System Design for the Effelsberg CryoPAF Backend

Niclas Esser
nesser@mpifr-bonn.mpg.de
Max-Planck Institute for Radio Astronomy (MPIfR)

Phased Array Feeds and Advanced Receivers
Sydney, Australia, November 2022

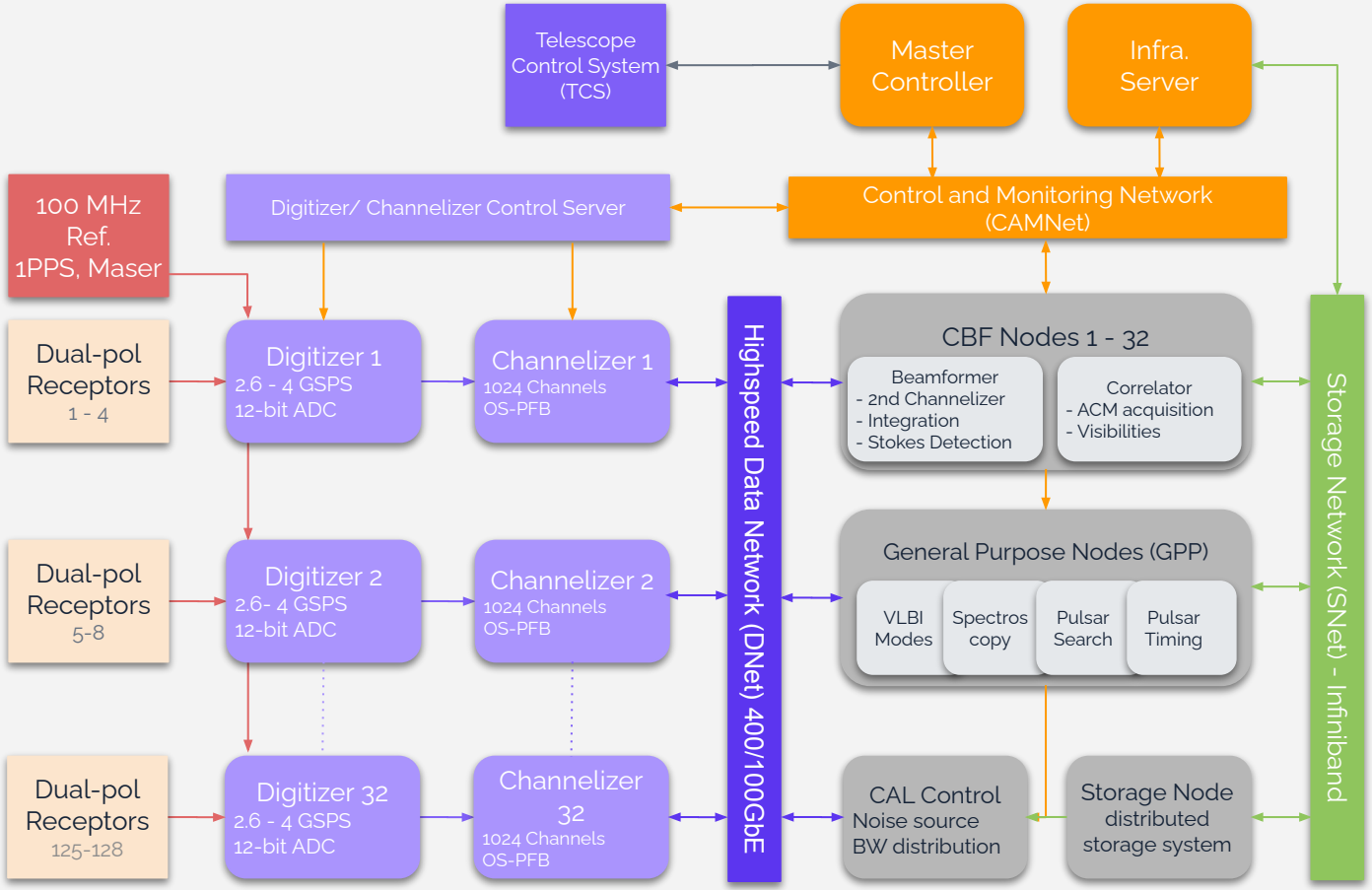
What is the Effelsberg CryoPAF Backend?

