SCAI
SuperComputing Application & Innovation

Sanzio Bassini – October 2017

The Consortium

- Private non for Profit Organization
- Founded in 1969 by Ministry of Public Education now under the control of Ministry for Education, University and Research
- Consortium structure
  - 70 Universities
  - Ministry of University and research
  - CNR, INFN, INAF, OGS, SZN
  - CREA, INDIRE, INVALSI

- Main premises in Bologna, point of presence in Milan and Rome

Current Computing systems

Logical Name	Tier 1 (December 2014)	PICO (big data) (October 2014)	MARCONI (June 2016)
Model	IBM / Lenovo NextScale	IBM / Lenovo NextScale	Lenovo NextScale
Architecture	Hybrid Linux Cluster	Linux Cluster Farm	Linux Cluster
Processor	Intel Haswell 8c 2.4 Ghz + Intel Xeon Phi 7120	Intel Ivy Bridge 8c/10c 2.5 Ghz + Nvidia K40	Intel BDW 18c 2.3 Ghz
# of core	8.192 (Haswell) + 46.848 (PHI)	1.400 (10c) + 128 (8c)	54.432
# of node	516 (384 with 2xPHI + 132)	78(66x10c + 4x10c with 2xK40 + 8x8c)	1.512
# of rack	14	3	25
RAM per node	128 GByte	128 Gbyte – 512 Gbyte	128 GByte
Interconnection	QDR Intel Truescale	Infiniband FDR	OPA 2:1
Operating System	RedHat	RedHat	RedHat
Total Power	~ 350 Kwatts	~ 70 Kwatts	~ 800 Kwatts
Peak Performance	~1 PFlops		~ 2.1 Pflops

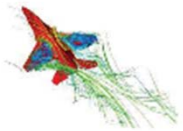
Computing systems – next upgrade

Logical Name	MARCONI T1 Tier 1 (November 2017)	MARCONI T0 Tier 0 (November 2017)
Model	Lenovo NextScale	Lenovo NextScale
Architecture	Linux Cluster	Linux Cluster
Processor	Intel BDW 18c 2.3 Ghz	Intel BDW 18c 2.3 Ghz + KNL 68c 1.4 Ghz + SKL 20c 2.3 Ghz
# of core	28.800	25.632 + 244.800 + 100.480
# of node	800	712 + 3.600 + 2.512
# of rack	14	12 + 50 + 21
RAM per node	128 Gbyte	128 Gbyte + 96 Gbyte +192 Gbyte
Interconnection	ETH	OPA 2:1
Operating System	RedHat	RedHat
Total Power	~ 400 Kwatts	~3.000 Kwatts
Peak Performance	~ 1 Pflops	~ 1 + 11 + 9 Pflops

Success stories

HPC - Scientific Computing and Technical Computing

117 Virtual Machines
on HPC Cloud



European data management infrastructure
552 Tb
13.700 K files
6.000 K user access

FORTISSIMO
HPC Cloud for technical computing
and innovation for industries.
14 projects ongoing
5 projects started
1.530 K cpu hours in started projects

Human Brain Project
70 Tb
1.800 K files

Weather forecast

	2015	2016
Model resolution:		
Step 1	7.0 Km →	5.0 Km
Step 2	2.8 Km →	2.2 Km
Day Nodes:	12 →	110
Area:	Italy →	Mediterranean

Big Data – Bioinformatics

15.000 genomic sample analysed
1.000 Whole Exome Sequencing
6.000 Chip-seq samples
132 users, 55 projects,
4.197 K cpu hours, 179 Tb

Big Data – Analytics

17.000 Fb posts 25.000 tweet 3.000 reviews
7.500 Fb users 10.500 Twitter users
43.000 interactions with social media (6 months)
→ 227.000 terms analyzed on Caserta Royal Palace

13.000 viewers
1.440 audience data (1 year)
→ 6.832.800 K audience data analyzed!

