



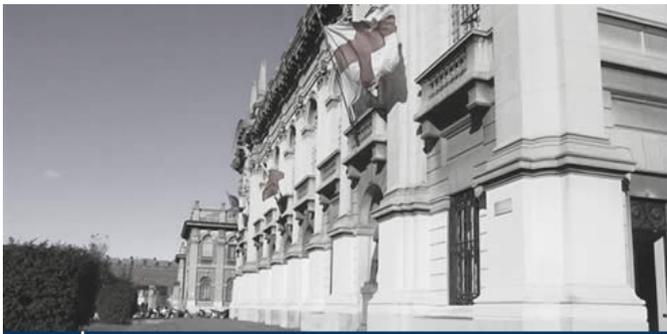
 POLITECNICO DI MILANO



Clothing/PoliCloud Lab



- Contesto: cloud computing e IoT
- Clothing/Policloud: sviluppo di un laboratorio cloud/IoT al Politecnico
- Applicazioni che usano PoliCloud, esempi
- Conclusioni



 POLITECNICO DI MILANO



Cosa è il cloud computing?



Utilizzatori di cloud

- **Accesso on-demand**
- **Modalità pay per use**

Cloud provider

- **massimizzazione dell'utilizzo delle risorse fisiche,**
- **minimizzazione dei costi energetici**
- **economie di scala**