- Processes PAF data to science-ready products
  - VLBI
  - Spectroscopy
  - Pulsar timing & search
- High-performance computing
  - Commercial off-the-shelf (COTS) hardware
  - Accelerators GPU & FPGA
  - High speed networking
- Follows the EDD backend approach
  - Containerized services
  - System deployment & provisioning

System Overview

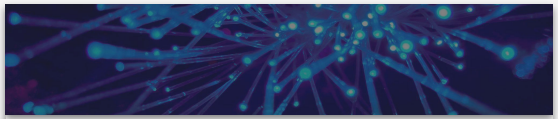
Key Specs

- 2x 128 PAF rezeptors
  - Dual-polarized
  - cryogenically cooled
  - 2.4-4 GHz (S-band)
- Digitization
  - Expected 2.6 GSPS
  - Maximum 4 GSPS
- Beamformer
  - 128 beams (integrated)
  - 2 x 16 dual-pol beams



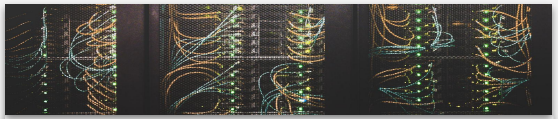
CAMNet & SNet

CAMNet



- Control and monitoring interactions between backend components
- Existing infrastructure (1 GbE)
- Used protocols
  - KATCP
  - SSH
  - HTTP
- System deployment
- System provisioning

