



NATIONAL RESEARCH CENTRE
"KURCHATOV INSTITUTE"

Development of Hybrid SQL/NoSQL PanDA Metadata Storage

Laboratory of BigData Technologies
for mega-science projects

Marina Golosova, Maria Grigorieva
Head of Laboratory: Alexei Klimentov

CERN, 2014



Laboratory of BigData Technologies for mega-science projects

- Laboratory was formed in "Kurchatov Institute" for research and development in the area of processing and analysis of extremely large data volumes.

<http://bigdatalab.nrcki.ru/>

- **Current project:** "Big Data Technologies for megaScience projects" supported by the Russian Federation Government grant.
- **Scientific program** is tightly coupled with LHC experiments priorities and address challenges we will meet in 2-3 years.



Project objectives:

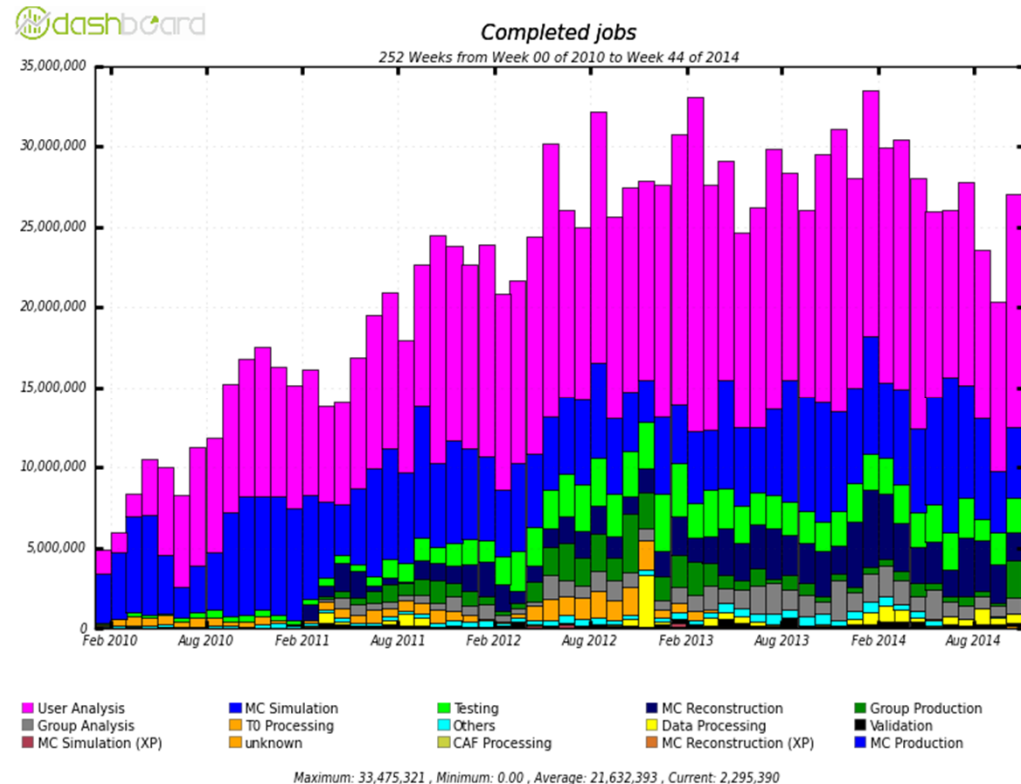
Development of the novel *Workload and Data Management System for Big Data*, based on PanDA (**MegaPanDA**):

- support for large-scale data handling
- HPC support
- Cloud and web-based computing services support

PanDA: metadata storage issue

- Archive: 900 M jobs (since 2006)
- Current rate: ~2M jobs per day
- RDBMS: Response time increases as the volume of stored metadata grows up
- Dividing metadata:
 - actual (read-write part): for the most recent and changing records (**ATLAS_PANDA**)
 - archive (read-only part): for all records since 2006 (**ATLAS_PANDAARCH**)
- Oracle

