Toward fast search and real-time inputs of big astronomical catalogs by the new generation relational database

Tadafumi Takata#, Hisanori Furusawa, Yoshihiko Yamada, Yuki Okura (National Astronomical Observatory of Japan),
Makoto Onizuka(Osaka University), Hidekazu Suga, Ryoji Kurosawa, Takashi Kambayashi (NAUTILUS Technologies, Inc.)
#E-mail: tadafumi.takata@nao.ac.jp

Background

- Until recently the application of database management system to astronomical data are limited to small to medium size data managements, at most several hundreds of millions rows with dozens columns even for large astronomical catalogs. The ongoing deep survey Subaru Strategic Program (SSP) using Hyper Suprime-Cam(HSC) (HSC-SSP) is aiming to promote scientific activities based on the archive of the data, which are produced through data processing pipeline. We are planning to provide the catalog database tables with more than 3,000 columns and ~30 billions rows as one of the final products of HSC-SSP. We are trying performance improvement using distribution of DB (e.g. w/Citus for PostgreSQL), however it is not sufficient.
- It is now the phase of revolution in database software due to growing demand of fast analysis of very large data stored in database. In Japan, there is an activity to develop a new generation RDBMS based on Open Source Software(OSS) by implementing the hybrid of OLTP (Online Transaction Processing) and OLAP(Online Analytical Processing) and realizing the fast data processing under the framework of industry-academia collaboration.
- We participate the project of developing the new generation RDBMS and try to implement the functions for fast astronomical queries on the final products of HSC-SSP and produce many scientific results.

Development Project of New RDB w/ OSS

劔

- **Project Tsurugi** : 5 years project funded by NEDO: start @ 2018)
- Joint Development under Industry-academia collaboration (NEC, NAUTILUS Technologies Inc., TITech, Osaka Univ., Nagoya Univ., Keio Univ., Tsukuba Univ. etc.)
- NAOJ will take a role in performance test of DB using big astronomical catalogs and demonstrating the usefulness of the DB in promoting *e-Science.*

Existing DB's bottleneck & new technologies in Big Data Era

- Limit on scalability of RDB
- HDD I/O speed problem
- Limited use-case of NoSQL
- Limits on OSS DB for Big Data
- Delay for shift to new technologies

- HTAP (Hybrid-Transaction/Analytics Processing)＝Fast RDB by Hybrid of OLTP and OLAP
- Streaming(Real-Time) Processing
- Many-cores/In-Memory Processing（Large capacity & Non-volatile)
- Extensibility of modules by using OSS

Astronomical Subjects in the Project Tsurugi

1. Fast ad-hoc astronomical catalog search

Enable very fast search of coadd-based and each exposure-based catalogs of HSC-SSP Public Data Release by injecting them into DBMS. (Figure 1)

2. Detection/Analysis of transient objects

Enable semi-realtime detection and measurement the objects in HSC image data, registering them into DBMS, then identifying variables/transients. (Figure 2)

Currently we are concentrating on the development of fast ad-hoc search. We are doing the following development using HSC-PDR1 and PDR2※.

※PDR: Public Data Release

- Pilot study of fast query by means of distributed DB technology.
- Recommendation of optimized schema by query log analysis.
  → Effectve use of materialized views, and search of optimal solution by integer planning method

Outlier detections using DBMS based on data distribution(w/ LOF etc.)
→ Basic proto-type development for detecting variables/transients etc.

