

Testing Oracle Database In-Memory for CERN applications

Artur Zygałło

CERN Database Services



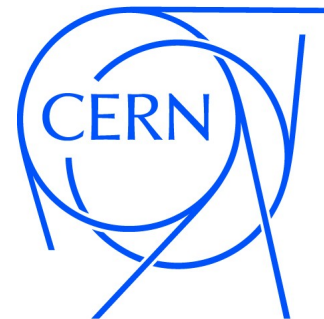
About CERN

CERN - European Organization for Nuclear Research

- Large Hadron Collider, Higgs boson, World Wide Web, ...

over 100 Oracle databases

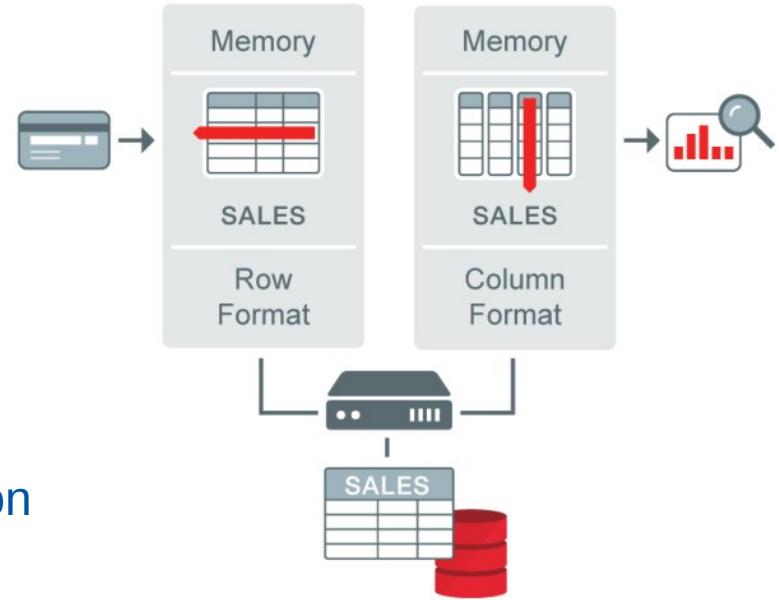
- running Oracle 11.2 and 12.1
- ~1 PB of production data files
- currently testing 12.2 (most recent)



Oracle Database In-Memory

goal: improve performance of analytic queries

- introduced in 12.1
- compressed **columnar format**
 - columns, not rows, stored contiguously
- data stored **in memory** (RAM)
 - no additional disk storage required
- automatic real-time synchronization after data modification



Testing CERN applications

names below: LHCb, CMS, ATLAS – CERN experiments

numbers below: total application data / data sent to In-Memory store / In-Memory compressed data

- **CERN experiments' databases**
 - **LHCb** – files and jobs tracking – 900 GB / 260 GB / 160 GB
 - **CMS** – data transfer between grid nodes – 100 GB / 40 GB / 35 GB
 - **ATLAS** – jobs tracking – testing in progress
- **administrative data warehouse** – 360 GB / 140 GB / 30 GB

