FairRoot

Status & Plans

M. Al-Turany

Status

- Works with C++11
- Ready for test beams and online analysis
- Re-engineering of some base classes is ongoing (Clean some historical stuff!)

Hot Topics

- Concurrency
- Grid /Distributed Computing
- FairRoot & ALICE O^2

Concurrency: Where we are now?

- Single threaded single process ROOT event loop
- User code is in Task hierarchy that runs sequentially
- Grid/batch jobs run embarrassingly parallel (one process/core)

What are the Problems

- C and C++ do not offer any support for concurrency!
- Embarrassingly parallel workload does not scale
  - Memory needed for each process ➔ expensive
  - How this scheme should work for the Online clusters?

Multi-processing vs. Multi-threading

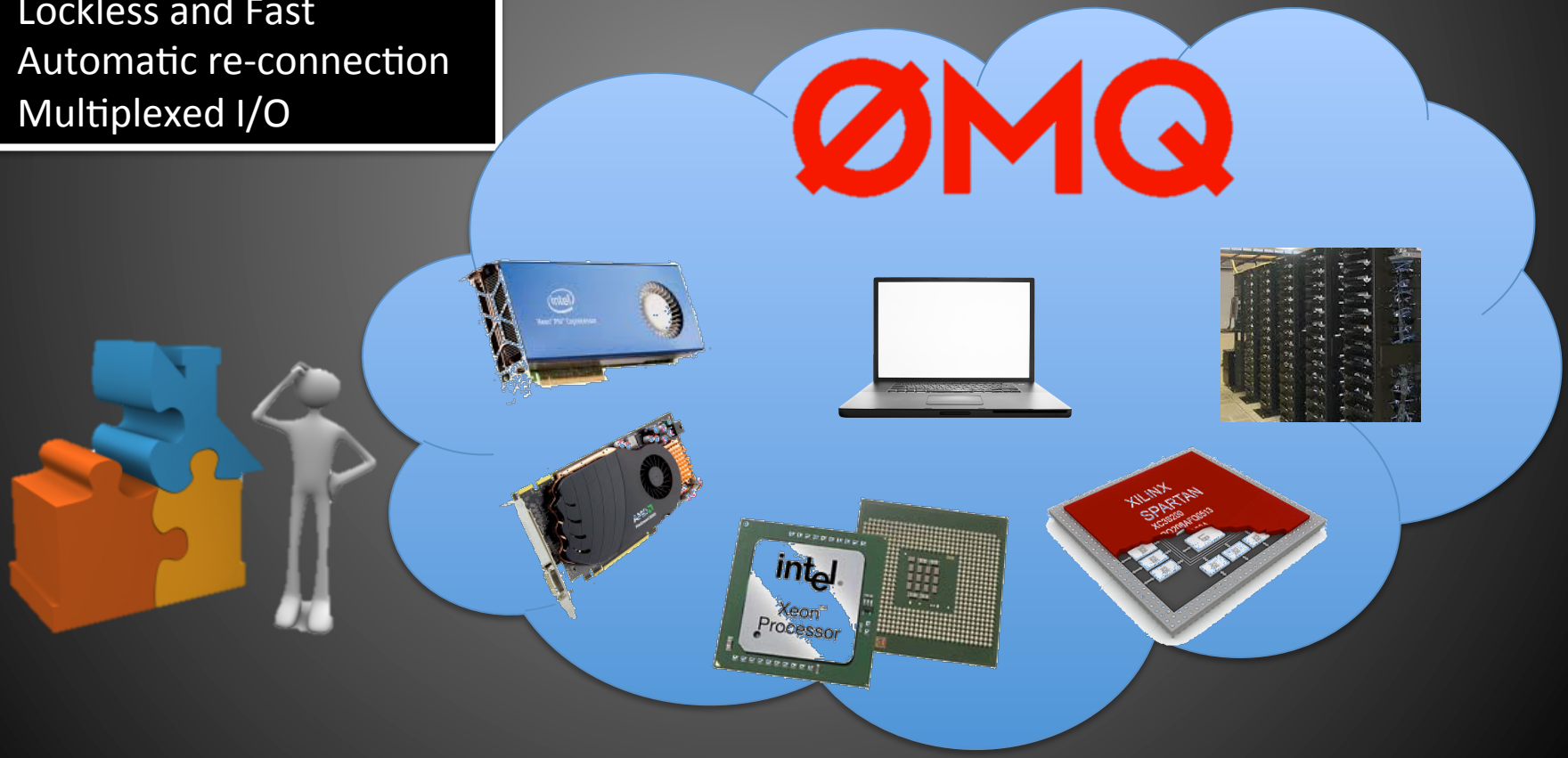
- Different processes are insulated from each other by the OS, an error in one process cannot bring down another process.
- Inter-process communication can be used across network
- Error in one thread can bring down all the threads in the process.
- Inter-thread communication is fast

