The current design and implementation of the AstroDS Data Aggregation Service
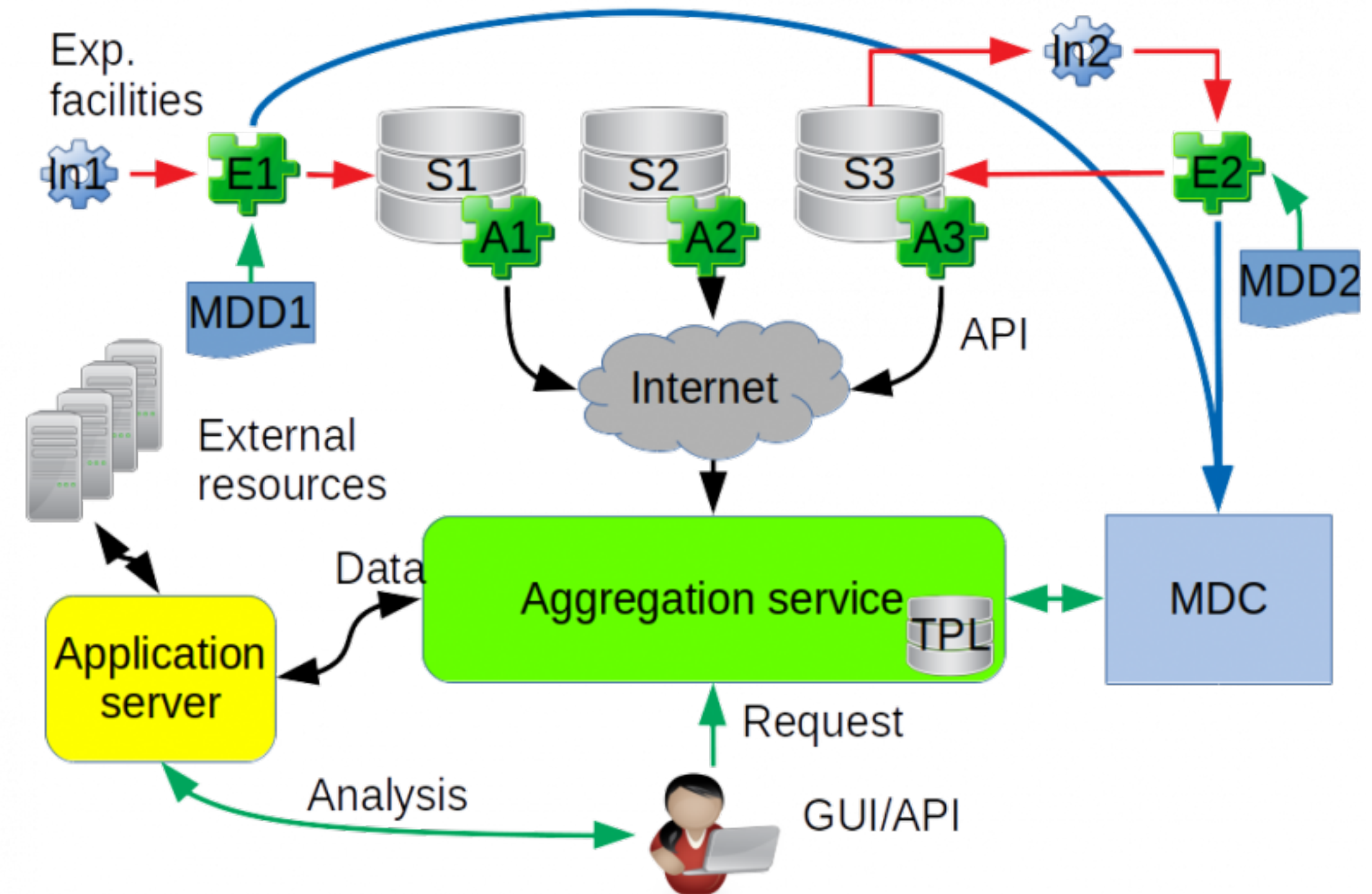
Minh-Duc Nguyen (nguyendmitri@gmail.com) (1), Alexander Kryukov (1), Andrey Mikhailov (2)

