

Developments on the ALFA Parameter manager

Charis Kouzinopoulos, CERN

CWG13 11.03.2016



Motivation



A new Parameter manager is under development for ALFA - the Condition and Calibration Data Base (CCDB)

Intended as a replacement for the Offline Conditions Database (OCDB) that is currently (Run 1 and 2) used to store calibration and alignment data

Why a new solution – requirements for the new DB

- Scalability - High number of processes/growing size of the DB
- Low latency – Faster storage/retrieval
- Redundancy – Data loss avoidance

OCDB is not a database - data consist of a set of entries in the AliEn file catalog that point to the actual ROOT files stored at the Grid that contain the data

Parameter Manager

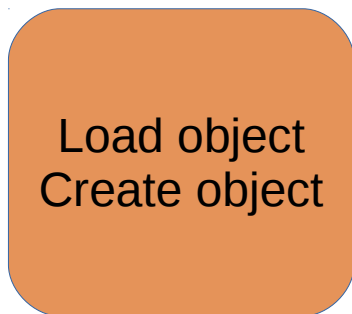
A continuation of the work that started by Tom Van Steenkiste

Tom Van Steenkiste 2016

An abstraction layer with different modules: data loading, serialization and DB communication

The modules offer a simple interface:

Data loading



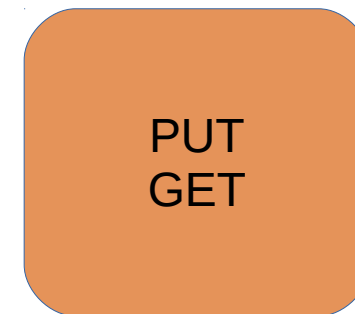
OCDB AliCDBEntry ROOT files

Serialization



messages

DB communication



key/value pairs

Additional modules are considered!



Parameter Manager



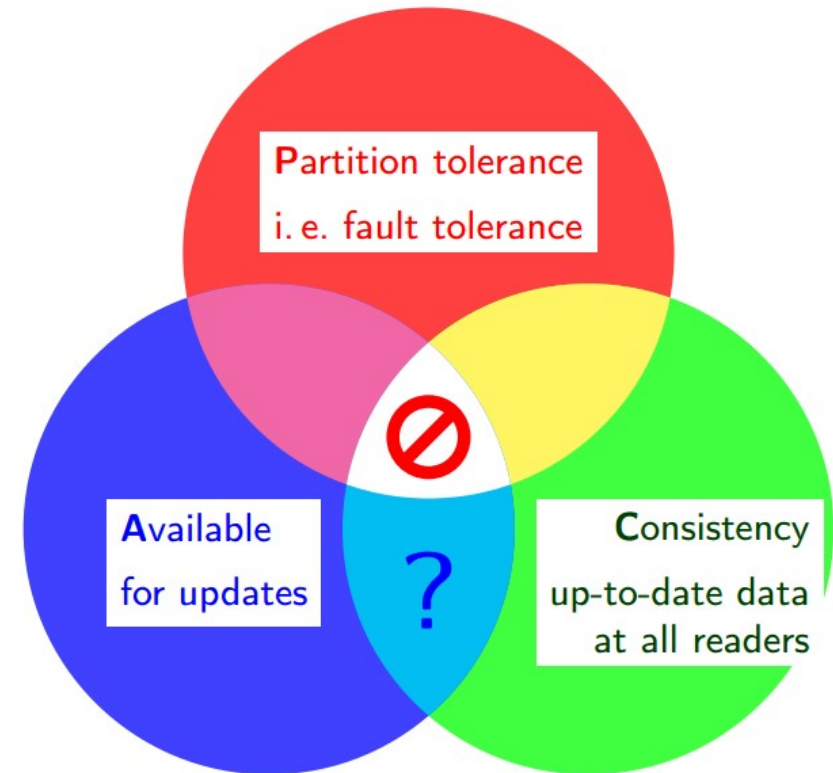
Why Riak?