Correct balance between reliability and performance

- Multi-process concept with message queues for data exchange
  - Each "Task" is a separate process, which can be also multithreaded, and the data exchange between the different tasks is done via messages.
  - Different topologies of tasks that can be adapted to the problem itself, and the hardware capabilities.

A cloud that let you connect different pieces together

- BSD sockets API
- Bindings for 30+ languages
- Lockless and Fast
- Automatic re-connection
- Multiplexed I/O

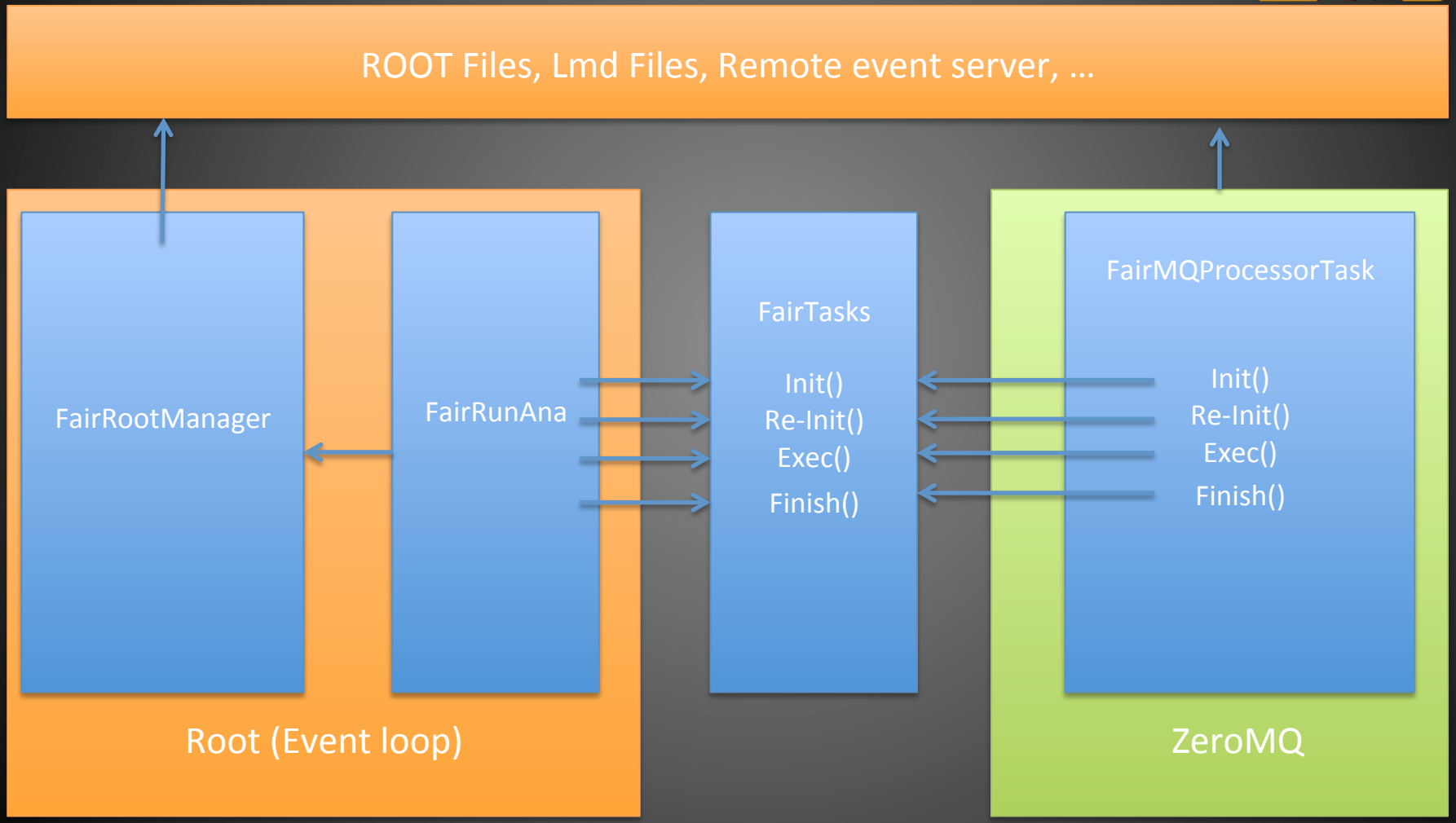


nanomsg is under development by the original author of ZeroMQ

nanomsg

- Pluggable Transports:
  - ZeroMQ has no formal API for adding new transports (Infiniband, WebSeockets, etc). nanomsg defines such API, which simplifies implementation of new transports.
- Zero-Copy:
  - Better zero-copy support with RDMA and shared memory, which will improve transfer rates for larger data for inter-process communication.
- Simpler interface:
  - simplifies some zeromq concepts and API, for example, it no longer needs Context class.
- Numerous other improvements, described here: http://nanomsg.org/documentation-zeromq.html
- FairRoot is independent from the transport library
  - Modular/Pluggable/Switchable transport libraries.

Integrating the existing software:



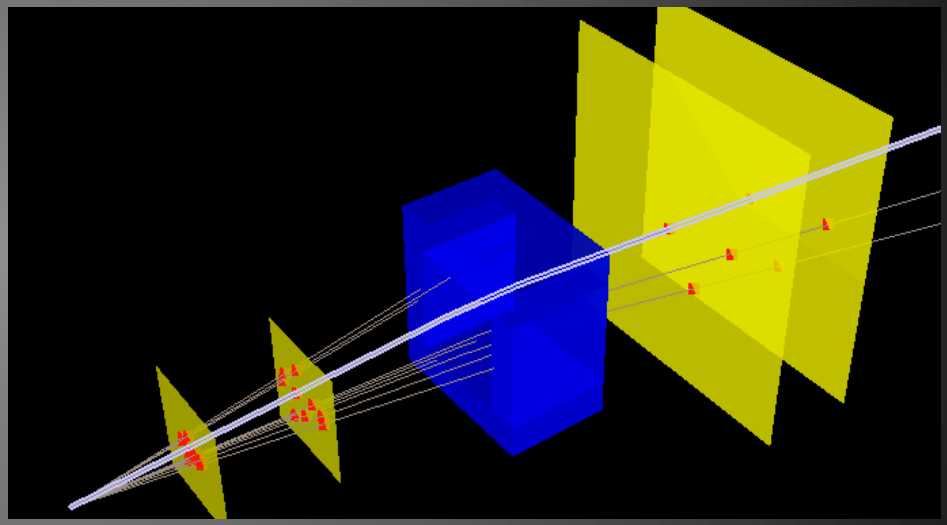
FairRoot: Example 3

4 -Tracking stations with
a dipole field

Simulation:
10k event: 300 Protons/ev

Digitization

