ESA Mission Operations Shared Storage Cluster Solution

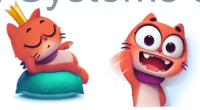
Klara Widegård, Alicia Yustos Rodriguez, Rui Santos

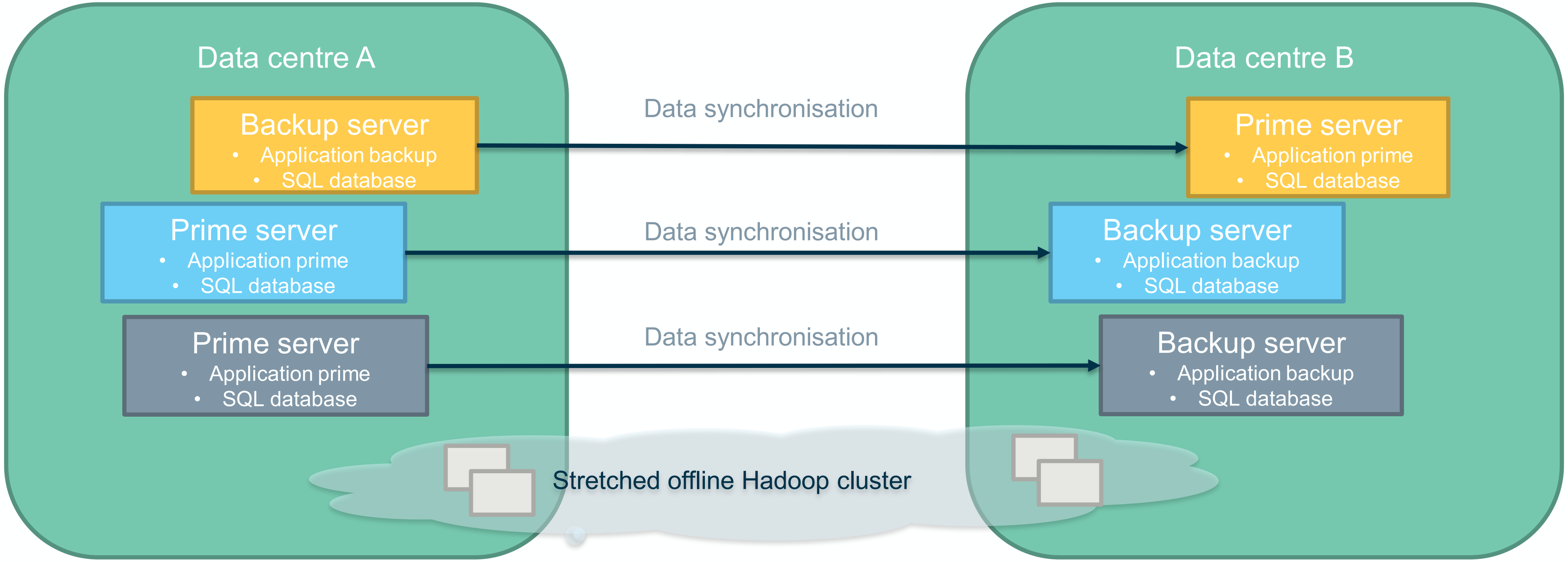
27/02/2023

Outline

1. Past - every mission is an island
2. Transition – moving to shared solutions implies moving to big data solutions
    - Technical solution
    - Concerns & mitigations
3. Future – new frontiers

