Data-Driven Scientific Computing

The goal of this area is to offer a simple and efficient data system. Currently we focus on big data scientific applications. Our main interests are: exploit novel hierarchical storage systems, design a data interface independent of the data system and support efficient multidimensional queries.

Summary

The goal of this area is to design resource management strategies for big data applications, defining policies that enable distributed data stores to meet high-level performance goals. We focus on scientific applications, like those from life science domain, which data generation and accesses are both bound precision and performance. Currently, the main threads of our work are:

- Propose resource management strategies that are able to exploit novel hierarchical storage systems
- Hecuba: this project aims to design and develop strategies to facilitate programmers the efficient usage of data stores for big data applications. For example, define simple interfaces that are independent from the particular type of data system.
- Qbeast: this project consists on designing and implementing a distributed system to provide to multidimensional queries an efficient and reliable indexing mechanism

Objectives

- To create a set of tools that boosts the performance of the storage system for big data scientific applications and that hides the internals of the data systems to programmers.