A popular key/value *distributed* database

Offers data availability: distributes data across multiple nodes

It handles synchronization between the nodes internally in a transparent way

It is an *eventually consistent* system – in a failure scenario data *can* be available but *potentially* not up to date



A distributed storage system can have at most two out of three desirable properties

See Jakob Blomer's primer on key/value databases

Jakob Blomer 2015

Parameter Manager

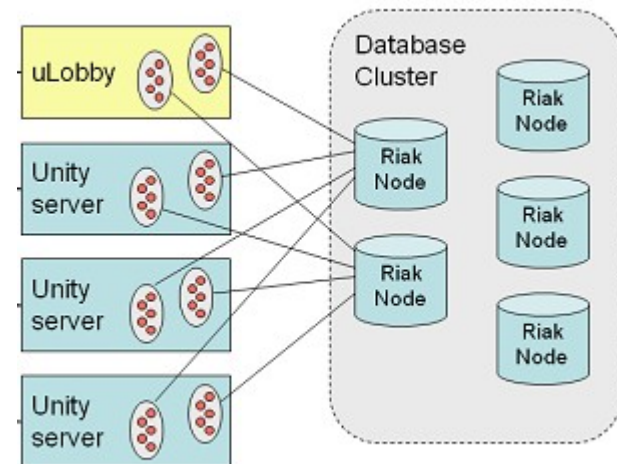
Why Riak?

Riak organizes data into Buckets, Keys and Values

Values are identified by a unique Key

Each Key/Value pair is stored in a bucket

Buckets offer a flat namespace - allow multiple keys with the same name to exist in the database and some per bucket configurability



See Jakob Blomer's primer on key/value databases

Jakob Blomer 2015

Parameter Manager



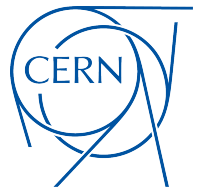
There is no C/C++ library available to interface with Riak with sufficient support

The components of the Parameter Manager use C++/Java

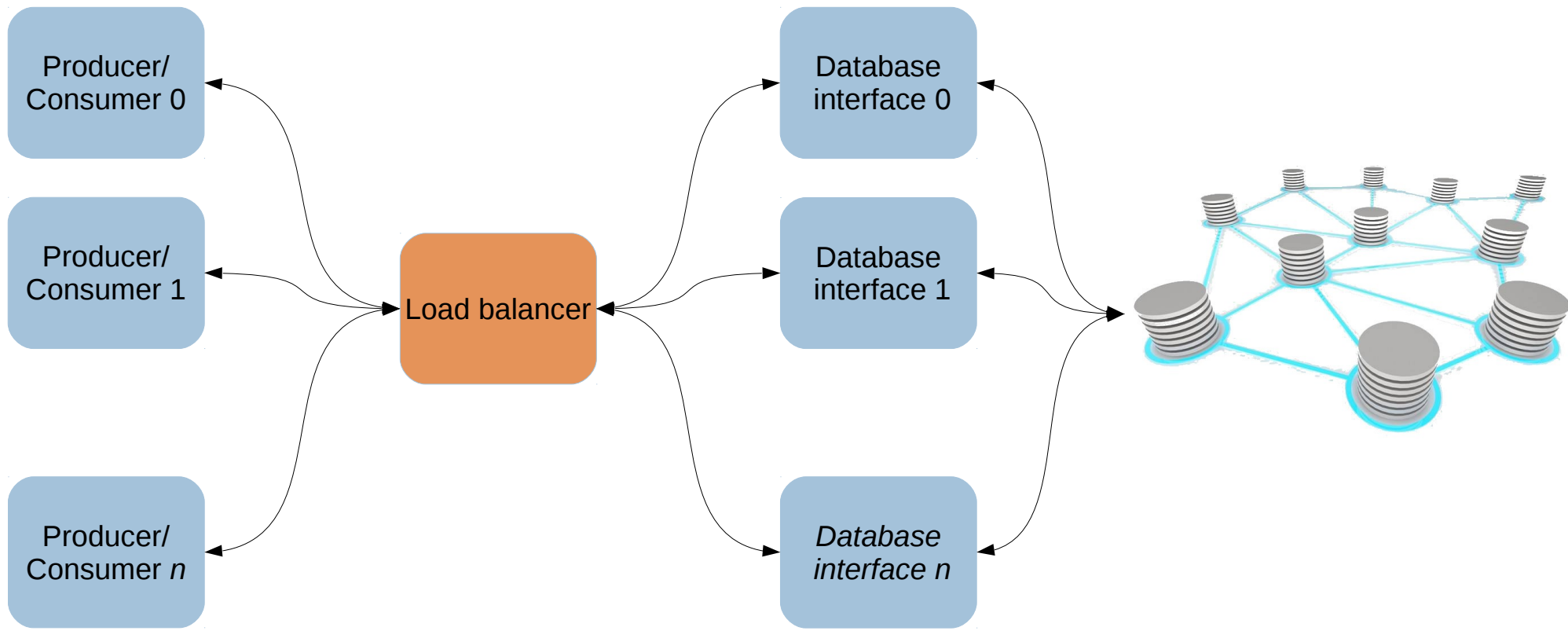
C++ repository: <https://github.com/kouzinopoulos/KeyValueClusterPerf>
Java repository: <https://github.com/kouzinopoulos/RiakJavaC>



