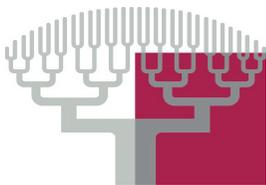


Data Series Analysis of Big Data

Discovering the Hidden Treasures

université
PARIS
DIDEROT
PARIS 7



Volker Beckmann
Sophie Bernard
Themis Palpanas
Cécile Cavet et al.

USPC

Université Sorbonne
Paris Cité



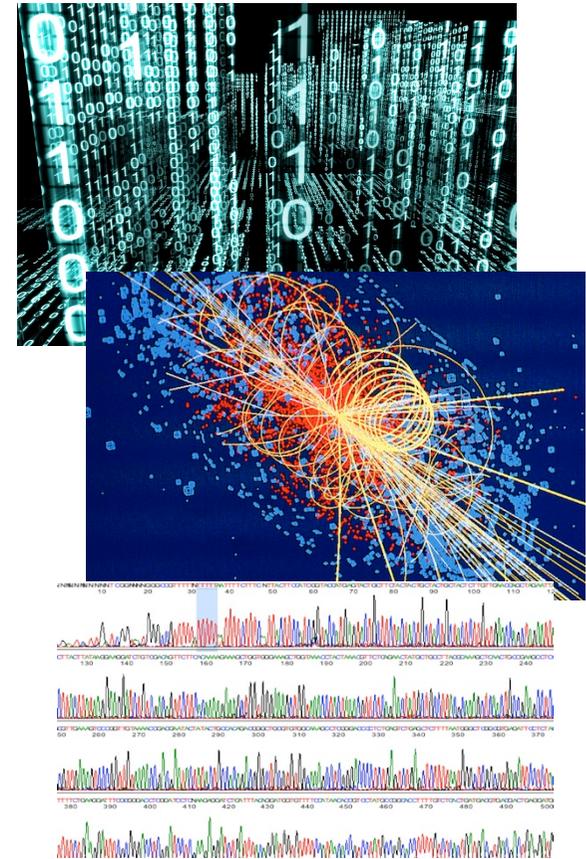
INSTITUT PARIS DESCARTES
NEUROSCIENCES
COGNITION



diNo

Motivation

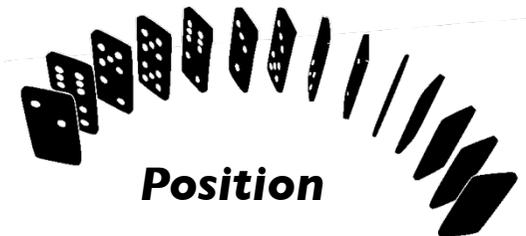
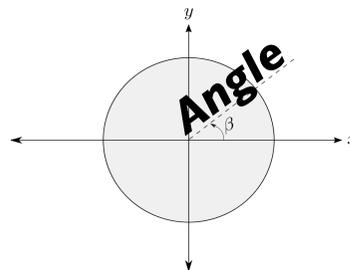
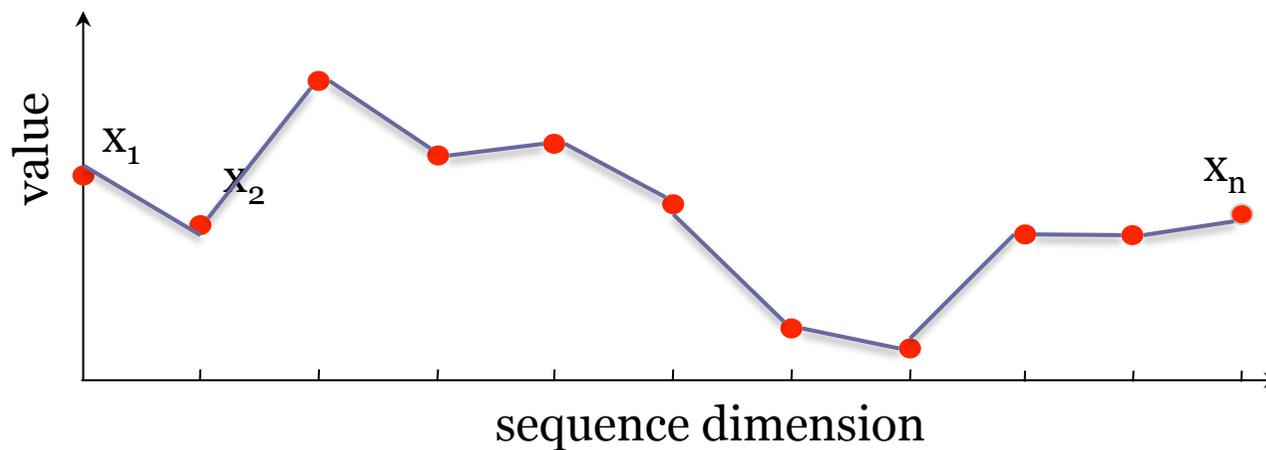
- data collected at unprecedented rates
- they enable data-driven scientific discovery
- lots of these data are sequences
 - takes **days-weeks** to analyze big sequence collections