Mission Operations in the past – single island approach

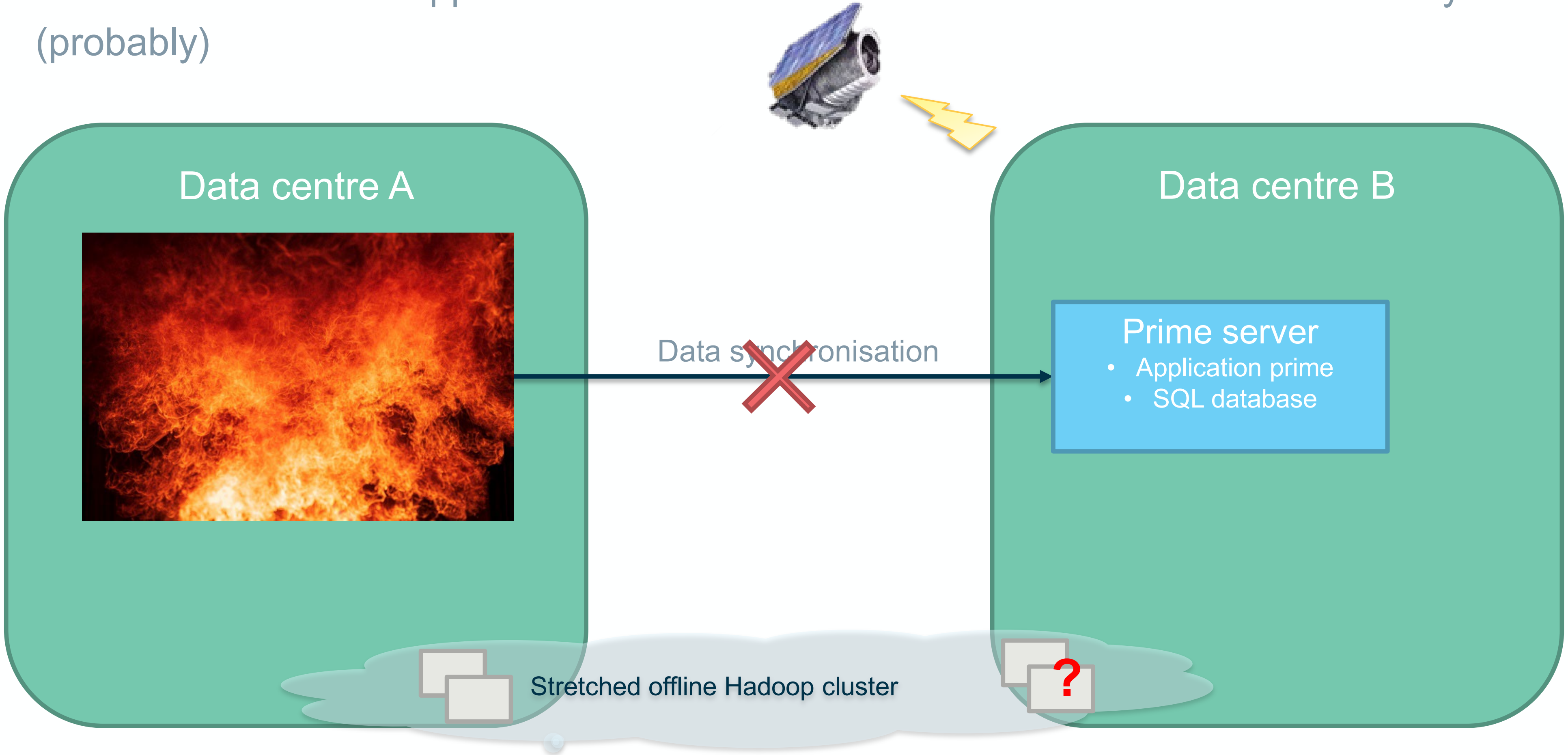
- The previous generation of Mission Control Systems at ESA relied on every mission
  - owning their own hardware (pets) 
  - setting up their own redundancy concept, typically identical prime & back-up servers distributed on two data centres
- The SQL data was in addition backed up by the IT department to tape
- For certain missions some data were also transferred to a single shared Hadoop cluster for further data analysis
- There were some incentives to use shared SQL database services, at least for testing



Storage volume
< TBs/year/mission

Mission Operations in the past – single island approach (2)

- In case of one data centre not being available, the mission switched to the backup server in the other data centre within a short time
- Unclear what would happen to the shared stretched cluster and if offline data analysis would be impacted (probably)



Mission Operations in Transition – towards common stores

We are now starting to build the solution for the future operational data storage cluster for mission operations:

- One shared solution for all missions
- MUST be performant
- MUST be scalable
- MUST cope with (ever-increasing) data volumes in the order of hundreds of TB
    - e.g. GAIA receives around 150 million unique HKTM samples per day
- MUST provide redundancy and be fault tolerant
- MUST be designed to prevent data loss
- The full solution does NOT NEED to support hot redundancy or hot failover
- Planned downtime MUST NOT impact all missions

The estimated storage volumes implies we are going Big Data! Do we really want that?