5.505 K vehicle operating parameters
analyzed form on-board diagnostic devices
1.542 K GPS data integrated in the analysis
(Fortissimo PRESERVE project)

20.000 queries on Agriturism
30.000 web sites
→ 20.000 K web pages analyzed

Big Data – Humanities

14.087 render hours
47.776 frames – 34 GB
(Isabella D'Este Virtual Studiolo)

469.293 Metadata records about Contributors
84.143 Metadata records about films
26.195 K Metadata records about Contributors from VIAF
(data coming from 9 European Film Heritage Institutions
imported in the the FORWARD catalogue)

6369 Open access datasets
3044 Closed access datasets
1290 Reserved access datasets
763 Under embargo datasets
(Opentesi: Datasets of PH.D Doctoral Thesis coming from 8 Italian Universities)

Main European funded R&D

HPC infrastructure

Digital infrastructure

Center of Excellence
Material Science

Fabric of the future
Industry 4.0

Environment

Multimedia / Cultural
Heritage

Energy efficiency

Life Science

Collaborations with qualified national players

- Eni E&P Research
  - HPC system management
  - Production management
  - Applications development
  - Innovation technology
- Protezione Civile / SMR Regione Emilia Romagna
  - Operation numerical weather forecast
  - Data Post processing
- ARPA Piemonte
  - Environmental numerical forecast
- Unipol
  - Risk management
  - Big data
- ISTAT
  - Web crawlers
  - Predictive analysis
- Telethon
  - Integrated peer review
  - Repository genomic data
- Human Technopole

PROTEZIONE CIVILE
Presidenza del Consiglio dei Ministri
Dipartimento della Protezione Civile