goal: **analyze big sequences** in **minutes/seconds**

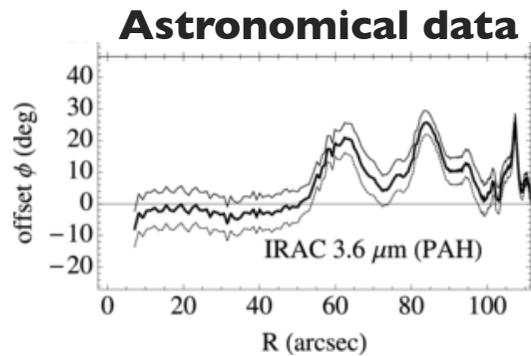
What are sequences?

- sequences are observations (measurements) ordered along some dimension (time, angle, mass, position, etc.)

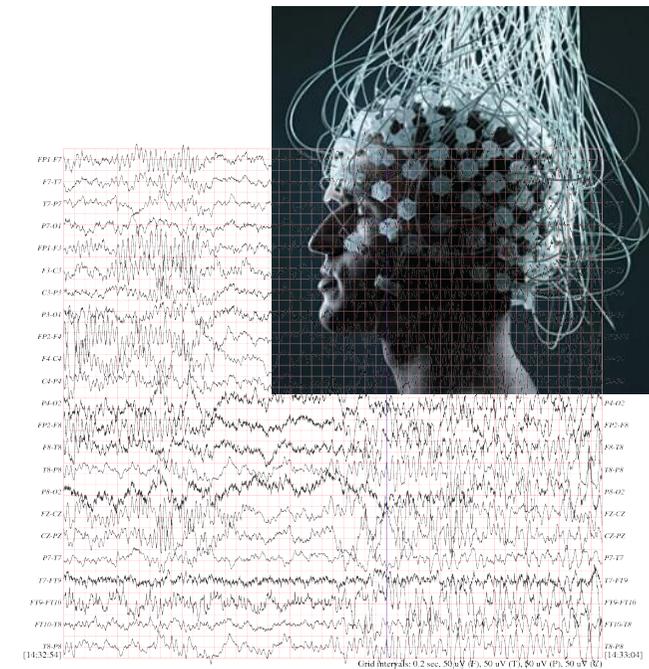
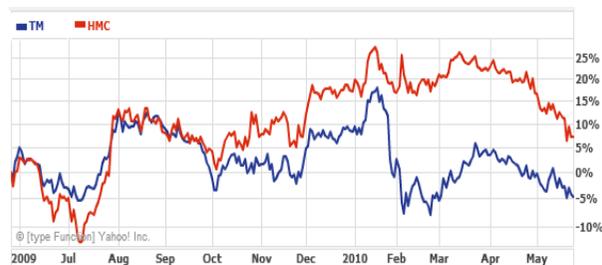


Why do we care about sequences?

