Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Importancia de los modelos de programación en la convergencia de la supercomputación y el Big Data

Rosa M Badia

Workflows and Distributed Computing Manager

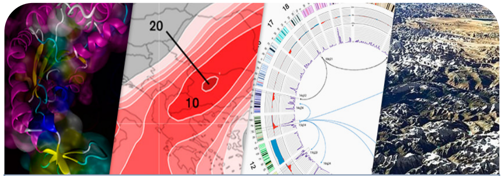
01/06/2017

11as Jornada SIG Libre, Girona

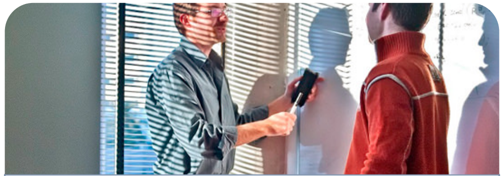
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

BSC-CNS objectives

Supercomputing services
to Spanish and
EU researchers



R&D in Computer,
Life, Earth and
Engineering Sciences



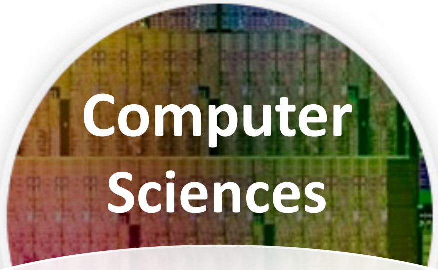
PhD programme,
technology transfer,
public engagement

BSC-CNS is a consortium that includes

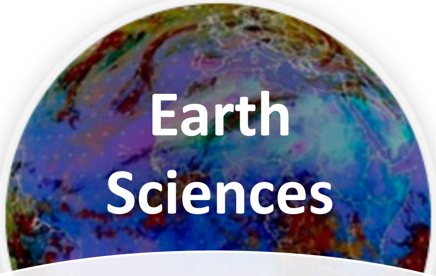
Spanish Government	60%	Gobierno de España – Ministerio de Economía, Industria y Competitividad
Catalonian Government	30%	Generalitat de Catalunya, Departament d'Empresa i Coneixement
Univ. Politècnica de Catalunya (UPC)	10%	Universitat Politècnica de Catalunya BarcelonaTech

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Mission of BSC Scientific Departments



To influence the way machines are built, programmed and used: programming models, performance tools, Big Data, computer architecture, energy efficiency



To develop and implement global and regional state-of-the-art models for short-term air quality forecast and long-term climate applications

To understand living organisms by means of theoretical and computational methods (molecular modeling, genomics, proteomics)

To develop scientific and engineering software to efficiently exploit super-computing capabilities (biomedical, geophysics, atmospheric, energy, social and economic simulations)

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

The MareNostrum 4 Supercomputer

Total peak performance

13.7 Pflops/s, 390TB

12 times more powerful than MareNostrum 3

Compute

General Purpose, for current BSC workload

More than 11 Pflops/s

3,456 nodes of Intel Xeon v5 processors

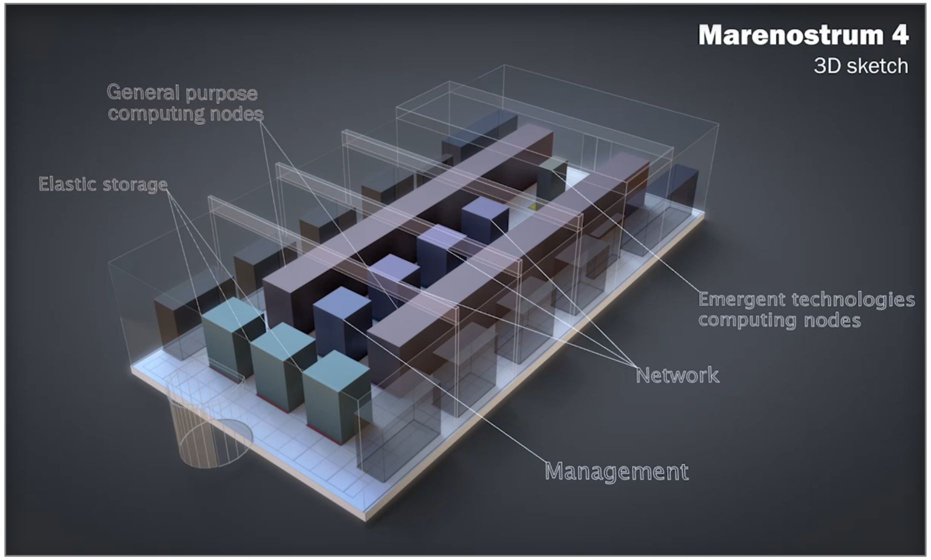
Emerging Technologies, for evaluation of 2020 Exascale systems

3 systems, each of more than 0.5 Pflops/s with KNL/KNH, Power9+NVIDIA, ARMv8

Storage

14 PB of GPFS

Storage System



Network

IB EDR/OPA

Ethernet

Operating System: SuSE

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Integration of Extreme Scale Computing and Big Data Management and Analytics