Istat
Istituto Nazionale di Statistica

TeleThon

Human Technopole ITALIA 2040

Unipol GRUPPO

PIEMONTE Arpa Agenzia Regionale per la Protezione Ambientale

Technology innovation: proof of concept

Manufacturing

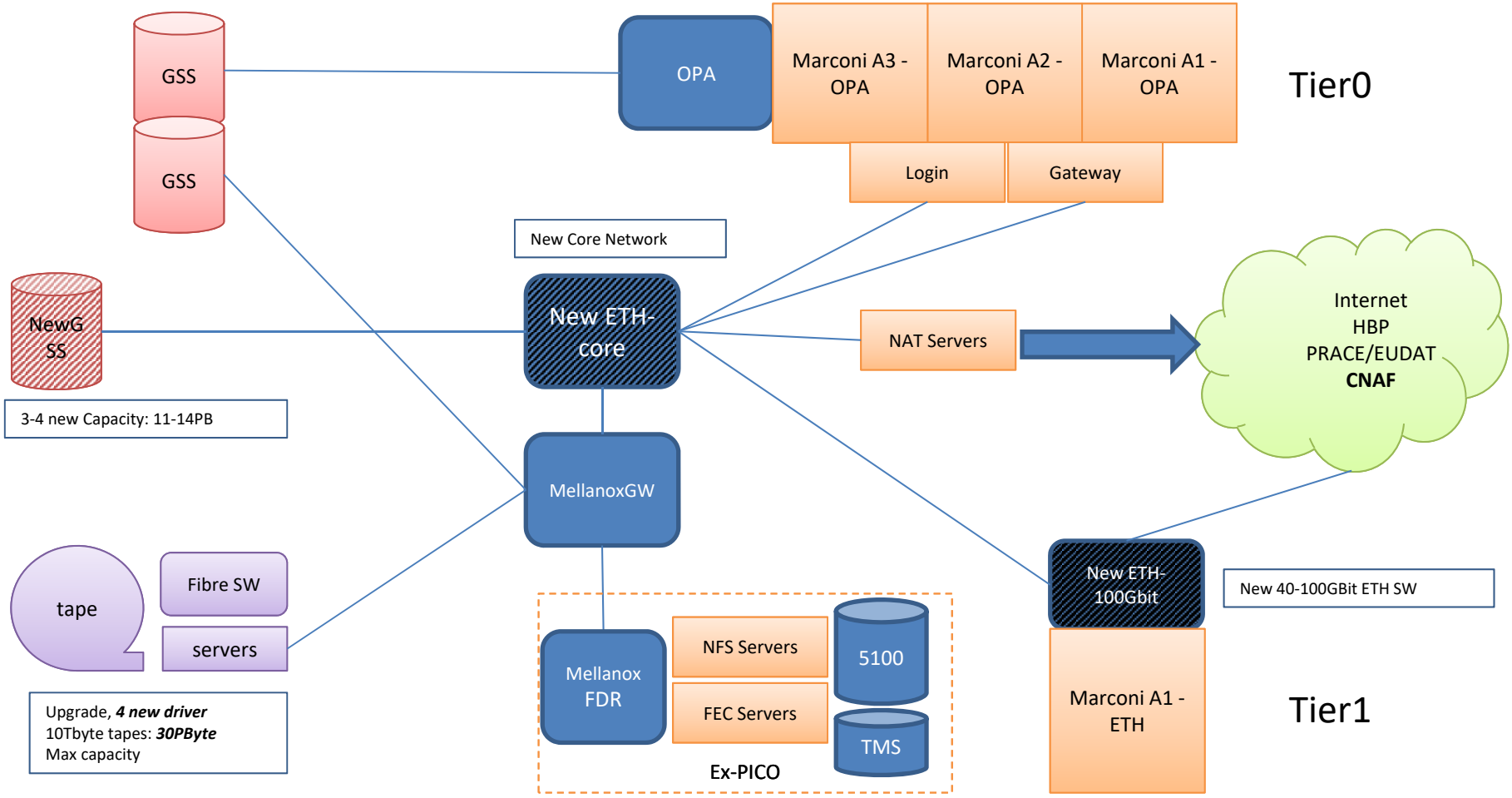
Engineering

Service

Middleware and Data Management

- **Deep Learning:**
  - Working with LENOVO and NVIDIA to benchmark and support DL common libraries on INTEL KNL and NVIDIA K80 & P100 architectures
  - AI Lab in collaboration with IBM & LENOVO
- **Cloud Computing**
  - Recondition of MARCONI A1 partition from OPA to Eth ROCE to enhance the Cloud Computing Platform → from 25 hypervisors on PICO to approximately 200 on Marconi (September 2017)
- **UNIPOL R&D**
  - Research Lab on Machine and Deep Learning topics
  - Long-term (3 years) framework agreement
- **Telethon/SIGU Data Repository**
  - Repository of genomic data and metadata
- **EVAR Planner**
  - Collaboration with “Humanitas Ospedale Milano” for the development of an application for Endovascular stent graft configuration
- **Collaboration with Microsoft & Amazon for testing Azure and AWS services**

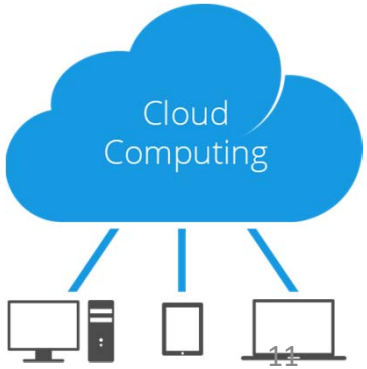
(New) Cineca HPC infrastructure



Systems evolution (HPC Cloud)

Logical Name	HPC Cloud (September 2017)
Model	NeXtScale Server
Architecture	Intel Broadwell
Processor	Intel E5-2697 v4 Broadwell 18 cores @ 2.3GHz
# of core	18
# of node	~ 400
# of rack	-
RAM per node	128 GB
Interconnection	Mellanox ETH 10/100 with ROCE
Operating System	GNU/Linux
Total Power	-
Peak Performance	-

