The SKAO Engineering Data Archive: From Basic Design to Prototype Deployments in Kubernetes

Thomas Juerges, SKA Observatory, Jodrell Bank, United Kingdom
Aditya Dange, Tata Consultancy Services, Pune, India

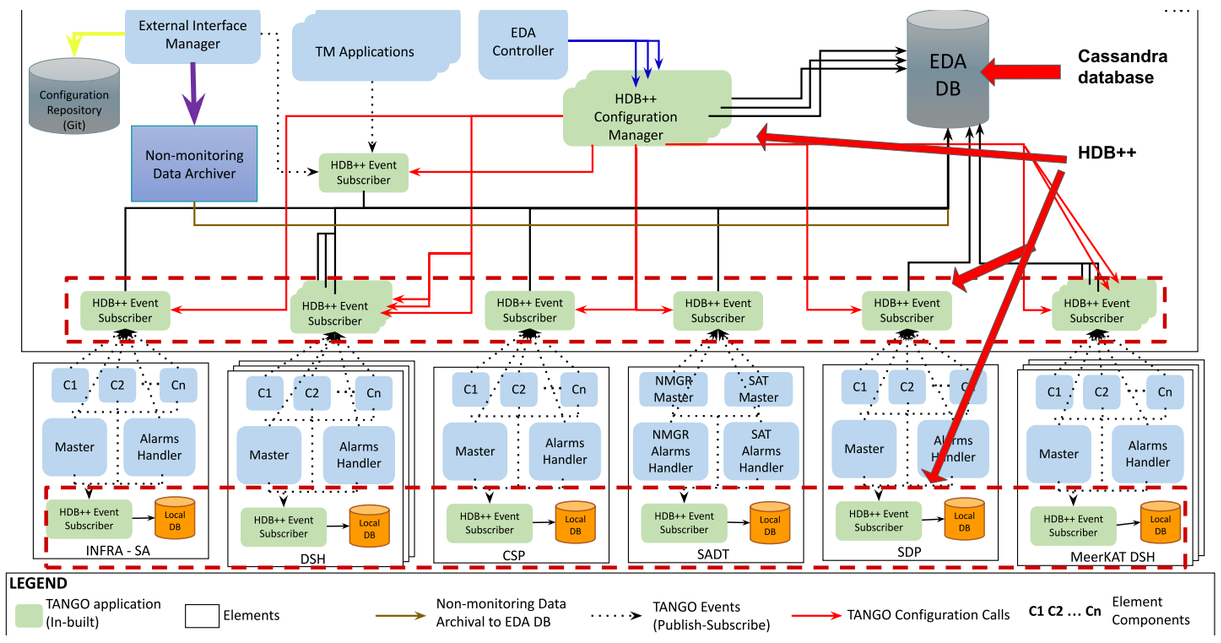
During its construction and production life cycles, the Square Kilometre Array Observatory (SKAO) will generate non-scientific, i.e. engineering, data. The sources of the engineering data are either hardware devices or software programs that generate this data. Thanks to the Tango Controls software framework, the engineering data can be automatically stored in a relational database, which SKAO refers to as the Engineering Data Archive (EDA). Making the data in the EDA accessible and available to engineers and users in the observatory is as important as storing the data itself. Possible use cases for the data are verification of systems under test, performance evaluation of systems under test, predictive maintenance and general performance monitoring over time. Therefore we tried to build on the knowledge that other research facilities in the Tango Controls collaboration already gained, when they designed, implemented, deployed and ran their engineering data archives. SKAO implemented a prototype for its EDA, that leverages several open-source software packages:

- with Tango Controls' HDB++
- the Timescale time series database
- and Kubernetes at its core.

In our overview we will answer the immediate question "But why do we not just do, what others are doing?" and explain the reasoning behind our choices in the design and in the implementation.
Keywords: TANGO, reporting, data visualisation, open source, SCADA

The Past

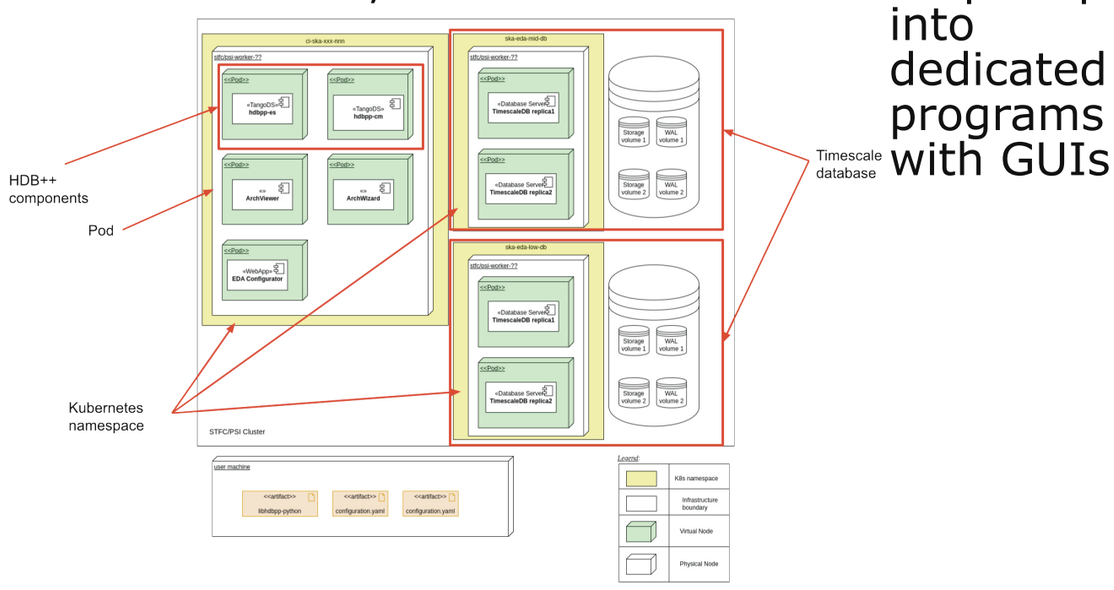
SKAO's first design for an Engineering Data Archive (EDA) focussed on addressing functional



