Effective Data Visualisations with Kibana Dashboards for CRAMS

Samitha Amarapathy
Team Leader Software Development | Senior Project Manager

Rafi Mohamed Feroze
Senior Analyst Programmer

Monash eResearch Centre

A Little bit on CRAMS....

Monash eResearch has developed CRAMS (Cloud Resource Allocation Management System) back in 2017

CRAMS provides an effective self service mechanism for researchers and research facilities to request cloud resources, monitor usage and manage own allocations.

CRAMS "Flavours"

Data dashboard - CRAMS offering for Research data storage allocation management and reporting.

HPC dashboard -   CRAMS offering for high performance computing (MASSIVE, MonArch and CVL) allocation management and reporting.

Cloud dashboard - CRAMS offering for compute cloud (Nectar) resource usage reporting.

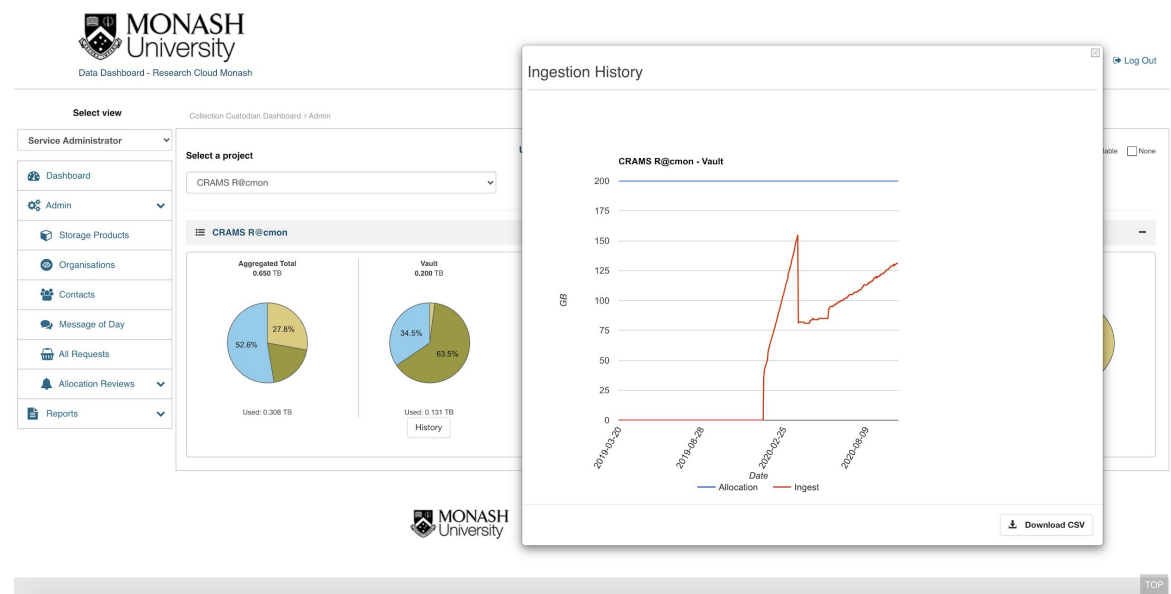# Some of the CRAMS "Programmed Dashboards"