Utilizzatori = provider



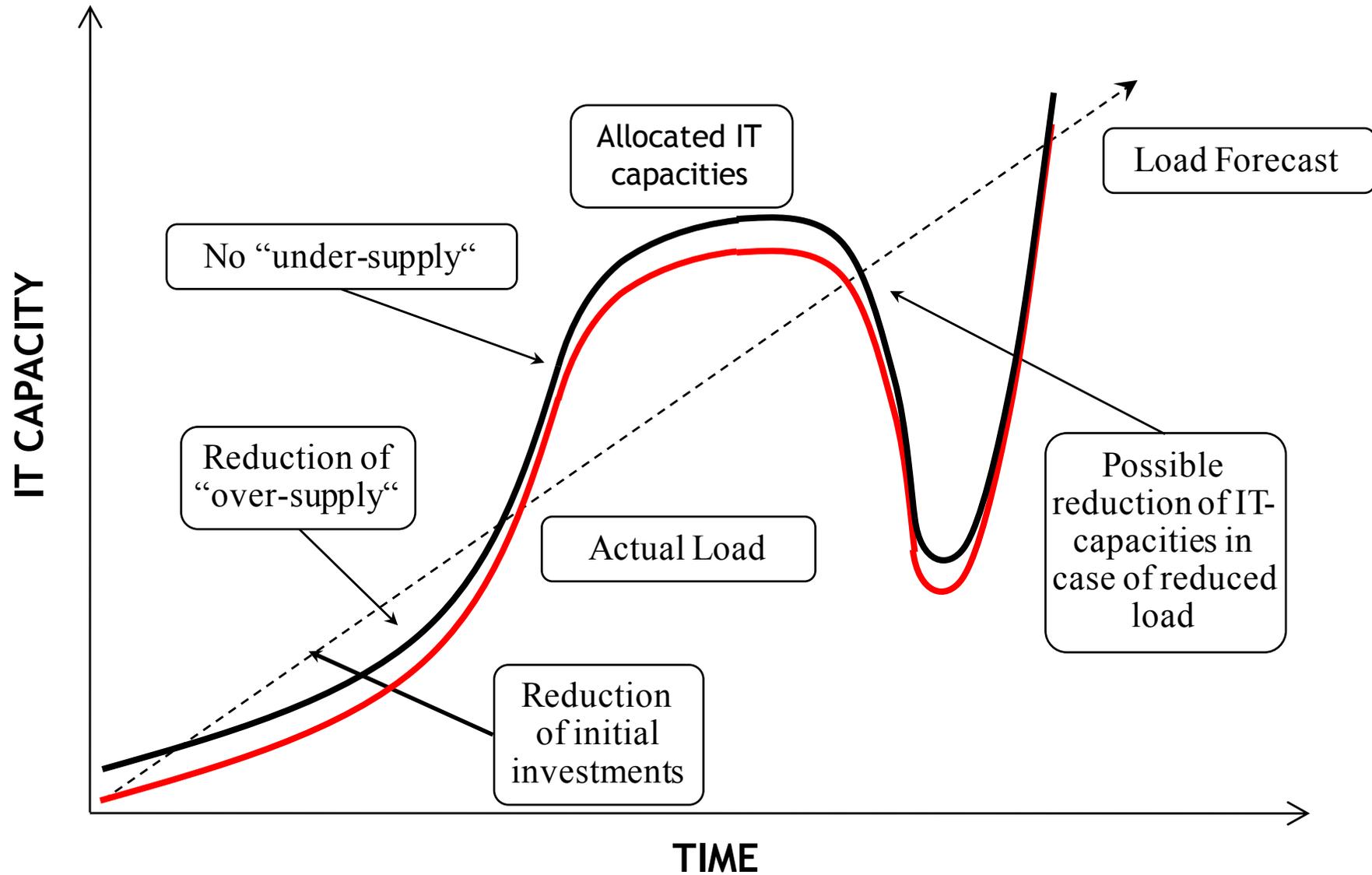
cloud private

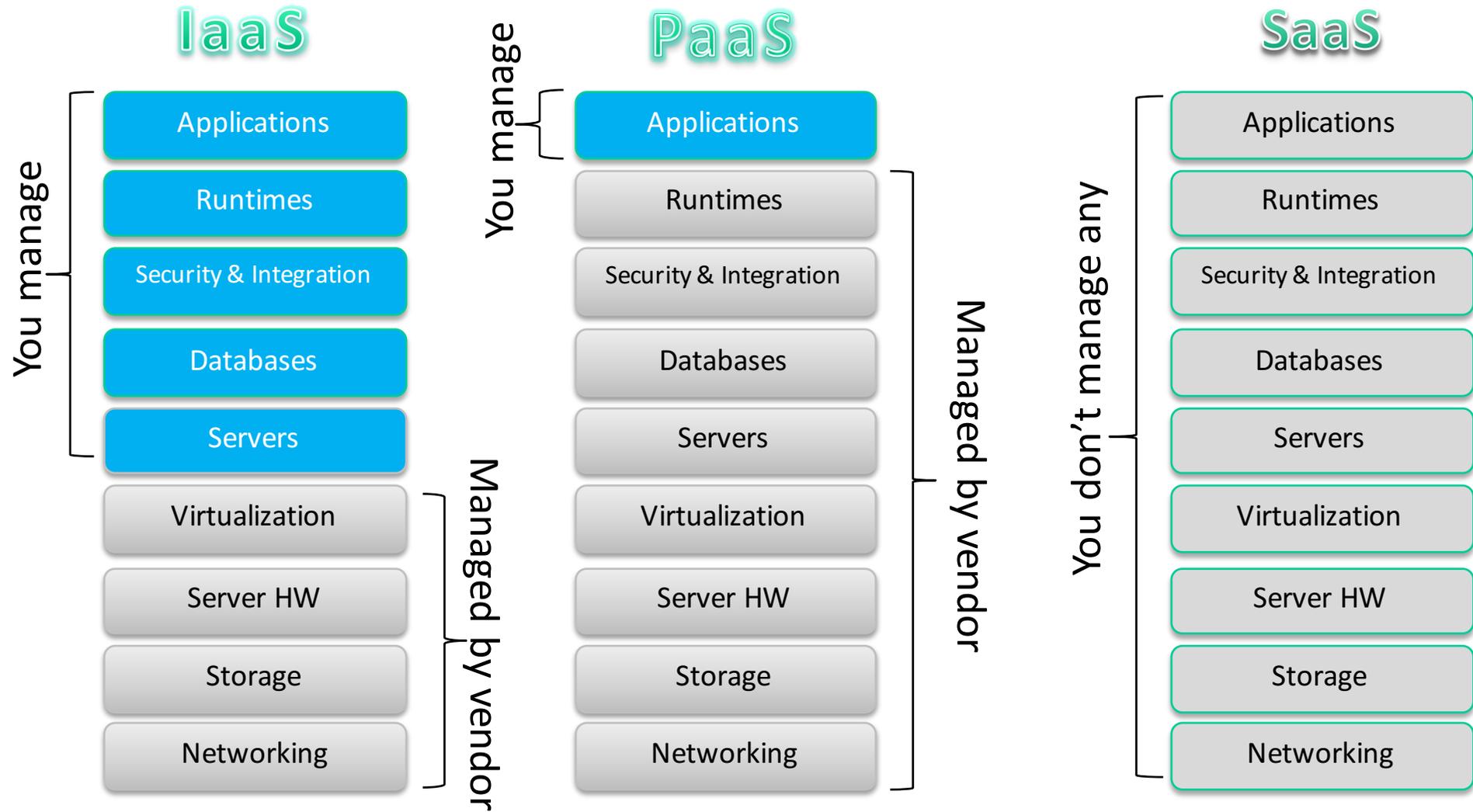


Dalla prospettiva degli utenti...

- Una soluzione attraente per l'installazione ed esecuzione di software
- La possibilità di offrire servizi utili alla programmazione che semplificano l'attività di sviluppo











Un data center cloud



Google google.com/datacenters



 POLITECNICO DI MILANO



Internet delle cose (IoT)?



Definizioni di internet delle cose

- The Internet of Things (IoT) is the network of physical objects—devices, vehicles, buildings and other items embedded with electronics, software, sensors, and network connectivity—that enables these objects to collect and exchange data [Internet of Things Global Standards Initiative] <http://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx>
- The Internet of Things allows objects to be sensed and controlled remotely across existing network infrastructure [Harvard Business Review, Nov 2014] https://hbr.org/resources/pdfs/comm/verizon/18980_HBR_Verizon_IoT_Nov_14.pdf



- Dispositivi dotati di intelligenza
- Identificabili e indirizzabili in modo univoco
- Geo-localizzati
- Architettura ad eventi
- Capacità della rete di scalare per indirizzare trilioni di oggetti e per spostare grandi quantità di dati