Completed Jobs 2009 – 2014 years



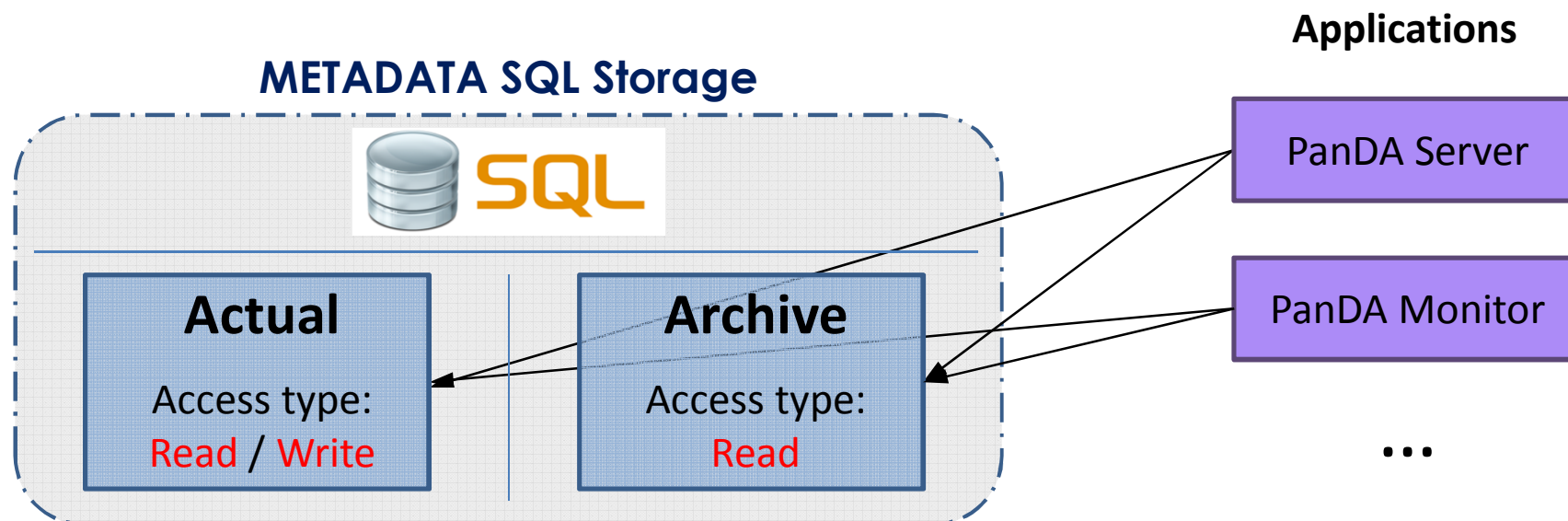
- 2015 (Run-2): current rate x5
- 2020 (Run-3): current rate x10
- ...?



RDBMS (SQL) Storage

SQL standard : ACID

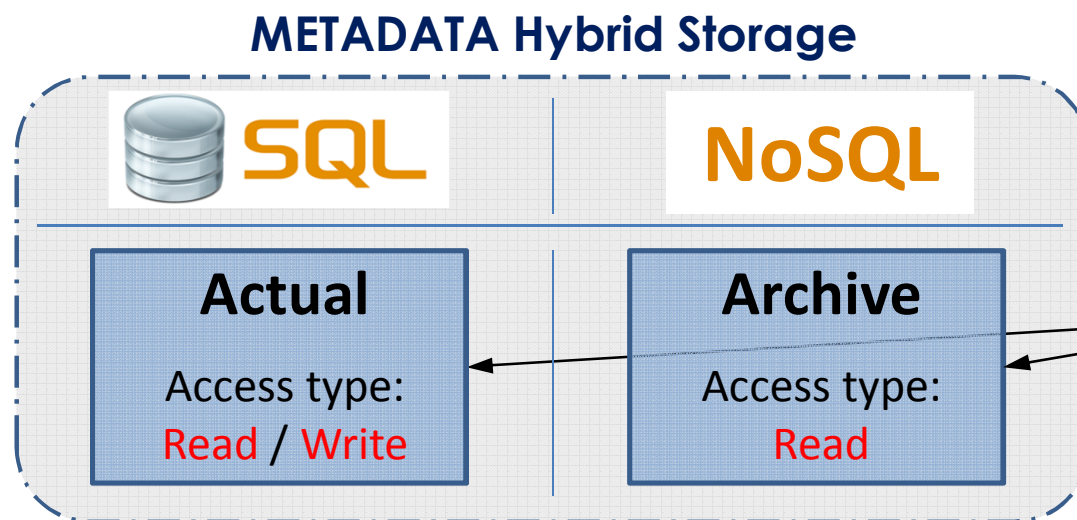
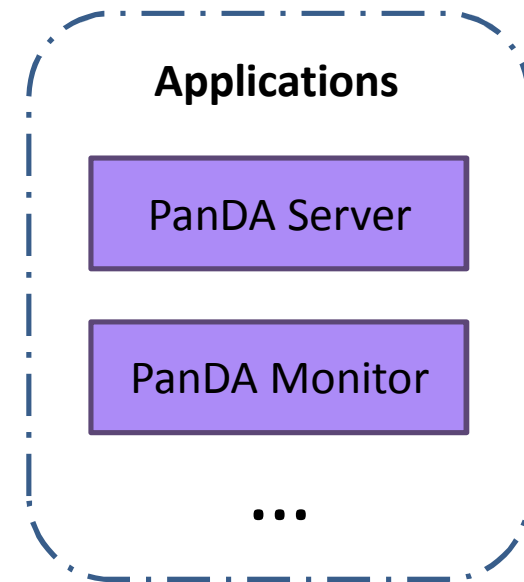
Atomicity
Consistency
Isolation
Durability