- Apparently two diverse worlds…

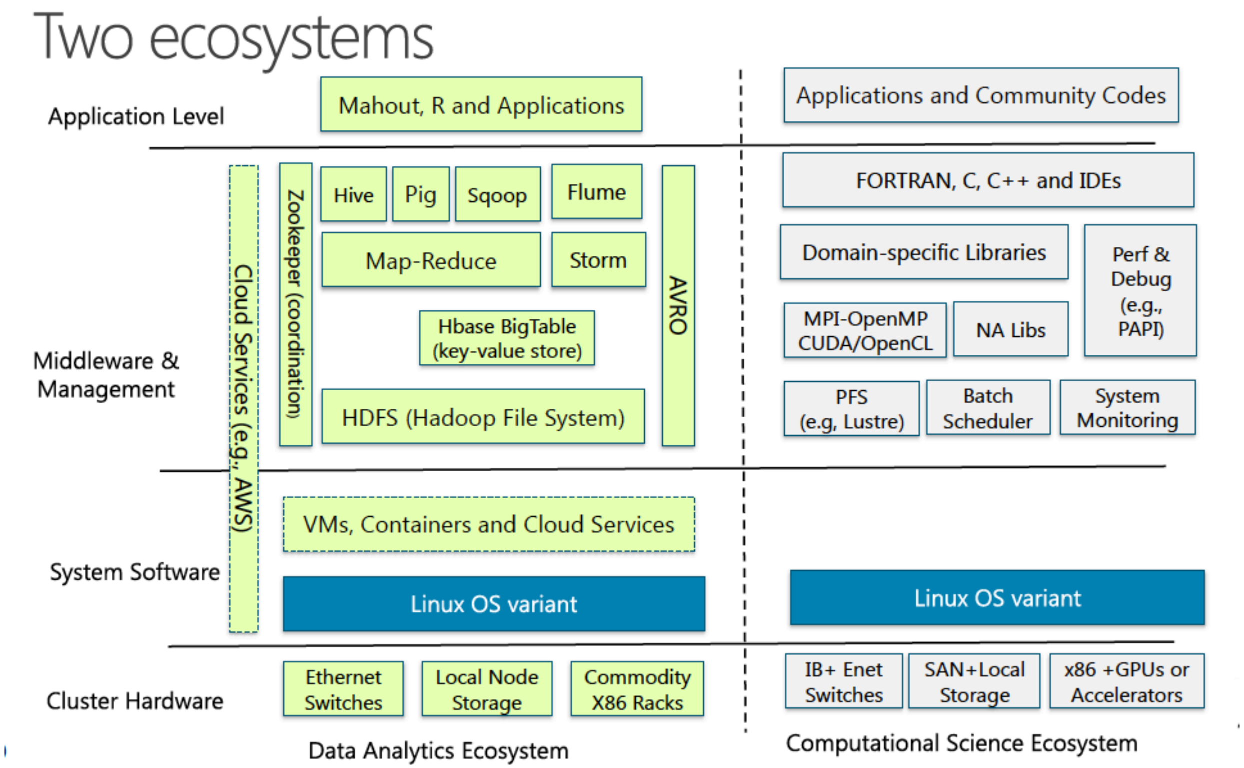
  ... with different needs, software stacks, ...

VS

- However
  - … convergence will enable to enhance both worlds
  - Big Data can leverage the results from HPC, and viceversa

Integration of Extreme Scale Computing and Big Data Management and Analytics

- Architecture looks similar
- Individual components are different
- Actual infrastructure, can be the same???
  - Vendors interest