Mission Operations in Transition – towards common stores

We have decided YES, because we want to

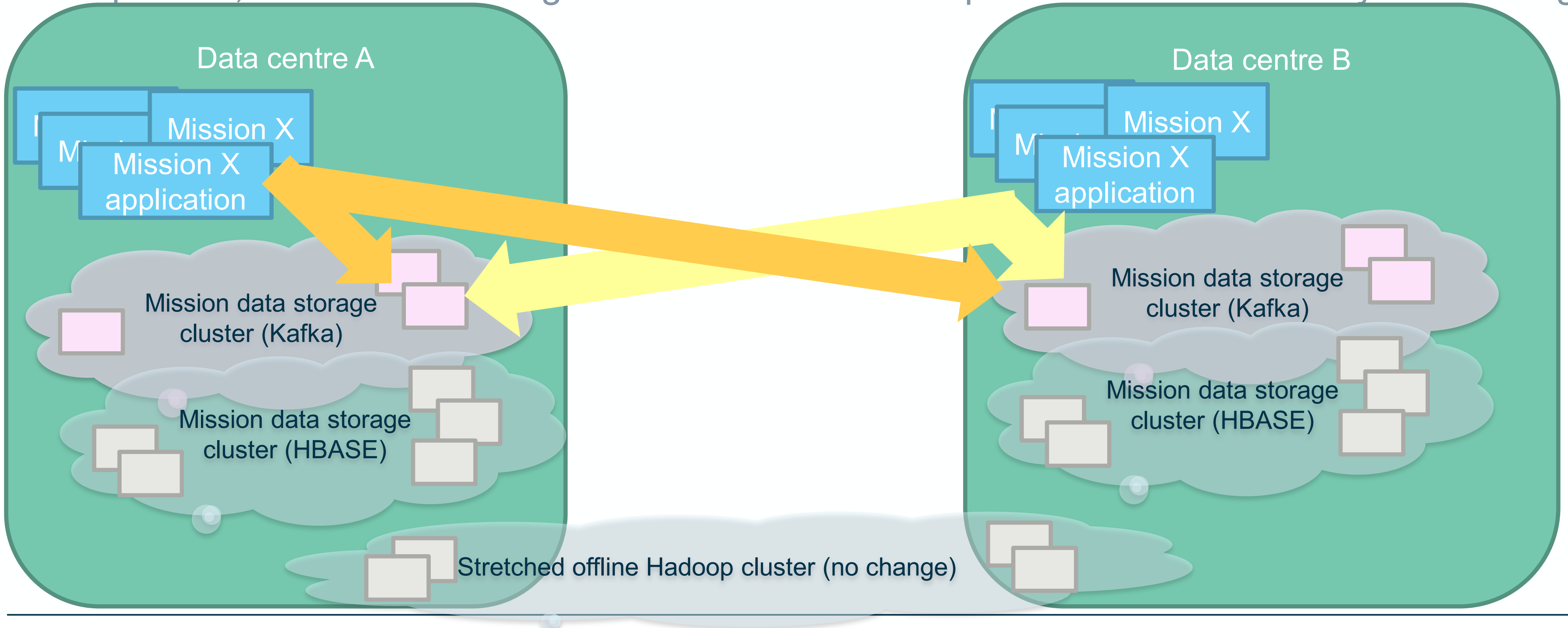
- Enable efficient data analysis and mining across different missions
- Provide a homogenous service and system evolution for all missions

- In line with the European Common Core requirements, HBASE was selected for the storage nodes.
- Kafka has been added on top to meet the performance requirements and allow for batchwise storing of data
- One cluster per data centre is envisaged to meet redundancy, disaster recovery and downtime requirements
- The previous offline data storage solution will be kept in parallel

Mission Operations in Transition – towards common stores (2)

With the new Mission Control System based on the European Common Core:

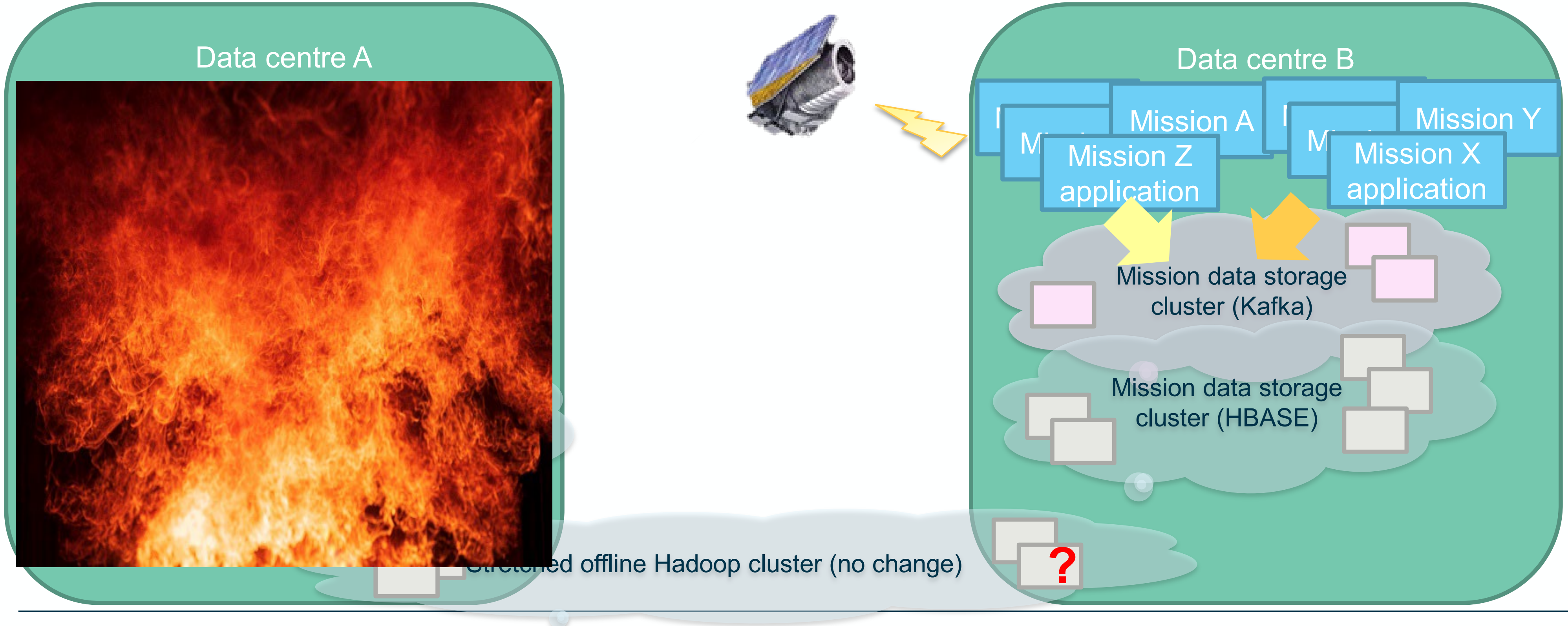
- missions will use shared infrastructure (transition from pets to cattle)
- rely on shared solutions for operational data storage with built-in high availability and redundancy
- data are fed simultaneously to two separate HBASE clusters in both data centres via Kafka databus
- In parallel, data are still being fed to the offline Hadoop cluster for offline analysis and long term storage



