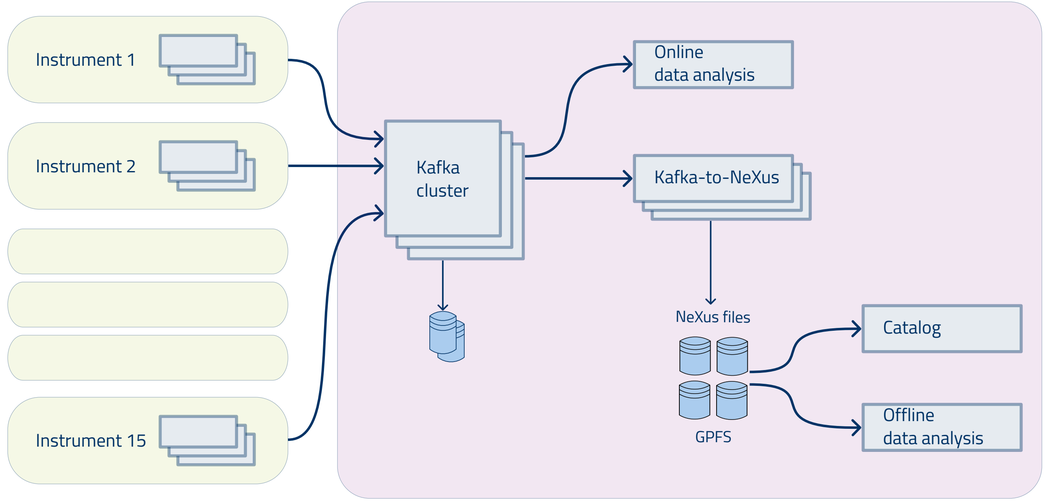
Experiment data streaming and aggregation at the European Spallation Source

Daniel Cacabelos, Afonso Mukai, Morten Christensen, Tobias Richter
European Spallation Source ERIC, Lund, Sweden

Architecture

Each of the fifteen instruments at the European Spallation Source will produce a high-throughput stream of neutron detection signals and additional data from comparatively slower sources like EPICS control devices.

The data aggregation and streaming system provides data collection, processing and persistence in a common platform shared by all instruments, with each instrument using its own set of adapters to access the system [1].



Data streaming architecture. Arrows represent the information flow.

The backbone of the system is composed of an Apache Kafka cluster, a distributed streaming platform that provides a consistent data processing interface, while decoupling data producers and consumers from each other.

Data analysis tasks can be performed either *offline* from NeXus files, or *online* by accessing the Kafka streams directly.

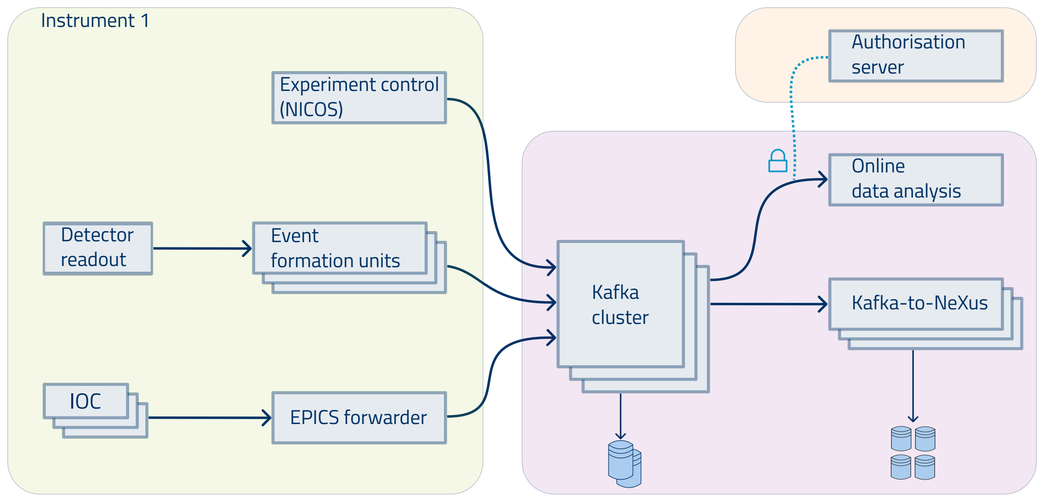
Components

Data collection

- **Event Formation Units** receive signals from detectors and consolidate them into discrete *neutron events*.
- **EPICS Forwarder** collects and forwards metrics from EPICS devices like choppers and motion controllers.
- **Experiment Control (NICOS)** is the experiment orchestrator, but also produces data like derived parameters and user-provided metadata.

Persistence

- **Kafka-to-NeXus** combines the separate data streams of an experiment into a single HDF5 file using the domain-specific NeXus format.



Data producers and consumers for each instrument

Access control

Authentication and authorisation mechanisms are used to grant access to individual streams within Kafka.

Critical applications make use of SASL/SCRAM, a challenge-response method to authenticate a client without the need of a third-party, as avoiding the additional third-party authoriser reduces their availability risk.

Other applications will make use of SASL/OAUTHBEARER, a method to authenticate users or applications via a third-party system trusted by Kafka.