NoSQL: not only SQL storage

SQL standard : ACID	NoSQL standard : BASE
Atomicity Consistency Isolation Durability	Basic Availability Soft-state Eventual consistency



Hybrid Storage project



NATIONAL RESEARCH CENTRE
"KURCHATOV INSTITUTE"

Objective:

Architecture and implementation of storage and access to PanDA metadata.

- ✓ **Stage 1:** Subject area research.
 - ✓ PanDA metadata
 - ✓ PanDA DB architecture structure
- ✓ **Stage 2:** Technology research.
 - ✓ NoSQL
- ✓ **Stage 3:** Storage schema.
 - ✓ Design
 - ✓ Implementation
 - ✓ Testing
- **Stage 4:** Storage software.
 - ✓ Design
 - Implementation
 - Testing
- **Stage 5:** Panda adaptation.

Stage 3: Data model for



1) Main table – JOBS

2) Helper tables for most popular queries

Jobs

PandaID	assignedPriority	atlasRelease	...
2038679208	1000	Atlas-17.2.7	...
2033030636

Primary key

- Partition key: *PandaID*
- Clustering keys: ---

pandaid	assig...	atlasrelease	atte...	batchid	bro...	brokera...	cloud	cmtconfig	com...	computingele...
2038679208	1000	Atlas-17.2.7	0	1337819.t2...	0	null	UK	x86_64-slc5-gc...	null	ANALY_OX_TE...
2216674414	830	Atlas-17.3.4	1	gridgk04.r...	0	<<null...	US	i686-slc5-gcc4...	<<nu...	BNL_PROD-co...
2033030636	140	Atlas-17.2.11	1	1779002.n...	0	null	FR	x86_64-slc5-gc...	null	GRIF-IRFU-no...
2036590233	800	Atlas-17.2.11	2	2465274.t1...	0	null	TW	i686-slc5-gcc4...	null	Taiwan-LCG2-...
2034895303	0	Atlas-17.2.6	0	condor.he...	0	null	CA	x86_64-slc5-gc...	null	NECTAR-clou...
2037942822	1000	Atlas-16.6.4	0	21437605.g...	0	null	CA	i686-slc5-gcc4...	null	ANALY_SCINET
2035006694	800	Atlas-17.2.11	2	8432280.ce...	0	null	CA	i686-slc5-gcc4...	null	TRIUMF
2239267905	1000	null	0	uct2-gk.m...	0	null	US	null	null	ANALY_MWT...
2032734374	950	Atlas-17.3.10	1	8396171.ce...	0	null	CA	x86_64-slc5-gc...	null	TRIUMF
2036709725	540	Atlas-17.2.1...	1	5661071.gr...	0	null	FR	i686-slc5-gcc4...	null	DESY-HH-all...
2237746159	120	Atlas-17.7.3	1	47606953.s...	0	null	UK	x86_64-slc6-gc...	null	UKI-SCOTGRI...
2032497040	950	Atlas-17.3.10	1	12563016.c...	0	null	CA	x86_64-slc5-gc...	null	TRIUMF
2237313667	120	Atlas-17.7.3	1	47591372.s...	0	null	ES	x86_64-slc6-gc...	null	UKI-SCOTGRI...
2237944282	140	Atlas-17.7.3	1	919475	0	null	CA	x86_64-slc5-gc...	null	SLACXRD-lsf
2035195946	150	Atlas-17.2.6	1	247525	0	null	US	x86_64-slc5-gc...	null	LRZ-LMU-all...
2035977211	140	Atlas-17.2.1...	1	834955.i2t...	0	null	UK	x86_64-slc5-gc...	null	UKI-SOUTHGR...
2031291870	540	Atlas-17.2.1...	1	486256026	0	null	FR	i686-slc5-gcc4...	null	CERN-PROD_S...
2036250012	140	Atlas-17.2.1...	1	7514700	0	null	US	x86_64-slc5-gc...	null	BU_ATLAS_Tie...
2035813604	800	Atlas-17.2.11	1	13289312	0	null	FR	i686-slc5-gcc4...	null	pikolit.ijs.si
2034160551	800	Atlas-17.2.11	1	487016820	0	null	IT	i686-slc5-gcc4...	null	CERN-PROD-a...
2031767879	140	Atlas-17.2.11	1	592187	0	null	US	x86_64-slc5-gc...	null	SLACXRD-lsf
2037871996	1000	Atlas-17.2.7	0	16328718	0	null	IT	i686-slc5-gcc4...	null	ANALY_INFN-...
2035753520	550	Atlas-17.2.11	1	2028015.gk...	0	null	US	i686-slc5-gcc4...	null	SWT2_CPB-pbs
2038438453	200540	null	1	gridgk06.r...	0	null	US	null	null	BNL_DDM-co...
2238630303	1000	Atlas-17.7.3	0	47619103.s...	0	null	UK	x86_64-slc5-gc...	null	ANALY_GLAS...
2033003418	550	Atlas-17.2.11	1	3652396.m...	0	null	UK	i686-slc5-gcc4...	null	UKI-LT2-RHUL...
2035685457	550	Atlas-17.2.11	1	tier2-01.oc...	0	null	US	i686-slc5-gcc4...	null	OU_OCHEP_S...
2035064396	1000	Atlas-17.2.7	0	null	0	null	FR	i686-slc5-gcc4...	null	tbit07.nipne.ro