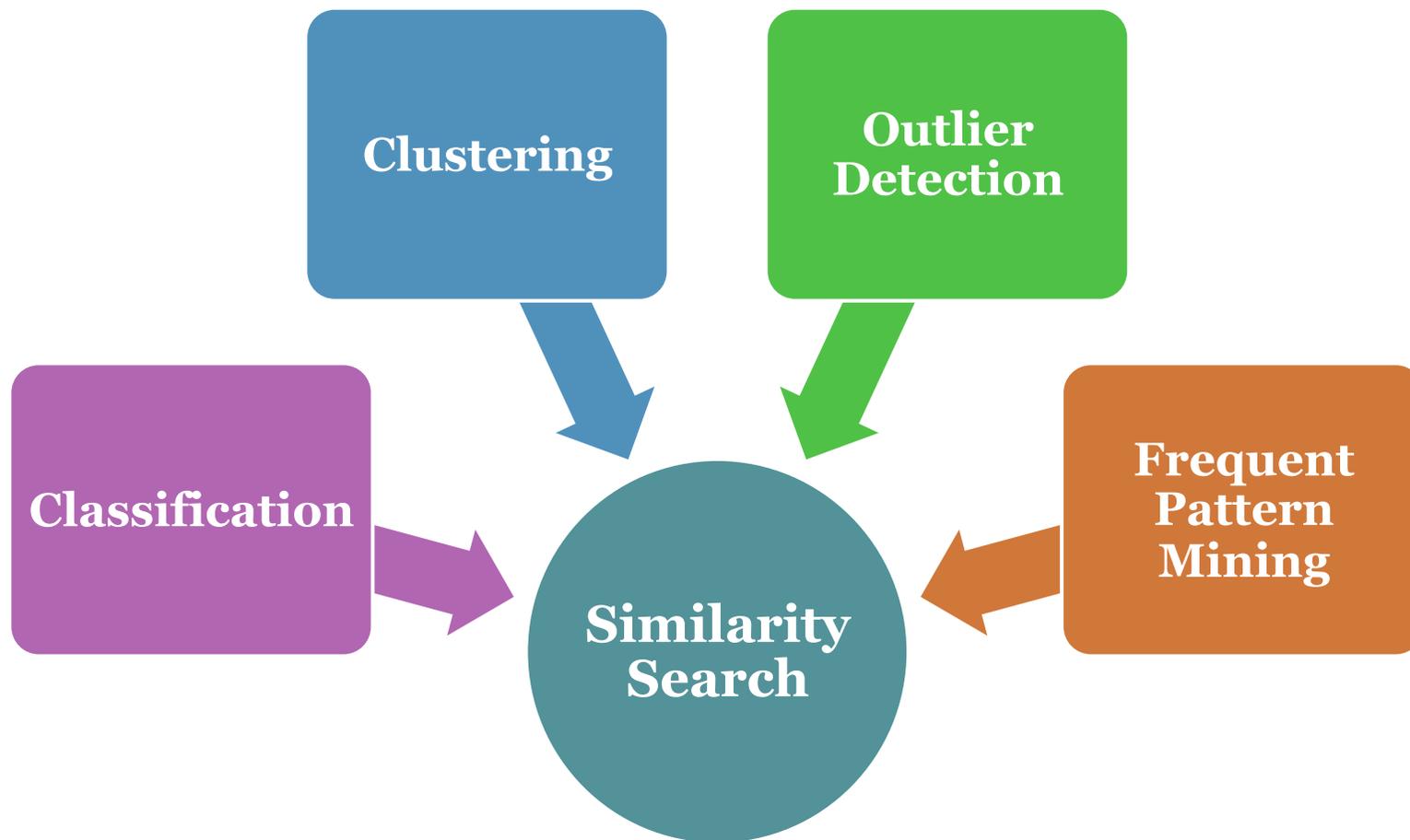
- occur in virtually every scientific, medical, business, and social domain
 - astronomy, smart cities, factories, finance, entomology, neuroscience, ...