System design



Processes of the Parameter Manager:



Data flow



Producer/
Consumer 0

Producer/
Consumer 1

Producer/
Consumer *n*

Loading of ROOT files containing AliCDBEntry objects

Streaming the objects to memory

Storing the object path

/alice/data/2011/OCDB/TPC/Calib/TimeDrift/

Run1_10_v0_s0.root

Run1_10_v0_s1.root

Run1_10_v0_s2.root

Data flow

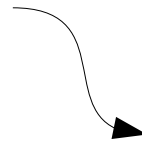


Producer/
Consumer 0

Producer/
Consumer 1

Producer/
Consumer *n*

value



Loading of ROOT files containing AliCDBEntry objects

Streaming the objects to memory

Storing the object path

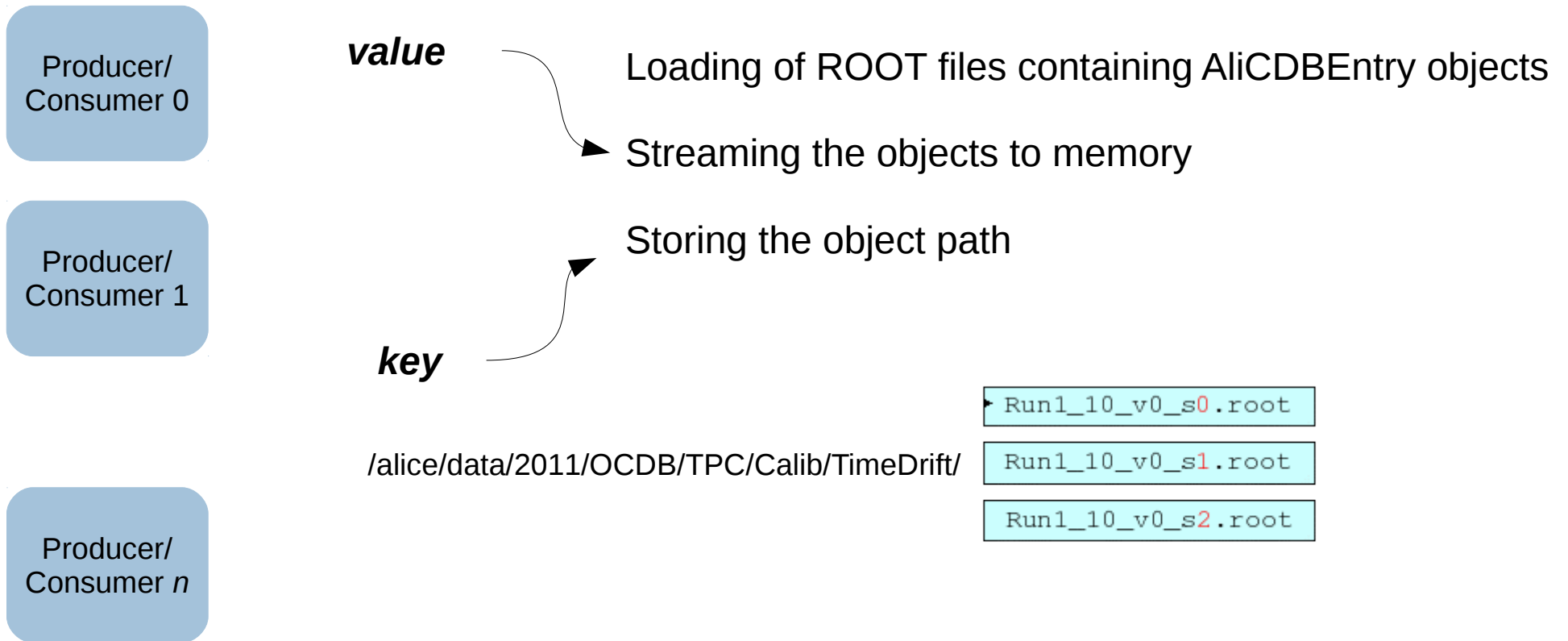
/alice/data/2011/OCDB/TPC/Calib/TimeDrift/