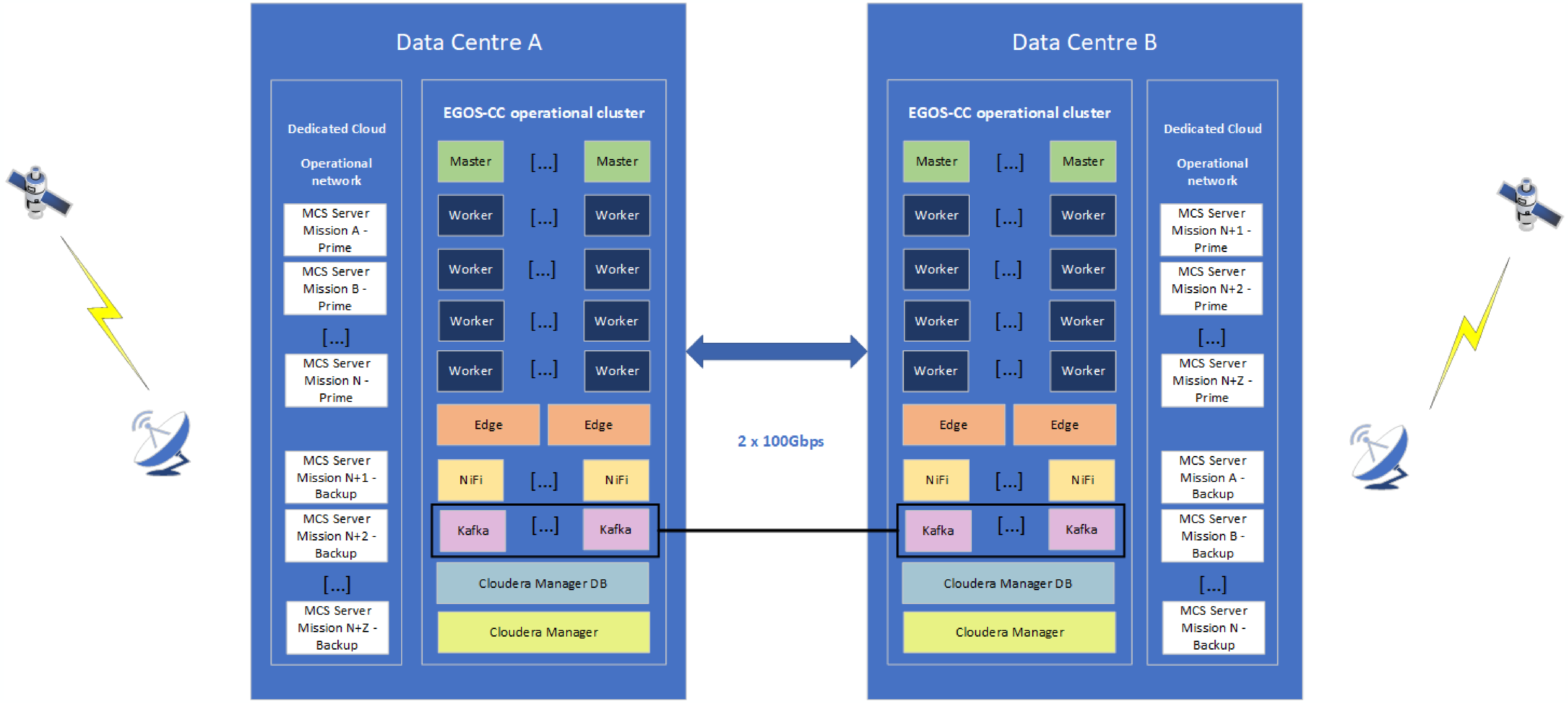
Up to 9 (!) replicas of some data

Mission Operations in Transition – towards common stores (3)

- Similar to the single island case, in case of one data centre not being available, the mission will start a new application instance in the other data centre within a short time
- Still unclear what would happen to the shared stretched cluster and if offline data analysis would be impacted (probably)