Results

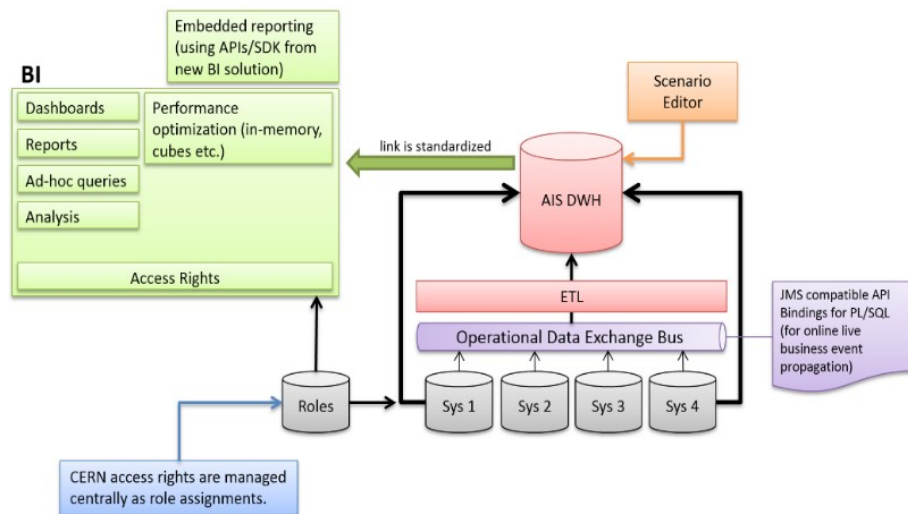
it all depends on your workload

- **LHCb** – files and jobs tracking – no improvement observed
- **CMS** – data transfer between grid nodes – slight improvement
- **administrative data warehouse** – significant improvement

Administrative Data Warehouse

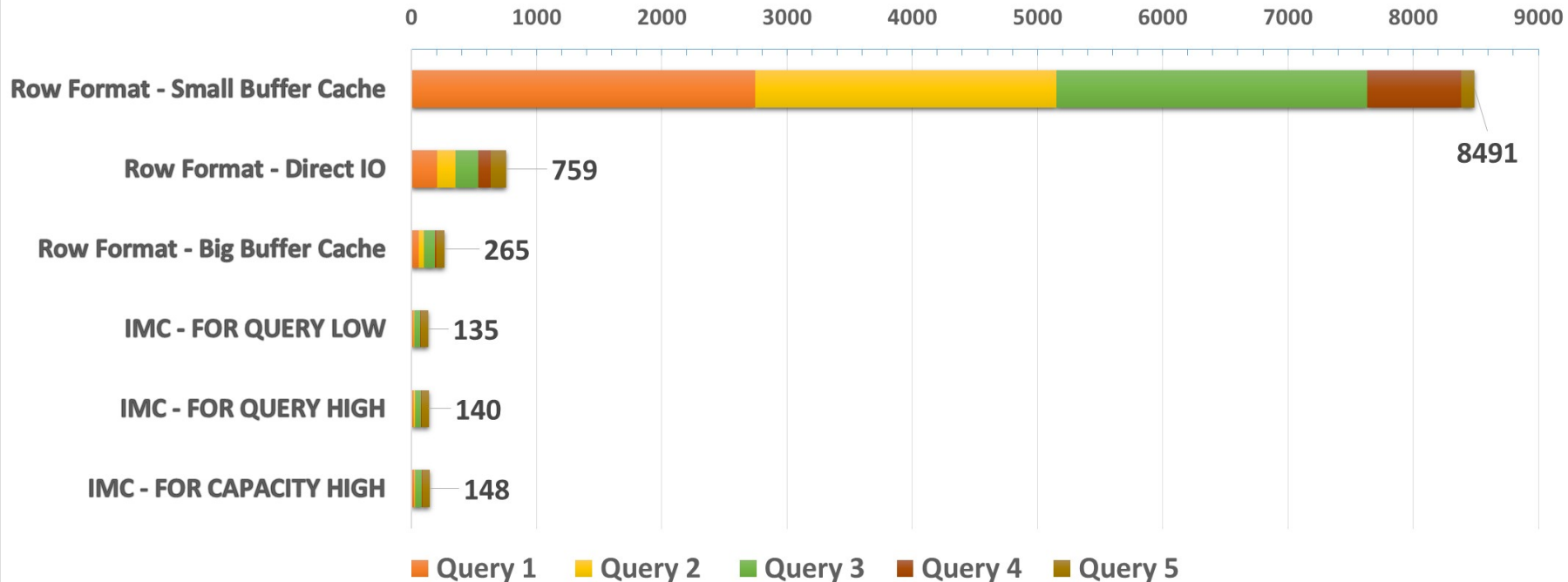
in production, using In-Memory feature since 2015