Run1_10_v0_s0.root

Run1_10_v0_s1.root

Run1_10_v0_s2.root

Data flow

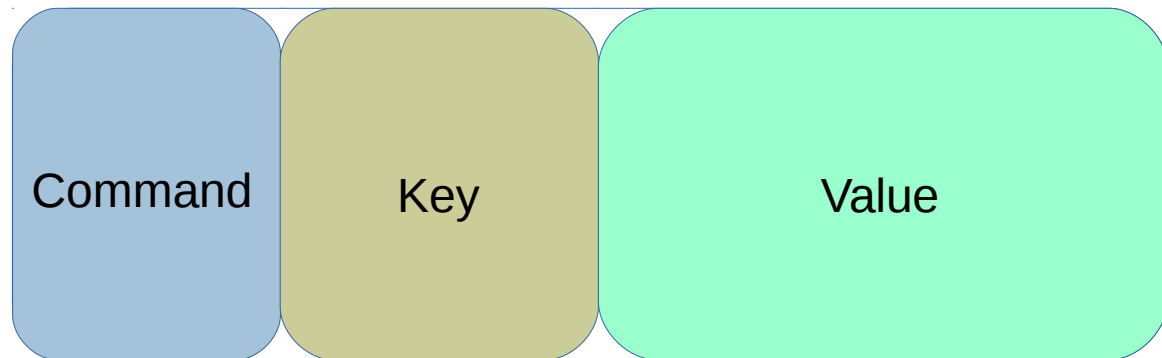


Data flow



Message crafting using the serialization module:

Producer/
Consumer 0



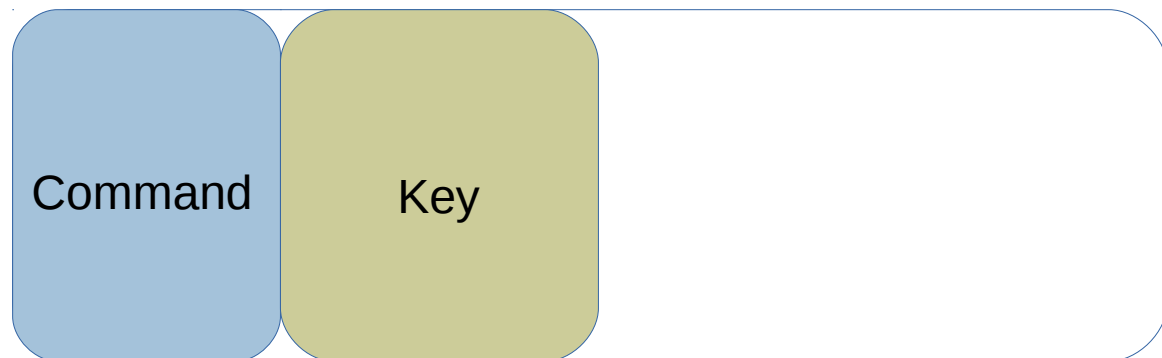
String
"PUT"