Financial Data

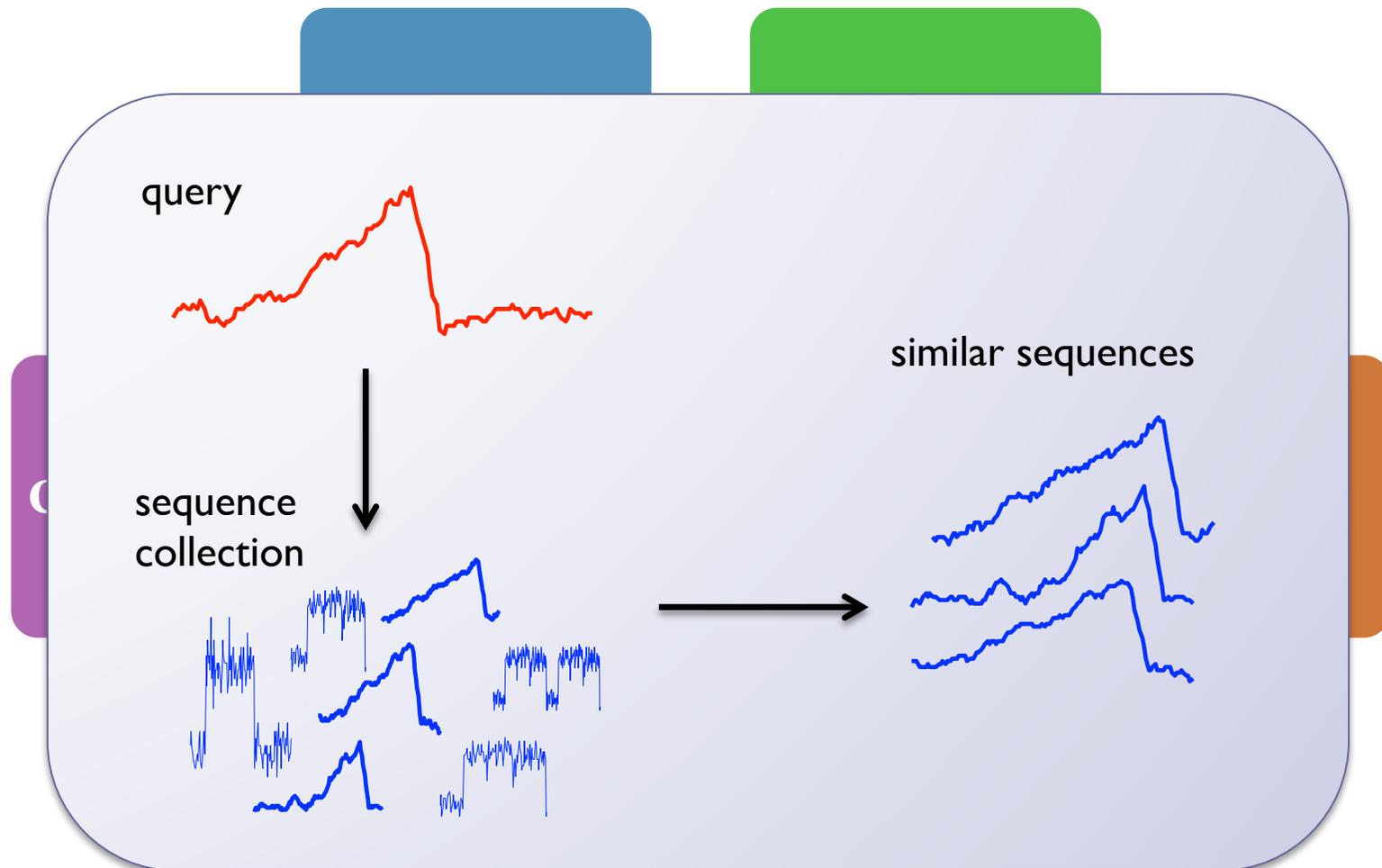


What do we want to do with them?
- complex analytics



What do we want to do with them?