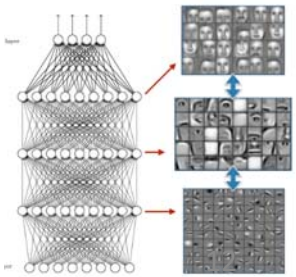
- Ensure high performance on single node
- Different workloads supported
- Interactive Computer
- Isolated environment for high-security environments



Systems evolution (Data Processing)

Logical Name	D.A.V.I.D.E. (August 2017)
Model	E4 Cluster Open Rack
Architecture	OpenPower NVIDIA NVLink
Processor	OpenPower 8 NVIDIA Tesla P100 SXM2
# of core	-
# of node	45 x (2 Power8 + 4 Tesla P100)
# of rack	-
RAM per node	-
Interconnection	Mellanox EDR
Operating System	GNU/Linux
Total Power	-
Peak Performance	~ 1 Pflops

- Result of a PCP (Pre-Commercial Procurement) commissioned by PRACE
- Based on OpenPOWER architecture, using IBM POWER8 processors with NVLink bus and the ultra performing GPGPU NVIDIA® TESLA® P100 SXM2.
- Low power consumption



The Roadmap

Logical Name	Tier 0 - FERMI (June 2012)	Tier 1 - GALILEO (December 2014)	Big data - PICO (October 2014)
Peak Performance	~ 2 Pflops; ~ 5 PByte	~ 0,5 PFlops	~ 0,3 Pflops; ~ 15 Pbyte

Logical Name	MARCONI T0 (2016 / 2017)	MARCONI T1 (2017)
Peak Performance	~ 20 Pflops; ~ 15 Pbyte	~1 Pflops; ~ 20 Pbyte

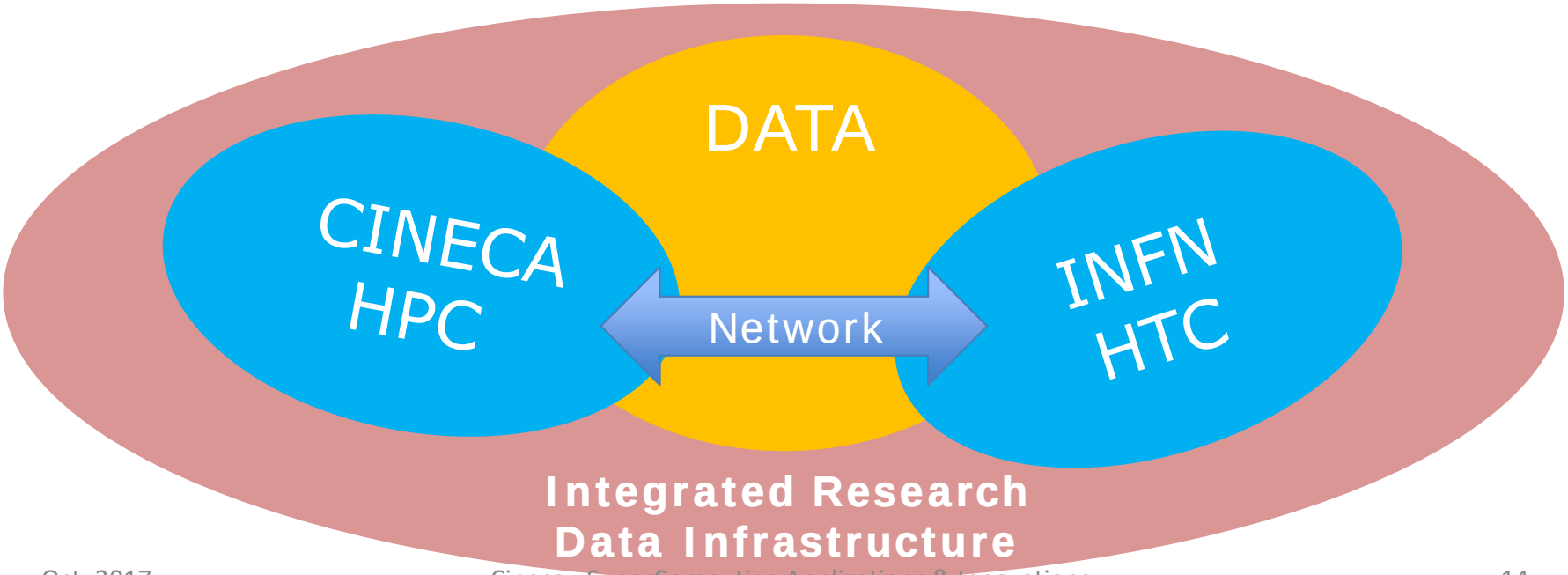
Logical Name	Tier 0 pre exascale (2019 – 2020)
Peak Performance	~ 50 Pflops; ~ 50 Pbyte on line storage; ~ 50 Pbyte repository

