Korean In-Kind Contribution to the LSST with the Lite Independent Data Access Center

Korea Astronomy and Space Science Institute

Min-Su Shin and **Chang Hee Ree**
Korea Astronomy and Space Science Institute
msshin@kasi.re.kr and chr@kasi.re.kr

Legacy Survey of Space and Time

ABSTRACT

The KASI plans to establish a Lite Independent Data Access Center (IDAC) serving the subsets of LSST data products (object lite catalog and image coadds) and computing resources, as a part of Korean contributions to the Rubin LSST in-kind program. We explain the required system and its operation plan.

What is the Lite IDAC?

The full IDACs will host the full LSST dataset including prompt data products, and the lite IDACs will host only subsets of the LSST data. The KASI Lite IDAC plans to host the so-called object lite catalog and some subsets of image coadds. The object lite catalog would contain 1840 bytes per row for the 40 x 10^9 objects, corresponding to about 74 TB. Two different sets of image coadds are expected in the LSST, and the expected volume will be about 300 TB. UK and Canada are considering the full IDACs, and other countries will host the lite IDACS.

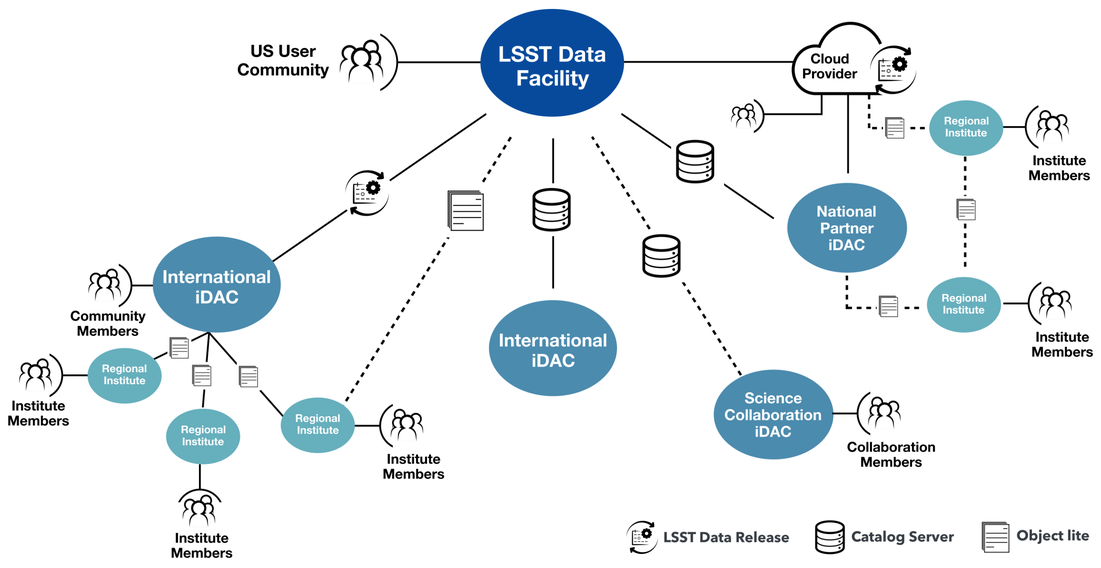


Figure. US data facility (DF) and IDAC network topology (LSST document RTN-003).

Hardware resources

The Lite IDAC will be hosted within the KASI Science Cloud platform already in operation. Current system configuration is :
- OpenStack cluster : 1,080 vCPU, 5.7 TB RAM, 620 TB storage,
- High-density object storages : 1,800 TB (raw capacity).

The KASI plans to add the following hardware components to build the IDAC:
- Intel Xeon CPU clusters,
- Solid state disks for database systems,
- Hard disks for image coadds.

The KASI already has a high-speed network access point and a possible dedicated data transfer node.

We also consider possible usage of domestic public clouds.