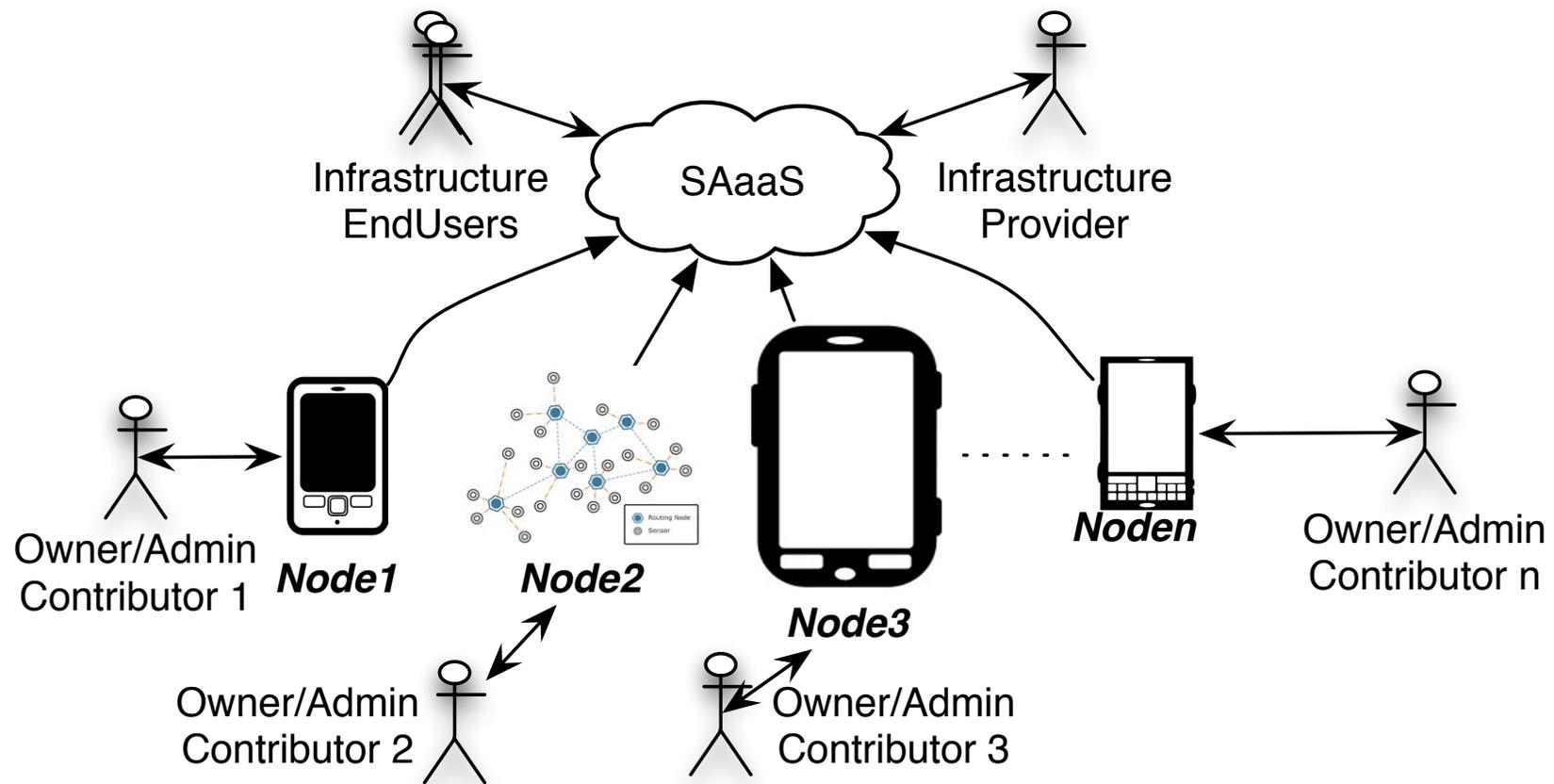
 POLITECNICO DI MILANO



Convergenza tra cloud computing e internet delle cose



Sensing and actuation as a Service - SAaaS





 POLITECNICO DI MILANO



PoliCloud

Obiettivi e Architettura



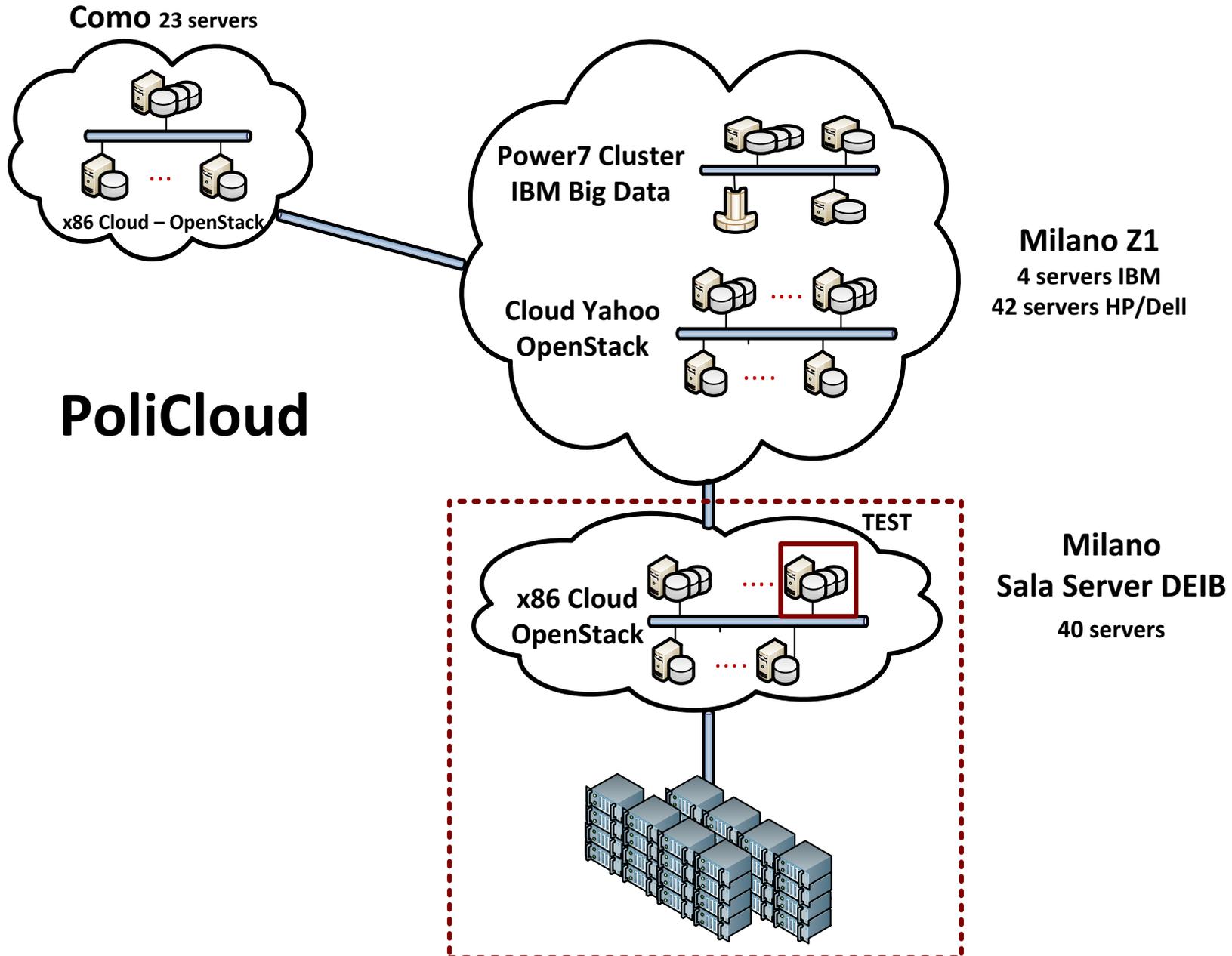
- Obiettivo: sviluppare un ambiente per la **sperimentazione** e **ricerca** nell'uso integrato di Cloud Computing e IoT
 - Un sistema che usa in modo **coordinato** Cloud e IoT è in grado di acquisire, processare e analizzare grandi quantità di dati, con un elevato livello di scalabilità e affidabilità.
- Indicatori di successo
 - Numero di pubblicazioni: 3 nel primo anno, 10 nel secondo e terzo anno, di cui metà a carattere multidisciplinare
 - Numero di proposte di progetto o collaborazioni con aziende esterne: 4 ogni anno di progetto



- Disponibilità di una piattaforma integrata Cloud e IoT a disposizione dei ricercatori
- Costituzione di gruppi di ricerca interdisciplinari
- Consolidamento dell'interazione con aziende esterne
- Incremento della partecipazione a bandi europei e nazionali che richiedano competenze interdisciplinari
- Formazione sul campo di dottorandi, tesisti e studenti di corsi avanzati



- Finanziamento del Politecnico 142.000 Euro
- Cofinanziamento da parte di gruppi di ricerca DEIB: finora 87.000 Euro + 8.000 Euro da parte del DEIB + 10.000 Euro Polo di COMO
- Donazione server Yahoo
- Dispositivi di proprietà dei singoli ricercatori



PoliCloud



Donazione Yahoo - Ott. 2013

220 server donati da Yahoo

- 2 processori quad-core,
- 16GB RAM
- Diskless

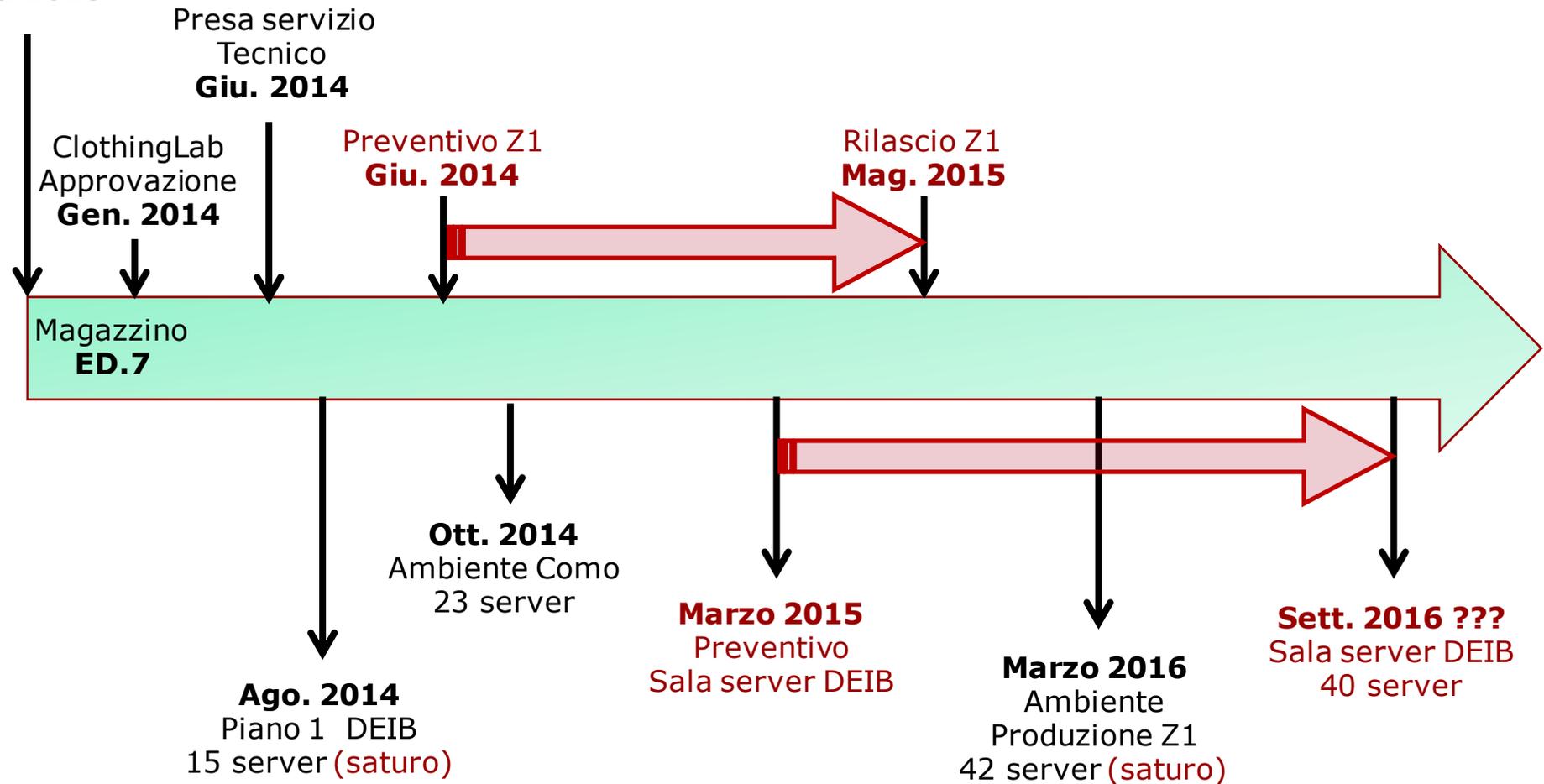
Alcuni dispositivi IoT di proprietà
dei singoli ricercatori