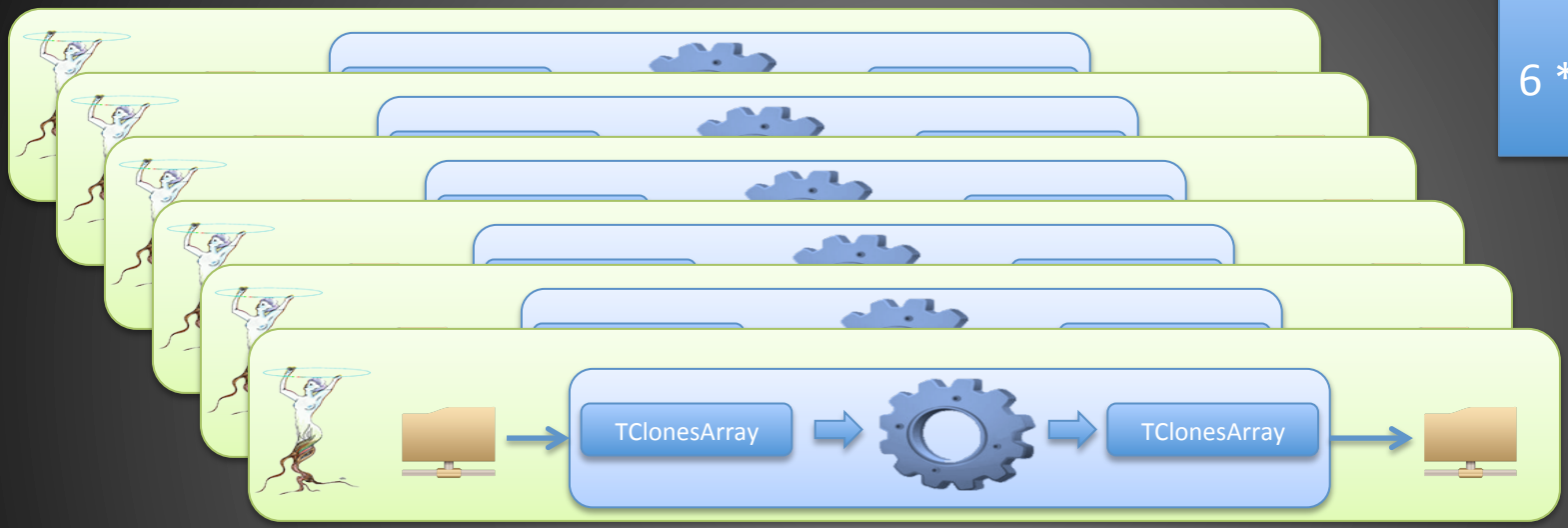
Reconstruction:
Hit/Cluster Finder



2 x 2.4 Xeon Quad core Intel Xeon
16 GB Memory

171 s

6 * 263 MB



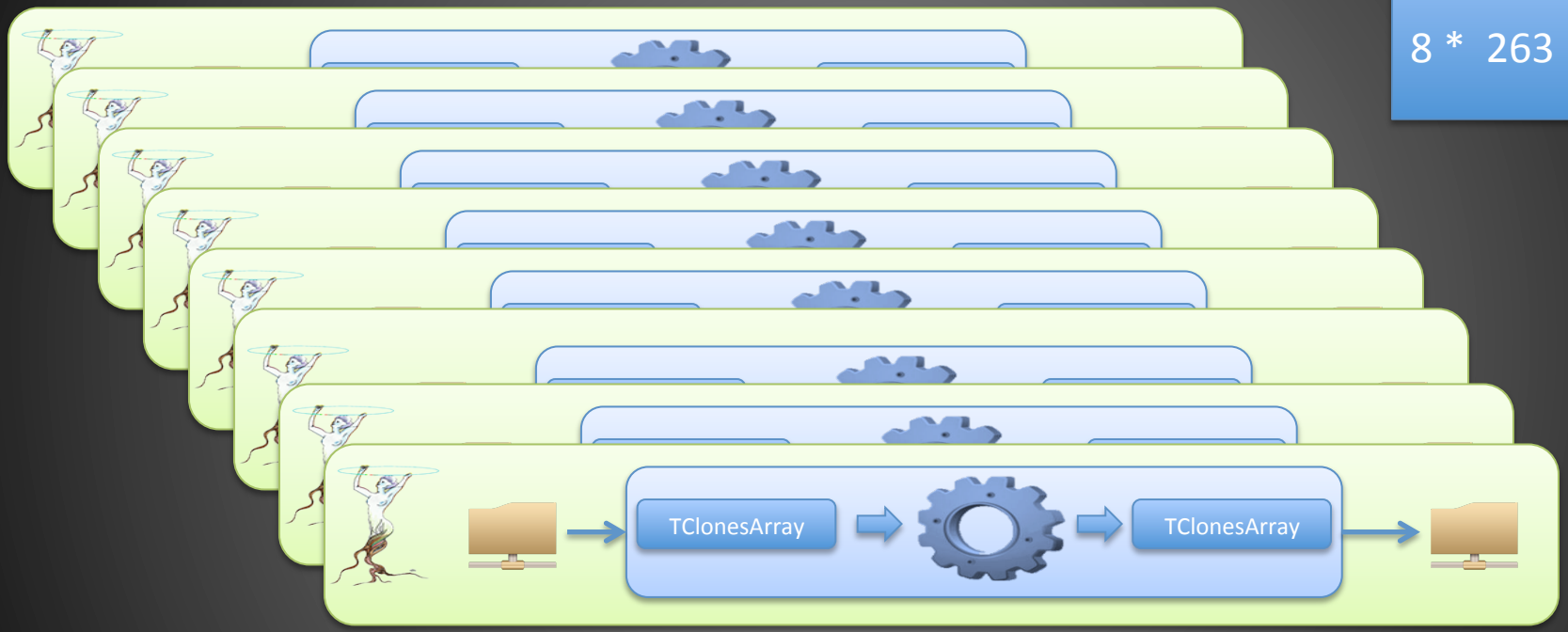
