KAI-Project

Self-Adaptive dCache

A collaborative project involving DESY und HAW Hamburg

What is KAI ?

- KAI stands for

  „Cooperation for Application and Innovation of HAW Hamburg and DESY“
- new strategic cooperation in research, teaching, innovation and technology transfer.
- joint research programs, synergetic use of complex scientific facilities
- increased cooperation in the academic training of the next generation of scientists and engineers as well as cooperation with industry
- is intended to help shape Hamburg's structural transformation from a port city to a city of science and innovation

Marketing texts stolen from: https://www.kai-hamburg.org/

What is the Self-Adaptive dCache project about ?

The problem statement:

- „Hot-Spots“ occur in dCache storage due to high demands of files and datasets

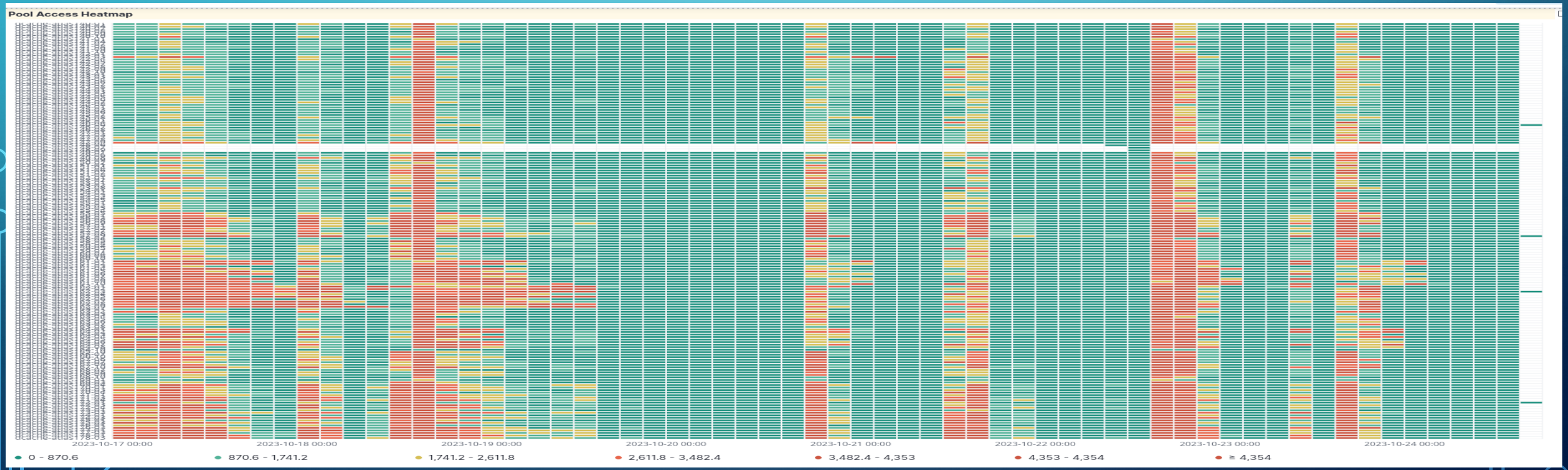


Image with courtesy of Tigran Mkrtchyan

What is the Self-Adaptive dCache project about ?

The Challenge:

➔ Define indicators for the dCache-system overload

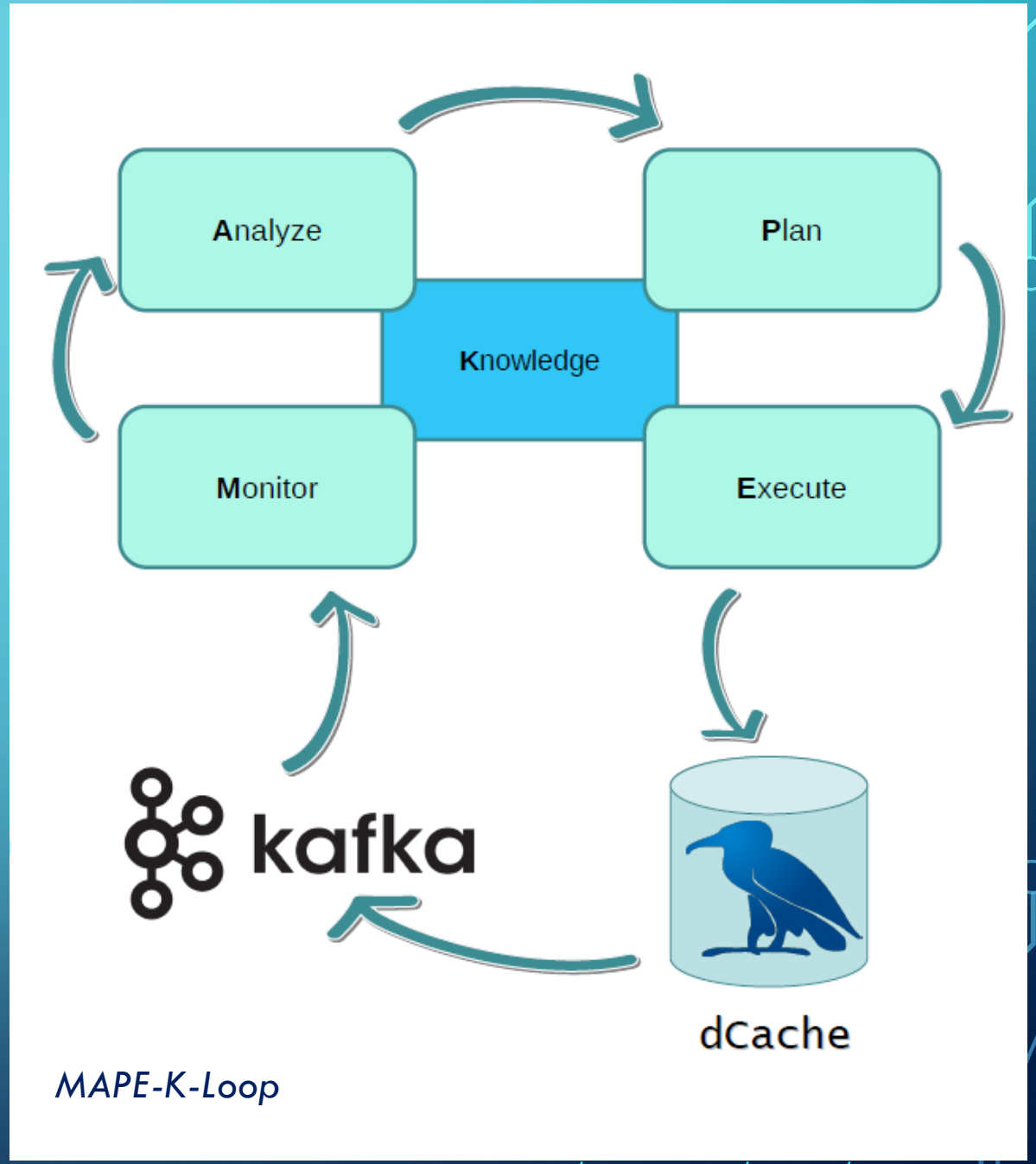
➔ Early pre-mortem Hotspot detection

➔ Definition of threshold values for the overload

➔ Detection of problematic job patterns and user behavior

Favorable outcomes and solution strategies:

- **Rebalance** the dCache-system, with the aim of equal distribution of data, while respecting **fairness** among user‘s access
- Enable dCache to **adapt** to different system states utilizing a **MAPE-Loop**
- **Relieve** and support administrative staff



MAPE-K-Loop

What was achieved so far:

- One already completed bachelor thesis deals with the topic of problematic user behavior and its consequences for the system.
- A soon-to-be-completed bachelor thesis is investigating the possibility of reading important software indicators before the system overload occurs.
- A REST-API for specific dCache operations has been implemented.
- A GAP analysis has been carried out and evaluated.

  The results were published in a white paper.
- A basic MAPE-Loop was implemented and made available.

Open threads

- A MAPE-Loop development and testing pipeline was setup and needs to be filled with useful content
- Extend the MAPE-Loop functionality
  - Execution phase triggers dCache cost module
  - Improve analysis strategies
- Include Machine Learning?
- . . .

https://gitlab.desy.de/dcache/KAI-Projekt/mape_loop

THANK YOU !

Questions ?