JIT query engine with adaptive indexing and partitioning

Keywords: In-situ query processing, LLVM, Adaptive Indexing

Problem: Scientists in many disciplines generate massive amounts of data. The amounts of data produce are already so big that they can barely be managed. In-situ query processing[1], offers flexibility and interactivity for scientists to query straight over raw data files and offers performance comparable to state-of-the-art relational DBMS. To further improve the performance of in-situ query processing and allow efficient query execution over multiple different file formats scientists have proposed just-in-time code generation of scan operators incorporating the logic of query execution[2]. In addition, by incorporating an adaptive partitioning and indexing approach over the raw data files the query engine can take advantage of the underlying data distribution without incurring any additional execution costs.

Project: This project has as goal to integrate adaptive partitioning and indexing over raw data files with the JIT execution engine to offer a full query engine offering indexing over any data format.


Plan:

1. Get acquainted with the literature
2. Understand the code of RAW as well as the LLVM framework
3. Implement the necessary data structures
4. Run thorough experimental evaluation to understand the performance gains.

Supervisor: Prof. Anastasia Ailamaki, anastasia.ailamaki@epfl.ch
Responsible collaborator(s): Matthaios Olma, matthaios.olma@epfl.ch
Manos Karpathiotakis, manos.karpathiotakis@epfl.ch

Duration: 3 months