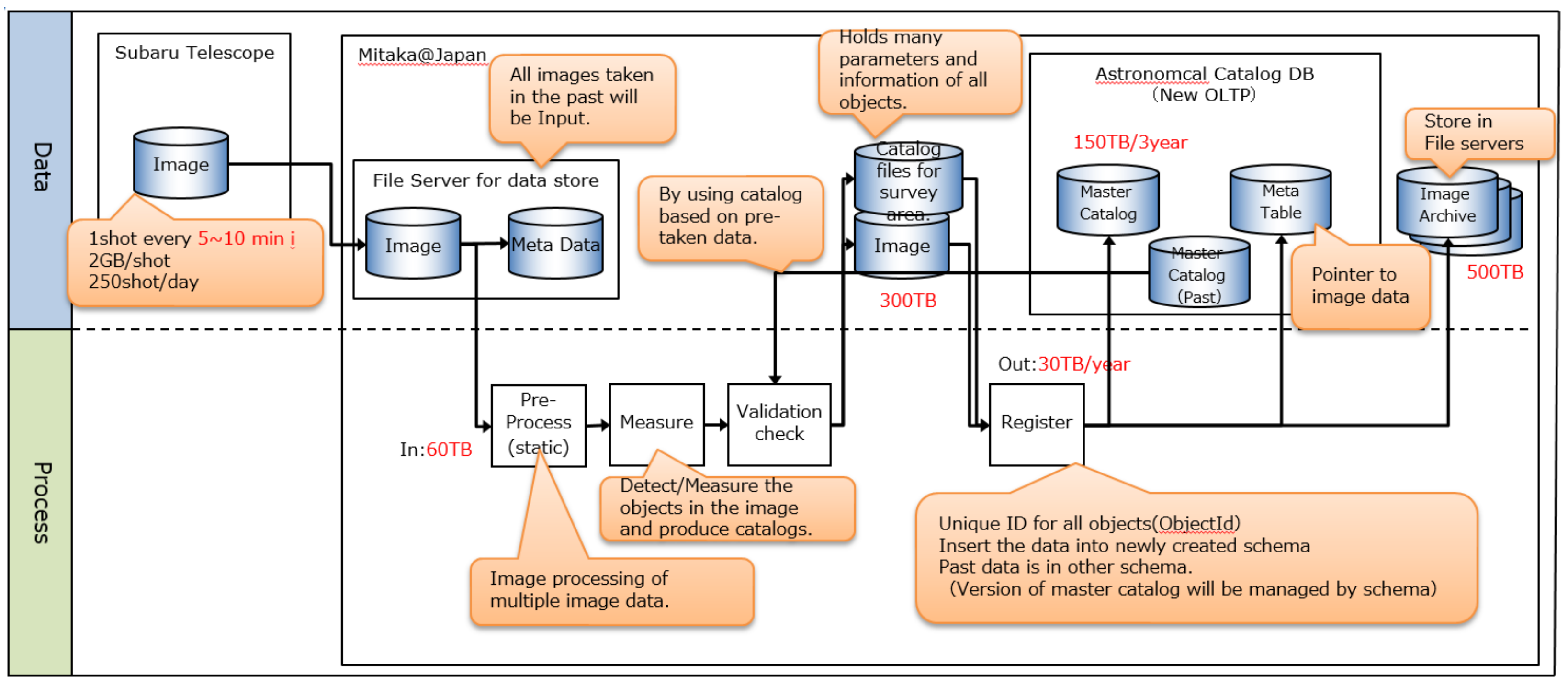
Data
Process
Subaru Telescope
Image
1shot every 5~10 min
2GB/shot
250shot/day
Mitaka@Japan
All images taken in the past will be Input.
File Server for data store
Image
Meta Data
By using catalog based on pre-taken data.
Holds many parameters and information of all objects.
Catalog files for survey area.
Image
300TB
Astronomcal Catalog DB (New OLTP)
150TB/3year
Master Catalog
Meta Table
Master Catalog (Past)
Store in File servers
Image Archive
500TB
Pointer to image data
In:60TB
Pre-Process (static)
Measure
Validation check
Out:30TB/year
Register
Detect/Measure the objects in the image and produce catalogs.
Image processing of multiple image data.
Unique ID for all objects(ObjectId)
Insert the data into newly created schema
Past data is in other schema.
(Version of master catalog will be managed by schema)

