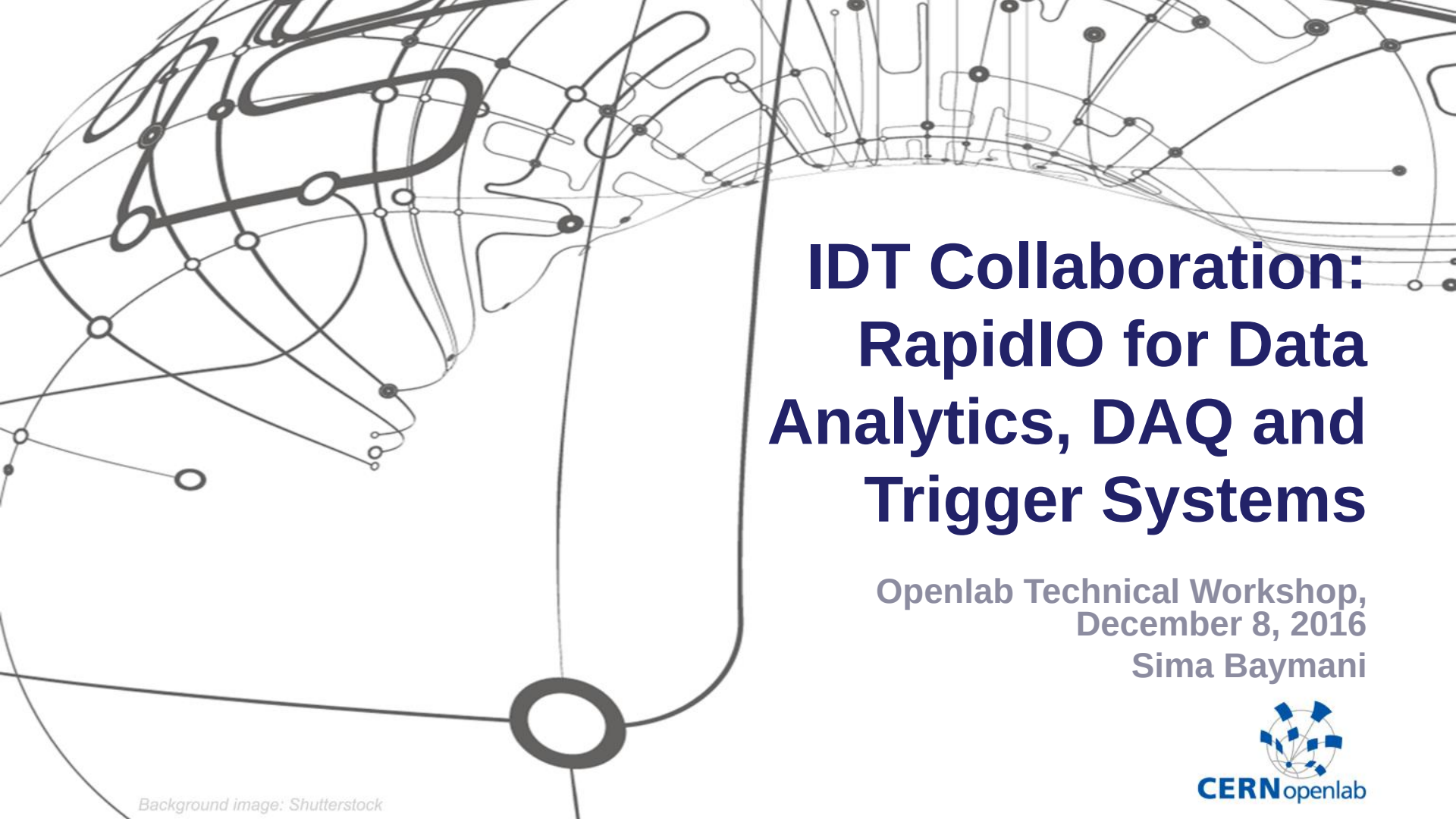


IDT Collaboration: RapidIO for Data Analytics, DAQ and Trigger Systems

Openlab Technical Workshop, December 8, 2016

Sima Baymani

CERN openlab

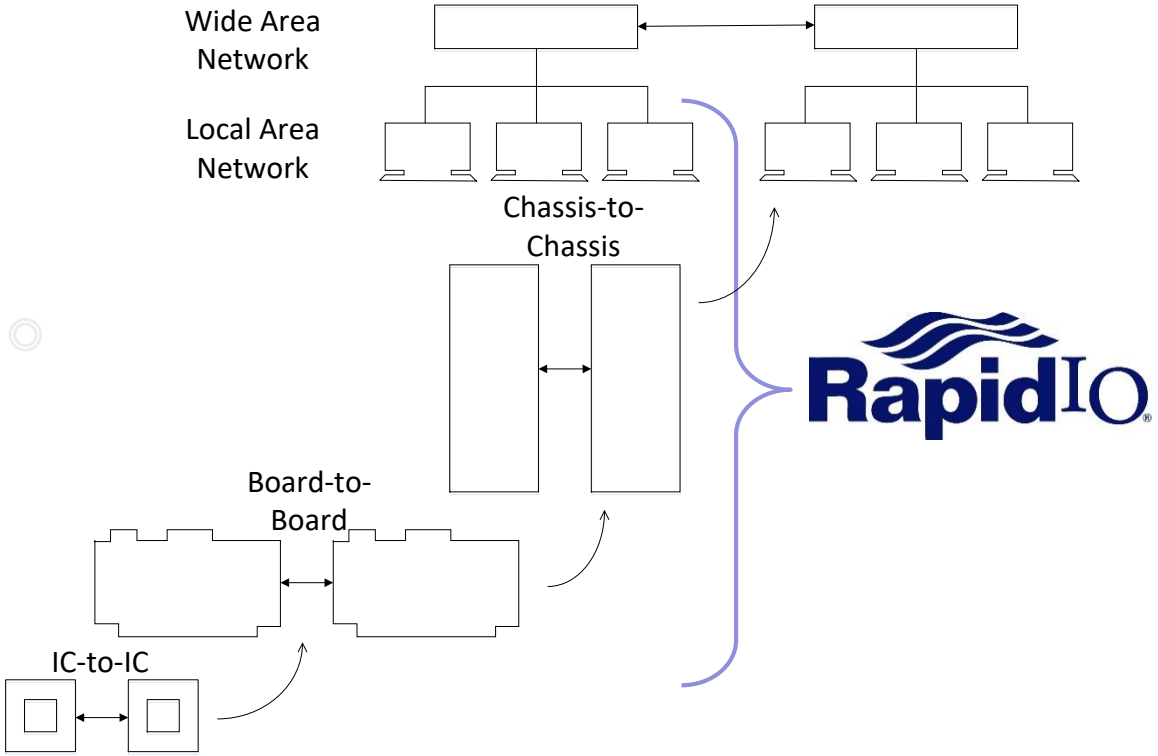
Background image: Shutterstock

CERN openlab Partner: IDT

- Integrated Device Technology
- ~1500 employees
- HQ in San Jose, CA
- Products target hard real time systems
- CERN openlab partner since 2015

RapidIO Introduction

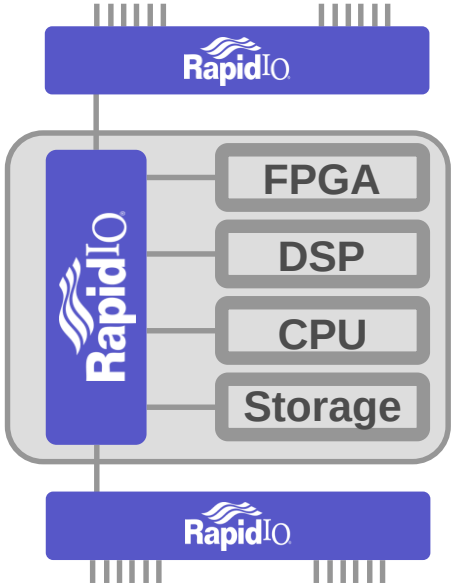
- Introduced in 1997
- From front side bus to system level interconnect
- Open standard – rapidio.org
- Meets real time needs as well as scalability