Evoluzione delle infrastrutture necessarie

227 server YAHOO
27 Dell, 200 HP
Ott. 2013







Collocazione dell'ambiente di test

23





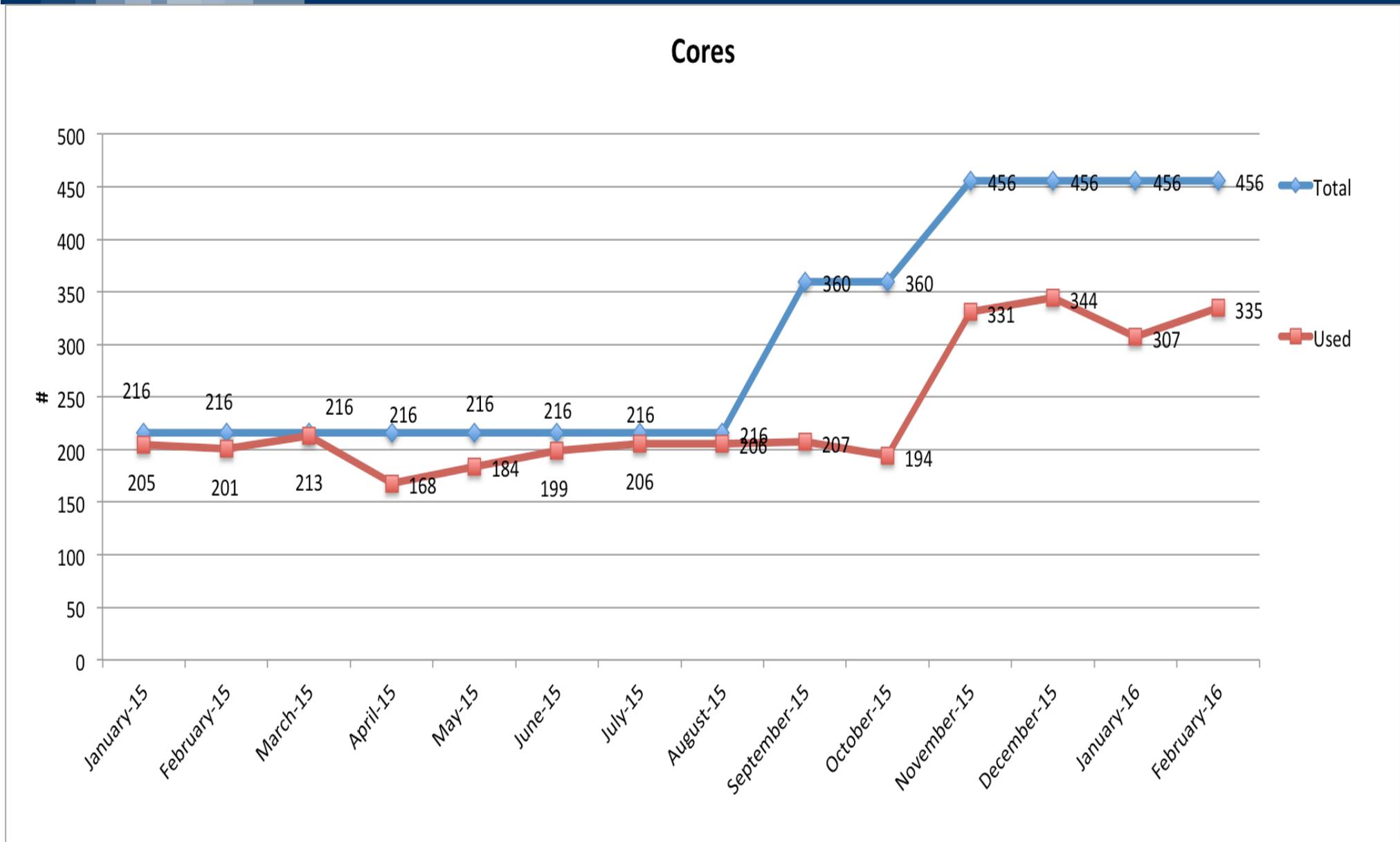
Collocazione dell'ambiente di test

