PUNCHLunch –

the online lunch seminar of the PUNCH4NFDI Consortium

Every second Thursday, 12:30-13:30 hrs: keynote and discussion

_____

Thursday, 30 Nov. 2023, 12:30

Tim Oelkers, Hermann Heßling (HTW Berlin)

Overview of methods and frameworks for storage of metadata in the petabyte range

Abstract

The huge data rates taken at large instruments, like the Square Kilometre Array Observatory (SKAO), need to be reduced significantly during the data acquisition phase. For an understanding and reproducibility of measurement results, it is necessary to store the multitude of details that define the (real-time) criteria for data selection. This is accompanied by a massive increase in metadata.

Dynamic Life Cycle Model

Traditional methods for storing metadata, such as SQL databases, are no longer suitable due to their poor scalability. Not only do they struggle to store petabytes, but they are also inflexible to schema changes, and their access time deteriorates as data size increases.
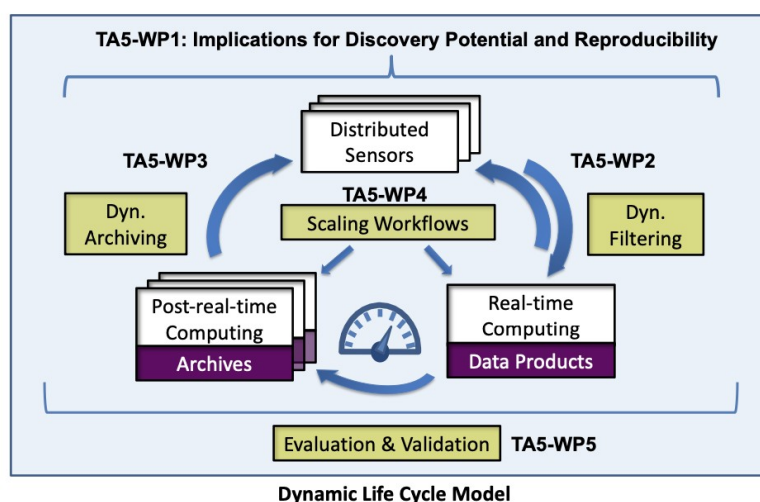
This talk provides an overview of selected challenges in metadata storage and analysis and discusses whether frameworks such as Lustre, Apache Parquet, Greenplum and Delta Lake could be useful.

Connection details:
ZOOM Meeting "PUNCHLunch seminar": https://indico.desy.de/event/42653/
Webinar ID: 919 1665 4877, passcode: 481572
Next event: News from PUNCH4NFDI: Education of students, December 14 2023

_____

Connect to PUNCH4NFDI:

Twitter: @PUNCH4NFDI   Mail: punch4nfdi@desy.de   Web: www.punch4nfdi.de