Features

- Combines scalability with low latency
  - Switch latency ~100 ns
  - Memory to memory latency < 1 µs
- CPU offload
  - Error recovery in physical layer
  - Protocol stack processed in HW
- Operations include
  - Read/write (remote DMA)
  - Messaging (4KB)
  - Doorbells (events)

- Heterogeneous Systems



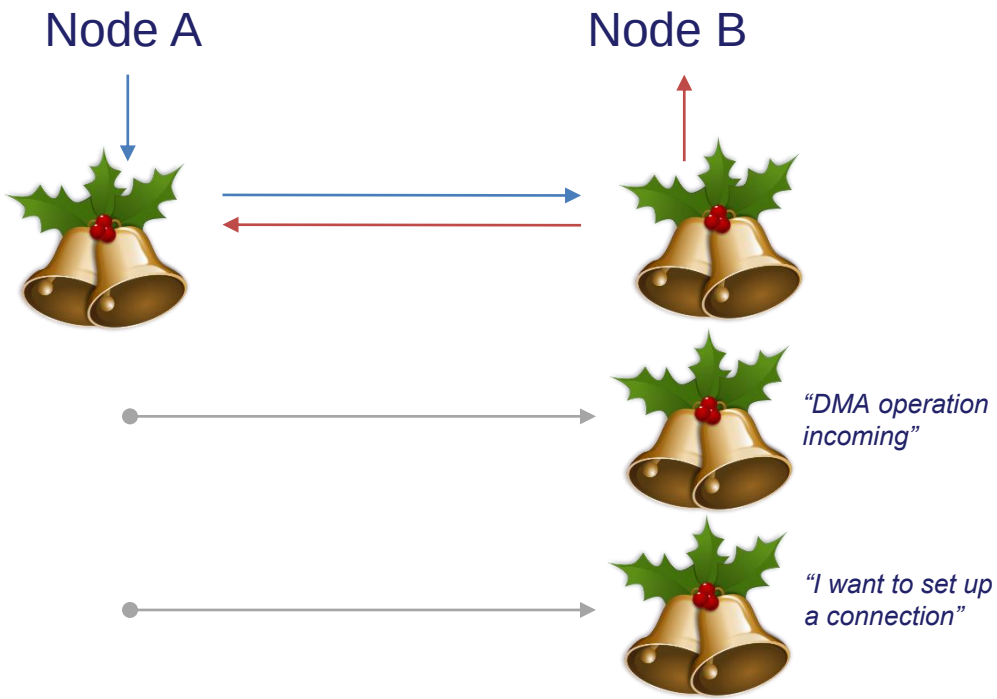
Setup

- **16 server nodes equipped with RapidIO-PCIe bridge cards**
  - Throughput 12 Gbps
- **38-port Top of Rack RapidIO switch**
- **RapidIO drivers for the Linux kernel**
- **User space libraries for Linux**

Image source: CERN, http://cds.cern.ch/record/2136852/files/DSC_2204.JPG?version=1