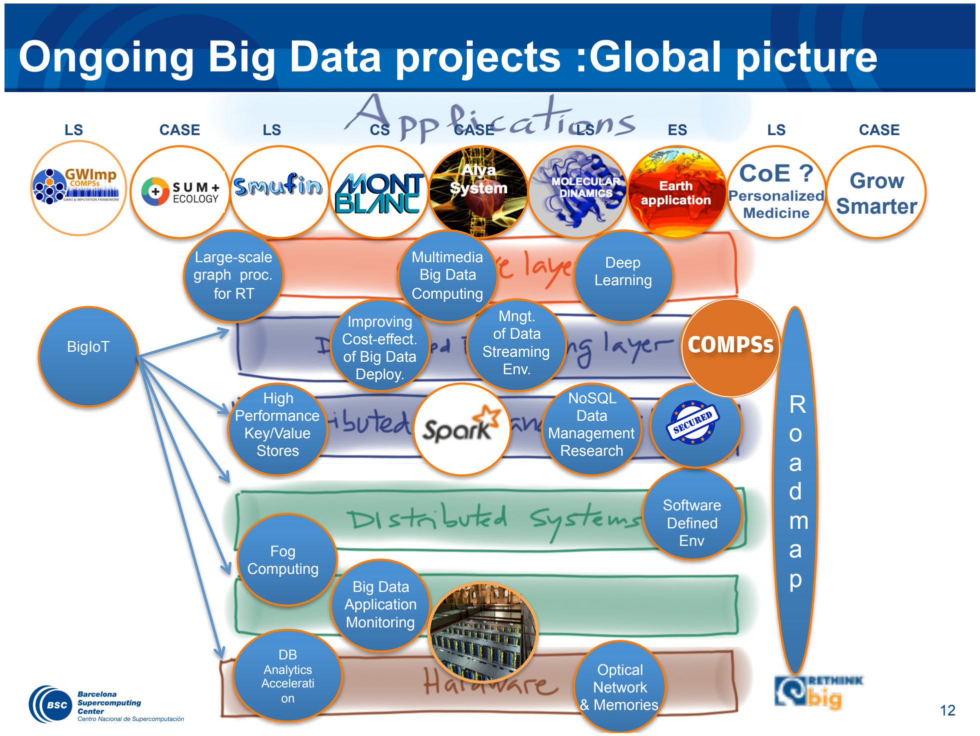


** Big Data Meets HPC, Dan Reed, BDEC 2015*

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

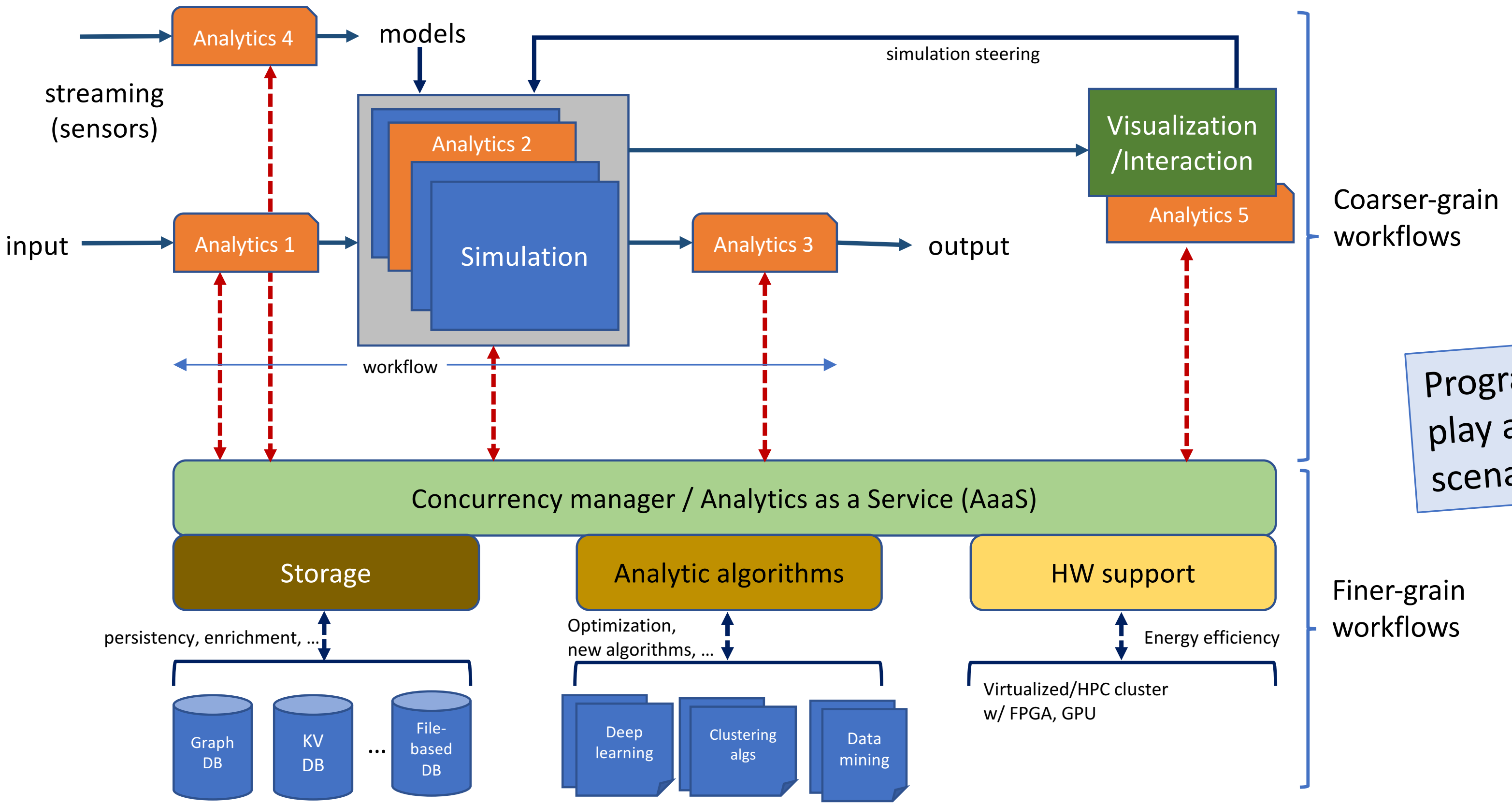
Big Data activities at BSC

- Research activities in broad topics
- Considerable cross-department collaborations
- Thematic applications
- EU funded projects
- Projects with industry
- Training activities
- Involvement in worldwide and European initiatives
  - BDVA & ETP4HPC
  - RDA
  - BDEC



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Future HPC-BigData workflows, a view from Barcelona*