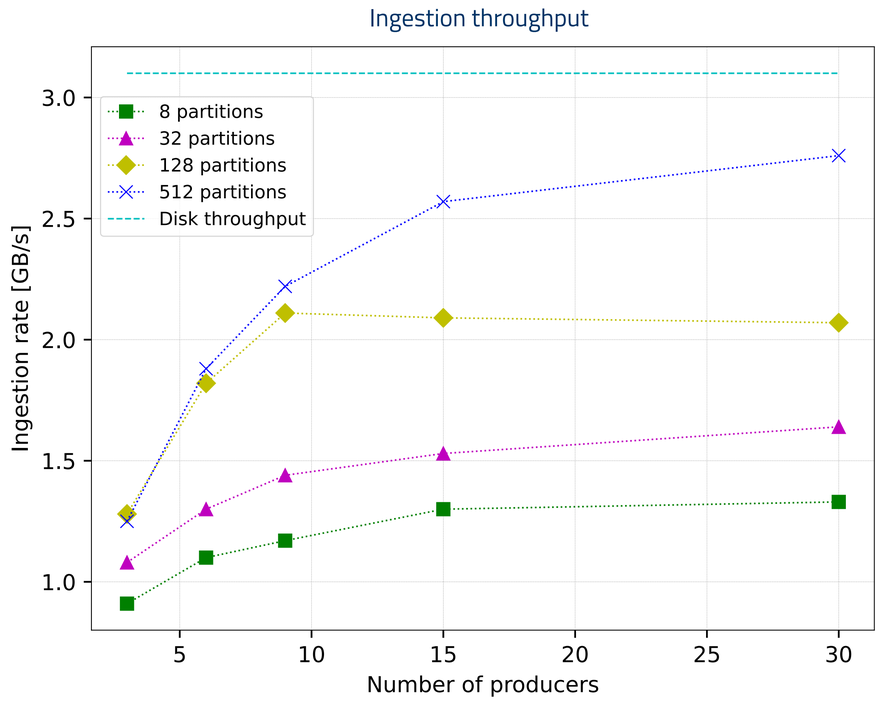
Neutron event ingestion

At each instrument, readouts from a detector are sent to one or more Event Formation Units (EFU) at a high rate via a 100 Gbit Ethernet link.

The EFUs consolidate separate signals into discrete *neutron events*, effectively reducing the data rate. Current estimations for the EFU output range between 0.1 and 4 Gbytes/s, depending on the instrument [2].

Benchmarks

Ingestion throughput was measured by generating synthetic data from the EFUs with different stream and producer configurations. The results are compared against the maximum available disk throughput.



The results highlight the large effect of stream partitioning. Large numbers of partitions for a single stream increase the average I/O utilisation per disk, resulting in higher overall throughput.

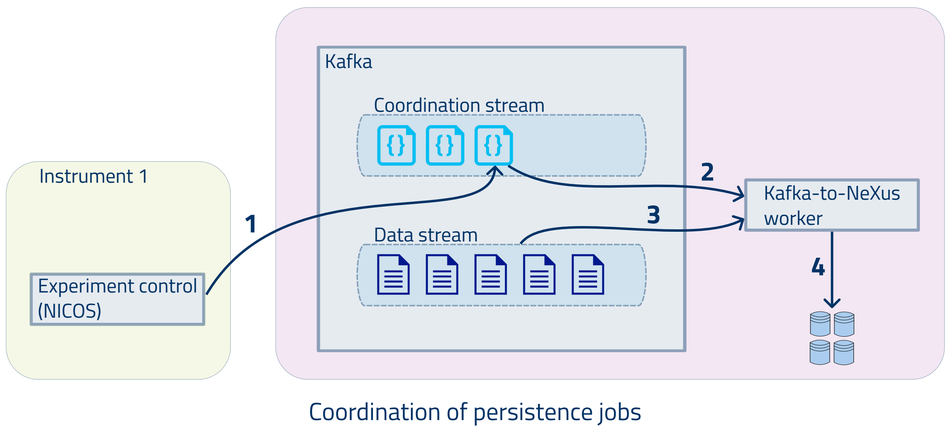
For large partition numbers and highly concurrent producer workloads, the ingestion rates reach values close to 90% of the available disk throughput, demonstrating the low overhead added by the processing layer.

Persistence orchestration

The long-term storage of experiment data is performed by combining several continuous data streams into a single HDF5 file in the NeXus format.

At each instrument, Experiment Control is responsible for initiating persistence jobs by specifying required parameters like stream identifiers, data schemas and the time window of the experiment.

The orchestration of this process is performed via dedicated Kafka streams where a pool of Kafka-to-NeXus workers pick persistence jobs in a coordinated manner.



Coordination of persistence jobs

References

[1] Mukai, A. H. C., et al. "Architecture of the data aggregation and streaming system for the European Spallation Source neutron instrument suite." Journal of Instrumentation 13.10 (2018): T10001
[2] Christensen, M. J., et al. "Software-based data acquisition and processing for neutron detectors at European Spallation Source - early experience from four detector designs." Journal of Instrumentation 13.11 (2018): T11002

Testing environment

The platform is currently operating at the Ymir test beamline, an end-to-end hardware and software deployment at ESS where components are integrated and tested.

6x Kafka brokers: 128 GB ram, 32 cores, 8 SATA 7200rpm (JBOD), 100 Gbit Ethernet.
3x Event Formation Units: 64 GB ram, 20 cores, 10 Gbit Ethernet.
Producer block size: 1 MB.