- complex analytics



What do we want to do with them?
- complex analytics

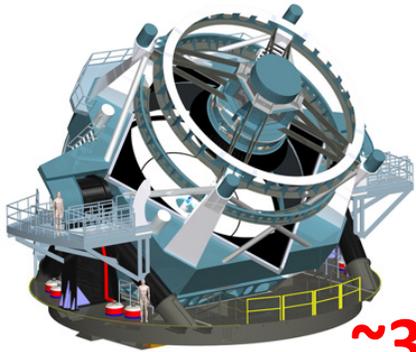
Clustering

Outlier
Detection

HARD, because of **very high dimensionality:
each data series has 100s-1000s of dimensions!**

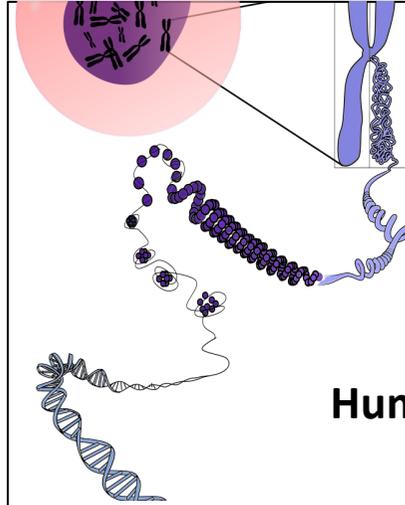
even HARDER, because of **very large size:
10s-100s of TBs of data series!**