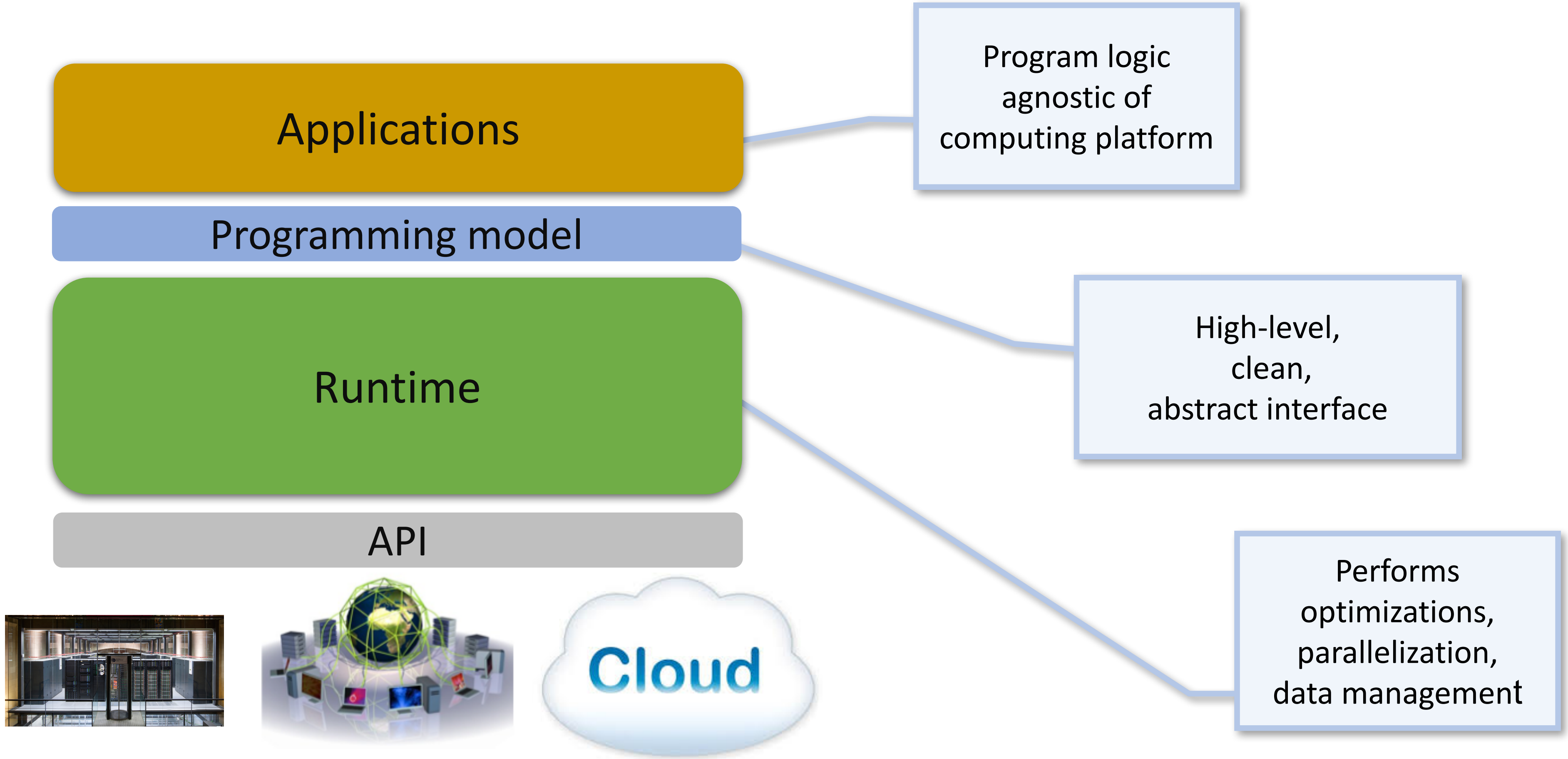


Programming models play a key role in this scenario

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Workflows for science: a challenge when facing the convergence of HPC and Big Data,
Rosa M. Badia, Eduard Ayguade, Jesus Labarta,
Journal Supercomputing Frontiers and Innovations, 2017

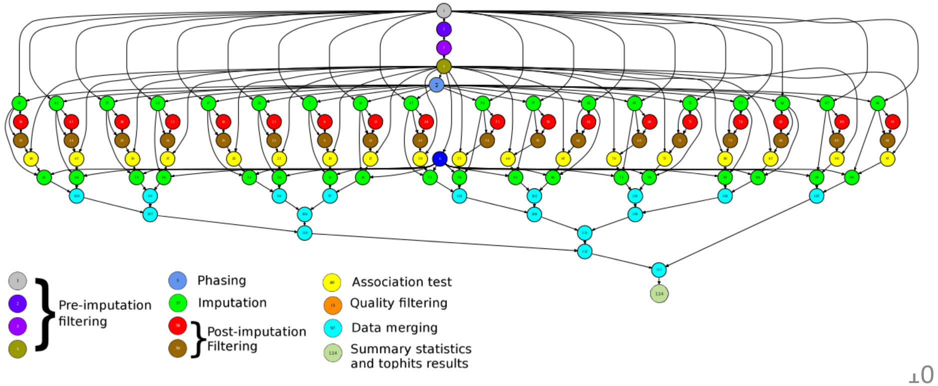
BSC vision on programming models



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Programming Model: PyCOMPSs/COMPSs

- Sequential programming
- Task based: task is the unit of work
- General purpose programming language + annotations/hints
  - To identify tasks and directionality of data
- Simple linear address space
- Builds a task graph at runtime that expresses potential concurrency
  - Implicit workflow
- Exploitation of parallelism
  - … and of distant parallelism
- Agnostic of computing platform
  - Enabled by the runtime for clusters, clouds and grids



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

PyCOMPSs

- Based on regular/sequential Python code
- Use of decorators to annotate tasks and indicate arguments directionality
- Other annotations: constraints
- Small API for data synchronization

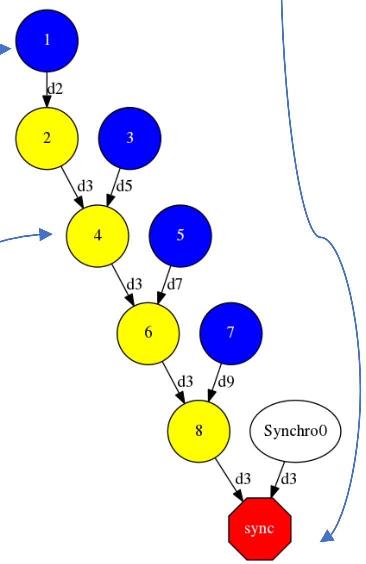
Main Program

```
Data = [block1, block2, …, blockN]
result=defaultdict(int)
for block in Data:
    presult = word_count(block)
    reduce_count(result, presult)
finalResult = compss_wait_on(result)
```

Tasks definition

```
@constraint(ProcessorCoreCount=mkl_threads)
@task(returns=dict)
def word_count(collection):
    ...
```

