Mips		200 nodes
		Total	**550 nodes**

Physics Analysis

physics production	10 groups 10^7 kMs/refinement/week		46 nodes
user analysis	100 10^7 kMs/job *2 /week		91 nodes
		Total	**137 nodes**

CPU Farms

- Setup low maintenance commodity processor farms
  - PCs running NT or Linux
  - central server, rack-mounted screenless nodes
- Automated massive installation, upgrade, booting
- Remote disk access by many satellites to central server
- Remote management and monitoring
  - centralised error logs, alarms
  - performance
  - node failure detection and recovery
- Tools for messaging
- Production facilities - batch, scripts, bookkeeping, tape management

Common LHC project - Filter Farms

Data management

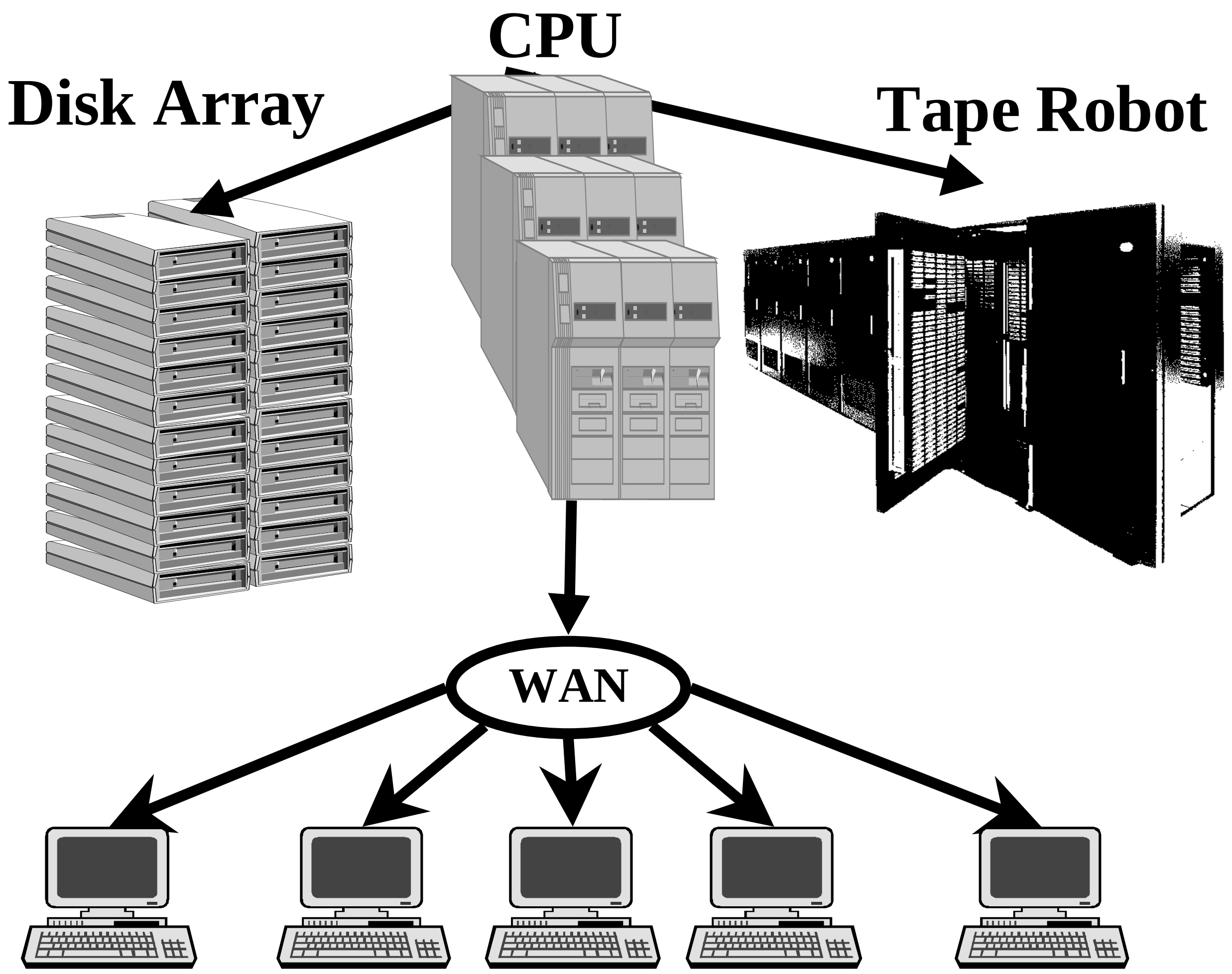
Technology

- Database Technology - ODBMS
- Mass Storage - HPSS

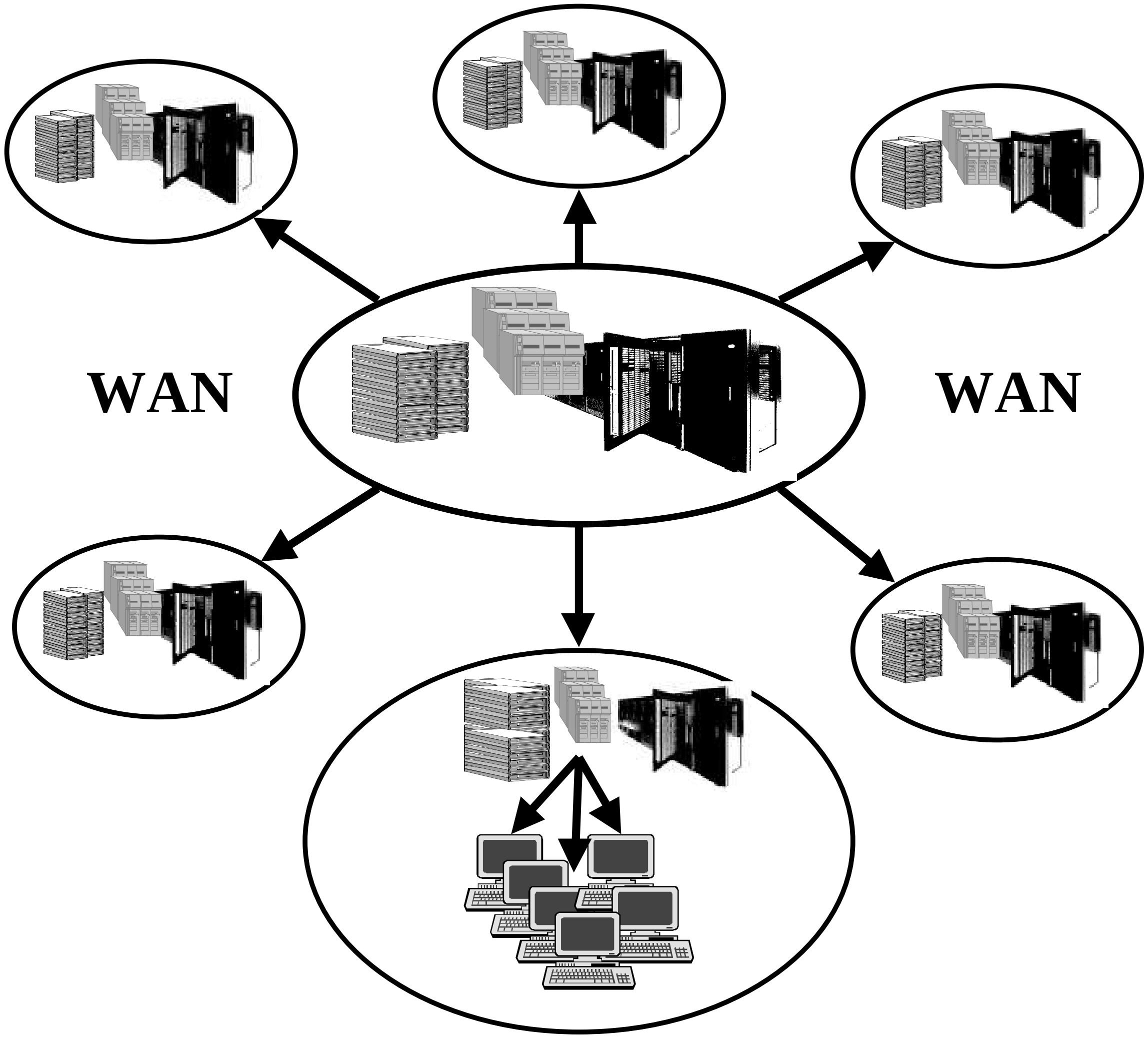
Datasets

- Event repository
- Parameter descriptions - detector,
- Secondary data - calibrations, alignment, output from quality checks