Logical Name	Tier 0 pre exascale (2021 – 2022)
Peak Performance	~ 250 Pflops; ~ 50 Pbyte on line storage; ~ 50 Pbyte repository

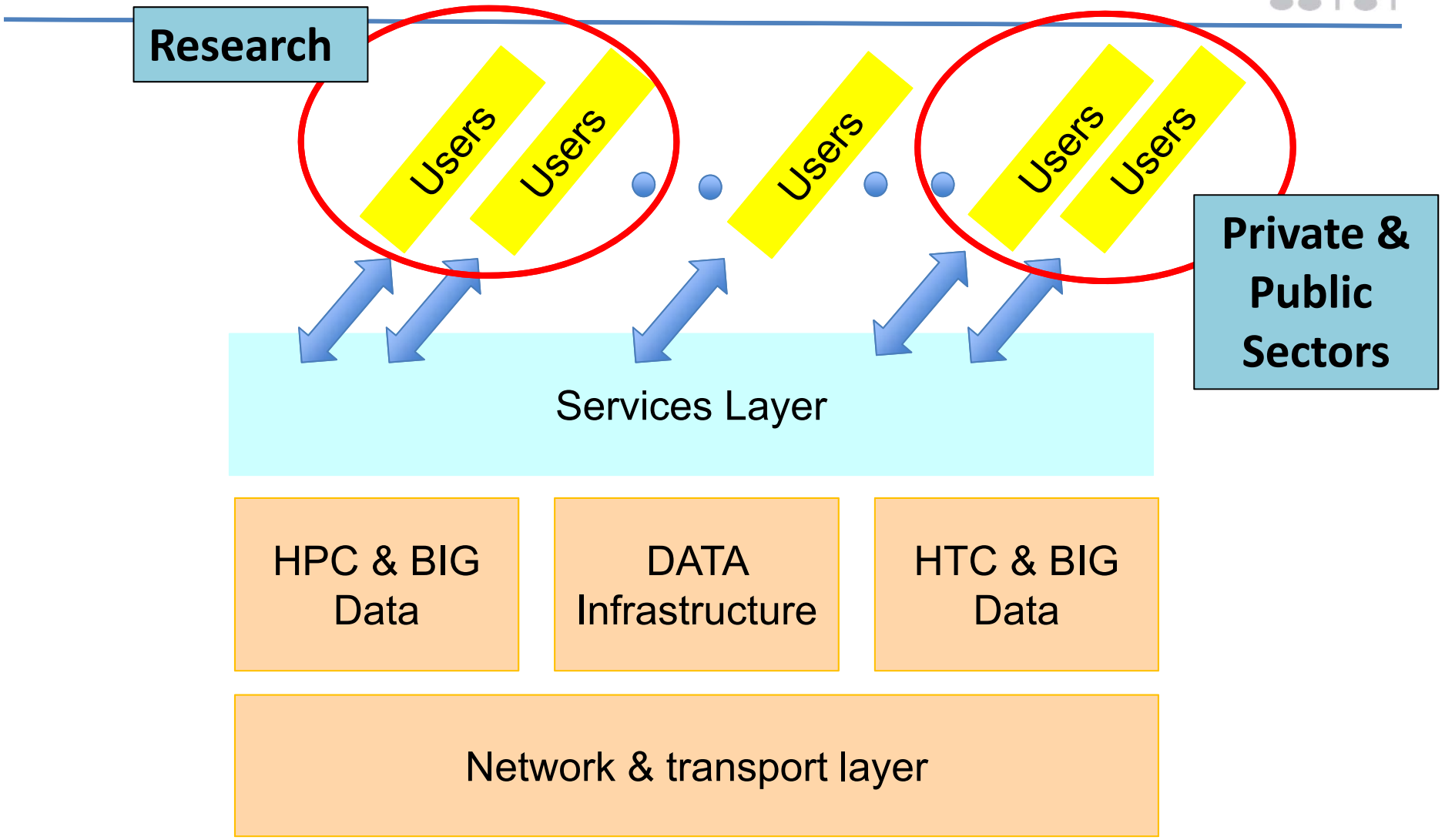
On going project towards Digital single market