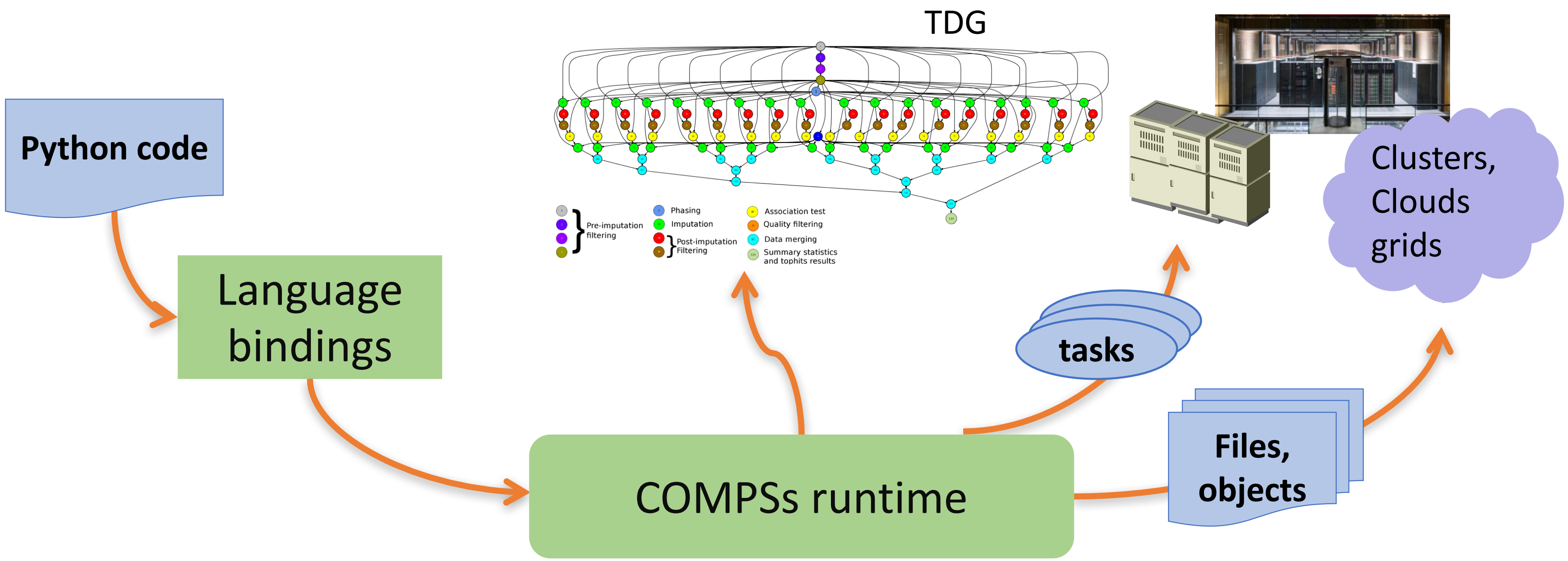
```
@task(dict_1=INOUT)
def reduce_count(dict_1, dict_2):
    ...
```



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

PyCOMPSs runtime

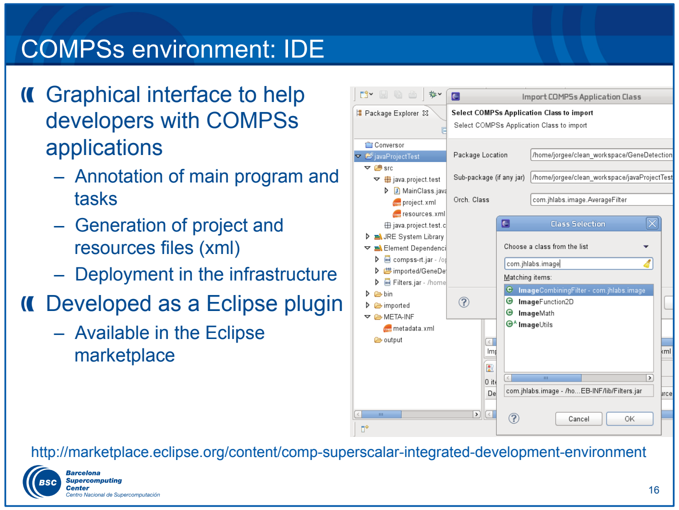
- Sequential execution starts in master node
- Tasks are offloaded to worker nodes
- All data scheduling decisions and data transfers performed by runtime