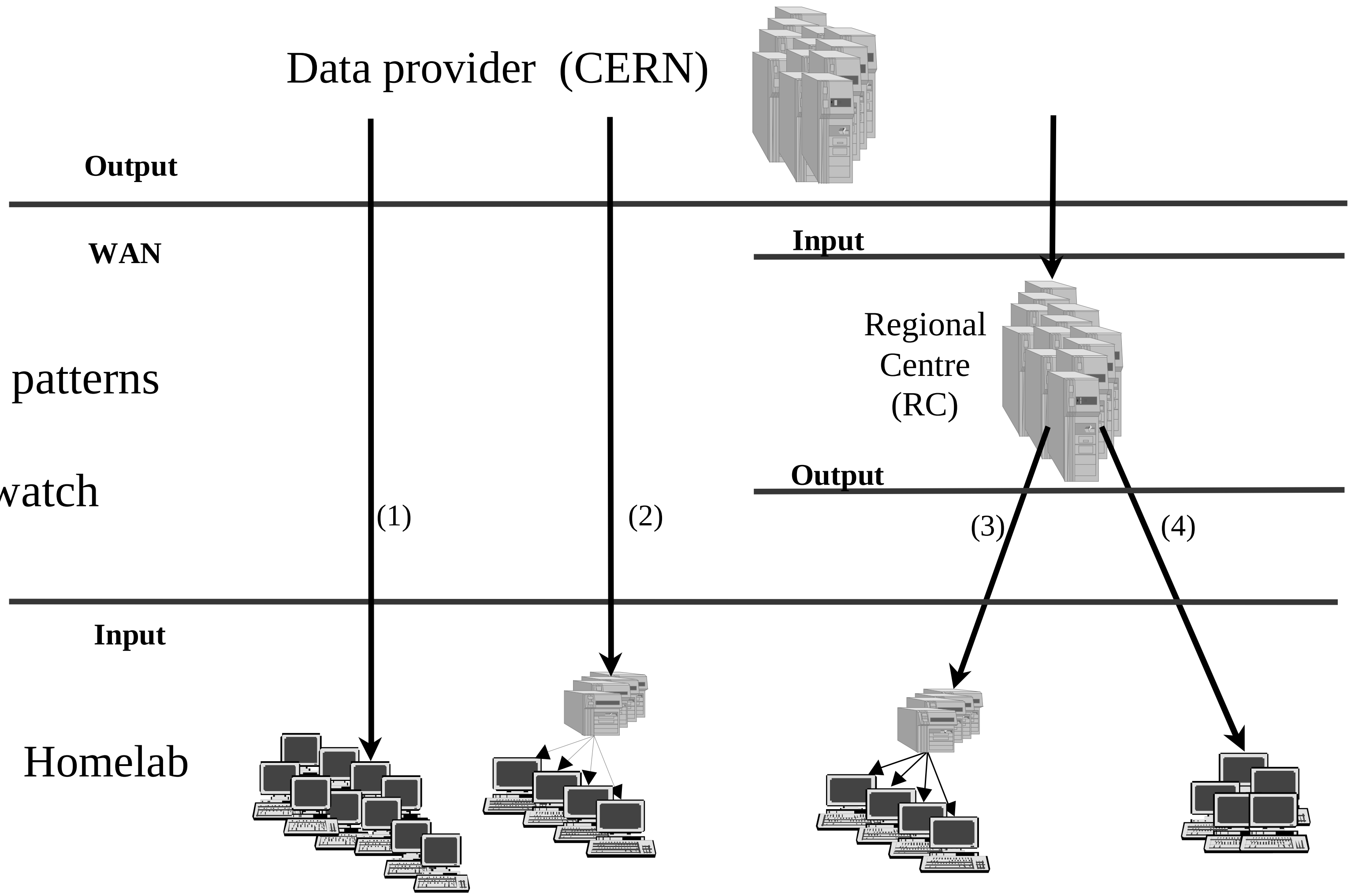
Computing Model - CE(R)NTRIC



Computing Model - DISTRIBUTED



Computing Model - To be studied



MONARC
- study access patterns
- simulation
- technology watch