(1) Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
(2) Matrosov Institute of System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia
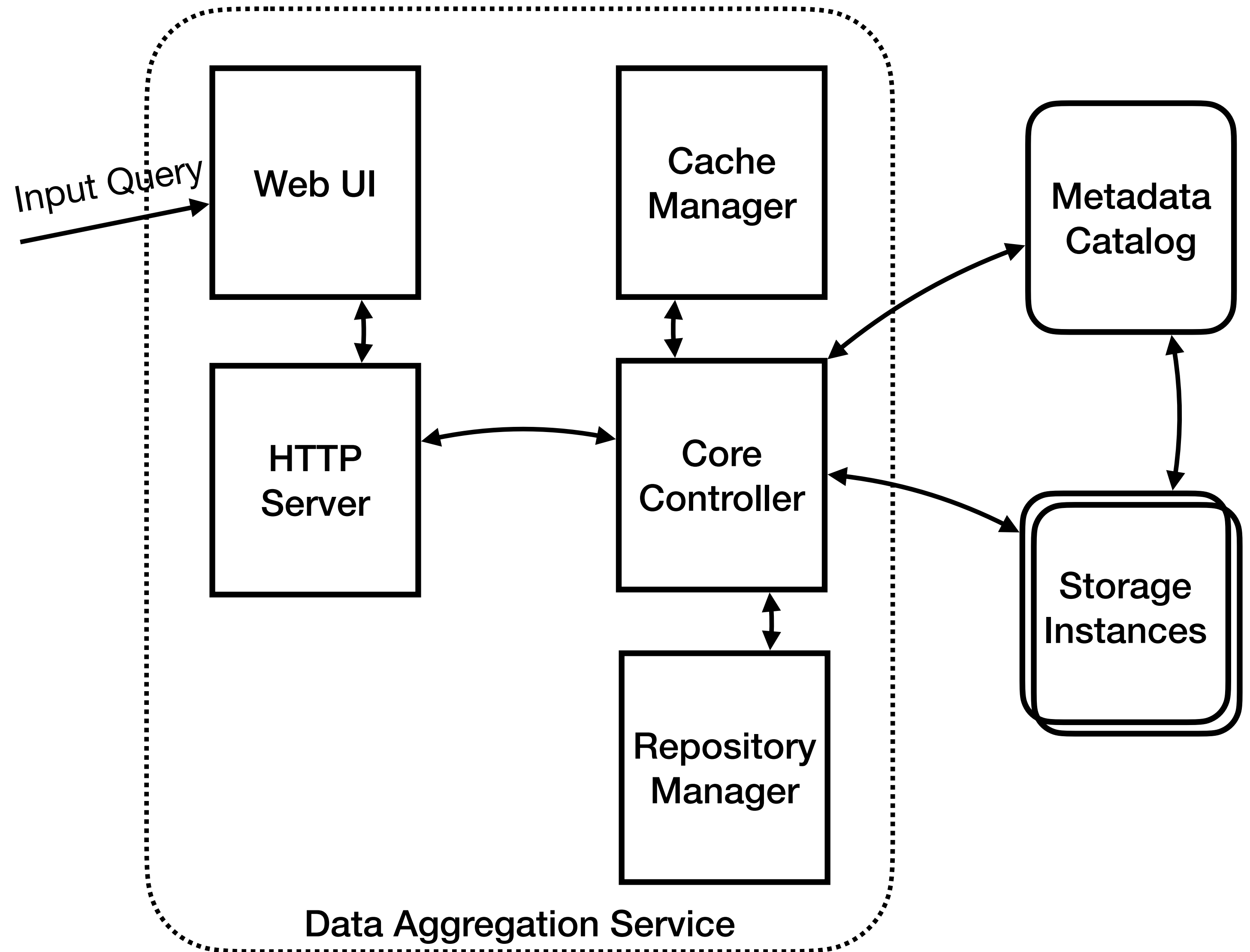
DLC-2020, June 8th, M.D. Nguyen

AstroDS Architecture

- Gather data measured by the instruments of various physical experiments (TAIGA, TUNKA, KASCADE, etc.) into global storage