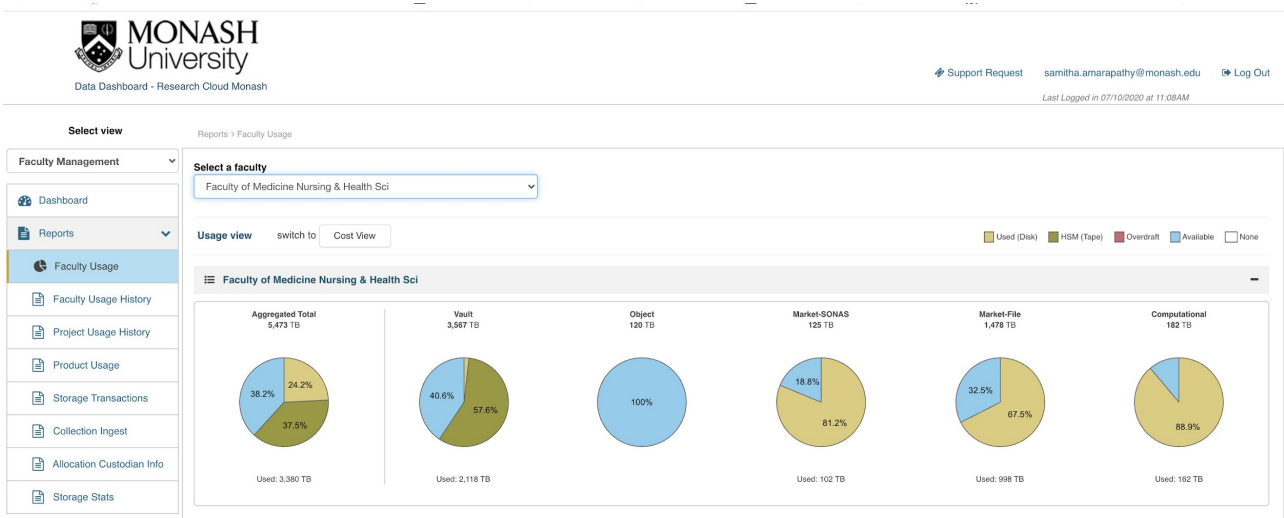Figure 1. Project View : Project level storage allocations and usage

Figure 2. Faculty View : Faculty level storage allocations and usage

"Programmed Dashboards" vs Kibana Dashboards

Motivation

- Ever increasing demand for more and more management reports and stats

- Cost of developing "Programmed Dashboards"

- Need for effective and easily customizable visualizations

- Integration of CRAMS data with some other Monash eResearch Centre data sets

- Need for more and more insights

- Sometimes management can't predict what they want next

Kibana

- Kibana is an open-source data visualization and exploration tool used for log and time-series analytics, application monitoring, and operational intelligence use cases.

- It offers powerful and easy-to-use features such as histograms, line graphs, pie charts, heat maps, and built-in geospatial support.

- Also, it provides tight integration with Elasticsearch, a popular analytics and search engine, which makes Kibana the default choice for visualizing data stored in Elasticsearch. (reference : https://kubedex.com/resource/kibana/)