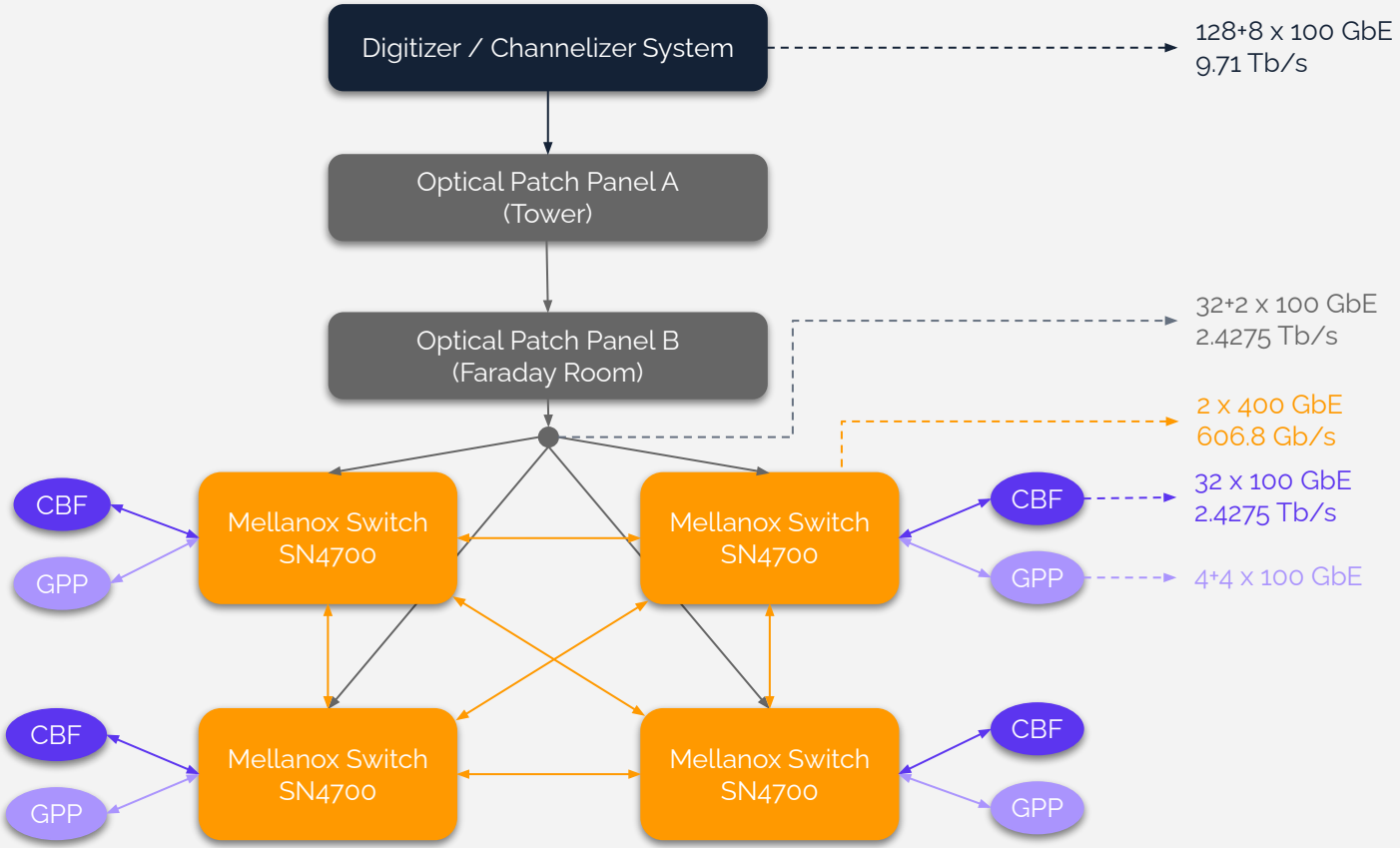
Storage Network & BeeGFS



- Fast mountable file system
- Access from all backend nodes
- Infiniband (200 Gb/s)
- Spinning disks
  - 400 MiB/s write speed
  - 18TB
- Up to 60 disks per node
  - 23.3 GiB/s w/r speed (IB bottleneck)
  - 1.08 PB

High Speed Data Network (Dnet)

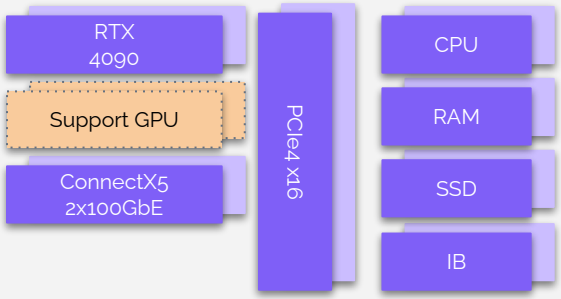
- 4 x Mellanox SN4700
  - 100/200/400GbE
  - 128/ 64 / 32 Ports
- SPEAD Protocol
  - RDMA with ibv
  - Multicasting
- Network Topology
  - fully-connected
  - flexible band subscription (CBF)
  - Stream MUXing



Correlator & Beamformer Node

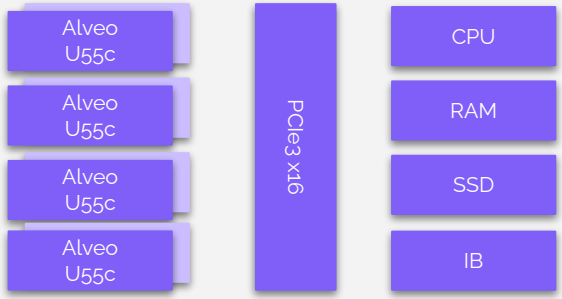
Server Setup

GPU Concept



- 2 NUMA node server system
- 2x100 Gb/s per NUMA
- Correlator uses Tensor Core Library [J. Romeln]
- Tensor Core based beamformer

FPGA Concept



- 8x Alveo U55c cards per server
- 2x 100 Gb/s interface per card
- Firmware written in HLS (Y. Men)
  - Beamformer in place
  - Correlator pending