requirements. Deployment aspects were intentionally not considered. Key components: Tango Controls HDB++ and Cassandra database backend.

The Present

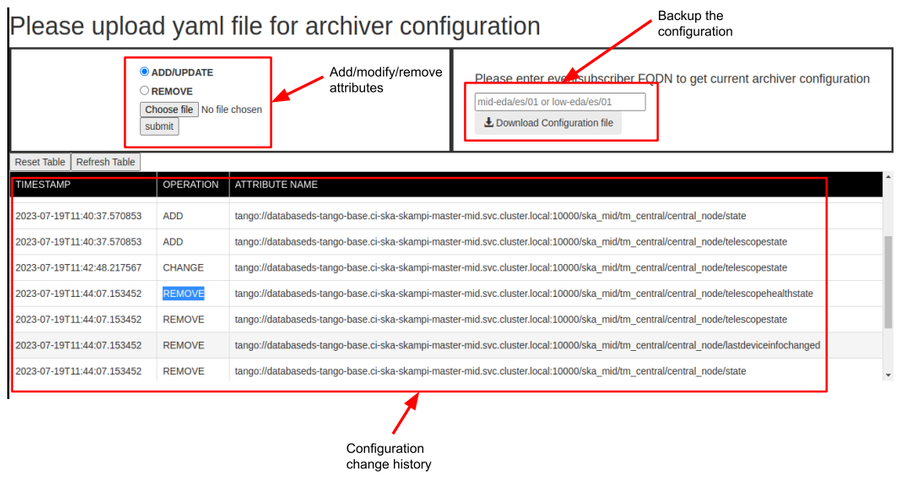
After consultation with other facilities (SKAO invited to "Learning from others" sessions) design changes were made. Timescale replaced Cassandra as the database backend, the EDA Controller was split up into dedicated programs with GUIs



(Configurator, ArchWizard, ArchViewer). The software would be deployed as containers in Kubernetes with Helm charts.

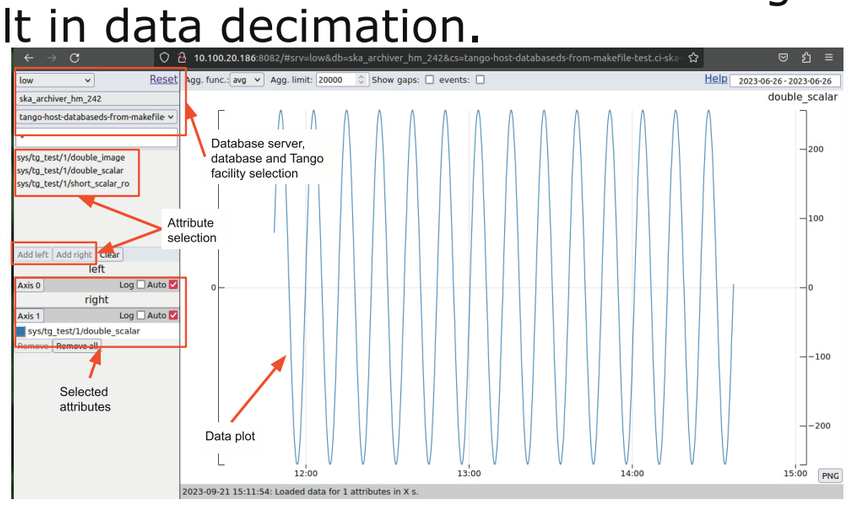
New tools

The Configurator allows to upload and download the configuration of which Tango Attributes get archived and how often that is done.



The addition of the ArchiveViewer GUI provides a web based and user friendly data access method that can leverage Timescale's built in data decimation.

Data retrieval on the command line is now possible with the pyhdbpp module.



User reactions after the first deployments:
No bigger issues when installing the EDA thanks to the Helm charts that are provided. But: Users struggle where the documentation is not explicit enough.

Lessons learned: Compare an initial design with how others have designed similar products. Look at technological advancements and do not be afraid to replace something. Users will always find weaknesses in the documentation.