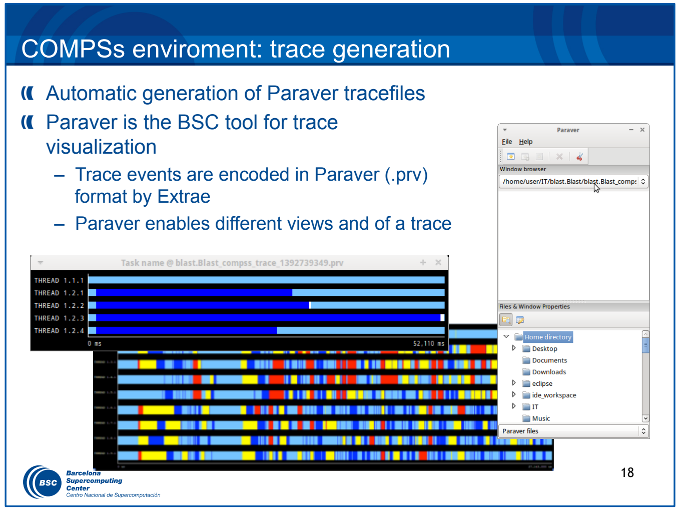


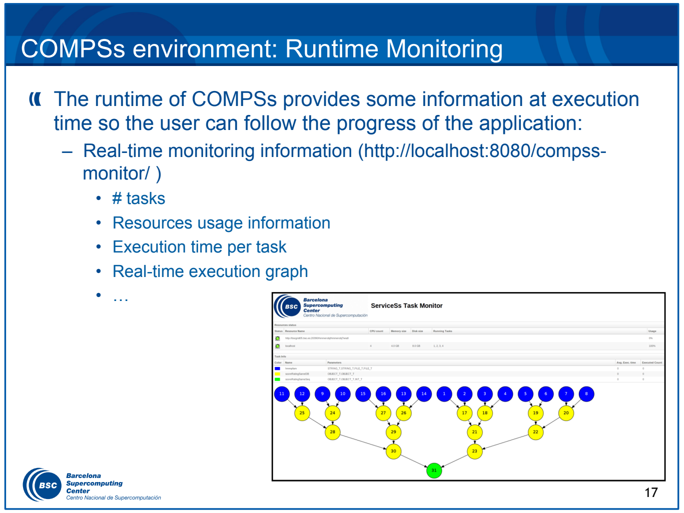
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

COMPSs development environment

- IDE graphical interface
- Runtime monitor
- Paraver traces



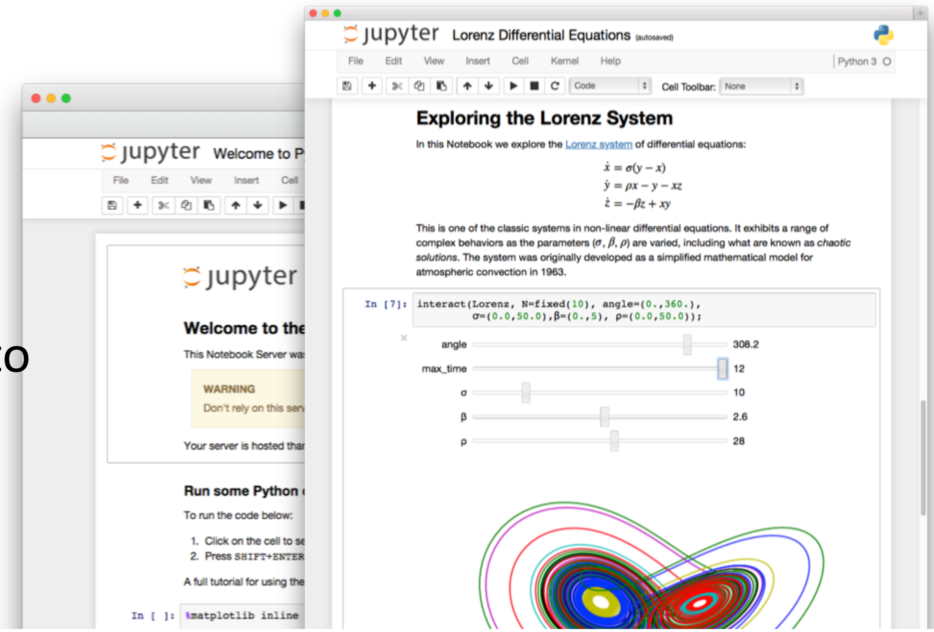




Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Integration with Jupyter notebook

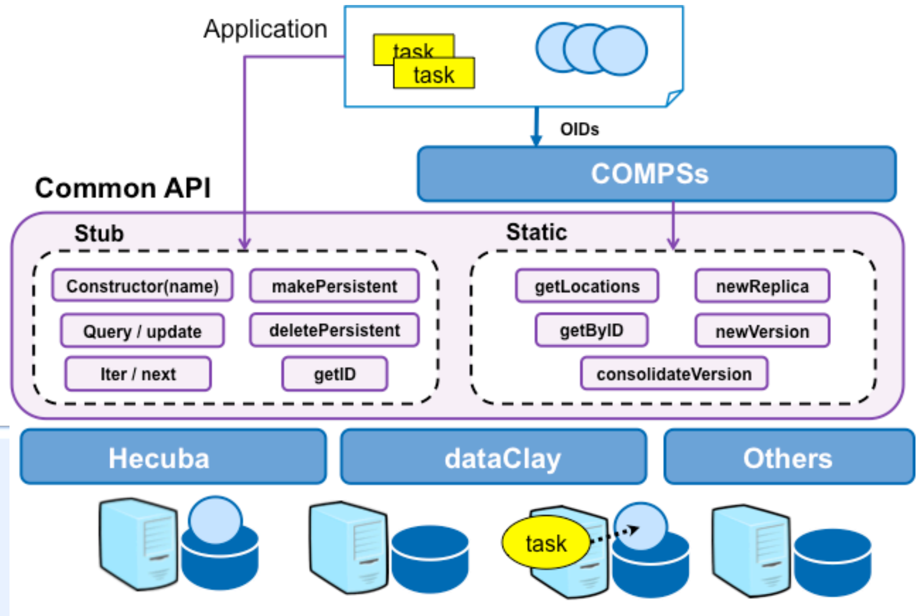
- The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text
- Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more
- Runs Python –sequential
- PyCOMPSs runtime integrated with Jupyter notebook
  - Runtime started from notebook
  - PyCOMPSs tasks registered and send to workers



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Integration with Storage: Storage API

- Integration of programming model with new storage management platforms
- Data made persistent, application agnostic of this persistency
- Producer-consumer
- In-situ