String

Binary payload

Producer/
Consumer 1

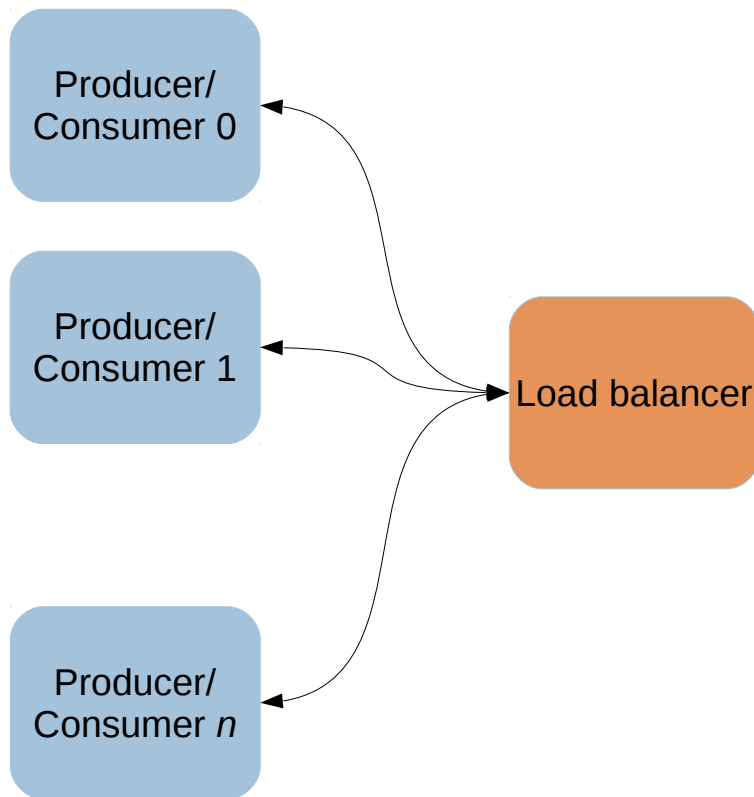
Producer/
Consumer *n*



String
"GET"

String

Data flow



The serialized message is transmitted to the Load balancer node using ZeroMQ (FairMQ)

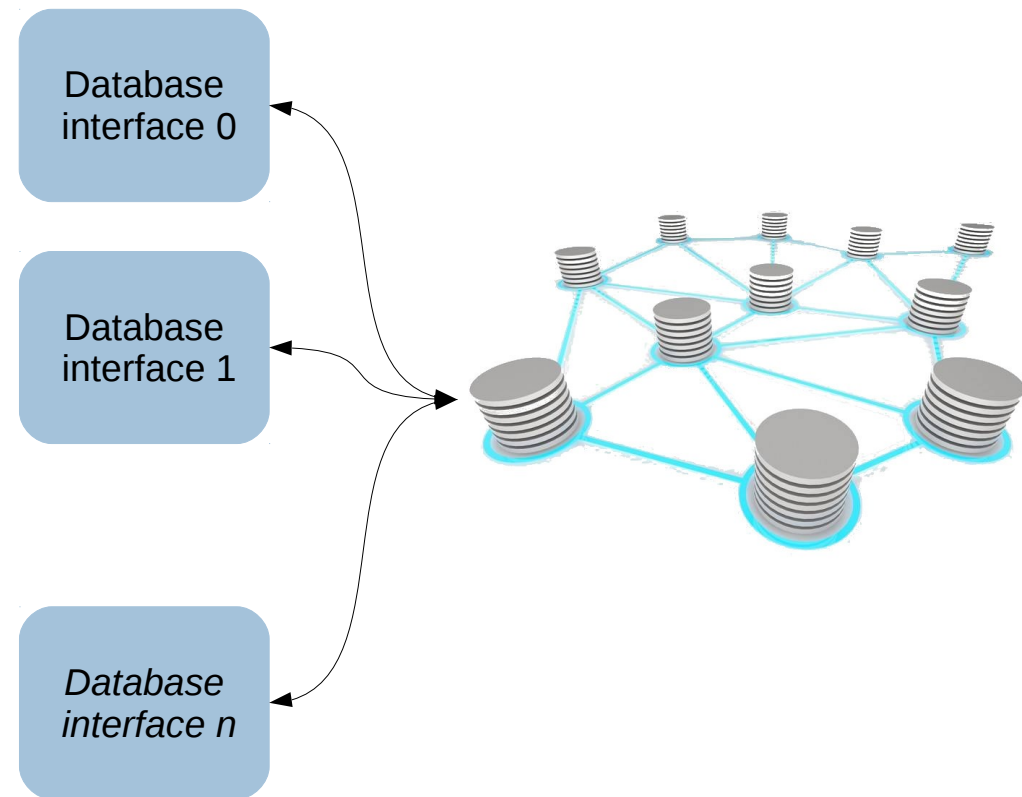
The Load balancer node forwards the message to a Database interface node on a round robin basis