24





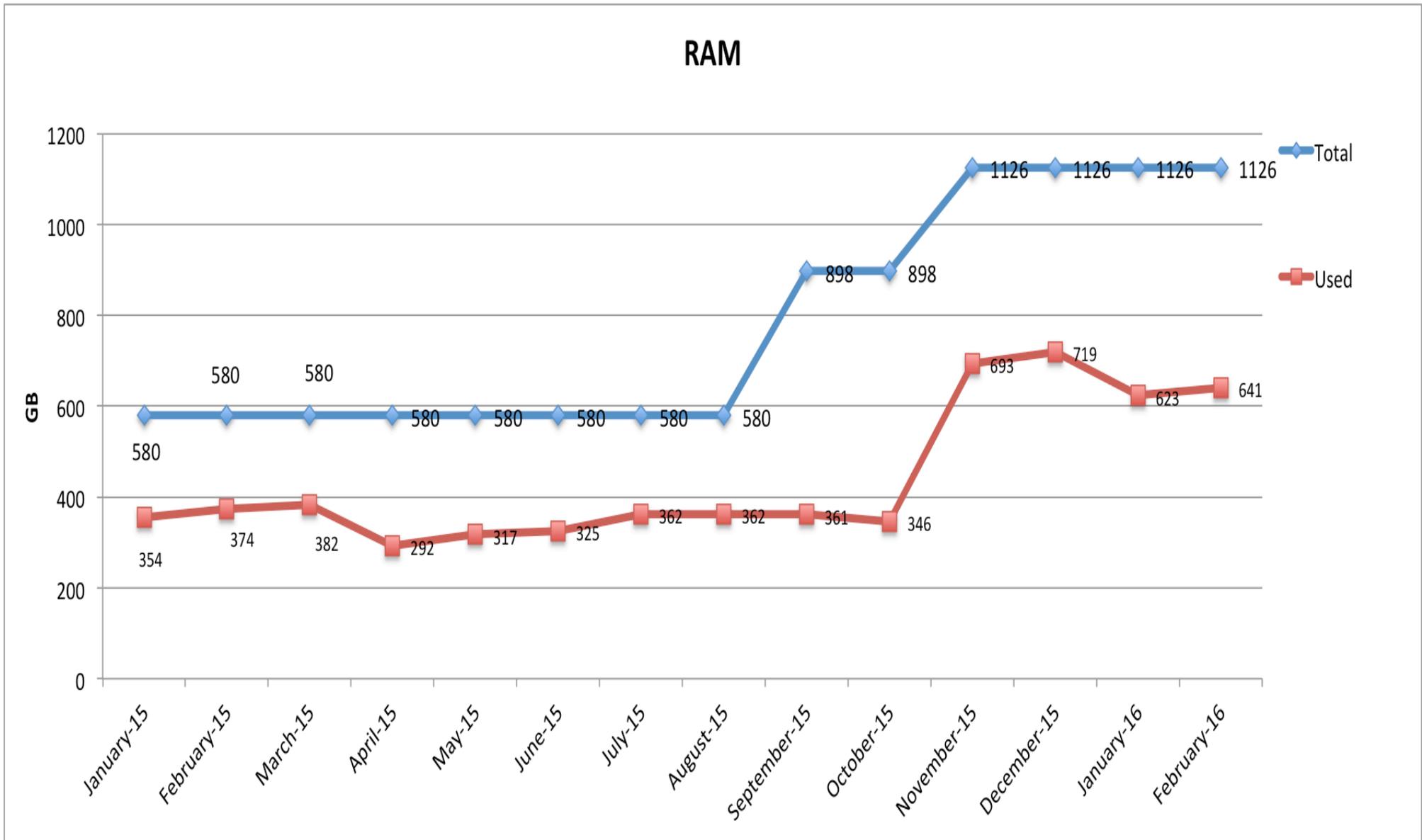






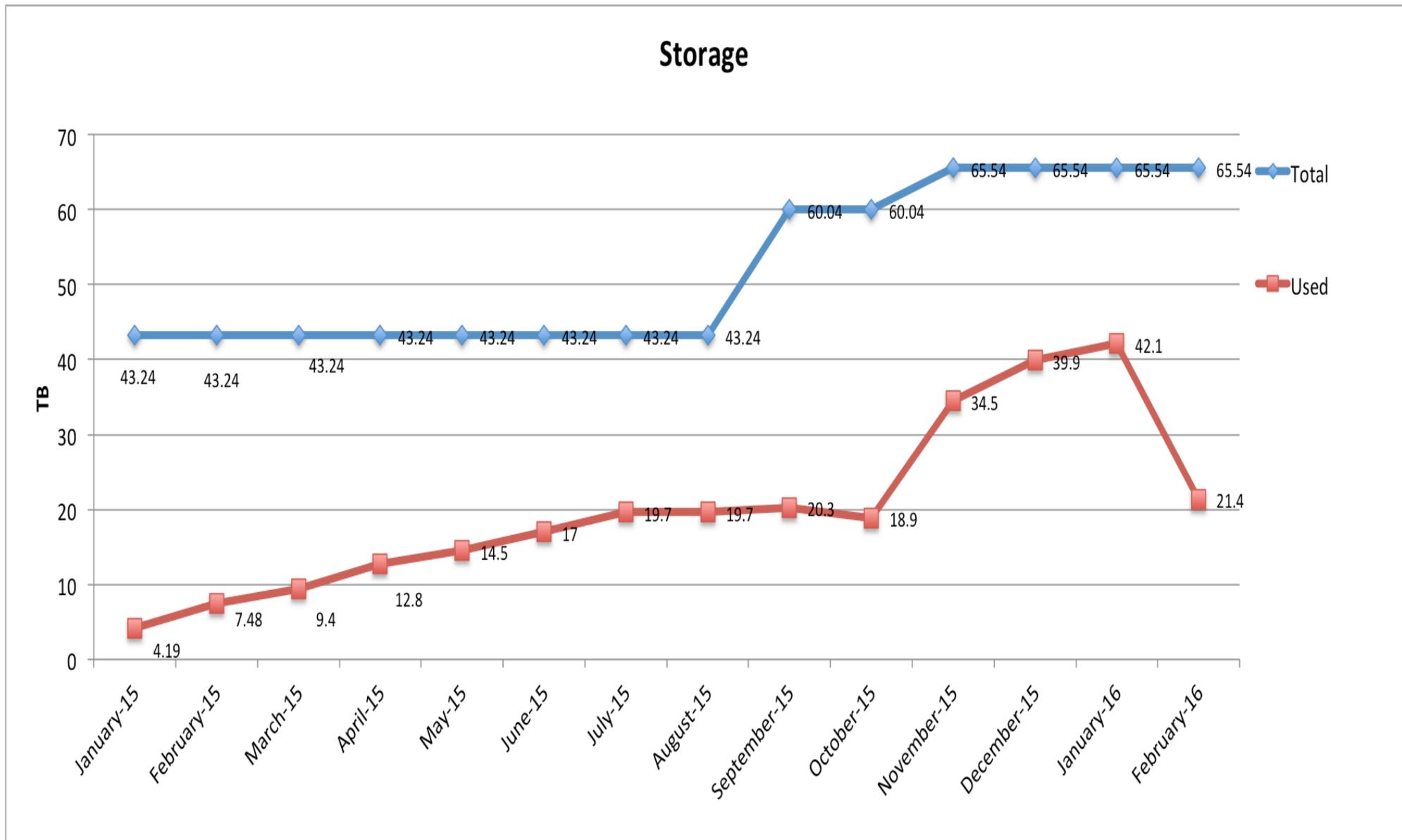
RAM (Gbytes)

RAM



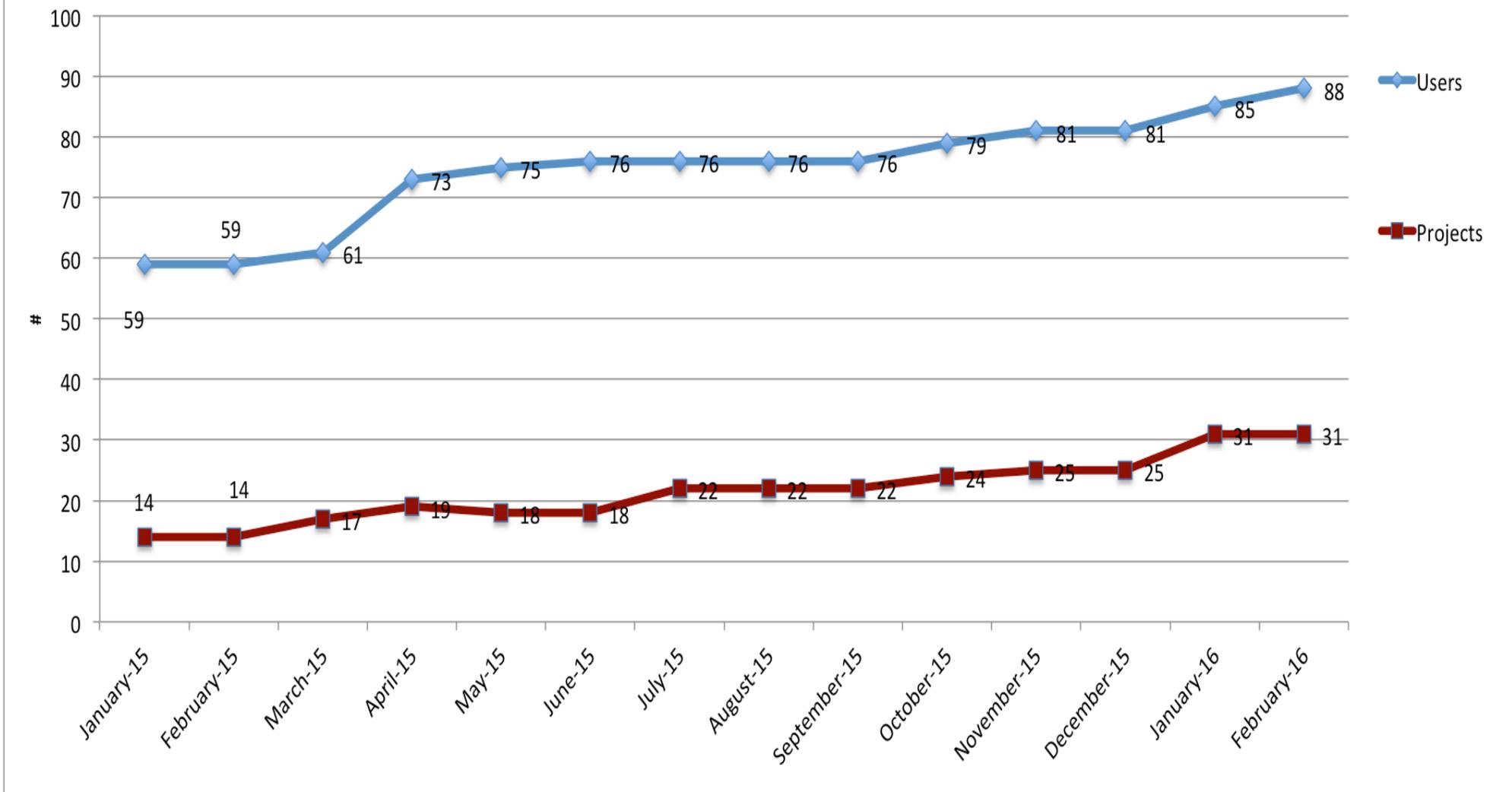


Storage (TBytes)