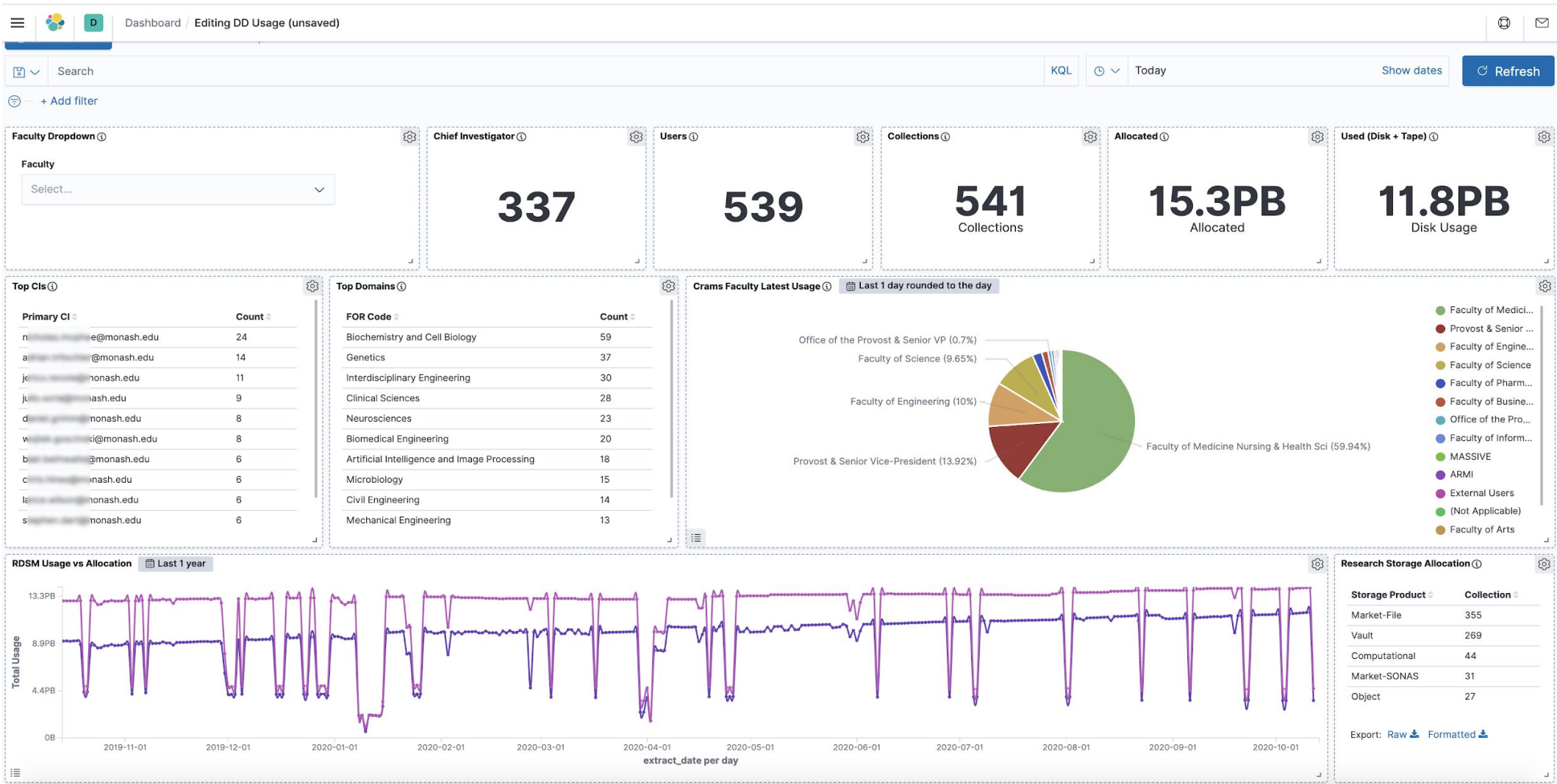# Kibana based Storage Stats Dashboard

Figure 3.