Throughput ~ 3500 ev/s

Wall time: 171 s
Total Event: 60k events

2 x 2.4 Xeon Quad core Intel Xeon 16 GB Memory

300 s

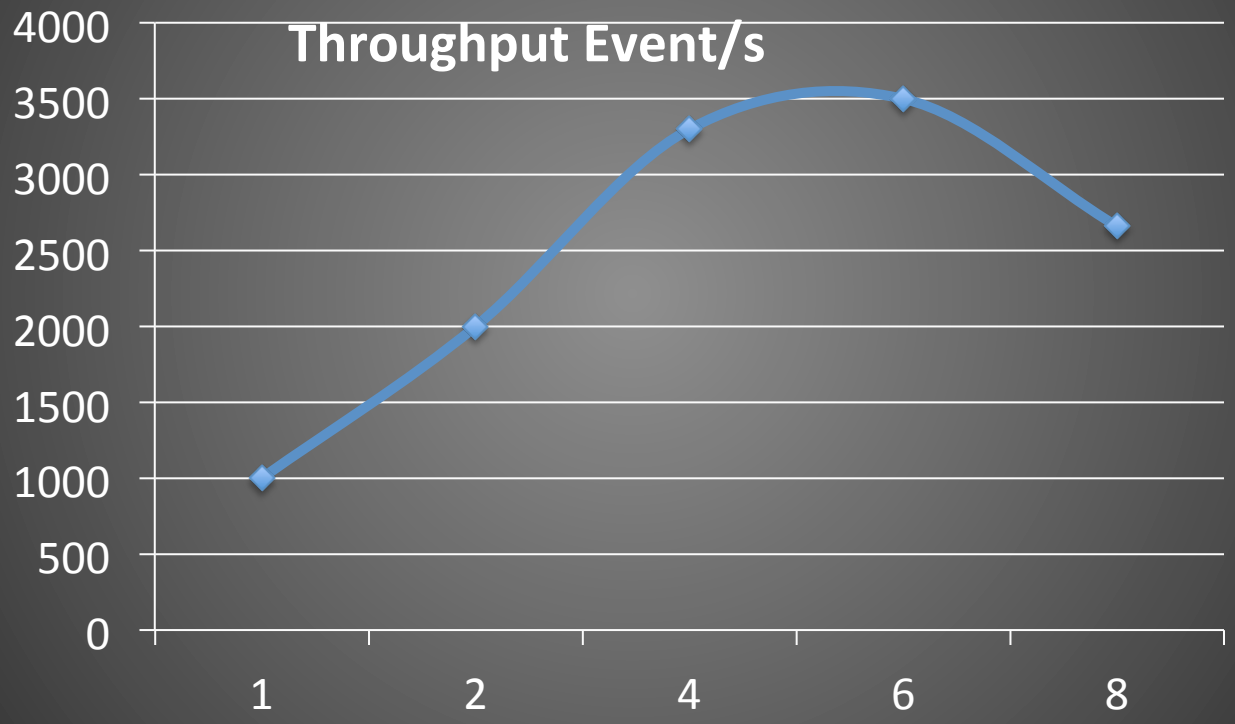
