Contribution ID: **98**   Type: **Poster**

Fusion Data Platform for HL-3

Wednesday, 17 July 2024 15:40 (1h 30m)

The HL-3 Fusion Big Data Platform is a system developed on the open-source Hadoop platform specifically tailored for processing tokamak experimental data. Unlike traditional big data platforms dealing with service data periodically, massive amounts of data are generated by tokamak experiments typically within seconds or minutes. And these data are mostly transmitted and stored in binary format.

In this context, the HL-3 team has researched and developed a big data platform suitable for handling fusion experiment data from tokamak devices. This platform seamlessly integrates with existing tokamak data acquisition systems and database systems, effectively parsing, cleaning, and converting binary data into formats readily processable by downstream applications, while meeting the time response requirements of tokamak researchers for data processing.[1]

The Data Source component is comprised of three parts: real-time experiment data collected during tokamak discharges (e.g., coil voltage, current), engineering-related data associated with the tokamak device (e.g., device dimensions, temperature variations of the tokamak walls during experiments), and video and audio data captured during the experiment (e.g., infrared camera data of the discharge process).

The Data Integration section primarily utilizes data acquisition tools to periodically retrieve data files from a file server or read real-time experimental data from a high-speed cache.

The Data Process stage utilizes batch computation engine MapReduce and stream processing engines Spark Streaming/Flink to process data according to various service logics, subsequently storing the processed data in HDFS or Ceph as per specified requirements.

The Data Service component currently serves two primary scenarios: calculating physical metrics for scientific research by physics data analysts, and deriving basic feature data for AI developers to use in AI model training.

Speaker's Affiliation

Southwestern Institute of Physics, Chengdu, China

Member State or IGO

China, People's Republic of

Primary author: SUN, Xiang (Southwestern Institute of Physics)

Presenter: SUN, Xiang (Southwestern Institute of Physics)

Session Classification: Poster Session

Track Classification: Data Storage and Retrieval, Distribution and Visulaization