Correlator & Beamformer Node

Comparison

Both Concepts are still in evaluation

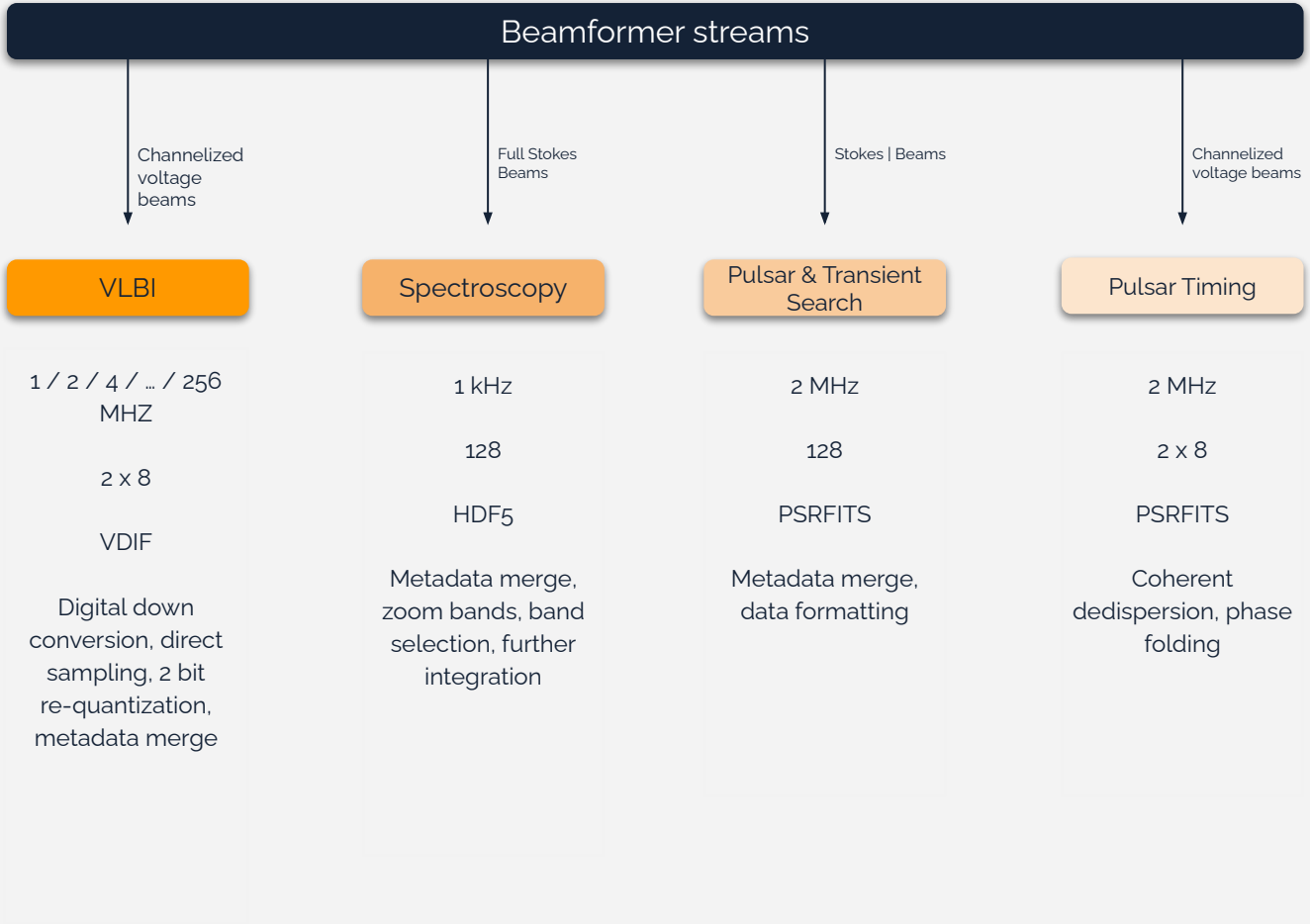
The FPGA concept is outstanding when a single card is able to process the data of both 100 GbE interfaces (not verified yet). Otherwise the GPU concept is preferred, because of its flexibility and general purpose.