8 * 263 MB



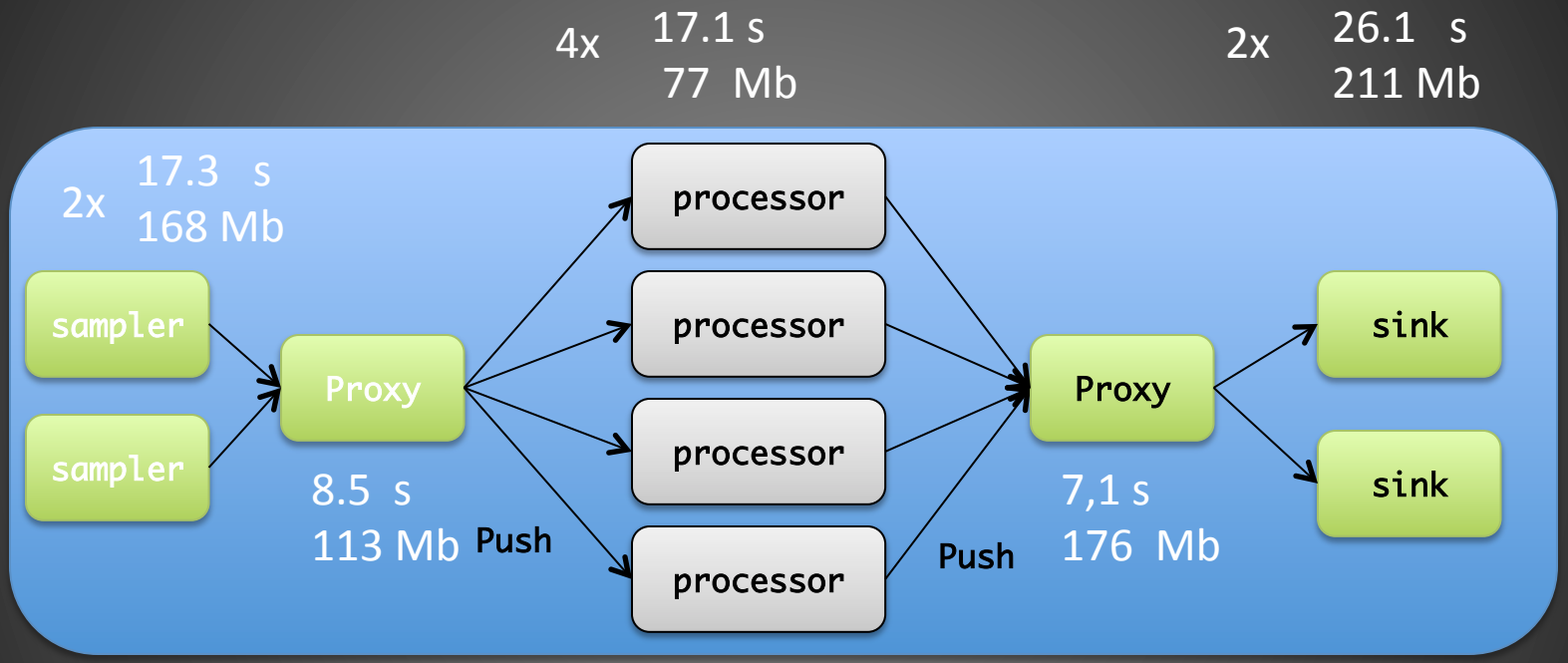
Throughput ~ 2660 ev/s

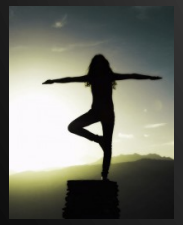
Wall time: 300 s
Total Event: 80k events