- supports CERN reports, dashboards and data analytics
- HR data, financial data, orders/purchases, electronic recruitment
- unique data source for all BI applications



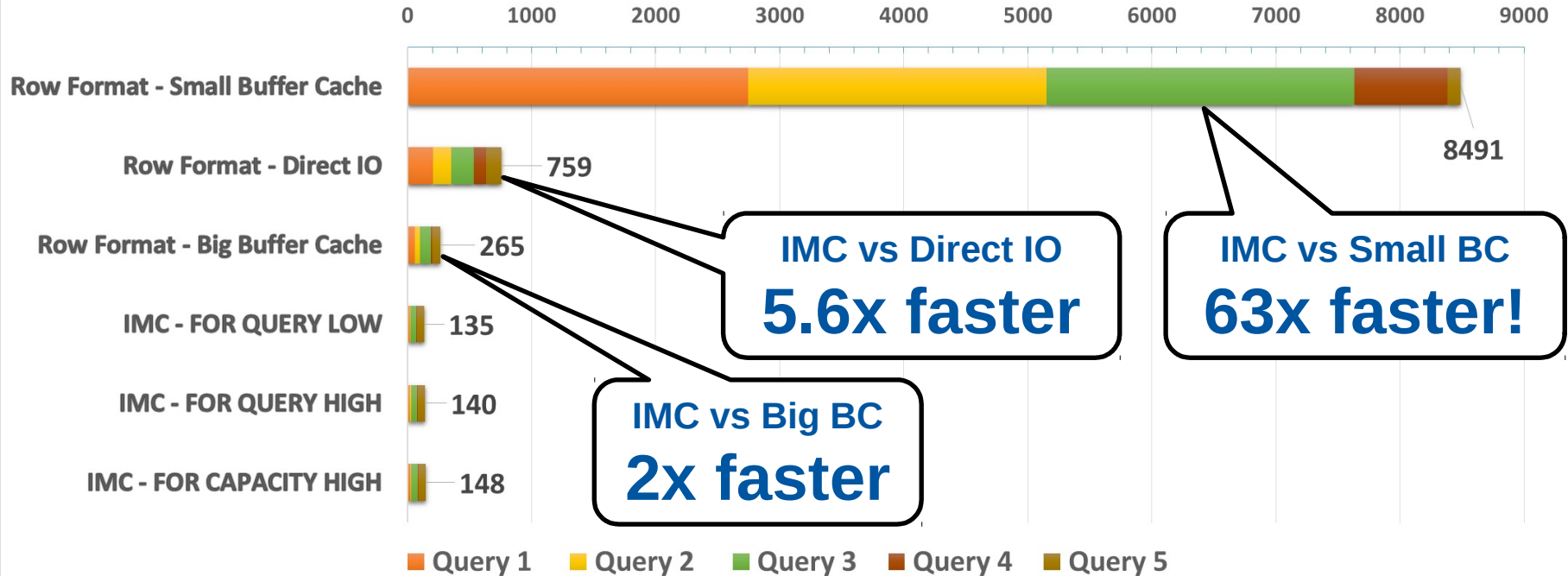
ADW In-memory benefits

Query response time (s)



ADW In-memory benefits

Query response time (s)



ADW In-memory benefits

Query response time (s)

0 1000 2000 3000 4000 5000 6000 7000 8000 9000

Row Format - Small Buffer Cache



8491

759

265

135

140

148

Realistic gain in
PRODUCTION
Queries on average
10x faster!

Query 1 Query 2 Query 3 Query 4 Query 5

Summary

conclusion: not a universal solution

- only if data fits entirely **in memory** (compressed)
- **best use cases:**
 - select a few columns from wide tables (with many columns)
 - full table scans on large tables
 - aggregations (sum, average, count, ...)
- **business intelligence / data warehousing / data analytics / reporting**

Thank you for your attention!

Artur Zygadło
artur.zygadlo@cern.ch





www.cern.ch