	XILINX	NVIDIA GEFORCE RTX
Pro	• Low latency • stand-alone device • Energy saving	• Fast prototyping (PoC) / General Purpose • Flexibility and less devel effort • Tensor Core promising technology for CBF processing
Contra	• High development effort • less flexibility • Software maintenance	• Batch processing -> latency • higher energy costs • Availability

	XILINX		NVIDIA GEFORCE RTX	
	Expected	**Full scale**	**Expected**	**Full scale**
Power	16 kW 17520 kWh/a (12.5%)	25.6 kW 28032 kWh/a	36 kW 39420 kWh/a (12.5%)	62 kW 67890 kWh/a
Size	40 HU; 1.78 m	64 HU; 2,85 m	72 HU; 3.20 m	124 HU; 5.52 m
Est. cost	€ 386.999	€ 619.198	€ 310.482	€ 534.719

Science Modes on the GPP Nodes

The EDD backend already supports various processing pipelines for different science modes.



EDD Pipeline Implementation

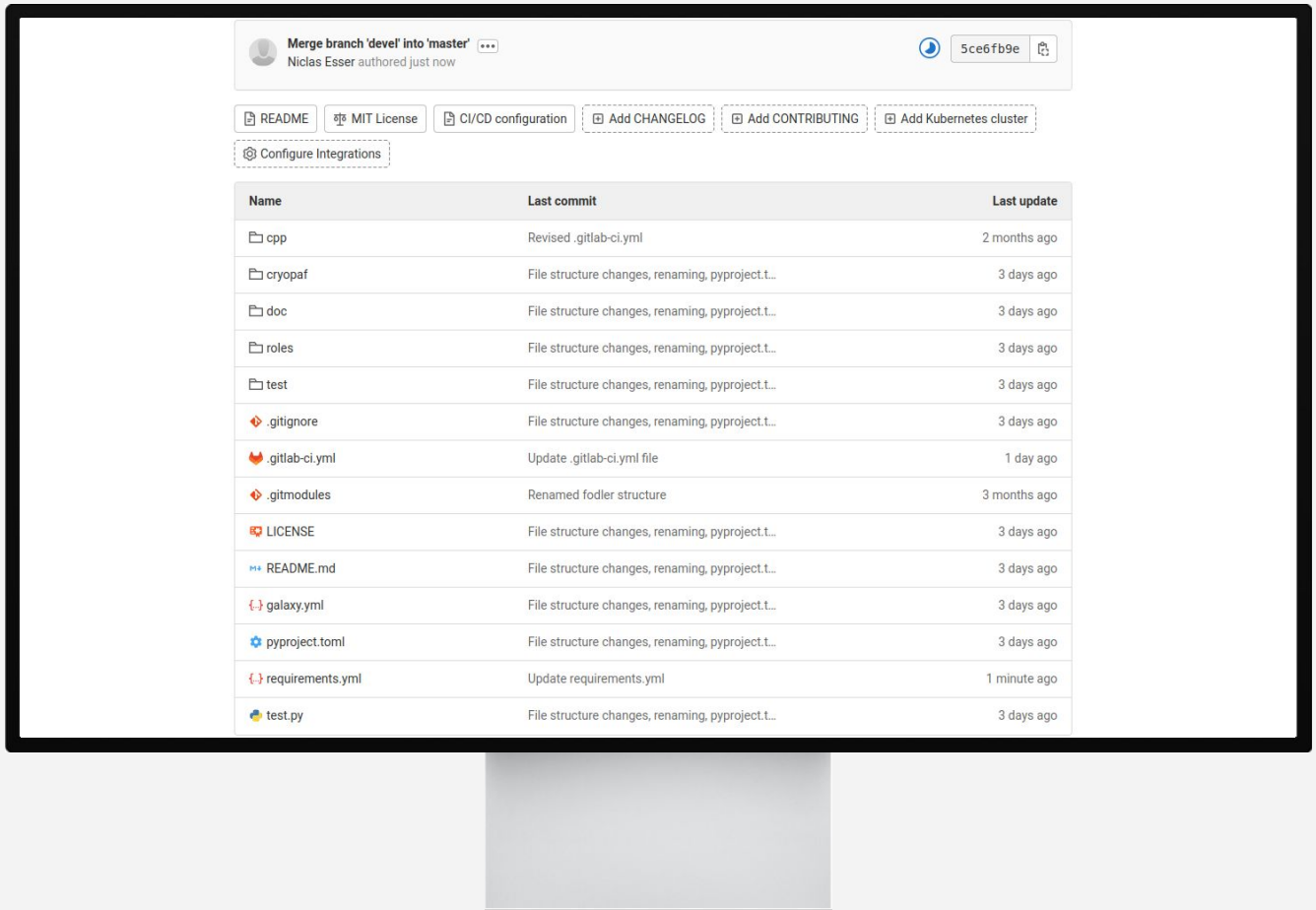
All shown pipelines are implementations of the EDDPipeline class and follow the EDD state model. They are also defined as Ansible roles to provide dockerized services.