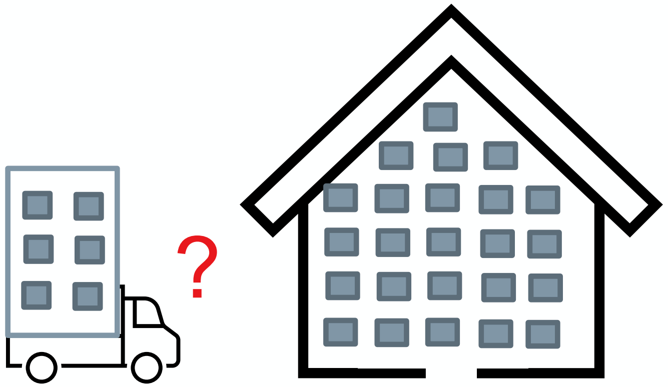


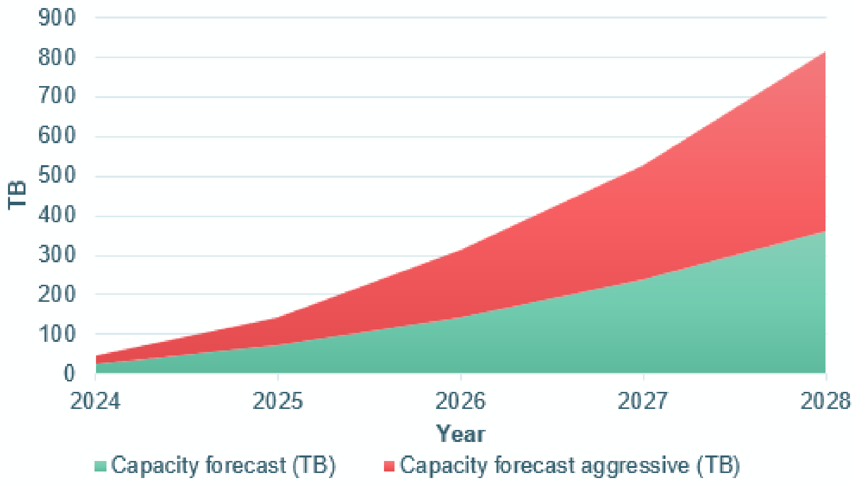
Shared solutions – Technical solution



Shared solutions - concerns & mitigations

- Loss of control for the individual mission
- **How to organise downtime** without impacting all/many missions? What about unplanned downtime?
- Data retention?
- Instead of fragmentation, we face big data problems, the storage solution needs to be scalable and support multiple missions without performance penalty
- Network capacity is becoming key
- **Cost concerns** – are shared solutions cost savers or rather the opposite?
- Big data storage solutions requires **lots of physical space** – can we go public cloud?
- **Do we need to keep everything?**





Shared solutions - concerns & mitigations (2)

- To reduce the storage footprint, it has been agreed to delete processed data on a rolling schedule from the operational storage clusters
  - Raw data will be kept for the duration of the mission, if needed, raw data can be reprocessed
  - Processed data (e.g. parameters) will be kept in the offline analysis cluster for the duration of the mission or longer
  - The shorter we need to keep processed data, the more missions we can support with the same number of nodes – can we agree on a sufficiently short retention period to avoid exploding the data centres?
- How can we justify the investment in hardware if we use the capacity only to store data? Do we foresee future exploitation which allows us to better use the processing capacity?
  - Interface/Federation with central data portal for enhanced data access and correlation with Science data
  - Augment capabilities of data processing layer (including machine learning techniques)
  - Introduce operationally validated AI and machine learning tools to operator teams
  - Create a curated "Anomaly Detection ESA Dataset" to allow to benchmark different anomaly detection approaches
  - …

New frontiers – data are kept not just for us …

We have mission needs to be able to perform deeper analysis at least during the mission, but beyond?