Storage Allocations Vs Usage Aggregate view
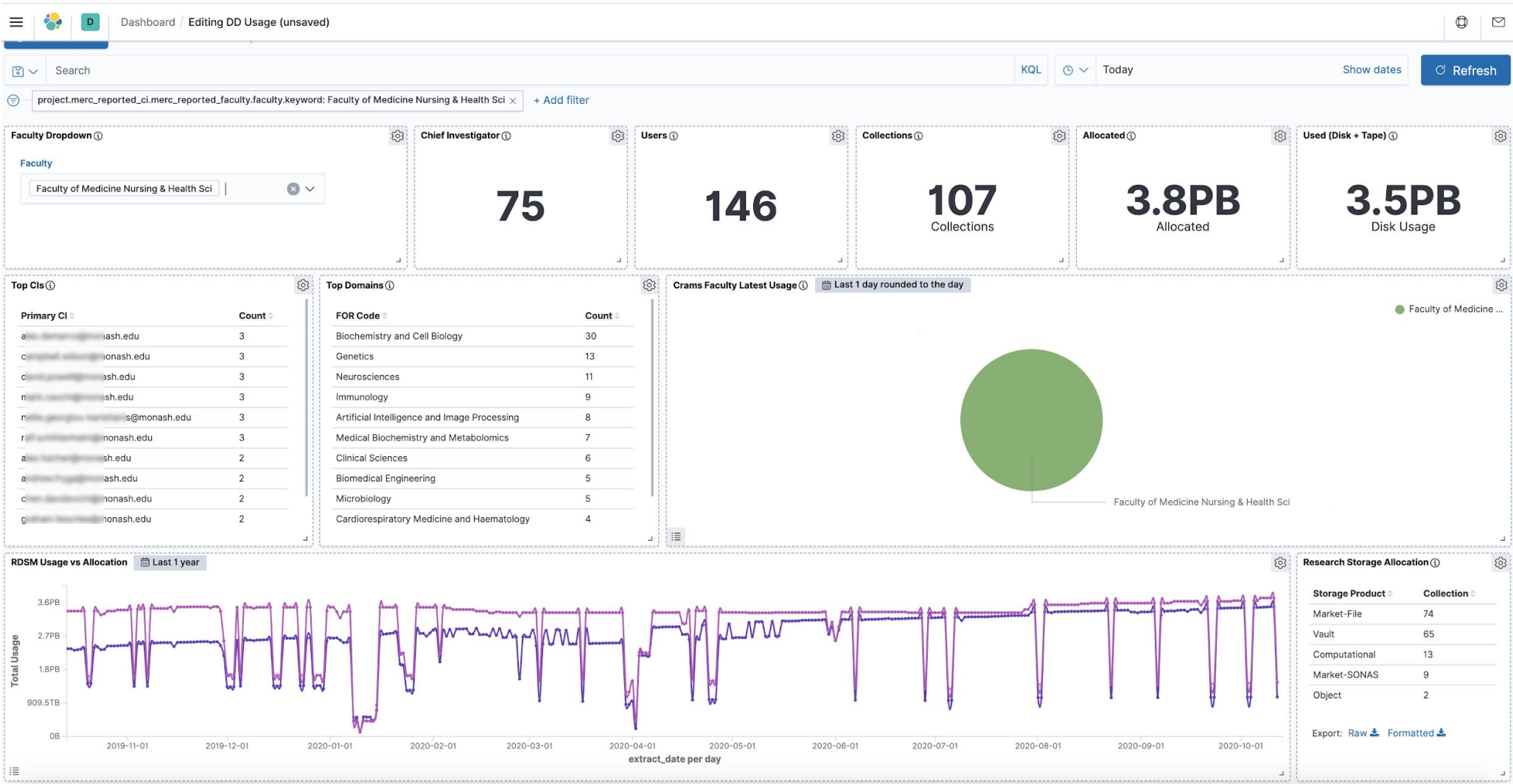
Kibana based Storage Stats Dashboard

Figure 4.