Data flow



The message is de-serialized using Protocol buffers

A "PUT" or "GET" command is executed to the Riak cluster



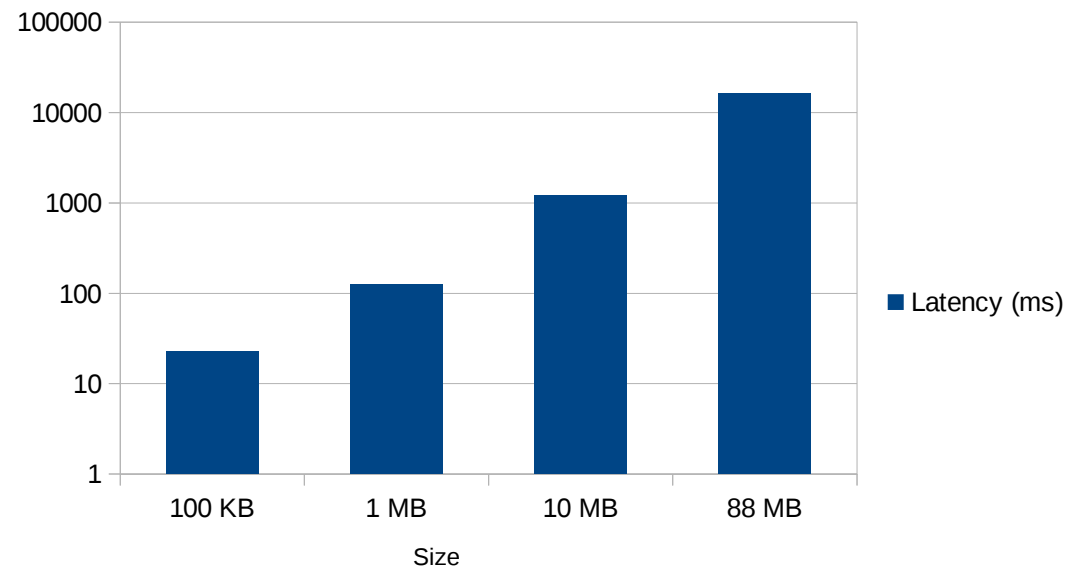
Preliminary results



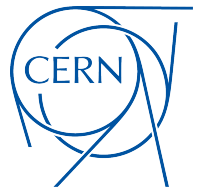
- /alice/data/2011/OCDB/TPC/Calib/TimeDrift/
- One producer/consumer process (cernvm14)
- Load balancer process (cernvm14)
- One DB interface process (cernvm14)
- One Riak node (cernvm14)

PUT/GET pair of commands on each object on a deterministic way

Size (KB)	Number of ROOT objects	Latency (ms)
100 KB	1	23
1 MB	9	125
10 MB	94	1222
88 MB	1648	16226