Under the hood the pipelines make use of CLI-programs

e.g. psrdada, mkrecv, mksend, dspsr

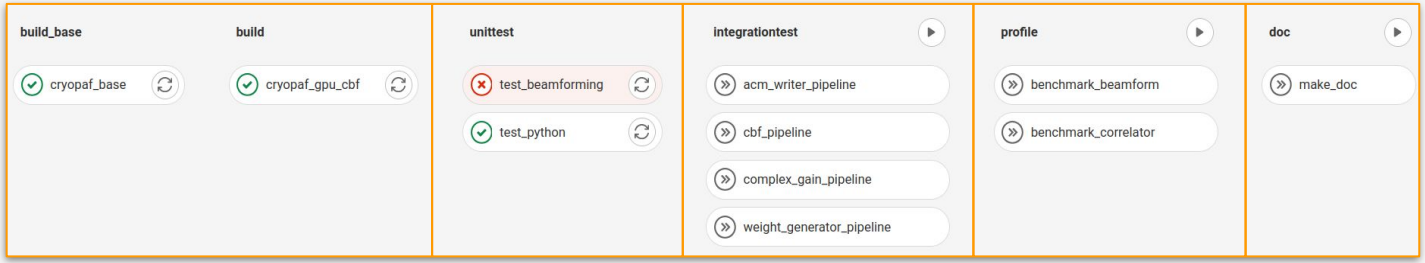
Just a plugin for the EDD framework

- HPC code / CLIs in `cpp/`
- EDD Pipeline implementations, utilities & scripts in `cryopaf/`
- Definition of the services in `roles/`
- Ansible collection defined by `galaxy.yml` and `requirements.yml`
- Makes use of other EDD plugins (e.g. VLBI)



Testing & CICD

- Using the GitLab runner
- Define the CI pipeline
- Executed on commits and on a daily basis
- Direct feedback with test driven development



Build docker images (with dependency)	testing units (e.g. GPU-kernels, utility module)	Test pipelines by injecting test data sets, compare against reference	Benchmarking of performance critical code	Create the documen-tation