Integration of CINECA-HPC and INFN-HTC computing infrastructure to provide services to:

- Institutional basic and applied research
- Enabling for Public administrations
- Proof of concept and innovation for private organizations and industries



Implementation model

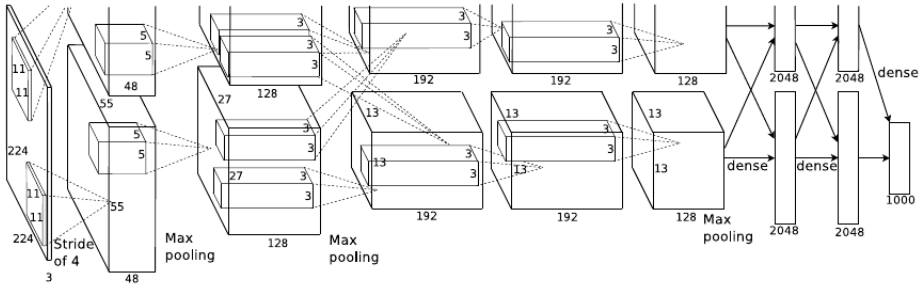


Data analytics software

- Apache Hadoop/Spark related applications (a, b)
  - Spark (Python, Scala and R shells)
  - Spark libraries (MLlib, Graphx, SQL, streaming)
  - Python Machine Learning libraries (Sci-Kit, Pylearn2, …)
  - R (SparkR, BigR)
  - H2O
  - Mahout
- Interfaces/Notebooks (a,b)
  - Jupyter/all-spark-notebook (Scala, R, Python)
  - Spark Notebook
  - Apache Zeppelin
  - Beaker Notebook
  - R Studio server
- High Performance tools
  - Google Tensorflow
  - Intel DAAL – Data Analytics Acceleration Library

Deep Learning at HPC

- Model size is important:
  - winning model of ILSVRC2012 classification task AlexNet: 5 conv. Layers + 3 fullyconn. layers



- Data size scaling
  - Larger datasets usually improve accuracy
- High Arithmetic intensity:
  - Multiple inputs, multiple outputs, batch: GEMM

- Dedicated facilities
  - data storage
  - parallel computation hardware (manycores, GPUs)
- Optimized primitives

Tested software frameworks

- Community based
  - you can clone/fork them from github

Caffe

- Common approach:
  - NN is built by defining it's computational graph
  - High level language is preferred (Python, protobuf, LUA)

- Intra-node multi-GPU parallelization
- Hardware-independent layer

theano

- Optimized backend engines:
  - MKL and MKL-DNN for Intel based CPUs and many-cores
  - cuBLAS and cuDNN for Nvidia GPUs
- Inter-node scaling:
  - inital support: Tensorflow (gRPC), Caffe (MPI)
  - dedicated communication library (Intel MLSL)