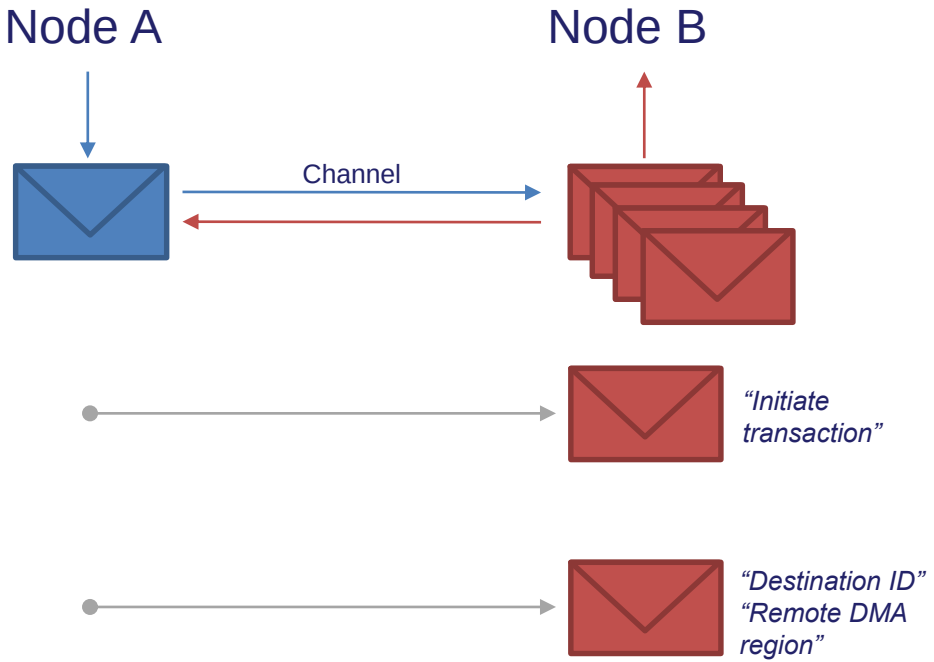
Doorbells

- Event-like messages
- Simple interface:
  - No connection required
  - Define a range of values
  - Send!
- Good for notifications
  - DMA transfer finished
  - Initiate connection