2 x 2.4 Xeon Quad core Intel Xeon 16 GB Memory



2 x 2.4 Xeon Quad core Intel Xeon
16 GB Memory

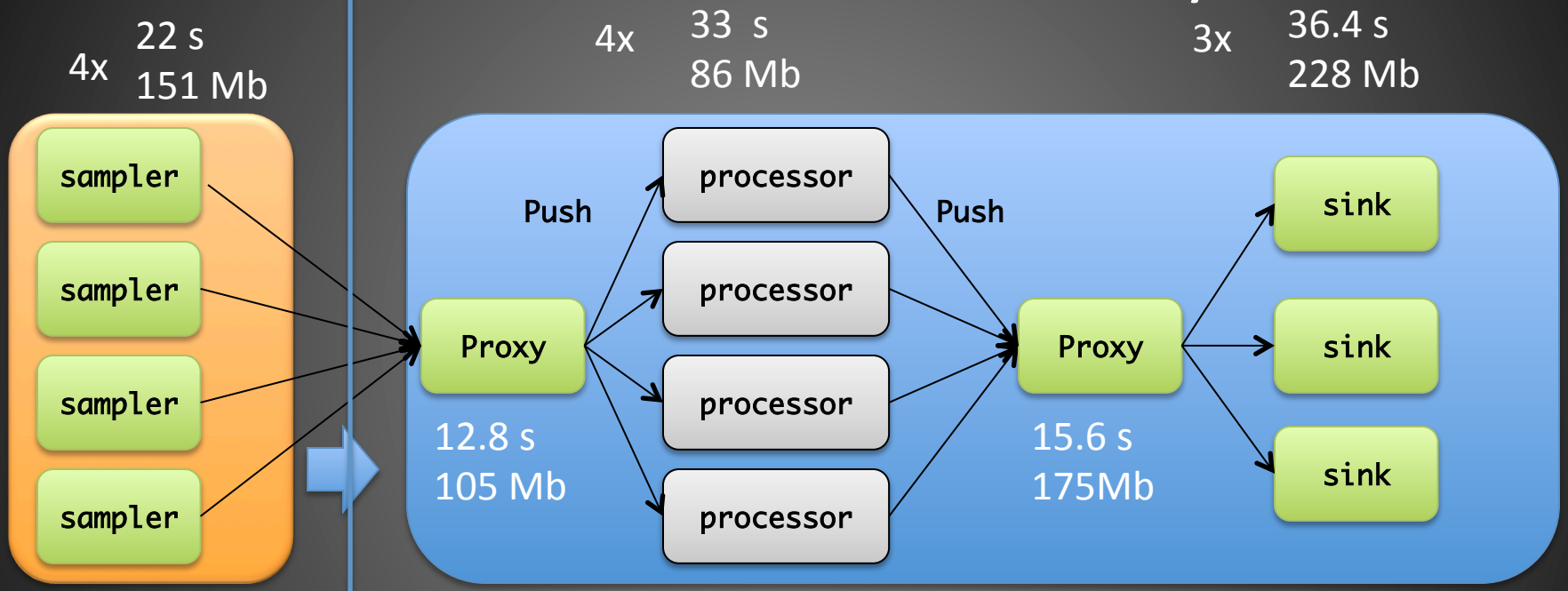




Throughput ~ 7400 ev/s
Total Memory 1355 Mb

Wall time: 26.1 s
Total Event: 20k events

2 x 2.4 Xeon Quad core Intel Xeon 16 GB Memory



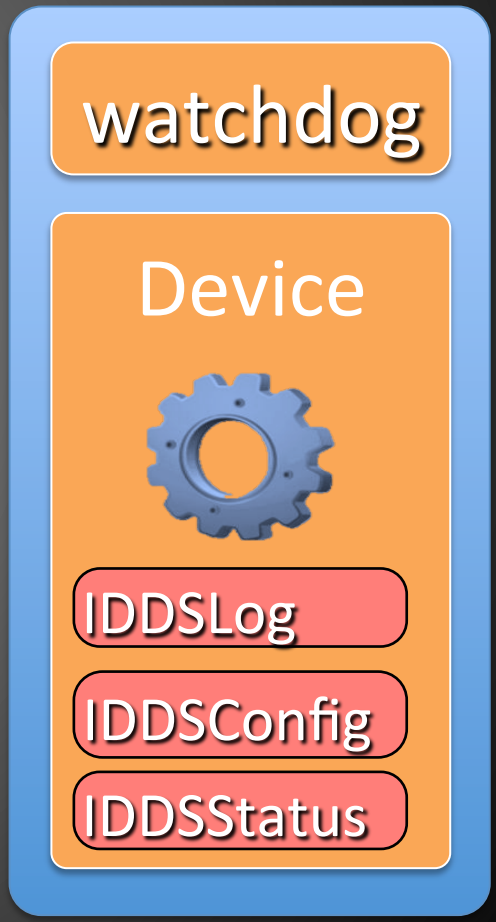
Throughput ~ 10990 ev/s

Wall time: 36.4 s
Total Event: 40k events

Gigabit Ethernet

New building blocks for FairRoot:

- Each entity has its own watchdog process
- The entity is statically inherited and implement 3 interfaces:
  - IDDSConfig
    - interface to configuration files
    - used by the watchdog and by the user process
  - IDDSStatus
    - High and low level status info used by the watchdog
  - IDDSLog



FairRoot worked with AliEn grid, But:

- The new FairRoot concept (Data driven framework) will NOT work with ALiEn as it is now.