- Search/Retrieve data via a standardized storage-independent user-friendly interface
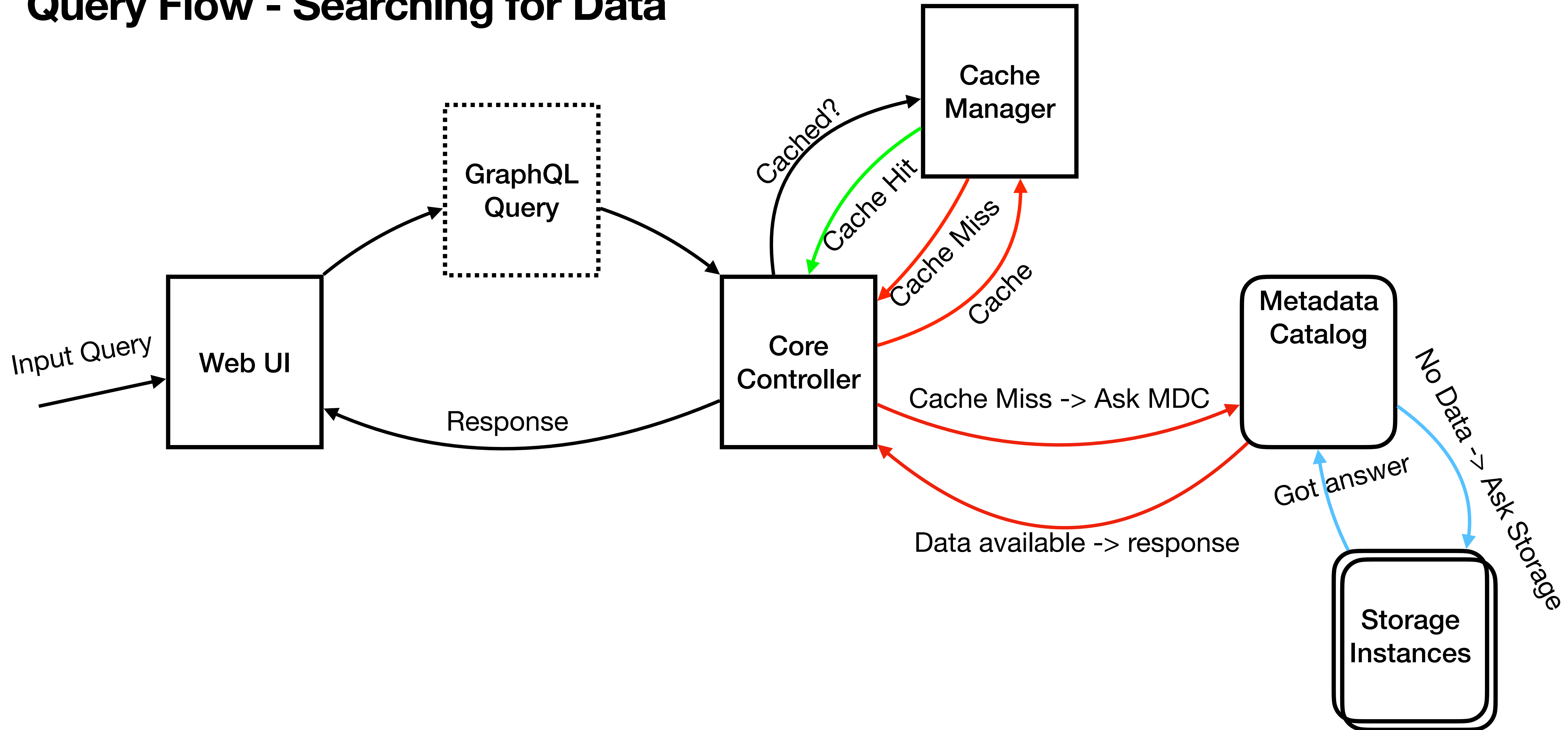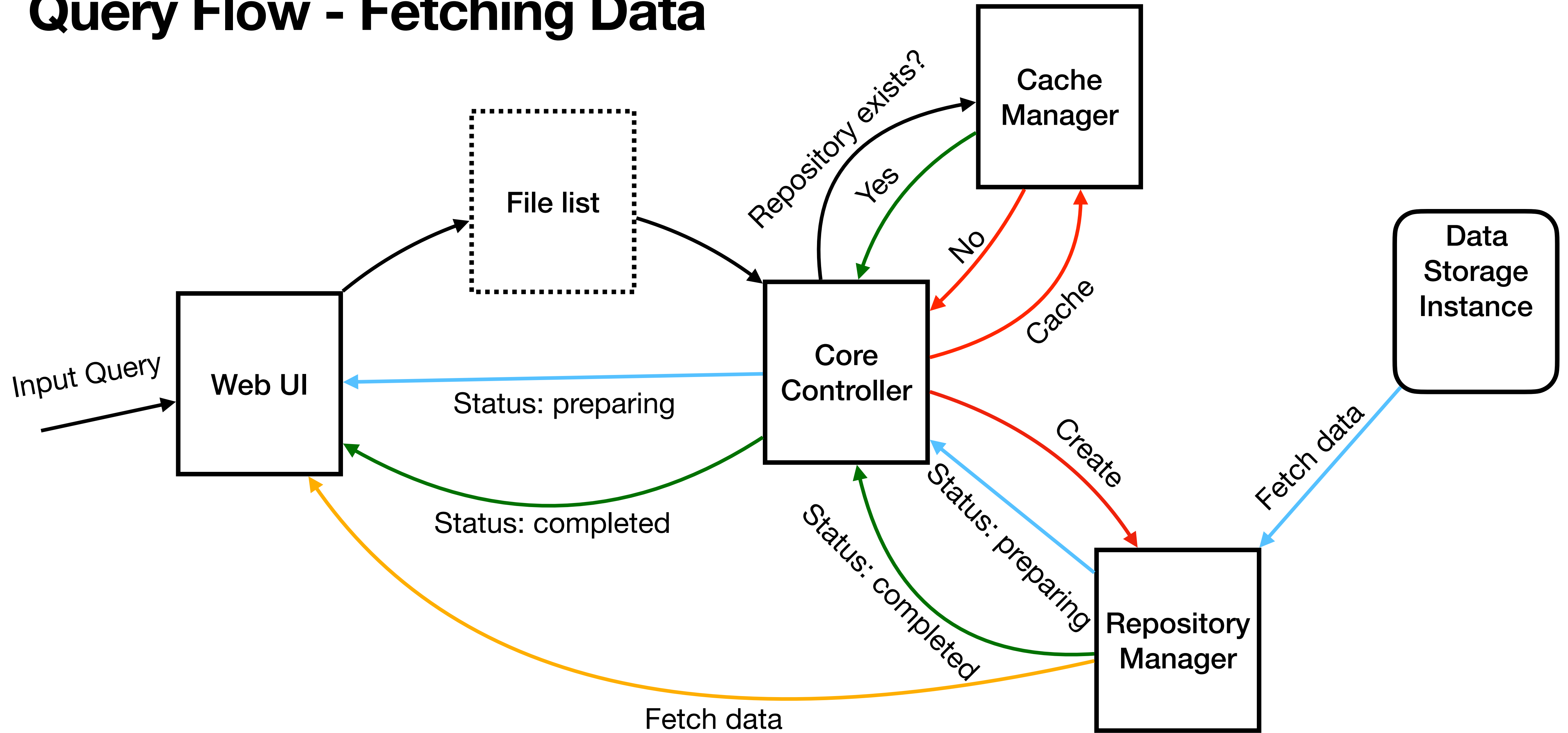
Data Aggregation Service - Design

