Contribution ID: **137**   Type: **Sectional**

Data streams processing in metadata integration system for HENP experiments

Friday, 4 October 2019 12:50 (15 minutes)

Nowadays, heterogeneous metadata integration has become a widespread objective. Whenever it is addressed, there are numerous tasks to be solved, such as data sources analysis and storage schema development. No less important one is the development of automated, configurable and highly manageable ETL (data Extraction, Transformation, and Load) processes, as well as the creation of tools for their automatization, scheduling, management, monitoring.

This work describes the Metadata Integration and Topology Management System, initially designed as a subsystem of the Data Knowledge Base (DKB) developed for the ATLAS experiment. The core idea of the subsystem is to separate the common features of the majority of ETL-processes from the implementation of particular tasks. It is implemented as standalone modules: supervisor and workers; a supervisor is responsible for data streams building through workers that implement a set of specific operations for a particular process. The system is intended to considerably facilitate the organizing of ongoing data integration operations with automated data stream processing.

Summary

This work describes the Metadata Integration and Topology Management System, initially designed as a subsystem of the Data Knowledge Base (DKB) developed for the ATLAS experiment.

Primary author: KAIDA, Anastasiia (National Research Tomsk Polytechnic University, School of Computer Science & Robotics)

Co-authors: Ms GRIGORIEVA, Maria (Lomonosov Moscow State University); Mrs GOLOSOVA, Marina (National Research Center "Kurchatov Institute"); Mr AULOV, Vasiliy (NRC Kurchatov Institute)

Presenter: KAIDA, Anastasiia (National Research Tomsk Polytechnic University, School of Computer Science & Robotics)

Session Classification: Computing for Large Scale Facilities (LHC, FAIR, NICA, SKA, PIC, XFEL, ELI, etc.)

Track Classification: Computing for Large Scale Facilities (LHC, FAIR, NICA, SKA, PIC, XFEL, ELI, etc.)