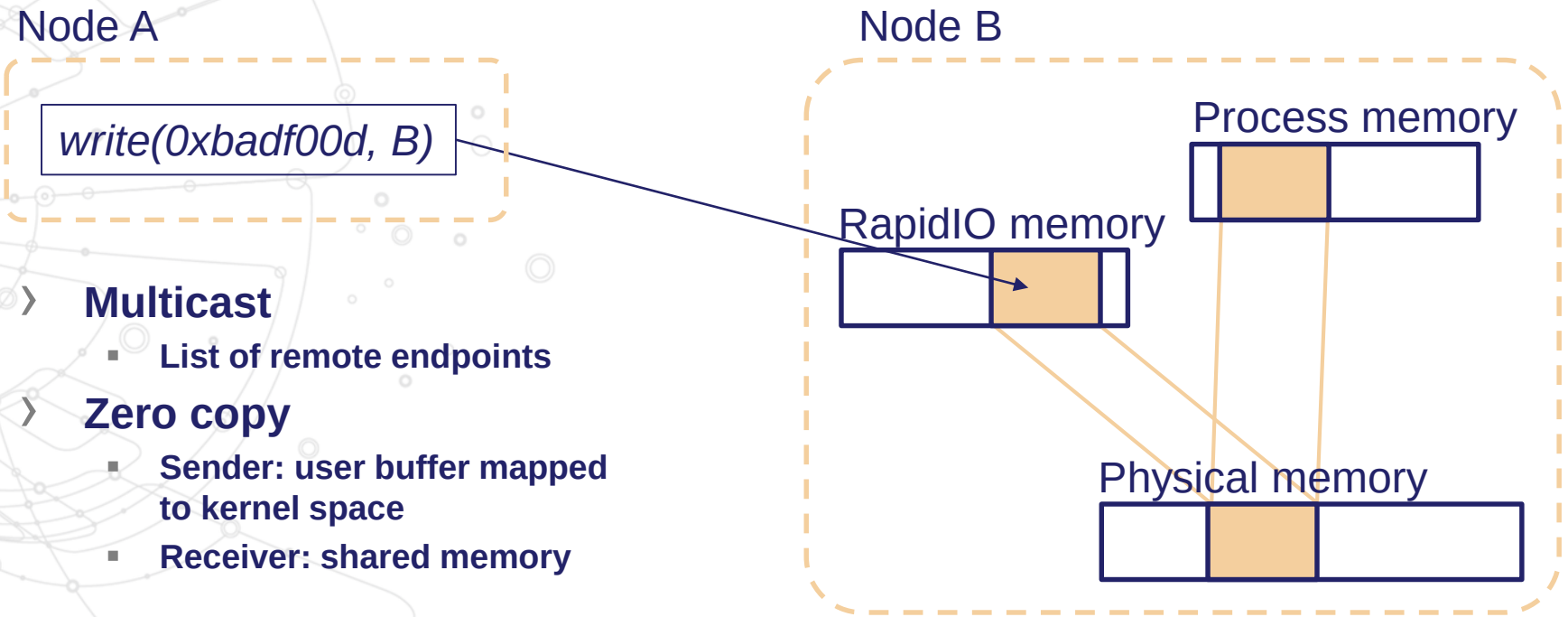


Channelized Messages

- Socket-like interface
- Up to 4 KB/message
- Good for orchestration
  - Initiate transactions
  - Exchange remote DMA information



Remote DMA Write



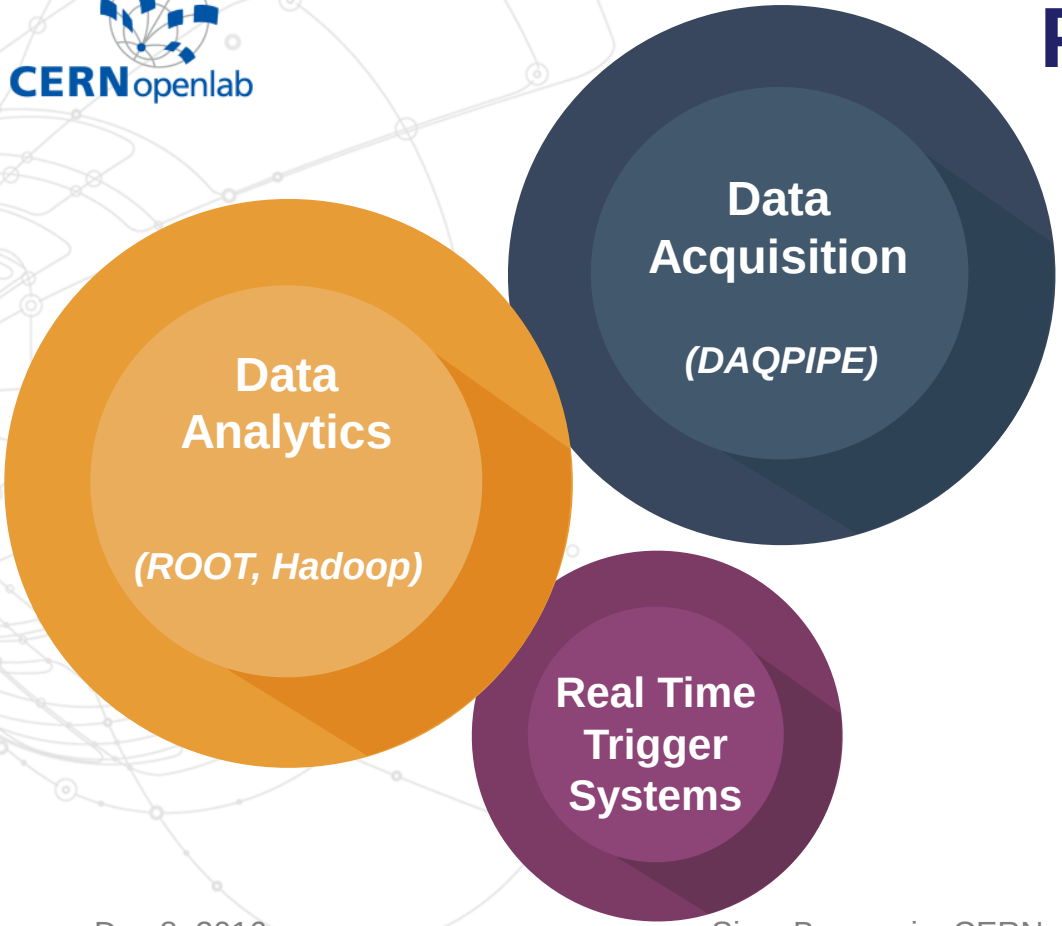
- **Multicast**
  - **List of remote endpoints**
- **Zero copy**
  - **Sender: user buffer mapped to kernel space**
  - **Receiver: shared memory**