Big Data Series



Large Synoptic
Survey Telescope

~30 TB per night



Human Genome project

130 TB

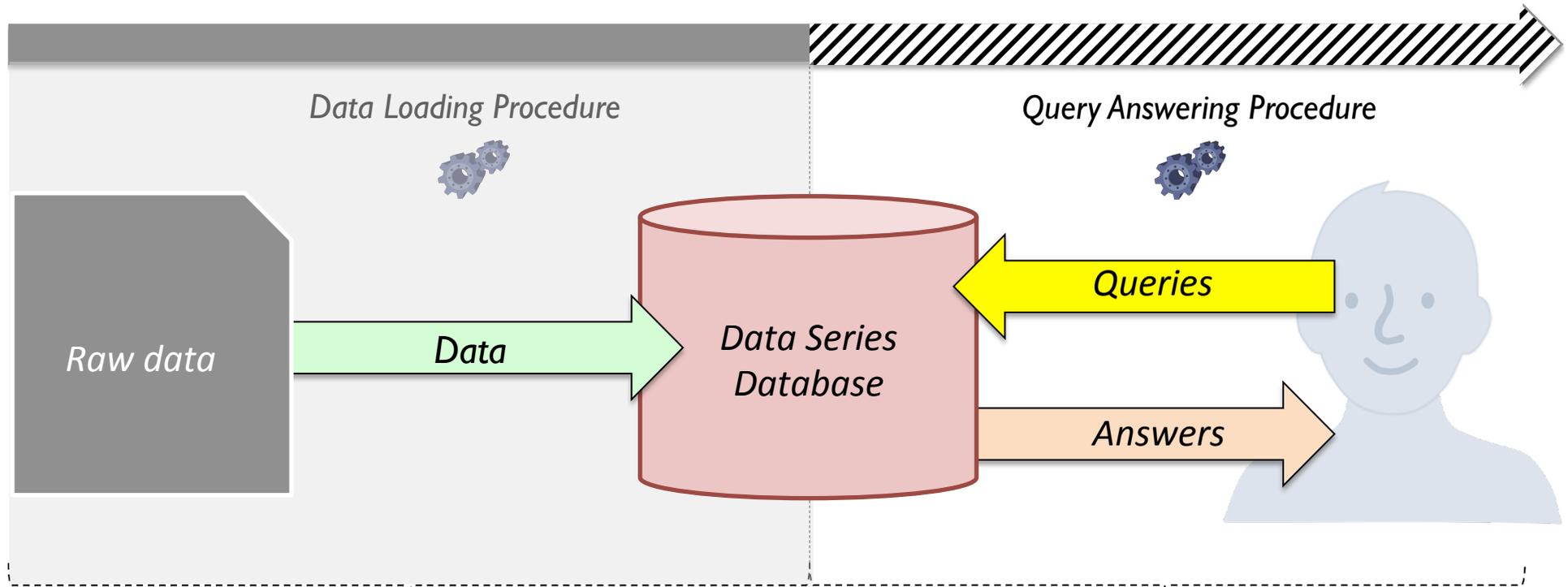


Boeing jet
20 TB per hour

data center and
services monitoring
2B data series
4M points/sec



Query answering process



data-to-query time

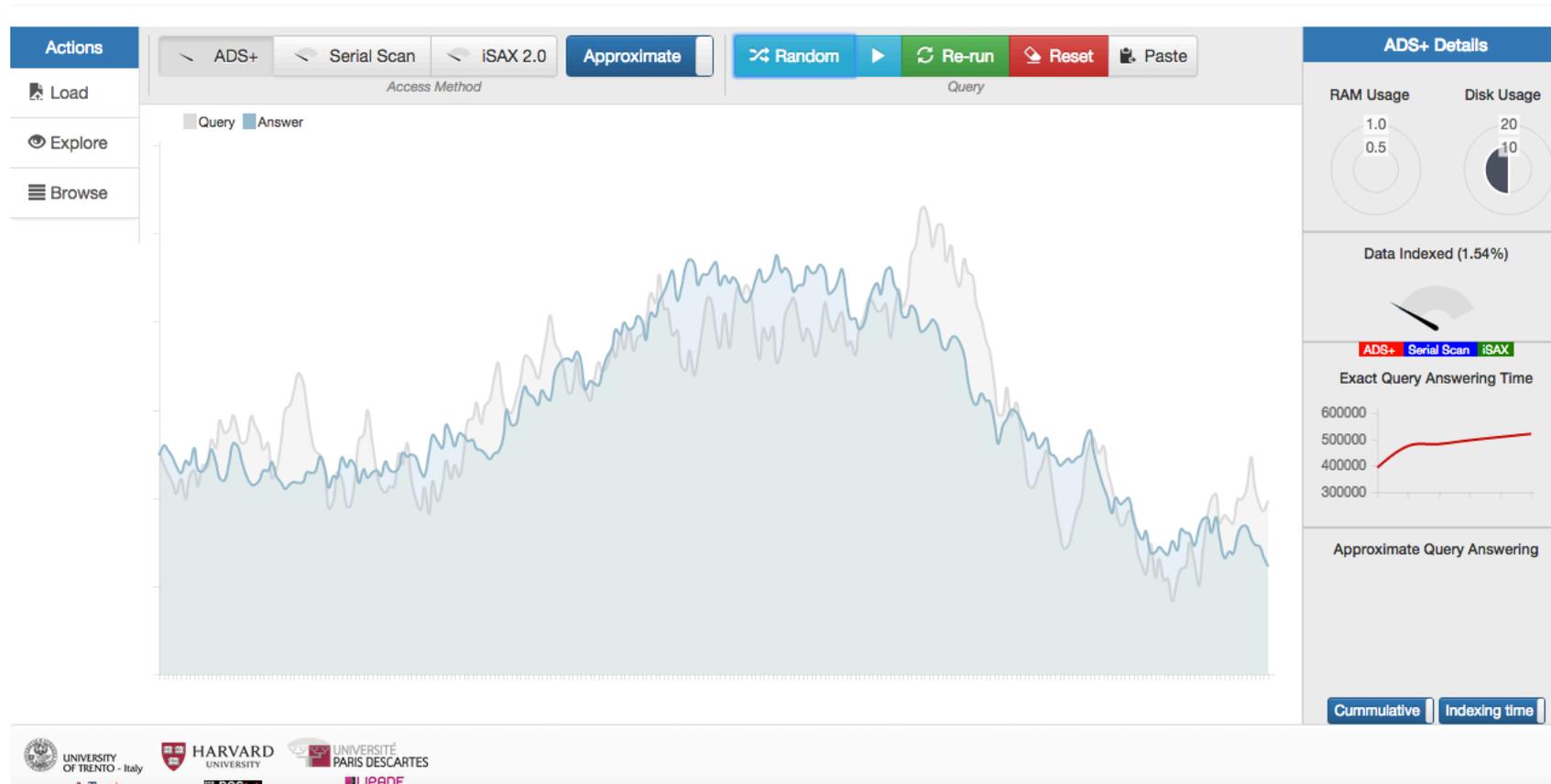
query answering time

these times are long!

**up to several days
for 1 Billion series!**

Prototype

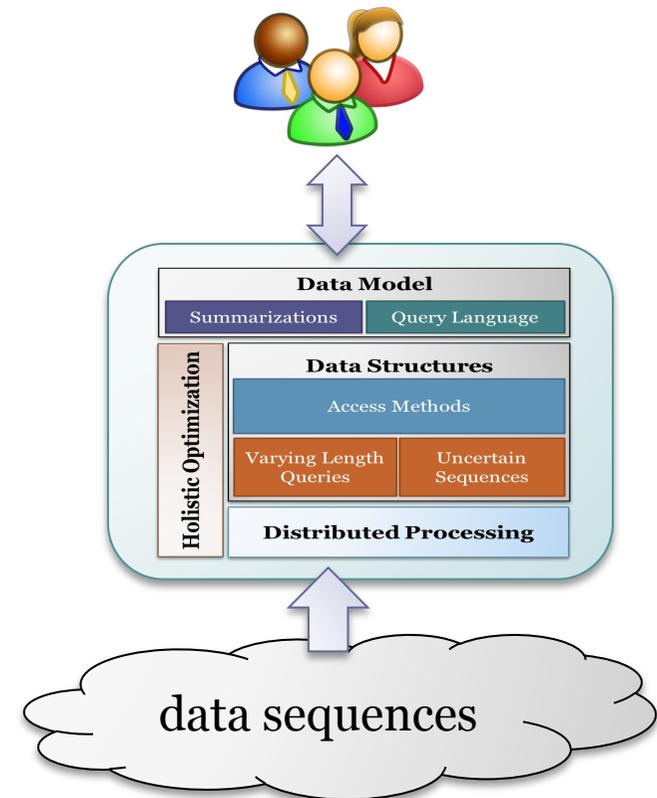
- <http://daslab.seas.harvard.edu/rinse/>



Our Goal

- **Big Sequence Management System**

- general purpose
- easy to use
- scalable



Who is Who

- Themis Palpanas
 - Paris Descartes University
 - computer science (data analytics)
- Cécile Cavet, Volker Beckmann
 - Paris Diderot University, APC
 - astrophysics
- Sophie Bernard
 - Paris Descartes University
 - neuroscience



Getting Involved

- If you
 - have large sequence collections
 - want to process and analyze them fast
 - want to share your expertise on data series analysis
- Then send us an email:
themis@mi.parisdescartes.fr

CNRS/MASTODONS workshop on data series:
~June 17, 2016