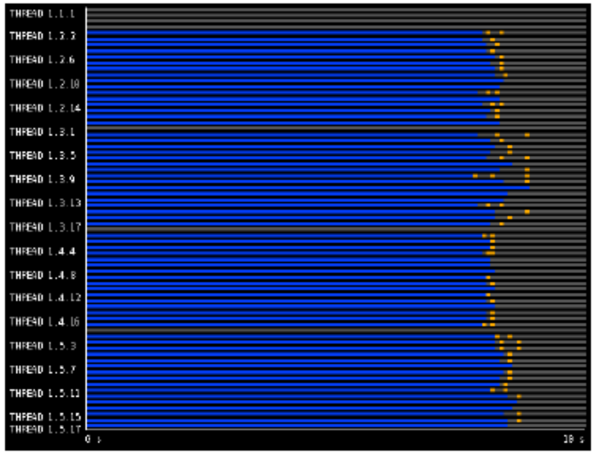



```
...
for i in range(n_points // self.points_per_fragment):
    np.random.seed(base_seed + i)
    fragment = Fragment(dim=self.dim,
                        points=np.random.random([self.points_per_
                        base_index=i * self.points_per_fragment)
    fragment.make_persistent(dest_stloc_id=storage_locations.next())
    self.fragments.append(fragment)
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

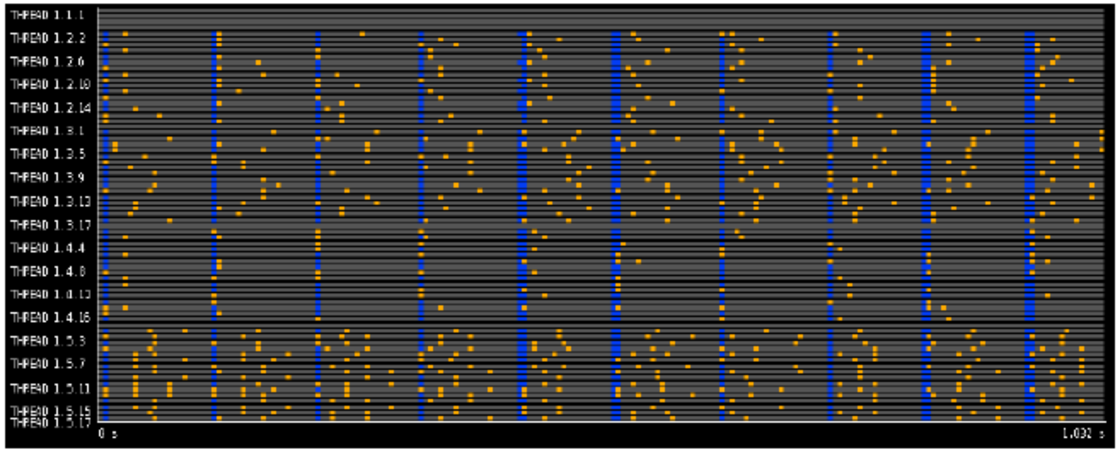
Integration with Storage: tests with dataClay

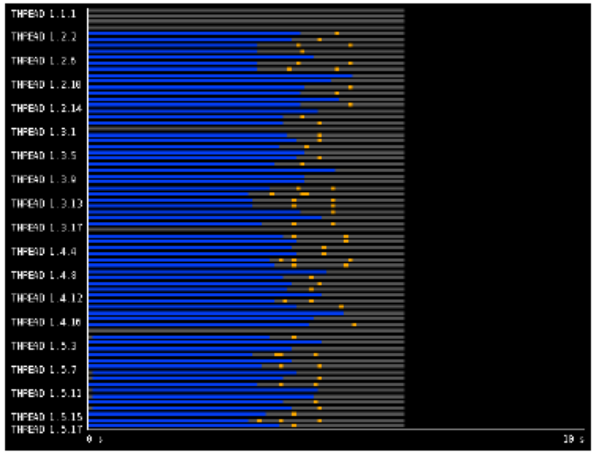
Wordcount



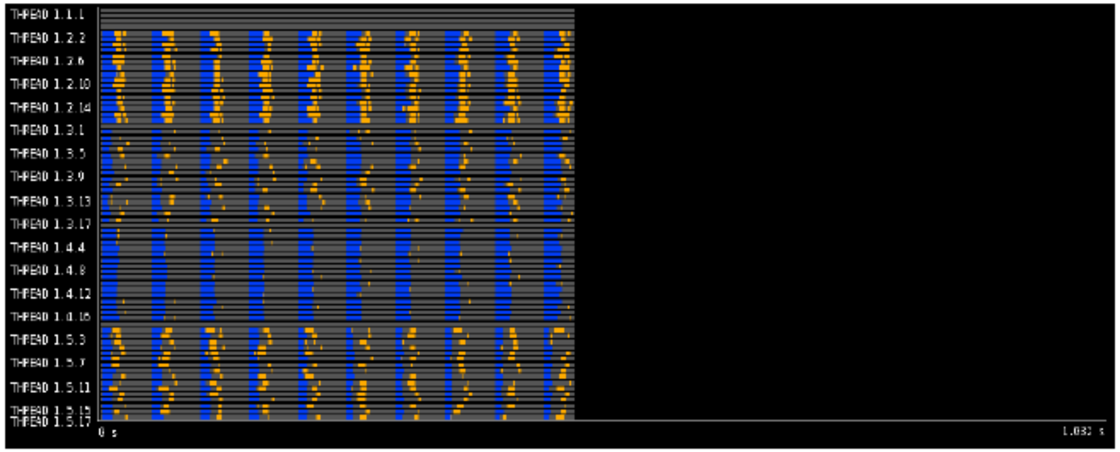
Kmeans

Files

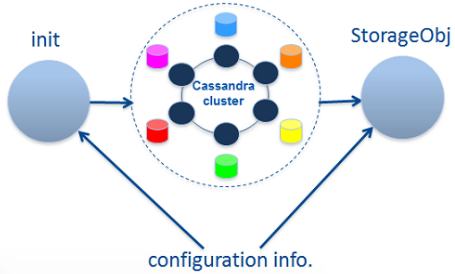




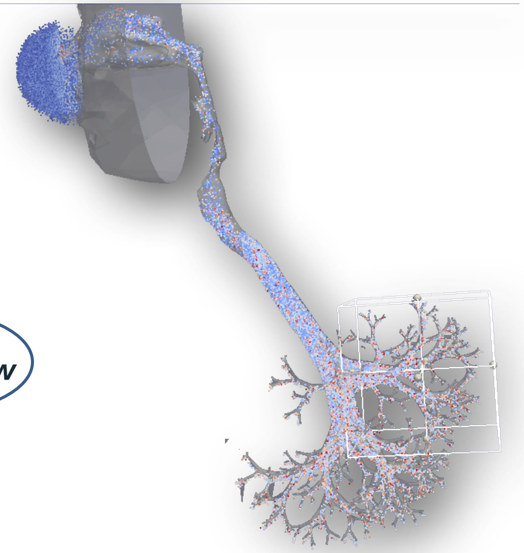
dataC

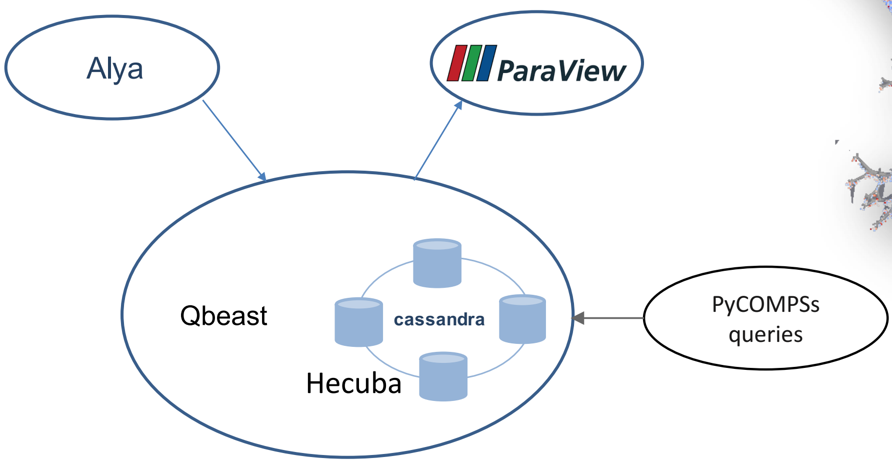


Case of study with Hecuba: Respiratory system simulator



- Alya simulation of the respiratory system
- Prototype demo implemented on top on key-value data store:
  - Particles generated by simulation stored in Cassandra
  - Managed by Hecuba
- Qbeast: D8-tree index distributed engine
  - Data access with linear scalability
- Queries parallelized with PyCOMPSs
- Visualization and queries simultaneous to simulation