- It seems that there will be no future for the current AliEn even by ALICE itself!

AliEn Grid

ALICE

Why change AliEn?

- While the system currently fulfills all the needs of ALICE users for reconstruction, simulation and analysis there are concerns about scalability of the file catalog beyond Run2
- Need to address the use for emerging cloud, volunteer as well as the opportunistic resources for ALICE
- In general, no manpower for maintenance and continuous development of the current system
- Adopt common solutions and tools where exist for a given use case

Predrag Buncic - ALICE Offline Coordinator

AliEn Grid

Predrag Buncic - ALICE Offline Coordinator

A Large Ion Collider Experiment

ALICE

Conclusions

- Run3+ will impose much higher computing requirements
  - ~~100x more events to handle~~
- Simply scaling the current Grid won't work
  - We need to reduce the complexity
- ~~Regional clouds may be an answer~~
  - We are **not** abandoning the grid
  - "Our grid on the cloud" model
- In order to complement these resource we might have to tap into flagship HPC installations
  - Not exactly grid friendly environment, needs some work(arounds)
- Looking for synergies and collaboration with other experiments and IT
  - Transition to clouds, CVMFS, location aware service, edge/proxy services, data management, release validation, monitoring, pilot jobs, distributed analysis, use of opportunistic resources, volunteer computing…

Lessons to learn from other LHC experiments!

Grids: what did we achieve?

And fail to achieve?

- Solved our problem of making effective use of distributed resources
- Made it work at huge scale
- Effective to ensure all collaborators have access to the data
- Networks are a significant resource
- Federation of trust and policies – important for future

- Cluster computing/grids not suitable/needed for many sciences
- Operational cost is high
- Very complex middleware was not (all) necessary
- Many tools were too HEP-specific

CERN 19th May 2013 ACAT 2013 Ian.Bird@cern.ch 5

Ian Bird
LHC Computing Grid Project Leader
CERN IT Department

Lessons to learn from other LHC experiments!

Some lessons

- Fear of the network was unfounded – remarkable success – technology evolved our problem away
  - Network is a resource, not a problem
- Initial computing models too rigid – and too hierarchical
- Service deployment was too complex
  - Driven by fears of unreliable networks
  - E.g. distributed databases; evolved to caches, and simple central instances
- Original data placement model was far too wasteful
- Data management was/is the main problem – but huge efforts invested in job management

Ian Bird
LHC Computing Grid Project Leader
CERN IT Department

Lessons to learn from other LHC experiments!

So what is the goal?

- Aid the research community that now requires computing and data resources:
- Provide an e-Infrastructure environment that:
  - Optimally provides computing and data services for science research communities and individuals
  - Can adaptably/flexibly/dynamically provide new (types of) services
- Ultimately this must be operationally sustainable with the science funding available
  - Must understand the business model

Ian Bird
LHC Computing Grid Project Leader
CERN IT Department

Fair Computing and Grid

- Do we need the current grid?
  - This is a question which the experiment should answer!
- FairRoot group will:
  - Support all reasonable environments available
  - Run on Laptop and Cluster
  - NOT participate in any middle ware development, we do not have the man power or the knowhow to do that.

FairRoot & ALICE O^2

- Long years of close collaboration with the Alice offline group
- The computing requirement for Alice after upgrade are very similar to FAIR experiments
- The new design suggested for FairRoot fulfill also the requirement of Alice O^2

FairRoot & ALICE O^2 (cont.)

- A common development with ALICE would be beneficial for all
- The HLT of ALICE could be seen as a proto type for FAIR online clusters
- We will benefit from the experience gained by supporting a running experiment!