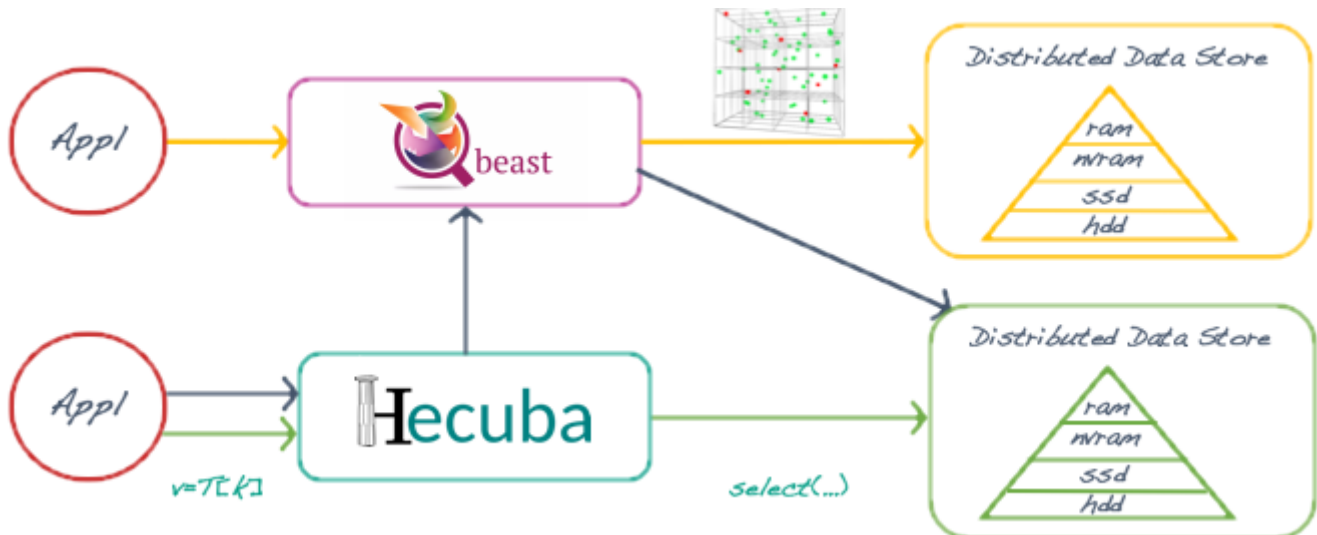


## Data-Driven Scientific Computing



Providing big data scientific applications with a simple and efficient data system

## Summary

The goal of this research line 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 bounded by 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 subproject aims to design and develop strategies to offer programmers a simple interface for efficient usage of data stores for big data applications. The interface implemented by Hecuba is independent of the underlying data stores; this way, Hecuba relieves the programmer from understanding the intrinsics of the data stores systems.

**Qbeast:** The goal of this subproject is to design and implement a highly scalable multidimensional indexing system that provides NoSQL databases with an efficient indexing and sampling mechanism. As part of this project, we are exploring new fields that can benefit from these features as, for example, machine learning applications.

**BD4HPC:** The goal of this subproject is to develop a set of tools that automatise the deployment and execution of BigData management software on HPC execution environments.

## 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.

**Source URL (retrieved on 15 ago 2020 - 00:55):** <https://www.bsc.es/ca/research-development/research-areas/big-data/data-driven-scientific-computing>