The status report of R&D to improve retrieval performance of HSC-SSP Database system

Takahiro Morishima*[1], Hisanori Furusawa[1], Hiroyuki Ikeda[1,2], Michitaro Koike[1], Sogo Mineo[1], Yuki Okura[1], Tadafumi Takata[1], Satoshi Takita[1,3], Masayuki Tanaka[1], Yoshihiko Yamada[1,4] and The HSC Software Team

* author (speaker)
[1] NAOJ
[2] National Institute of Technology, Wakayama College
[3] Department of Astronomy, The University of Tokyo
[4] Japan Space Forum

Abstract

The Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) is a large imaging survey of the sky aimed to address a wide array of astrophysical questions. A massive data set from the survey is served to the user through a dedicated catalog database (DB), which contains various metadata such as properties of the sources. The catalog DB is implemented by means of an open-source relational database management system (PostgreSQL). With typical/standard SQL commands, the user can flexibly query the catalog DB to retrieve a subset of the data for detailed analyses. The catalog DB system only has catalogs (i.e., no images), but the data volume exceeds 19 TB in latest S21A release. Due to this large volume, some queries take long time to execute. In order to resolve this issue, we, the HSC Software Team, are improving the catalog DB in two ways. One is to make a distributed DB system (Citus), and the second is to allow parallel execution using GPU (PG-Strom). In both cases, we succeed in improvement of DB retrieval performance. It is 10～300 times faster than original (normal) DB system ! Those results and technical knowledge are fed back to current DB system to be served to the users. pdr2-citus, pdr3-citus, dr3-citus, dr4-citus correspond to them. In order for further improvement, we are tuning and optimizing those parameters of the distributed DB system (citus). In addition, we are preparing to serve a next-generation type citus DB system (SATA SSD → NVMe SSD). And also we are preparing to setup and to serve the GPU based parallel executing DB system (PG-Strom). We already verified that GPU based DB is effective for improvement of retrieval performance by using a prototype machine. We are planning to serve an online Jupyter notebook environment for users. In such an interactive use case, our current achievement in the DB performance will no longer be sufficient because even a slight delay of a response from the DB will interrupt your thinking and disturb your good idea. The DB performance will be a primary factor that determines your total working performance. An even faster DB system will be required. The HSC Software team proceeds with further R&D of the catalog DB in order to provide useful and comfortable user services.

HSC-SSP DATA

It is comparatively small.
But the data volume size is absolutely large enough !

HSC-SSP DATA (main)

FITS products / Release NAME	DR3 S20A	DR4 S21A
Total	574 TB	650 TB

too massive !

HSC-SSP catalog DB (only metadata, no image)

Catalog products / Release NAME (DATA BASE)	DR3 S20A	DR4 S21A
Total	14 TB	19 TB

The catalog DB system is a dedicated metadata set, which contains various metadata as properties of the sources.

Improvement history of DB performance (SQL benchmark)

Execute Time : 12735 [s]
(03 h 32 m 25 s)

Machine (DB) \ SQL test	example-01	example-01 (non-boxSearch)	example-02	example-03	example-04	example-06	example-07	example-08	option (block_size)
dr3_s20a (original)	972	1193	4480	11160	1301	60	12735	66	PostgreSQL-11.8 (heap)
dr4_s21a (original)	973	687	1362	6694	730	13	10100	50	PostgreSQL-13.2 (heap)
dr3_s20a (citus / heap)	5	90	116	588	90	1	1316	1	PostgreSQL-13.3 + Citus-10.0.3 (heap)
dr4_s21a (citus / heap)	5	67	128	650	67	1	3687	4	PostgreSQL-13.3 + Citus-10.0.3 (heap)
dr3_s20a (citus / columnar)	4	2	33	59	13	1	1626	4	PostgreSQL-14.1 + Citus-10.2.3 (columnar)

fast

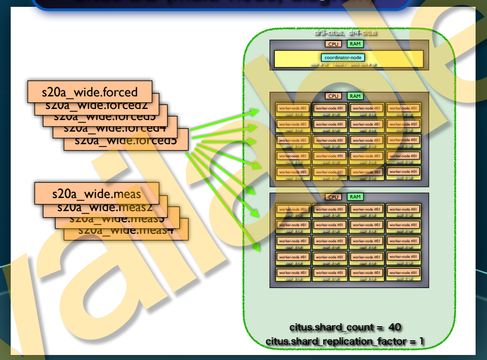
fast

It is 10～300 times faster than original !

CPU based distributed DB system (Citus)

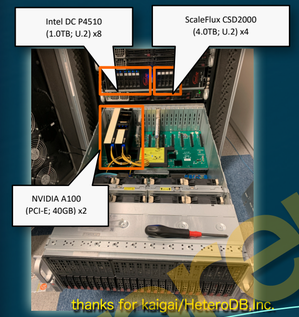
https://www.citusdata.com

Citus DB (multi-node, diagram)



GPU based parallele executing DB system (PG-Strom)

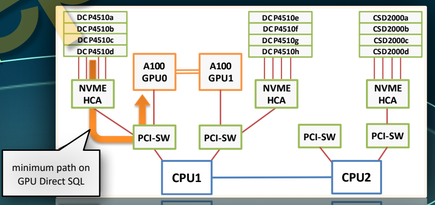
prototype test machine.



thanks for kaigai/HeteroDB,Inc.

System parts list:

Chassis	Supermicro SYS-4029GP-TRT	1
CPU	Intel Xeon Gold 6226 (2.7GHz/12C)	2
RAM	16GB DIMM (DDR4-2933; ECC)	12
GPU	NVIDIA Tesla V100 (40GB; PCI-E)	2
SSD1	Intel DC P4510 (1.0TB; U.2)	8
SSD2	ScaleFlux CSD2000 (4.0TB; U.2) - working area	4
HDD	SATA HDD (2.0TB; 72krpm) x3	3
N/W	build-in Gigabit Ethernet x2	2
OS	Ubuntu 18.04 (kernel-5.4.0-52-generic x86_64) CUDA Toolkit 11.1 (driver: 455.32.00) GPUDirect Storage 0.9beta	---
DB	PostgreSQL v12.5 + PG-Strom v3.0devel	---



minimum path on GPU Direct SQL

https://heterodb.github.io/pg-strom/

available in Reception