Figure 1. Process/Data flow of master catalog registration

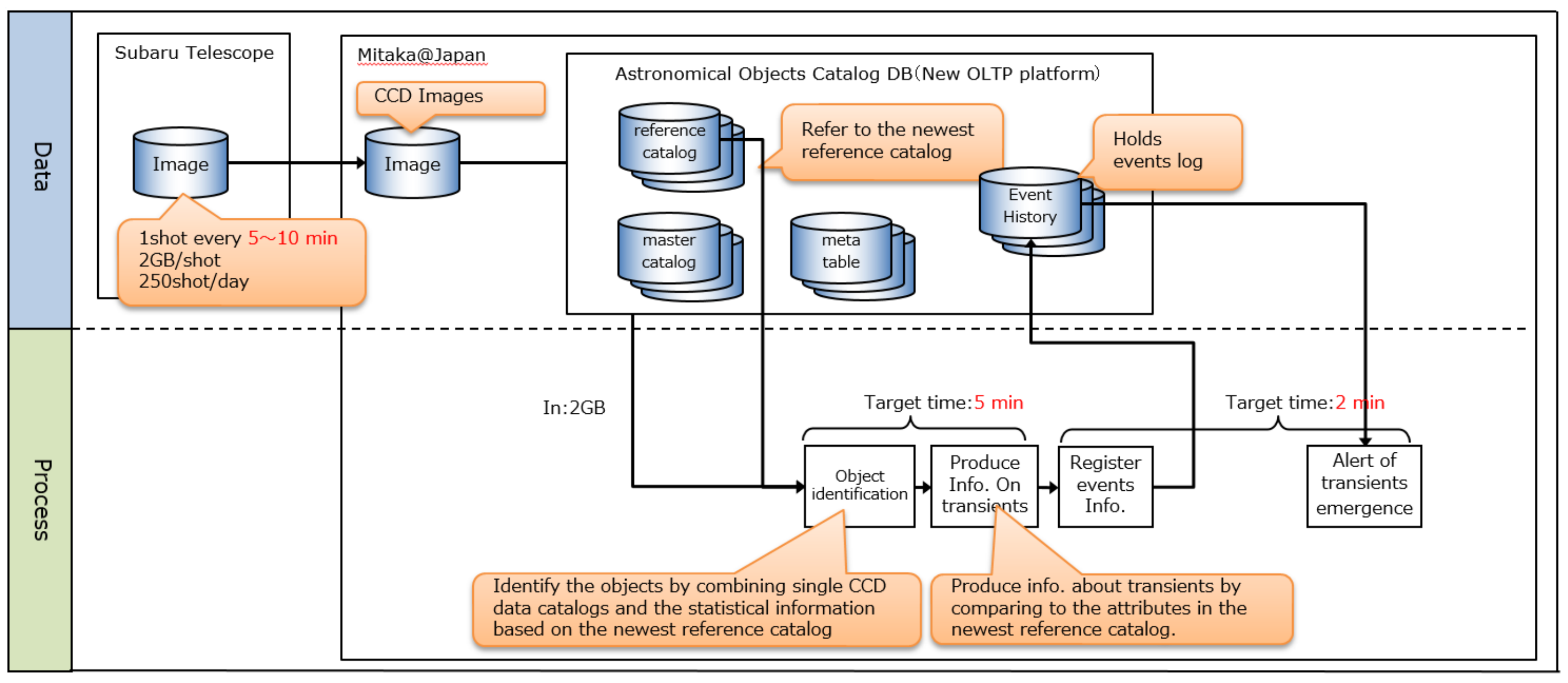
Data
Process
Subaru Telescope
Image
1shot every 5~10 min
2GB/shot
250shot/day
Mitaka@Japan
CCD Images
Image
Astronomical Objects Catalog DB(New OLTP platform)
reference catalog
Refer to the newest reference catalog
Holds events log
Event History
master catalog
meta table
In:2GB
Target time:5 min
Target time:2 min
Object identification
Produce Info. On transients
Register events Info.
Alert of transients emergence
Identify the objects by combining single CCD data catalogs and the statistical information based on the newest reference catalog
Produce info. about transients by comparing to the attributes in the newest reference catalog.

Figure 2. Process/Data flow of detecting transients/variables.

Speeding-up test of catalog search

We tested and confirmed the improvement of typical queries' speed by distributing the queries for HSC-PDR1 catalogs on the constructed system in Sakura Internet data center @ Ishikari(Hokkaido).

- **Environment**：65 servers（CPU: 64core/x86_64, RAM:1.5TB, Disk:10TBx7)
- **Setting**：Hadoop Cluster, No partitioning(No optimization)
- **3 types of slaves** w/ 10, 20, and 60 servers
- **Used query constraints**：

Magnitudes, Mag Errors, Colors, Sky area(Tract), flags w/ table 'JOIN'

"**select** object_id··· **from** pdr_wide.forced as main **LEFT join** pdr_wide.forced as meas using object_id **where** main.tract in (10054······) and main.gcmodel > 0.0 and main.gcmodel < 25.0 ··· and main.flags_pixel_edge = 'f' ·········" (~100 output columns and constraints w/ ~100 parameters (~50 flags))