Task

TaskID	JobStatus	ModificationTime	PandaID	...
769	failed	2014-01-06	2037208385	...
			2037208386	...
		
[769]		finished	2014-01-01	2032493594
		2014-01-06	2037208384	...
...

Primary key

- Partition key: *(TaskID [, JobStatus])*
- Clustering keys: *[JobStatus,] ModificationTime, PandaID*

taskID: jobStatus	modificationTime: pandaID	modificationTime: pandaID	modificationTime: pandaID
15:cancelled	2014-01-01:2032221606	2014-01-01:2032221616	2014-01-02:2031265463
15:finished	2014-01-01:2031265476	2014-06-01:2031265519	
175:finished	2014-01-01:2031265476	2014-06-01:2031265519	2014-06-01:2031265519



Stage 3: Testing Results

Test №1

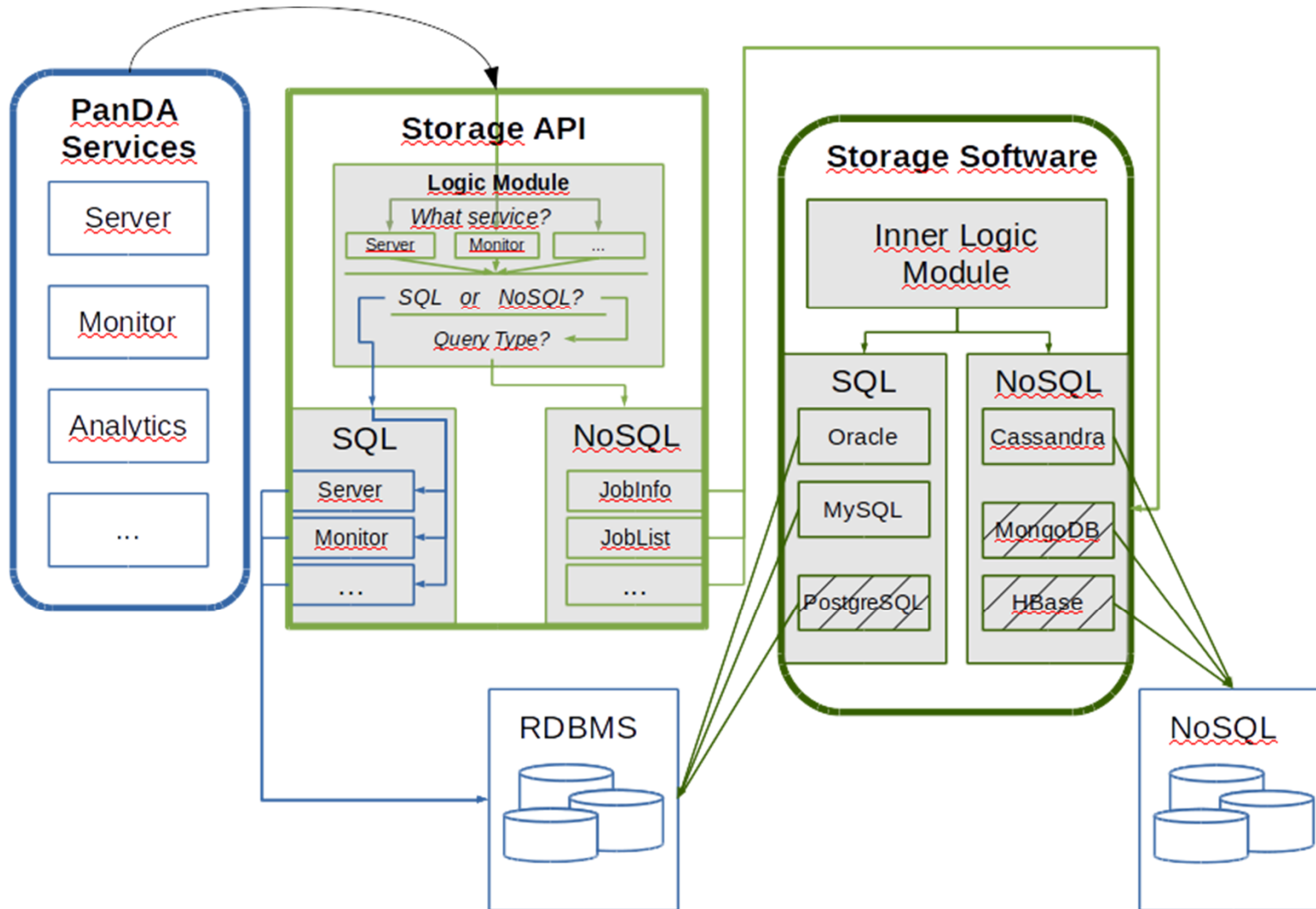
Single Query average response time (ms)



QUERY conditions	ORACLE®	Data Model #1	Data Model #2