riosockets

- Emulated TCP/IP over RapidIO
- Standard network interface

iperf setup	Speed
Half duplex 1-to-1	11 Gbps
Duplex 1-to-1	11 Gbps in both directions

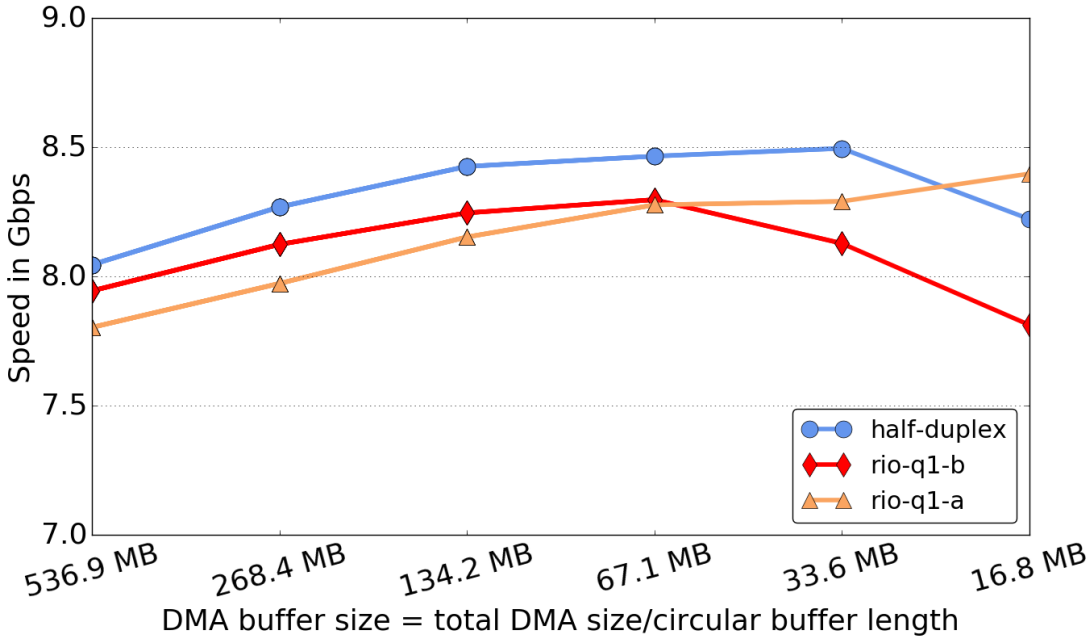
Project Use Cases

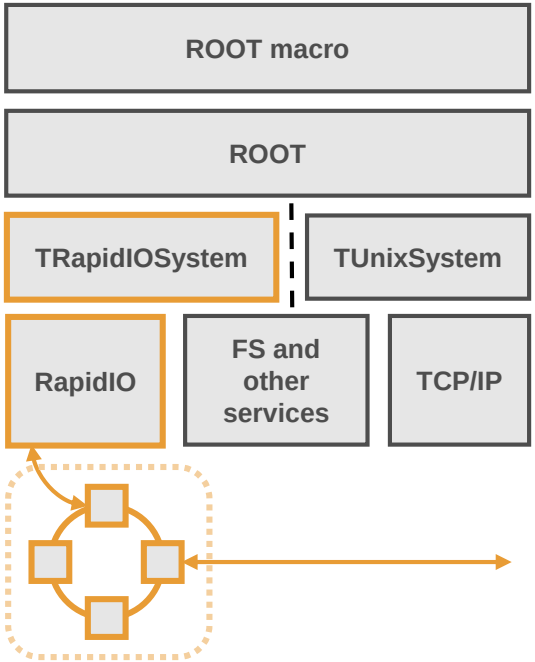


- Explore RapidIO
- Focus on different areas
- Evaluate suitability

Use Case 1: ROOT

Half-duplex vs Full duplex





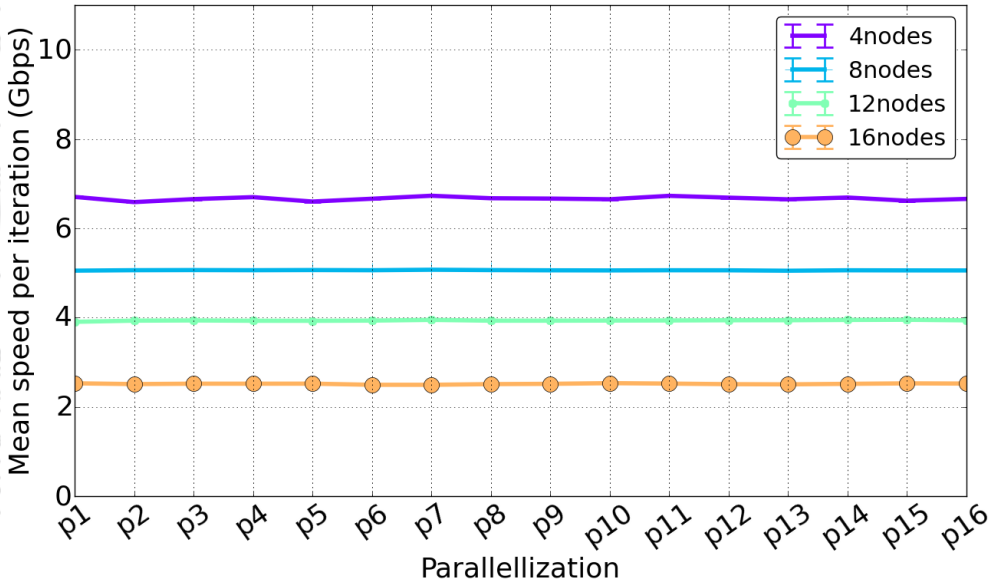
Use Case 1: Hadoop

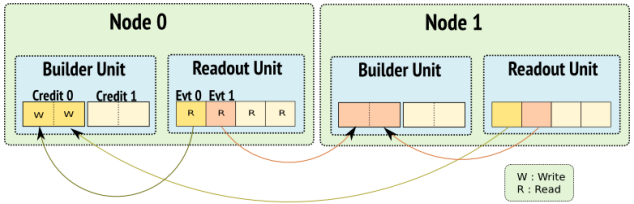
- Set up Hadoop configuration to use riosockets
- No porting work needed!
- Suitable benchmarks hard to find
  - Existing ones don't exercise the network enough

Image source: http://insidebigdata.com/2015/07/21/hadoop-for-hpc-it-just-makes-sense/

Use Case 2: LHCb DAQPIPE

Parallelization scan, credits = 1 (work in progress)





- LHCb benchmark for DAQ network
- Event data spread across network
- Nodes collect data for one event (credit)
- All-to-all communication!

Porting Experience

- **Template Openlab project**
- **Porting or providing?**
  - Standard APIs valuable
  - If not, abstraction layers in application architecture
  - Or specialized (self-owned) software

Next Steps

- Develop test suite for different communication patterns
- Further investigation on Hadoop
- Use case 3: Triggering systems
- Explore multicast

Thank you!

Extra: Latency on Skylake

Write Size	Avg (us)
1	2.087
2	2.056
4	2.626
8	2.16
16	2.092
32	2.779
64	2.832
128	3.18
256	3.312
512	3.672
1024	3.818

Write Size	Avg (us)
2048	4.426
4096	5.685
8192	8.089
16384	13.002
32768	22.791
65536	42.37
131072	81.54
262144	159.907
524288	316.681
1048576	630.093
2097152	1257.119
4194304	2510.724