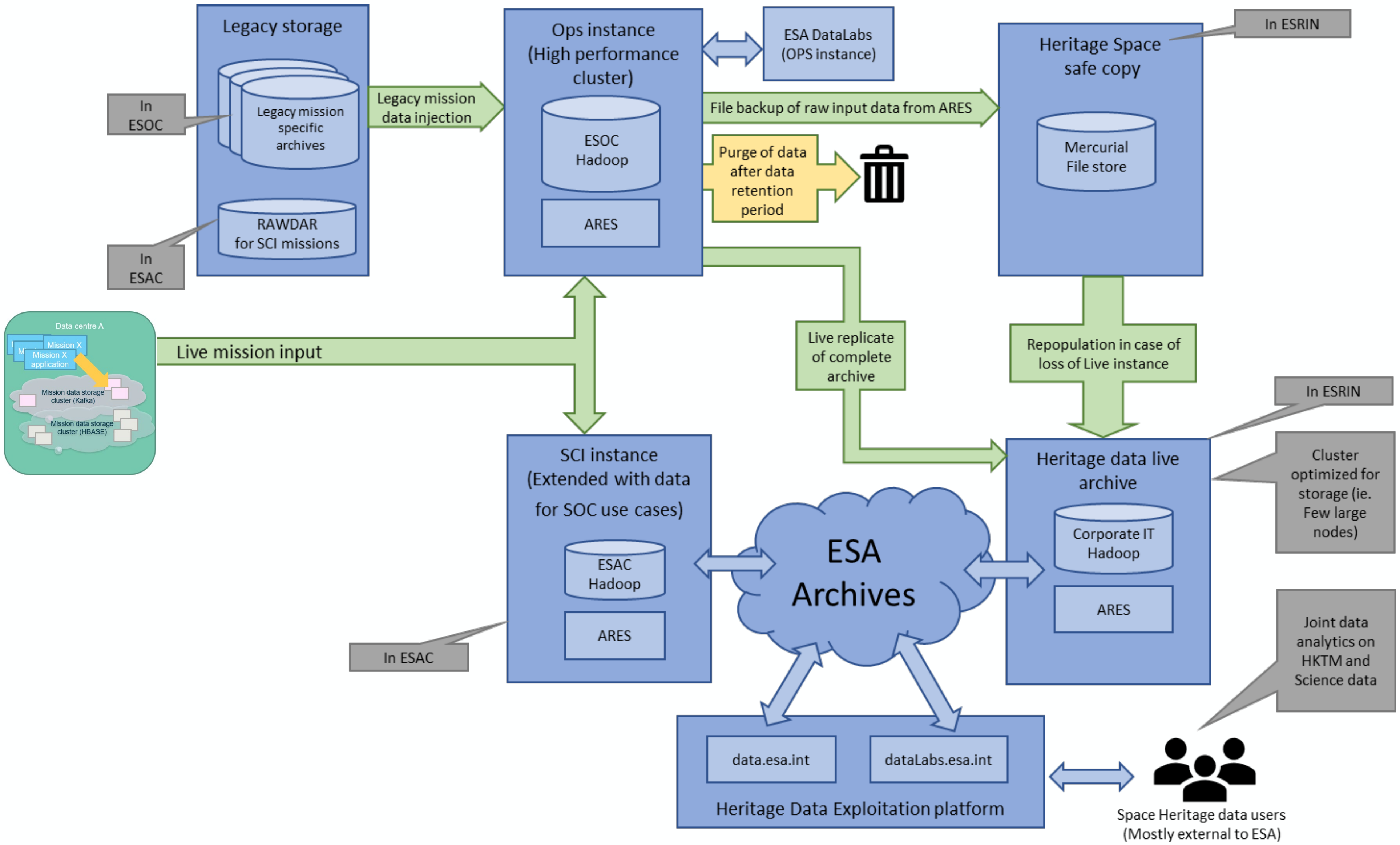
The fact is, we don’t know if all data we keep will be useful to someone tomorrow

We don’t even know to whom

But we see it as our heritage and we want to make it possible for others to decide for themselves, hence:

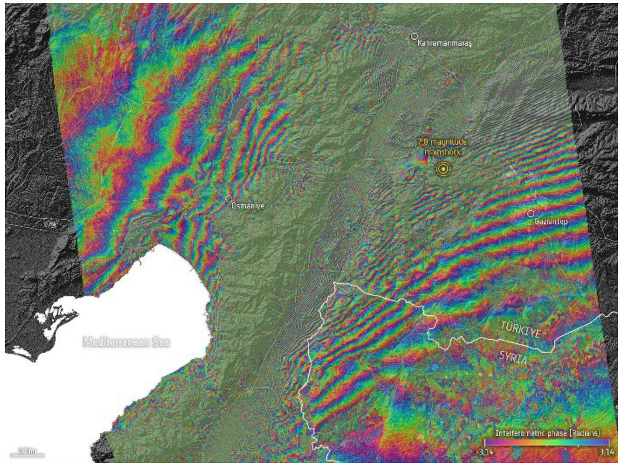
The Space Heritage Programme

New frontiers – data heritage beyond the centre



Summary

- Big data is inevitable, and we need to cope with it
- We need to improve how we store the data and where, to avoid running out of physical space and explode costs
- We need to improve in the areas of processing, value extraction and decision making to better benefit from our Big Data
- We need to make our data accessible for future users otherwise we keep it for nothing



Thanks for your attention!

Klara Widegård – European Space Agency