Plan: development and operation

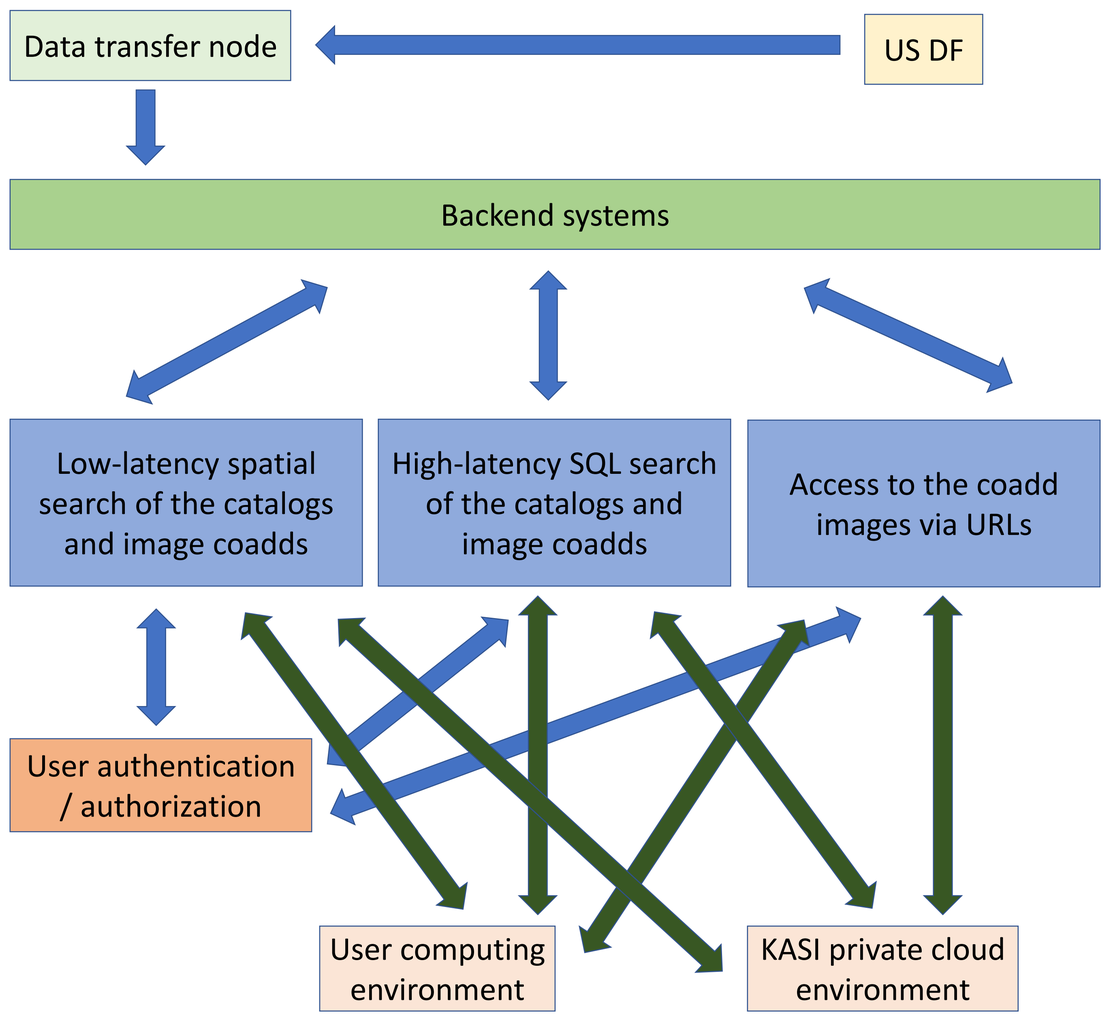
We plan to conduct necessary development activities in 2023 – 2024 before the first data release of the LSST data products.

The following development items are in our to-do list.
- Tests of the relevant software stacks in the KASI private cloud,
- Preparation of the required extra hardware components,
- Development of user access interface under security measure,
- Evaluation of long-term usage of the public clouds,
- Writing user documents and tutorials to help LSST community use the KASI Lite IDAC.

We like to find well-defined analysis applications by Korean community members, and the KASI Lite IDAC will consider the deployment and user-support for the chosen analysis applications. Example applications include regular updates of photometric redshifts and cross-matching between the lite object catalog and community-common catalogs.

The KASI Lite IDAC needs to be maintained during the whole project period (DR1 to DR13). The minimum requirement of dedicated system resources is 20M total CPU-hours plus 500 TB of continuous storage, in exchange for 1 LSST PI data right.

System architecture



Software stacks

Our software stacks consist of open-source softwares, and they are deployed with Kubernetes.

Authentication/authorization: Keycloak (https://www.keycloak.org/) as open source identity and access management with OpenID connect service from the US data facility.

Low-latency spatial or object ID search:
- Aerospike (https://aerospike.com/) as distributed, scalable in-memory/SSD NoSQL database with spatial indexing and search,
- Redis (https://redis.io/) as distributed, scalable in-memory NoSQL database,
- Apache Kvrocks (https://kvrocks.apache.org/) as distributed key value NoSQL database using RocksDB as storage engine in the SSD storage.

General SQL search:
- MySQL (https://www.mysql.com/) as default SQL database engine,
- Vitess (https://vitess.io/) as scalable, reliable, MySQL-compatible, cloud-native database.

Storage for image coadds:
MinIO (https://min.io/) as high-performance, S3 compatible object storage.

Monitoring: Prometheus (https://prometheus.io/) as monitoring system and time series database.

Orchestration: Kubernetes (https://kubernetes.io/) as system for automating deployment, scaling, and management of containerized applications.

Other softwares will be chosen by following the guides from the US DF.

If you are interested in development for software stacks and operation of the KASI IDAC, please, contact C. H. Ree for possible contribution and collaboration.

VERA C. RUBIN OBSERVATORY