Users and projects





Snow watch



FACE

Formal Avenue for Chasing malwarE



PROtezione del territorio con infrAstrutture ICT avanzate, cIttadinanza attIVA e rEti sociali

Droydseuss

Ransom

Admire

EIT StreetSmart Retail

Mise-en-scène



- Competenze
 - Amministratore di sistemi cloud
 - Approcci e tecniche per big data e per ambienti di virtualizzazione
 - Dimensionamento di risorse
 - Sviluppo di esperimenti su larga scala
- Servizi Offerti
 - Possibilità di ottenere VM secondo il paradigma IaaS
 - Sviluppo di esperimenti controllati (uso e stress test)
 - Possibilità di acquisire l'hw del laboratorio per sperimentazioni a livello meccanismi di virtualizzazione
- Stiamo studiando l'attivazione di un billing tipico dei cloud



 POLITECNICO DI MILANO



PoliCloud Applicazioni



Utilizzo didattico

PoliCloud è stato utilizzato dagli **studenti** del corso di **Algorithms and Parallel Computing (Laurea Magistrale in Ing. Matematica)** come piattaforma su cui esercitarsi per mettere in pratica le nozioni apprese sul calcolo ad alte prestazioni

- AA 2014/15 e 2015/16
- titolare del corso: Paolo Cremonesi, DEIB

È stato inoltre utilizzato in altri corsi, titolari Plebani, Ardagna



- Sistemi di personalizzazione nell'ambito fashion utilizzati durante la fashion week a Milano.
- PoliCloud ha permesso di analizzare il comportamento di 1 milione di clienti dello shopping retail a Milano
 - coordinatori dei task Polimi: Paolo Cremonesi e Franca Garzotto, DEIB
 - altro personale DEIB coinvolto: William Fornaciari, Antonella di Rienzo, Mona Nasery



- Raccomandare film, video e musica, basandosi sugli aspetti stilistici ed artistici del contenuto.
- Il progetto ha richiesto l'analisi dei contenuti multimediali e di grossi volumi di dati (più di 13000 film per circa 26 TB di spazio) e non sarebbe stato possibile senza l'utilizzo di PoliCloud
 - coordinatore: Paolo Cremonesi, DEIB
 - altro personale DEIB coinvolto: Massimo Quadrana, Yashar Daldo, Mehdi Elhai
- Un team di 6 studenti Polimi utilizzano la recente installazione di PoliCloud in Z1 per partecipare alla RecSys competition 2016 organizzata dal MIT
 - coordinatore: Roberto Pagano, DEIB



SCUOLa - Smart Campus as Urgan Open Labs

Main goal: a platform to maximize energy efficiency in large communities

SCUOLa Project (2 years / 800 k€)
DEIB/DE/ABC

Highlighted Components
are deployed in Polycloud

Energy Information Base (prof. Bolchini)

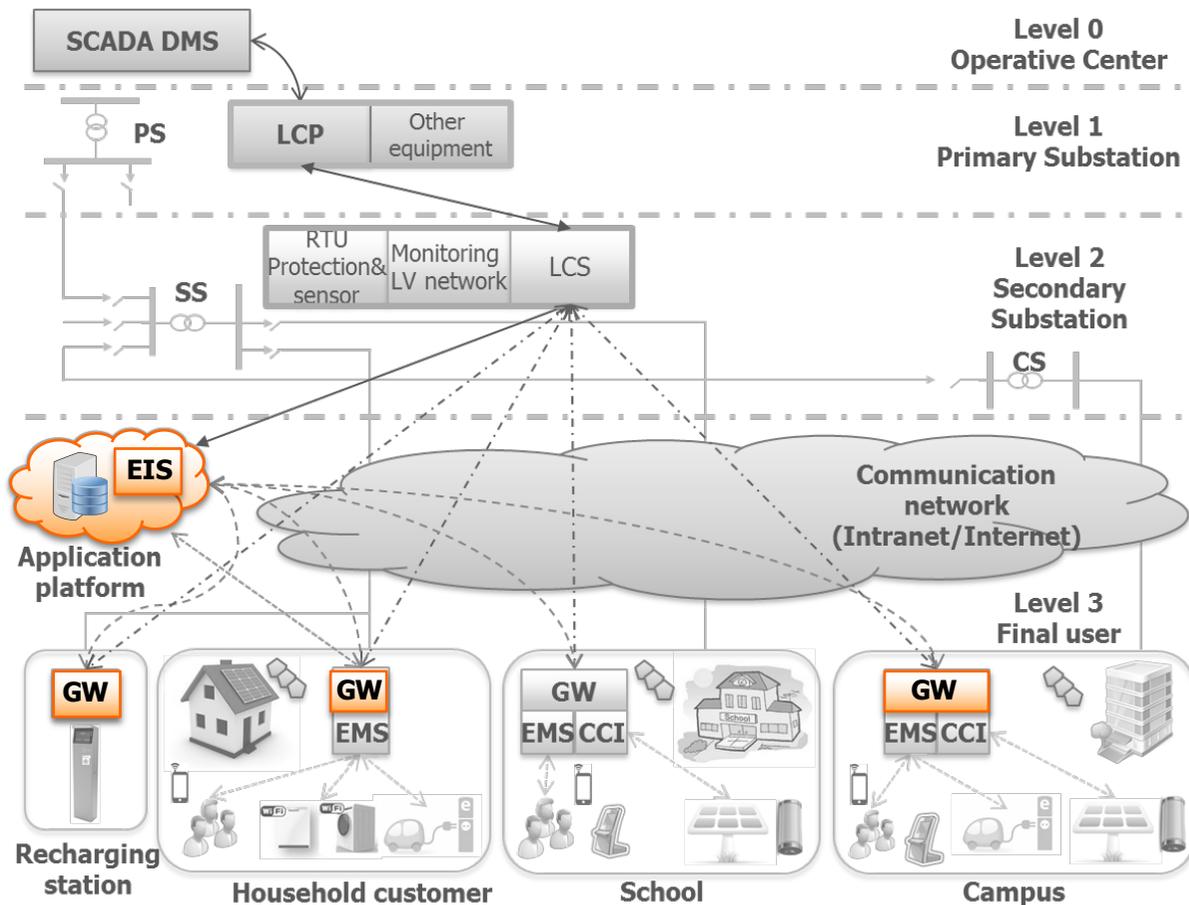
GW (proff. Capone/Verticale) to

- Electric Vehicles Recharging Stations
- PV and trigenerator controls

People from prof. Capone group
involved:

- 2 faculty
- 2 postdoc
- 2 MSc students
 - 1 INF-MI
 - 1 Energy-PC

1 journal paper (+1 submitted)
2 conf papers (+1 submitted)





Performance profiling of big data applications

By Danilo Ardagna

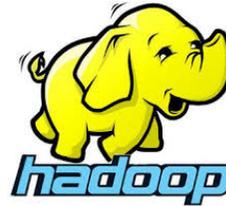
- Research motivations and goals
 - Many sectors of our economy are guided by data-driven decision
 - Data-driven modeling and hypothesis generation have a key role to understanding complex system behavior and interactions
 - Big Data Frameworks:
 - Hadoop Map Reduce: “Legacy” for large volume unstructured data
 - Spark: New “Map Reduce” engine 3-10x faster than Hadoop
 - Predicting the performance of jobs in such environments is very complex
 - Determining the optimal number of nodes in a cluster is an important and challenging problem