Storage Allocations Vs Usage Faculty view

Some of the dips indicate missing infrastructure level reporting

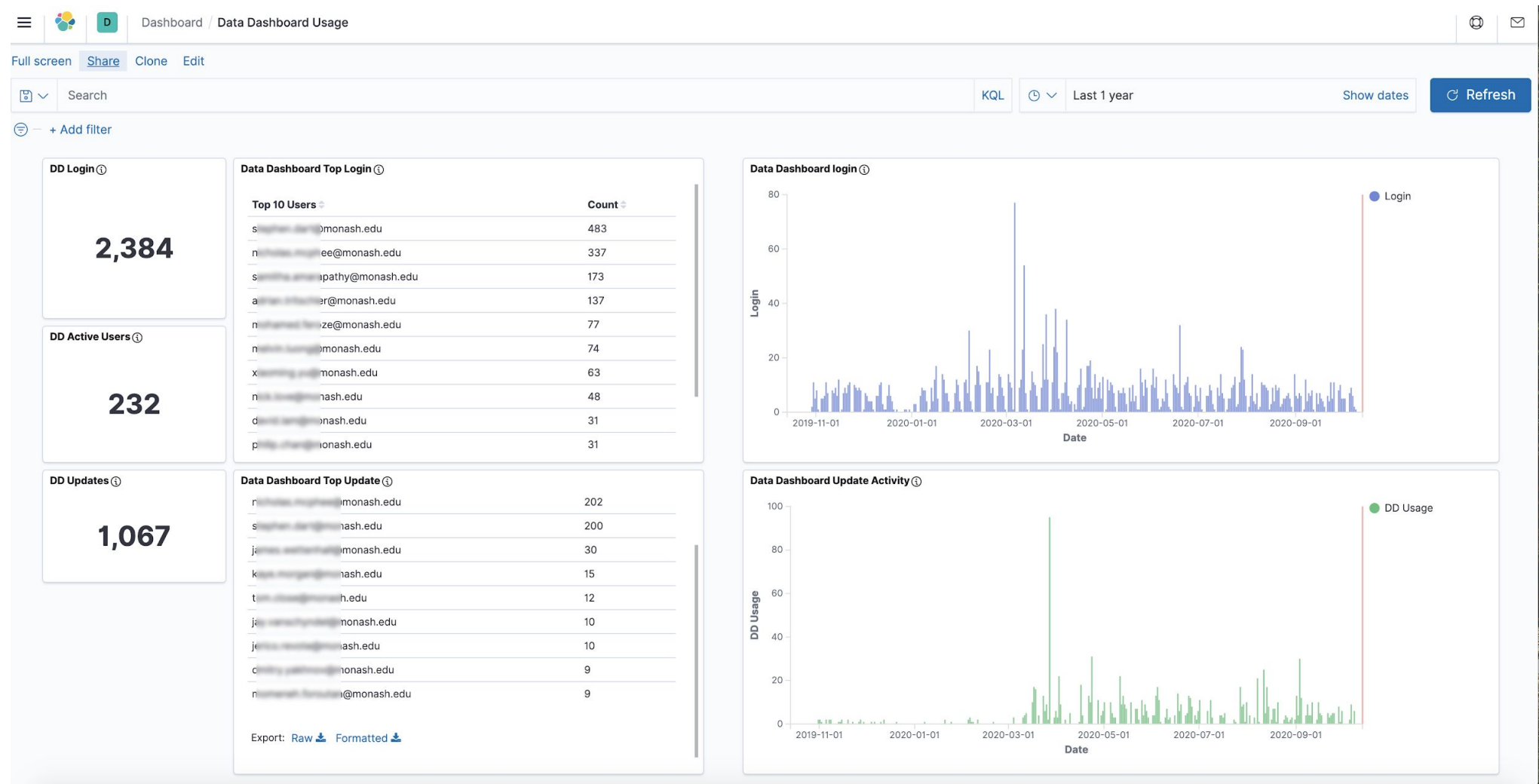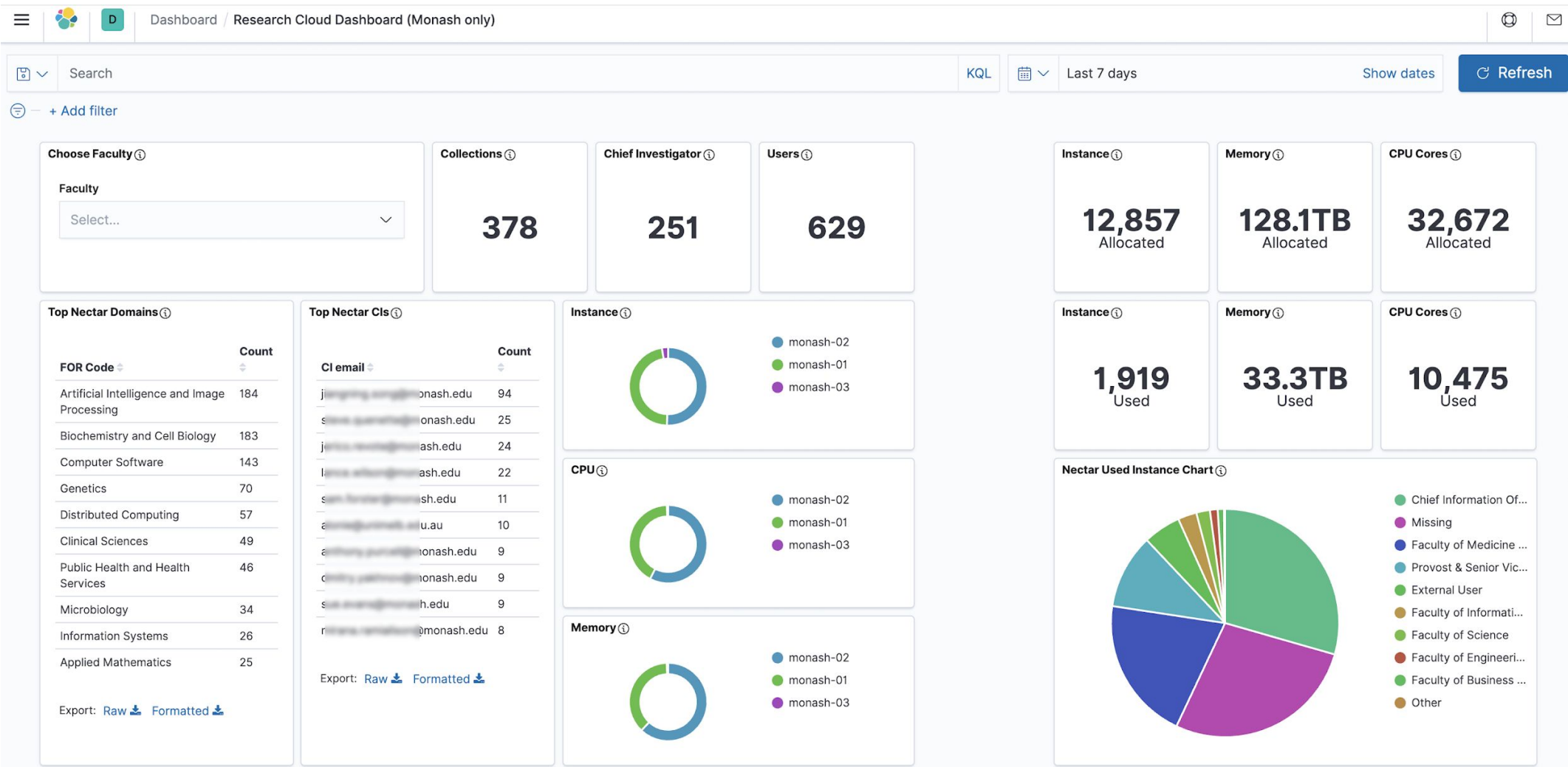# Kibana based User Interactions Dashboard

