IDAC Knowledge Base and Frequently Asked Questions

General

Handbook for Proposal Teams v1.0 - Rubin LSST International In-Kind Contribution Program
Guidelines for Rubin Independent Data Access Centers
Rubin Observatory Documentation site
2021 Rubin PCW IDACs breakout

LSST Data

LSST Data Product Categories
LSST Data Products Definition Document

Data Management

Data Management Middleware Design
Data Management Database Design
DM sizing model and purchase plan for the remainder of construction.
Spreadsheet with new hardware sizing model

Science Platform

LSE-319: LSST Science Platform Vision Document
Data Management LSST Science Platform Requirements
Science Platform Design
Video recording of Rubin Boot Camp session on How to Deploy Kubernetes Services with ArgoCD
George Beckett's notes on deploying the Rubin Science Platform at the UK IDAC

Authentication and Authorization

SQR-044: Science Platform identity management requirements
Science Platform Authentication Frontend: Gafaelfawr — Gafaelfawr v0.1.dev1+g3e3fd0a
RSP identity management implementation strategy: https://dmtn-224.lsst.io/

Policy

LSST Data Policy

FAQs

Q: We plan to operate a Lite IDAC, and would like to use the Rubin Science Platform as an interface. Can we do this?

A: Lite IDACs **should not** count on the Rubin Science Platform being available as an interface to their hardware resources. The resources provided by IDACs must have a service layer in order for them to be useful to users. It is expected that Lite IDACs will use existing service layers, or develop them themselves, to support usage of their hardware. Full IDACs are required to adopt the Rubin Science Platform.

Q: Are the values for number of cores and storage listed in the templates suggestions or requirements? Do they need to be up and running by 2024 according to the new time table?

A: Please consult the tech note RTN-003 (https://rtn-003.lsst.io) referenced in the handbook; there are some numbers in the Handbook provided as examples.

Q: The value of IDAC CPU resources are based on CPU hours. But what is the reference CPU type/clock speed/memory per core?

A: The current fiducial CPU used for sizing at the US Data Facility is an AMD Rome processor with 128 cores and 512 GB total RAM. As specified in the Handbook, IDAC proposers should report both CPU hours and FLOPS provided, so that a fair assessment of the computational resources can be made.

Q: Can requirements for Object Lite DAC be negotiable if it is a science specific Object Lite DAC (e.g. star catalogs) and how will such a science specific Object Lite DAC be valued?

A: The storage and computational resources for an Object Lite DAC could be allowed to differ somewhat from the requirements listed in RTN-003, depending on the volume of catalog data hosted. The requirements on Rubin-compatible Authentication and Authorization, the agreement to make the data broadly available to Data Rights Holders, the need for sufficient connectivity to support users, the need for TAP, MyDB, and CAOM, and the requirement for support may not be negotiated. A science-specific Object Lite DAC should consult with the appropriate Science Collaborations for an endorsement of their value.

Q: For an IDAC or IDAC Lite proposal: what kind of relation to the SCs is expected? A direct relation/embedding in one SC? More SCs? Or independent of SCs? Must an IDAC (Lite) have

a well-defined scientific scope/profile?

A: IDAC proposals do not need to have a direct relationship to the Science Collaborations. Indeed, IDACs that provide significant data, storage, and computational resources, **supported by a service layer**, to the general LSST community will be among the most valuable In-Kind contributions. If an IDAC aims to have a specific scientific purpose, however, it should seek an endorsement from the appropriate Science Collaborations. Please contact the In-Kind Helpdesk (in-kind-team@lists.lsst.org) for guidance on how to proceed.

Q: One promising option for running an IDAC is to use a GPU-based supercomputer. I know that it is technically feasible to run a database on GPUs and this works even faster than on CPUs. However, can you see any potential caveats, any cons ? For example, can the Rubin Science Platform be run on GPUs ? Also, is there any party of the LSST Consortium which considers launching IDACs on GPU machines ?

A: There are no plans on the part of Rubin to use GPUs in the science pipelines. However, GPUs could be beneficial for certain computationally bound use cases, such as crossmatching large tables or all-sky visualizations. Operating a database on GPU-based hardware could offer speed benefits if data access is not I/O-bound, but would need development work on the part of the proposers to demonstrate the utility. Note that the Rubin Science Platform should not be assumed to be available for deployment for an IDAC; please contact the In-Kind Helpdesk (in-kind-team@lists.lsst.org) for questions about it.

Q: What sets the minimum connection speed of 20 Gb/s for a Full IDAC? What sort of connection requirement is there for Lite IDACs?

A: The requirement of sustained speed of 20 Gb/s for a Full IDAC is not negotiable, and is needed in order to support the transfer of pixel data in a timely way. For Lite IDACs, which typically will serve slimmed-down catalog data such as Object Lite, the requirement on connection speed is only that it be sufficient to support users of the Lite IDAC in a performant way. Lite IDAC proposals could use past user experience over their existing connection to justify that this requirement is met.

Q: What are the requirements for the authentication and authorization procedure? Do you plan to have one central authorization server and then distribute users among the IDACs, or will users directly access each IDAC individually?

A: Work on authentication and authorization will be performed during summer 2022, so exact details are subject to change. However, the expectation is that IDACs will authenticate against a service.

Q: When should IDACs expect to begin operations?

A: It is expected that, in general, IDAC operations will start with the first data release of the LSST catalog. The date for this release is still uncertain given delays in construction as a result of the pandemic, but is likely to be sometime in 2025.