Define performance models and techniques for capacity planning



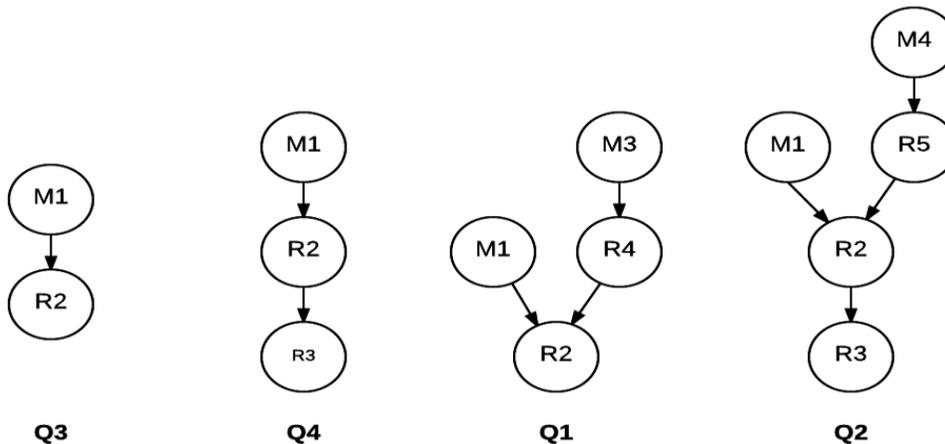
Hadoop performance modelling



Single querie

Colored Petri Nets

TPC-DS (3-40-250-500GB) benchmark, 4 ad-hoc Hive-Tez queries:



Test case	Avg error
3GB - 5 nodes	3.5%
40GB - 5 nodes	10%
40GB - 30 nodes	12.5%
250GB - 30 nodes	5.3%
500GB - 30 nodes	8.6%

Concurrently running queries

Test case	Avg error
40GB - 30 nodes	17.25%
250GB - 30 nodes	18%
500GB - 30 nodes	26.3%

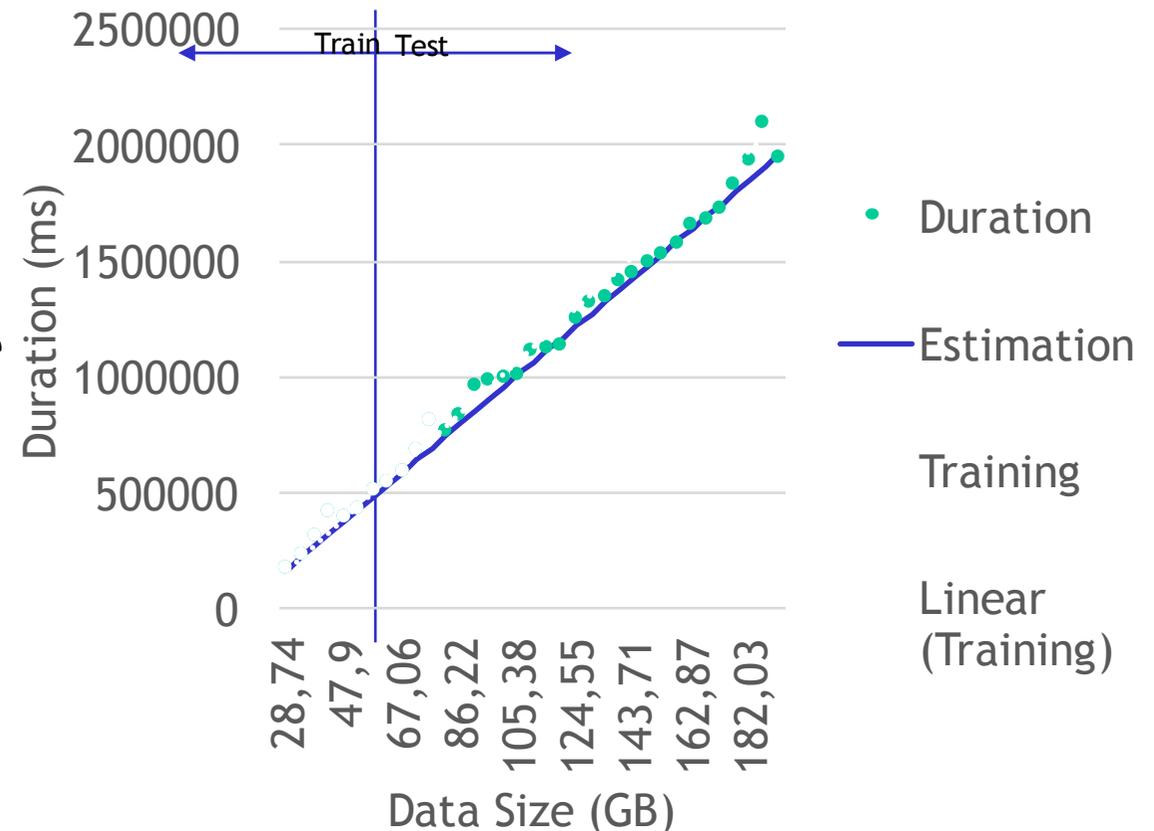
39



Spark performance modelling



- Gray-box approach based on machine learning
- SparkBench benchmarks iterative algorithms (linear and logistic regression, page rank, k-means)
- Generalization capabilities
- Average error 2%



Joint work with IBM T.J. Watson Research Center



SI projects on Cloud Computing

Ongoing projects on Policloud:

- Tweet analysis for veracity evaluation (thesis, Cappiello)

Planned projects to use Policloud

- EUBrazil-BIGSEA the information system group is involved in the definition of Data Quality-aware Service (DQaS) architecture (Cappiello, Vitali)
- PhD thesis: evaluation of energy efficiency adaptive application management techniques (Nguyen Ho, sup. Pernici)
- Proposal H2020 Marie Skłodowska Curie on (under evaluation in 2016, Pernici, Vitali)
- Teaching support using Policloud (Plebani)

Previous EU FP7 project on Cloud computing

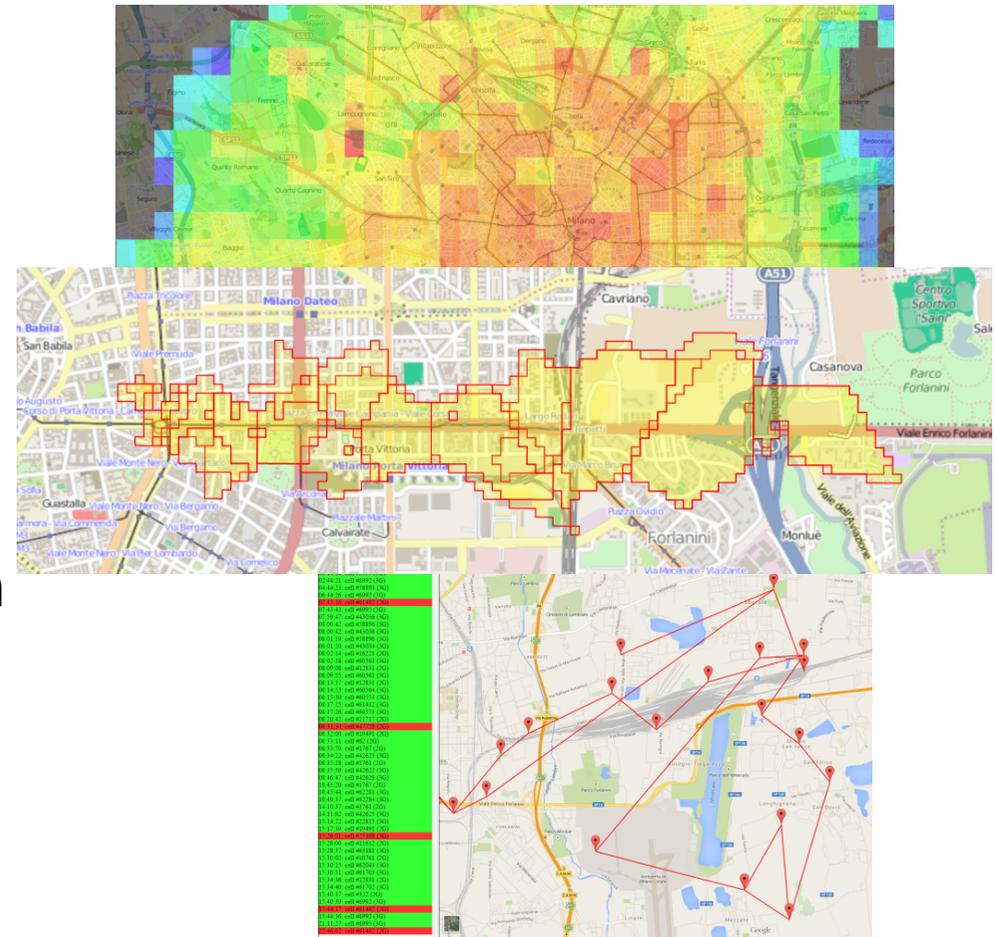
ECO2Clouds <http://eco2clouds.eu>

- Environmental impact of cloud applications
- All group participants involved, including Fugini