Third party software
EDD infrastructure tooling

Orchestration Software for EDD Backend Deployment

Versioning System, CI/CD Pipelines, Documentation, Software Management

Large, fast & distributed storage system used for high speed file IO-operations

Containerized & isolated EDD services e.g. processing pipelines, controllers

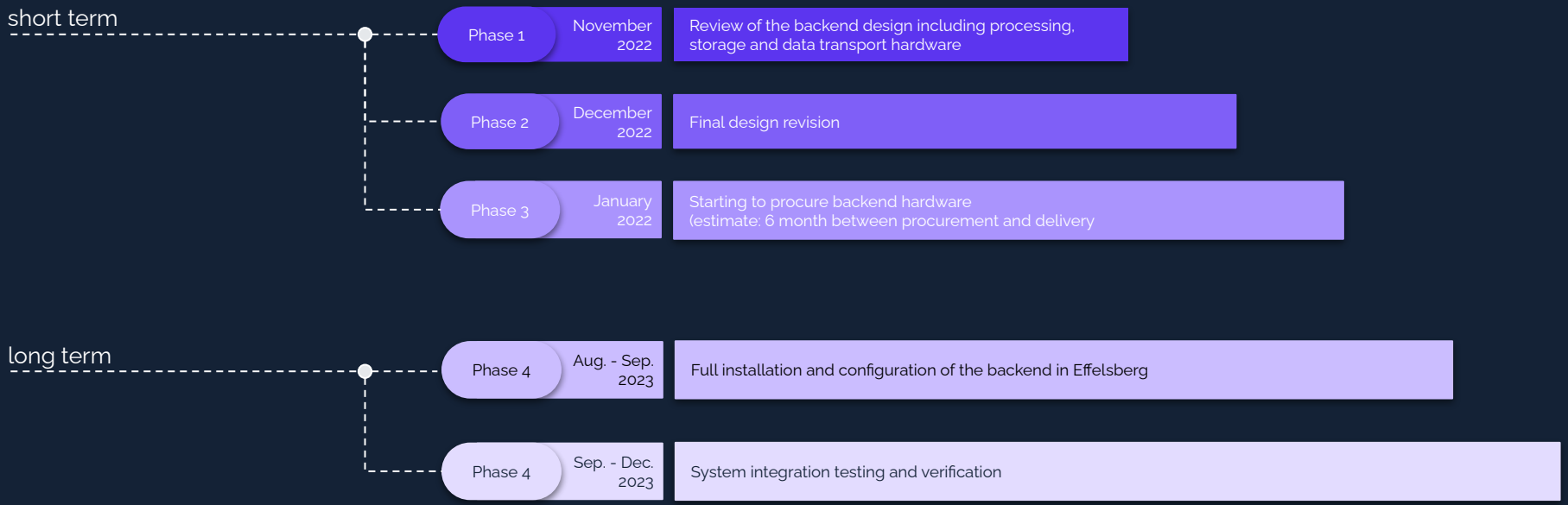
Long-term storage database for sensor data of individual EDD components

Database for volatile metadata e.g. product information, telescope data

Alerting system

Monitoring system for sensor data, system status, configuration

Outlook


short term
Phase 1
November 2022
Review of the backend design including processing, storage and data transport hardware
Phase 2
December 2022
Final design revision
Phase 3
January 2022
Starting to procure backend hardware (estimate: 6 month between procurement and delivery
long term
Phase 4
Aug. - Sep. 2023
Full installation and configuration of the backend in Effelsberg
Phase 4
Sep. - Dec. 2023
System integration testing and verification

Summary

- Proposed the system design for the CryoPAF backend @ Effelsberg
  - Networking
  - Processing nodes
  - Science modes and data products

- We do not reinvent the wheel, but we improve it
  - reuse of existing EDD soft- & hardware
  - comprehensive third party tooling
  - test driven development

References

[1] https://www.mpifr-bonn.mpg.de/7081564/digitizer-and-channelizer-for-the-cryopaf
[2] https://gitlab.mpcdf.mpg.de/nesser/paf_edd/
[3] https://gitlab.mpcdf.mpg.de/mpifr-bdg/mpikat
[4] https://gitlab.mpcdf.mpg.de/mpifr-bdg/edd_provisioning
[5] https://git.astron.nl/RD/tensor-core-correlator

Thank you
for your attention

Niclas Esser
nesser@mpifr-bonn.mpg.de
Max-Planck Institute for Radio Astronomy (MPIfR)

Phased Array Feeds and Advanced Receivers
Sydney, Australia, November 2022