

Databases for large-scale science

Thursday, August 30, 2018 1:00 PM (5 hours)

In this workshop, the students will (a) learn how to efficiently use relational and non-relational databases for modern large-scale scientific experiments, and (b) how to create database workflows suitable for analytics and machine learning.

First, the focus of the workshop is to teach efficient, safe, and fault-tolerant principles when dealing with high-volume and high-throughput database scenarios. This includes, but is not limited to, systems such as PostgreSQL, Redis, or Elasticsearch. Topics include query planning and performance analysis, transactional safety, SQL injection, and competitive locking.

Second, we focus on how to actually prepare data from these databases to be usable for analytics and machine learning frameworks such as Keras.

An intermediate understanding of Python, SQL, and Linux shell scripting is recommended to follow this course. An understanding of machine learning principles is not required.

Summary

Presenter: LASSNIG, Mario (CERN)

Session Classification: Tutorials