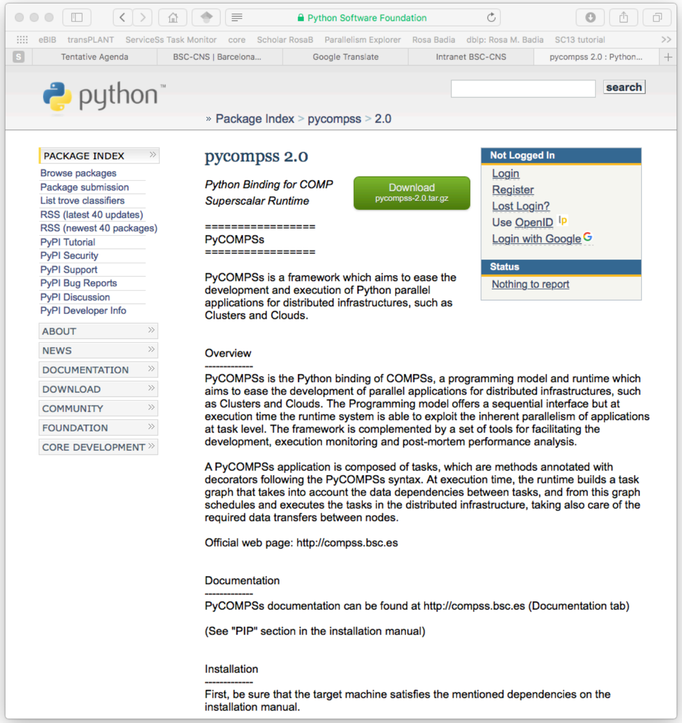
PyCOMPSs/COMPSs status

- Periodical releases – every 6 months
  - Next release end of june 2017
- Open Source – Apache v2
- Distribution of linux packages
- Virtual image with code, environment and tutorial examples
- Documentation
  - Installation manual
  - Application execution manual
  - Application developer manual
  - MareNostrum manual

Available at: compss.bsc.es

PyCOMPSs - PIP install

- January 2017
- Release of PyCOMPSs pip package to enable automatic installation with "pip install".
- Documentation for the package



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions

- Convergence between HPC and BigData is necessary and benefits both worlds
- Different worlds, but not irreconcilable
  - Need to find gaps and overlaps
- Part of the new ecosystem should deal with programming models
  - Task based programming models offer tools for application development at different level
- BSC roadmap on workflows
  - Considers traditional parallel simulations
  - Integrates with new storage technologies
  - Provides means to integrate all components

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Thank you

rosa.m.badia@bsc.es

01/06/2017