- Distributed components talking to each other via REST API

- Caching query results at the service's level

- Actual data fetching on reads

Input Query

Web UI

Cache Manager

Metadata Catalog

HTTP Server

Core Controller

Storage Instances

Repository Manager

Data Aggregation Service
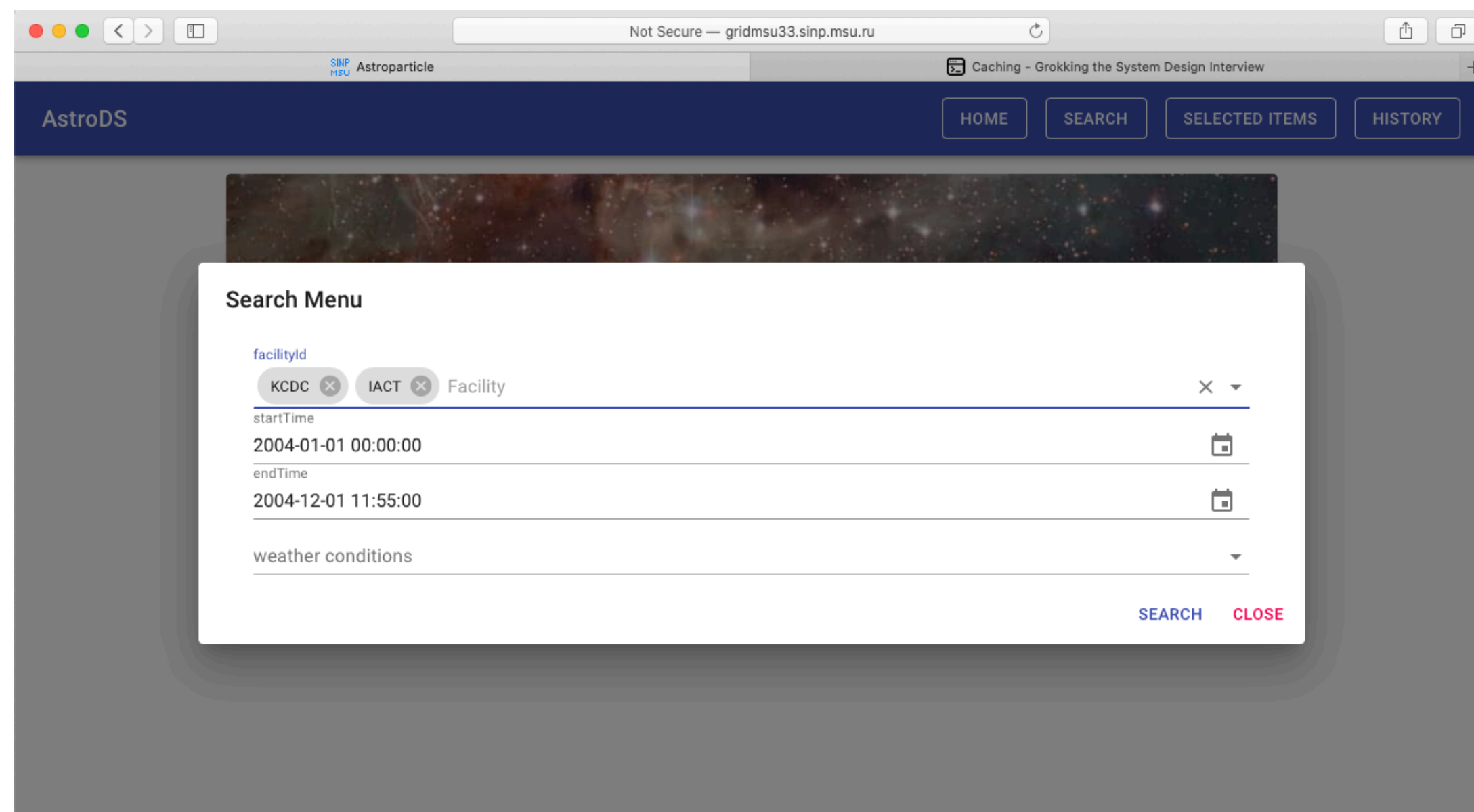
Query Flow - Searching for Data

Query Flow - Fetching Data

Data Filters

- Grouped by facilities

- Types: number, datetime, list

- Conditions: "=", ">", "<", ">=", "<=", "!="

```json
{
  "id": "1",
  "name": "IACT",
  "version": "1",
  "codename": "iact",
  "filters": [
    {
      "name": "startTime",
      "type": "datetime",
      "conditions": ["="]
    },
    {
      "name": "endTime",
      "type": "datetime",
      "conditions": ["="]
    },
    {
      "name": "weather",
      "type": "list",
      "options": {
        "table": "weather",
        "request": "query{weather{id wScale}}",
        "fields": {
          "id": { "type": "integer" },
          "wScale": {"type": "string"}
        },
        "conditions": ["="]
      }
    }
  ]
}
```

Cache Manager

- **Hash table:**
  - key: query hash (MD5)
  - value: response and metadata
- **Write-back cache**
- **Least Frequently Used (LFU):** Counts how often an item is needed. Those that are used least often are discarded first
- **Expiration time:** 7 days

Repository Manager

- **Least Frequently Used (LFU):** Counts how often a repository is needed. Those that are used least often are discarded first
- **Expiration time:** 3 days

Current Status

Implemented

✓ Available data sources: IACT, HiScore, KCDC

✓ Multiple data source selection

✓ Dynamic filters: time ranges, weather conditions

✓ Cached query result

✓ Pagination

✓ Repository generation

TODOs

- Authorization

- Other filter types

- Asynchronous repository preparation

- Cache invalidation

- Production deployment

Demo

Thank you :)

AstroDS Data Aggregation Service: http://gridmsu33.sinp.msu.ru

Minh-Duc Nguyen (nguyendmitri@gmail.com)