Figure 5.

User interactions view for Data Dashboard

Kibana based Compute Cloud Stats Dashboard

Kibana based Research Cloud Stats

CRAMS Data Extract

Daily extract
- Time-series data since last extract timestamp: used in time graphs
    - user activity logs, infrastructure reported usage information
- Snapshots
    - allocations, contact information, collection metadata, faculty information

Data Format
- newline delimited JSON
- flat and nested data
- convert numerical data to base units, for example bytes instead of GB or MB

Data Integrity
- use unique date field 'extract_date' to identify data (can be different from ingest date)
- generate unique id at time of ingest using elastic pipeline from data fields
    - helps prevent accidental duplication of data
    - for example: db_id + extract_date will ensure the same record is not re-ingested for any day

Data Ingest & Visualisation

Use python Scripts

- Both data extract and ingest into ElasticSearch are managed by python scripts
- Extracted data is stored in files for reloading ElasticSearch, if needed
- Use python elasticsearch client
    - https://elasticsearch-py.readthedocs.io/en/master/

Kibana Index Patterns

- Identify Unit conversion fields
    - identifying which fields are in bytes allows Kibana to display it in GB/TB/PB as required

Build Base Visualisation models

- provide a starting point for management
    - to investigate data, find answers to questions related to business decisions
- provides a data sandboxed playground
    - restricts data to selected index and ensures integrity (to an extent)
        - nested data comes with its own set of problems

What am I interested to know...