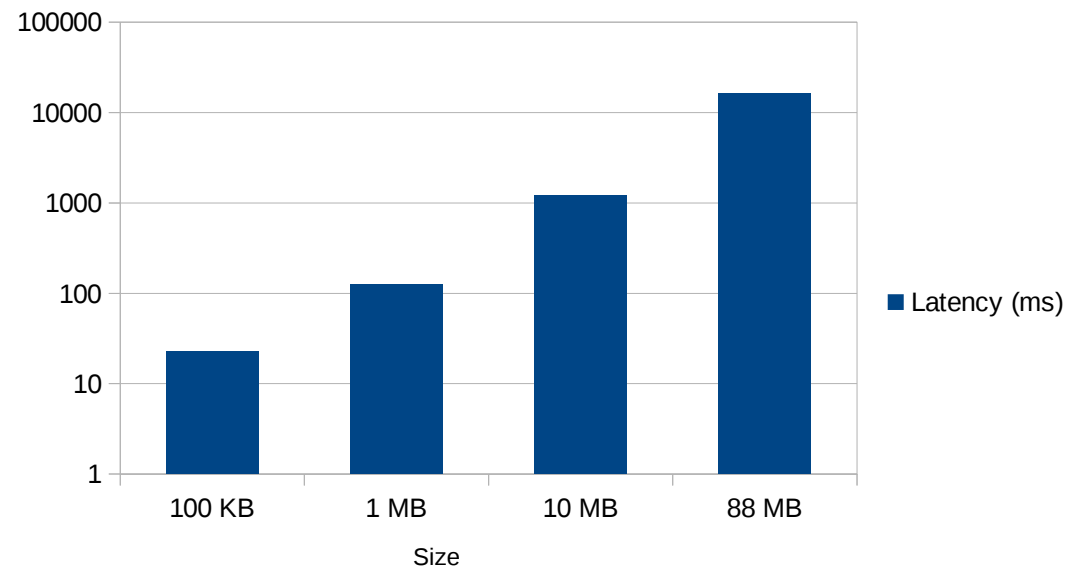


Preliminary results

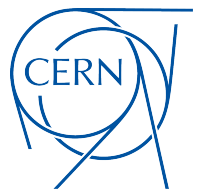


- /alice/data/2011/OCDB/TPC/Calib/TimeDrift/
- One producer/consumer process (cernvm14)
- Load balancer process (cernvm14)
- One DB interface process (cernvm14)
- Two Riak nodes (cernvm14 and cernvm13)

Size (KB)	Number of ROOT objects	Latency (ms)
100 KB	1	24
1 MB	9	151
10 MB	94	1330
88 MB	1648	17204

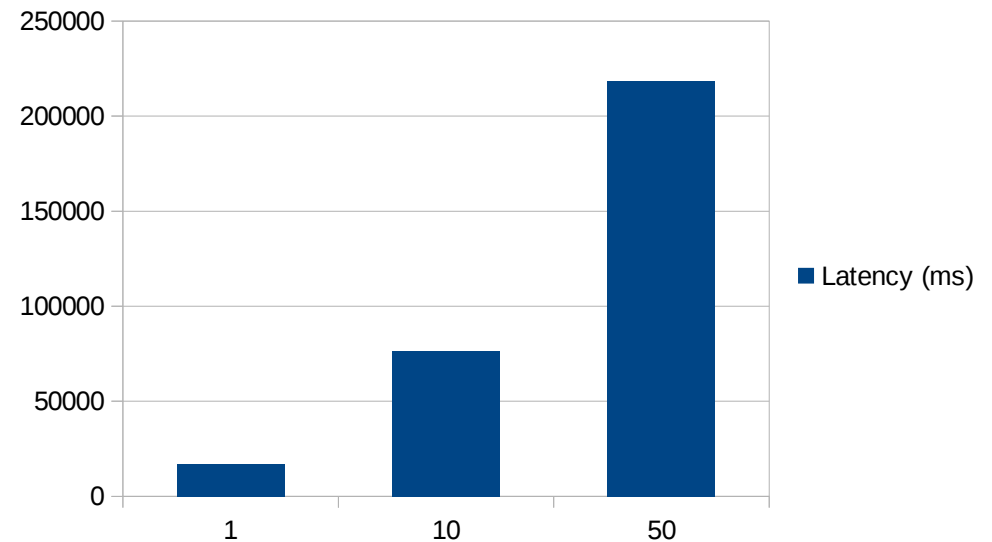


Preliminary results



- /alice/data/2011/OCDB/TPC/Calib/TimeDrift/
- 1, 10 ,50 producer/consumer processes (cernvm14)
- Load balancer process (cernvm14)
- One DB interface process (cernvm14)
- Two Riak nodes (cernvm14 and cernvm13)

Number of producers	Latency (ms)
1	17204
10	76266
50	218543



Questions?