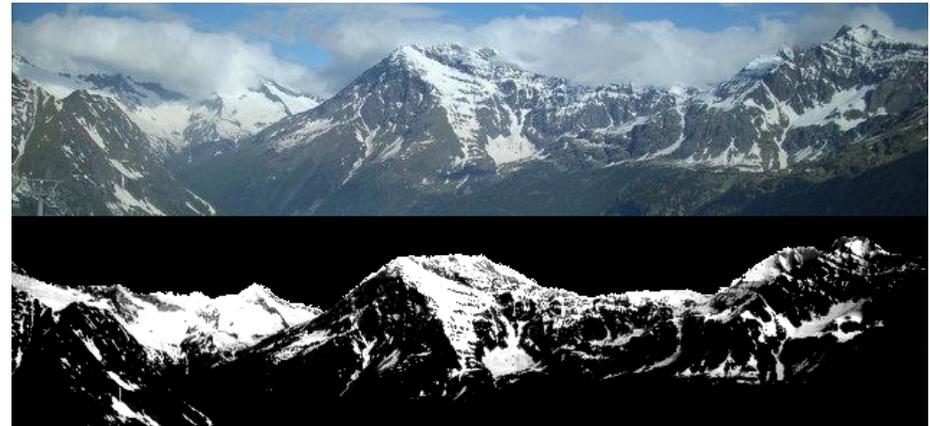
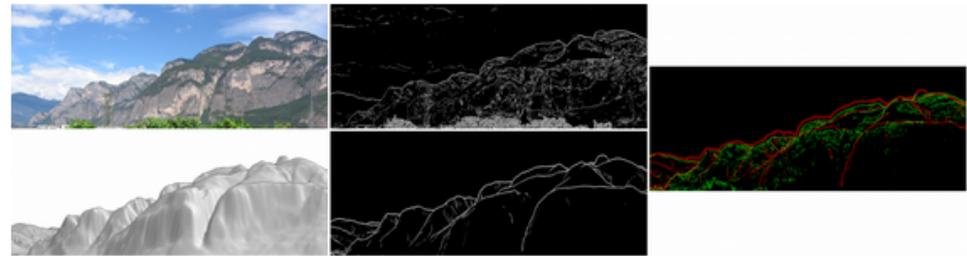
PeopleOnTheMove

- Project:
<http://www.proactiveproject.eu/>
- Data: cellular phone traces, metropolitan area of Milan
- Goal: predicting the presence of people in space and time
- Use of cloud: data collection and processing.
- ~250.000.000 of cellular events every day



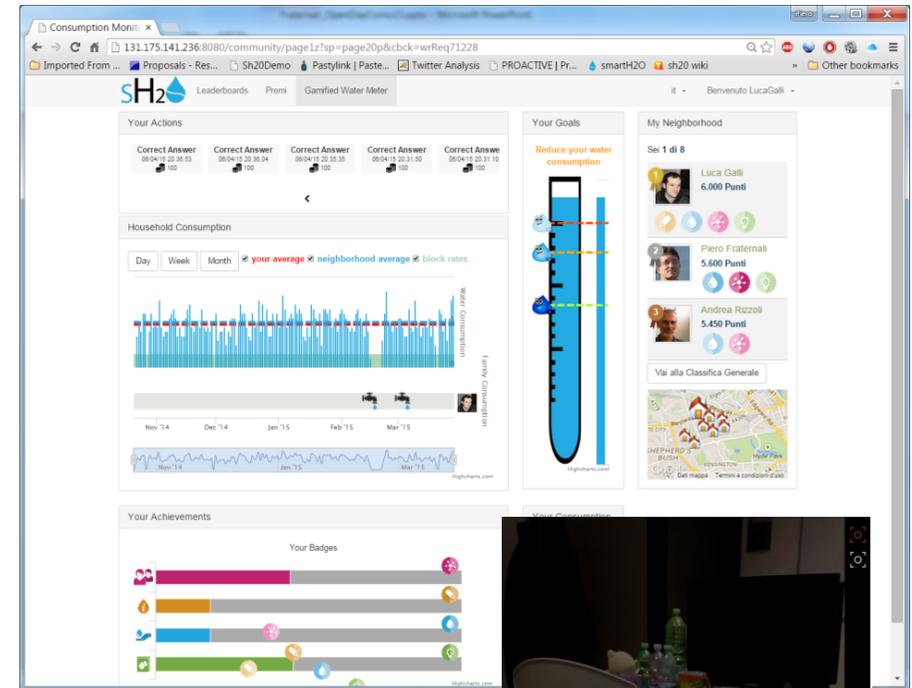
SnowWatch

- Project:
<http://www.proactiveproject.eu/>
- Data: images crawled from social media and web cams
- Goal: analyzing the status of snow cover to predict water availability
- Use of cloud: massive image crawling and storage, image processing (classification, peak recognition, snow extraction) predictive modeling.
- ~2.000 webcams crawled at maximum frequency, 10GB of data downloaded and stored daily



SmartH2O

- Project: <http://smarth2o-fp7.eu/>
- Data: smart meter readings, user generated content, game play traces
- Goal: monitoring user's actions and recommend sustainable behaviors
- Use of cloud: IoT integration, data fusion, user modeling, demand prediction, gamification of actions, action recommendation





 POLITECNICO DI MILANO



Conclusioni



Risultati conseguiti rispetto agli indicatori di successo

- Numero di pubblicazioni: 3 nel primo anno, 10 nel secondo e terzo anno, di cui metà a carattere multidisciplinare
 - Primo anno: 10 pubblicazioni
 - Secondo anno: 23 pubblicazioni di cui 4 a carattere multidisciplinare
- Numero di proposte di progetto o collaborazioni con aziende esterne: 4 ogni anno di progetto
 - Primo e secondo anno: 12



- Disponibilità di una piattaforma integrata Cloud e IoT a disposizione dei ricercatori
 - Stiamo lavorando su questo punto
- Costituzione di gruppi di ricerca interdisciplinari
 - Dipartimenti e sezioni coinvolte in attività svolte in PoliCloud: DICA, ABC, DESIGN, MAT, BIO, ELET, AUT, TLC
- Consolidamento dell'interazione con aziende esterne
 - Aziende interessate nell'utilizzo di PoliCloud: ATOS, CA, IBM
- Incremento della partecipazione a bandi europei e nazionali che richiedano competenze interdisciplinari
 - 12 proposte presentate
- Formazione sul campo di dottorandi, tesisti e studenti di corsi avanzati
 - Degli 88 utenti di PoliCloud la maggior parte sono studenti a vari livelli



- Incremento del numero di utenti
- Attivazione del meccanismo di billing
- Rinnovo dei server Yahoo
- Nuova installazione nella sala server DEIB



- Fondi privati
 - 95.000 Euro spesi
 - 10.000 in cassa
- Fondi di ateneo
 - 45.000 spesi
 - 97.000 in cassa



- Sale macchine adeguate
- Staff tecnico
- Configurazione di rete
- Connessioni di backbone Milano - Como, DEIB - Z1



 POLITECNICO DI MILANO



Discussione