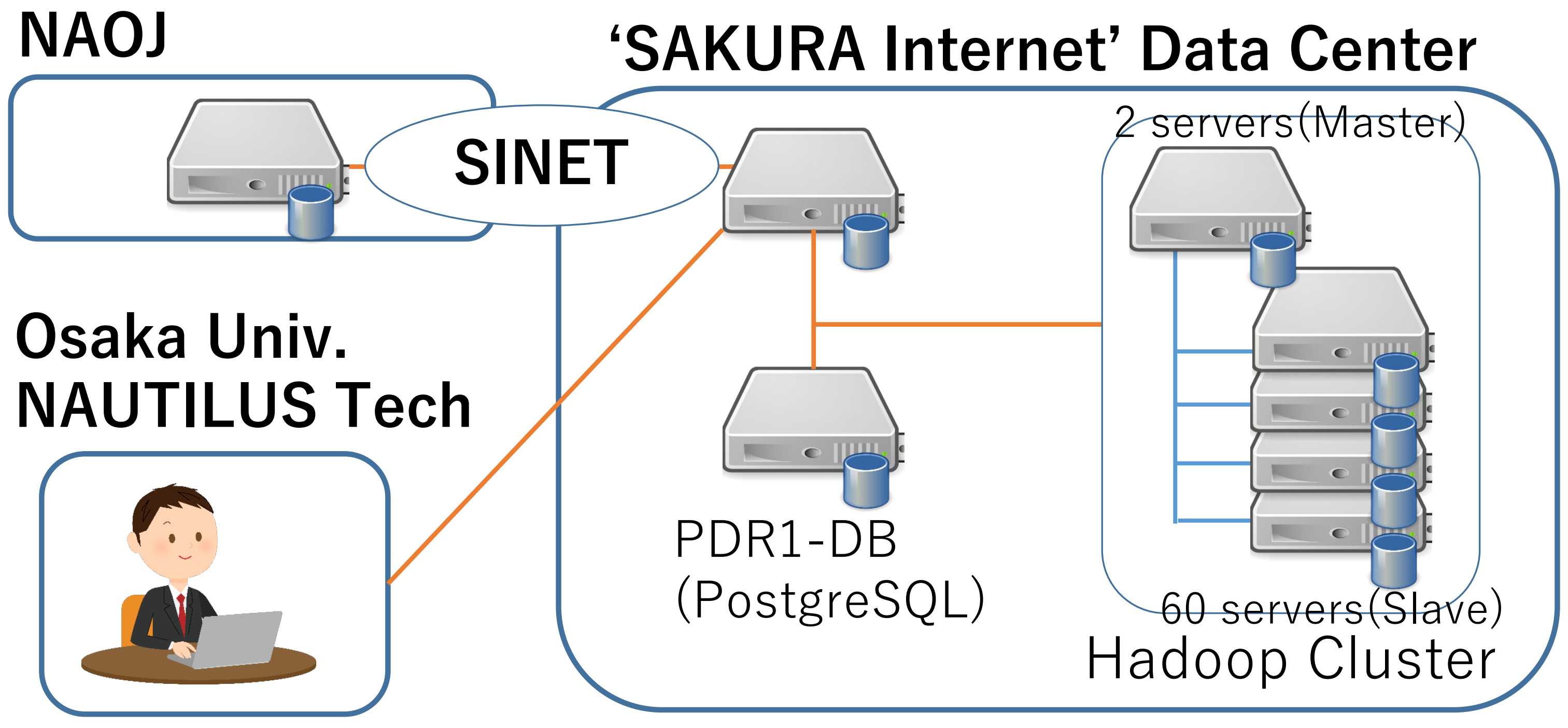

Figure 3: Environment for query performance test

※SINET: Science Information NETwork (https://www.sinet.ad.jp/en/top-en)

●Information of used catalogs for the query test (1,455 records output)

Input	Data Amount	# of records	# of columns
pdr1_wide.forced	721GB	84,017,247	1,068
pdr1_wide.meas	659GB	84,042,022	1,086

●Elapsed times for completing the test query

Execute Engine	Data Format	Elapsed Time w/ 10 servers(sec)	Elapsed Time w/ 60 servers(sec)
Hive on MR	TEXT (CSV)	342	112
	Parquet	175	72
SparkSQL	TEXT (CSV)	138	52
	Parquet	40	37
Impala	TEXT (CSV)	19	13
	Parquet	12	10

※We used 3 engines for the query test. Parquet is one of the columnar formats.

The elapsed time for the query we tested was more than 12 hours (it could not be completed by time-out) for HSC-SSP PDR1 data using single PostgreSQL server. On the other hand we've got the good response time in our tests, by the advantage thanks to columnar format and distributed SQL engine. In the case we did not use JOIN, combining 2 tables into 1 table, response time was shorten to 8 seconds. Response times were not scalable to number of servers, as there were some reasons, like overhead for information transfer and so on. We have a plan to realize the fast queries of HSC-SSP catalogs and promote researches by finding rare objects and/or variable objects, and also improve the calibrations enough for performing 'precise' astronomy and astroinformatics by investigating carefully the information of multiple measurements.