Questions

- Which faculty storage usage increased during covid-19 ?

- Who were the top users of our system in the last three months ?

- Can I look at the global view and then drill down to a small subset or individual item ?

Investigations

- What are those sudden dips in the Usage reporting
    are they related to reporting failures ?

- Has research user behaviour/practices changed in the last three months?

- Which collections/users will be better-off moving from one Product to another ?
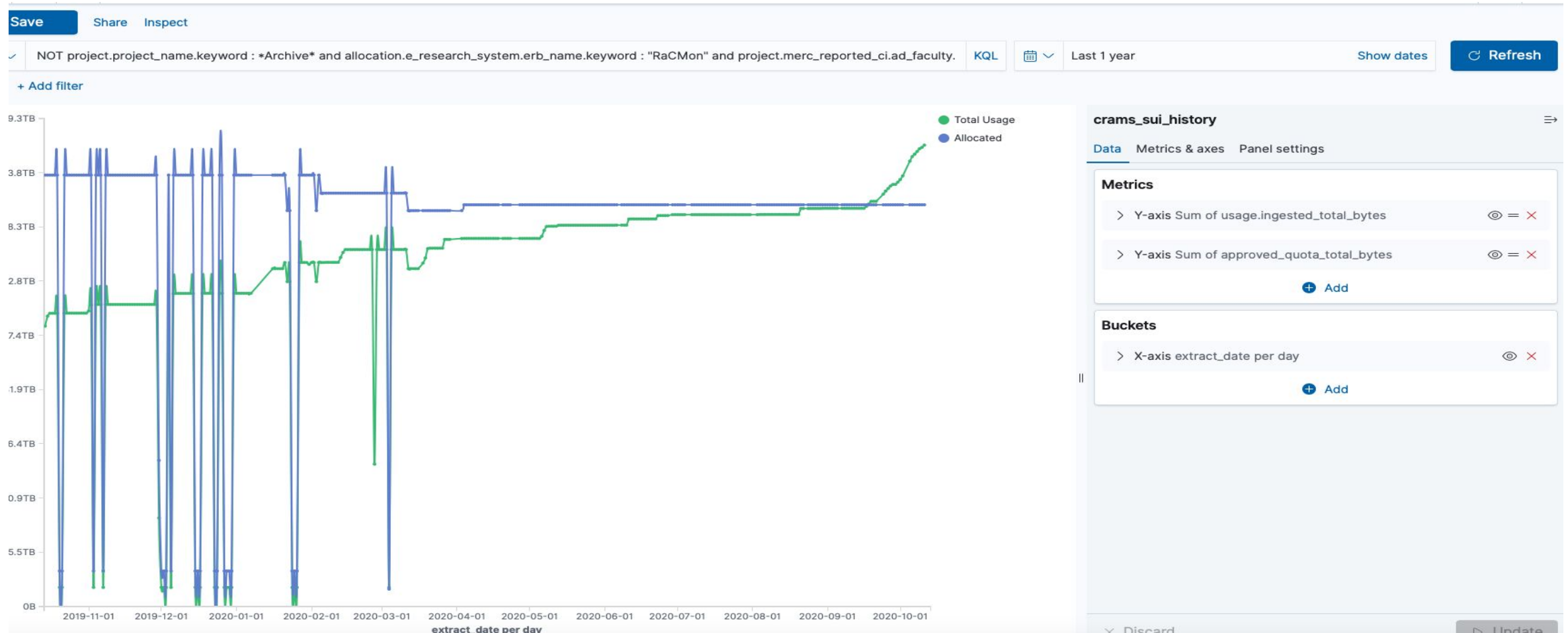
MONASH University

Faculty usage during Covid-19

Faculty usage during Covid-19

Faculty usage during Covid-19

Thank You