pandaID = ?	45	19	4 405
taskID = ?	40	13	6 478
JEDITaskID = ?	26	5	11 257
taskID = ? jobStatus = ? ?<= modificationTime < ?	19	12	---
taskID = ? ?<= modificationTime < ?	11	24	---



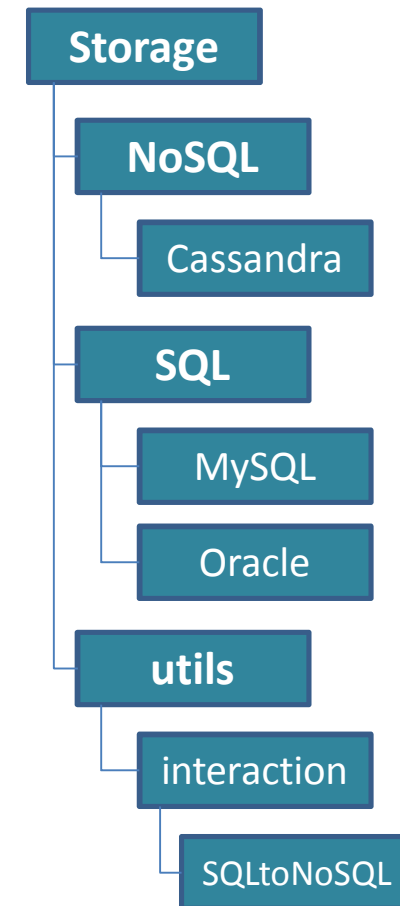
Stage 4: Storage architecture





Hybrid Storage: current state

- ✓ Development of NoSQL schema
- ✓ Creating test bed for schema testing
- ✓ Loading a two weeks slice of ATLAS archive data into both Cassandra cluster and Oracle DB
- ✓ NoSQL schema testing
-
- ✓ Storage software design
- ✓ Basic functionality implementation:
 - unified wrappers: Cassandra, Oracle, MySQL
 - data export (Oracle)
 - data import (Cassandra)
 - full copy (export-import) from SQL to NoSQL





NATIONAL RESEARCH CENTRE
"KURCHATOV INSTITUTE"



Thank you!