

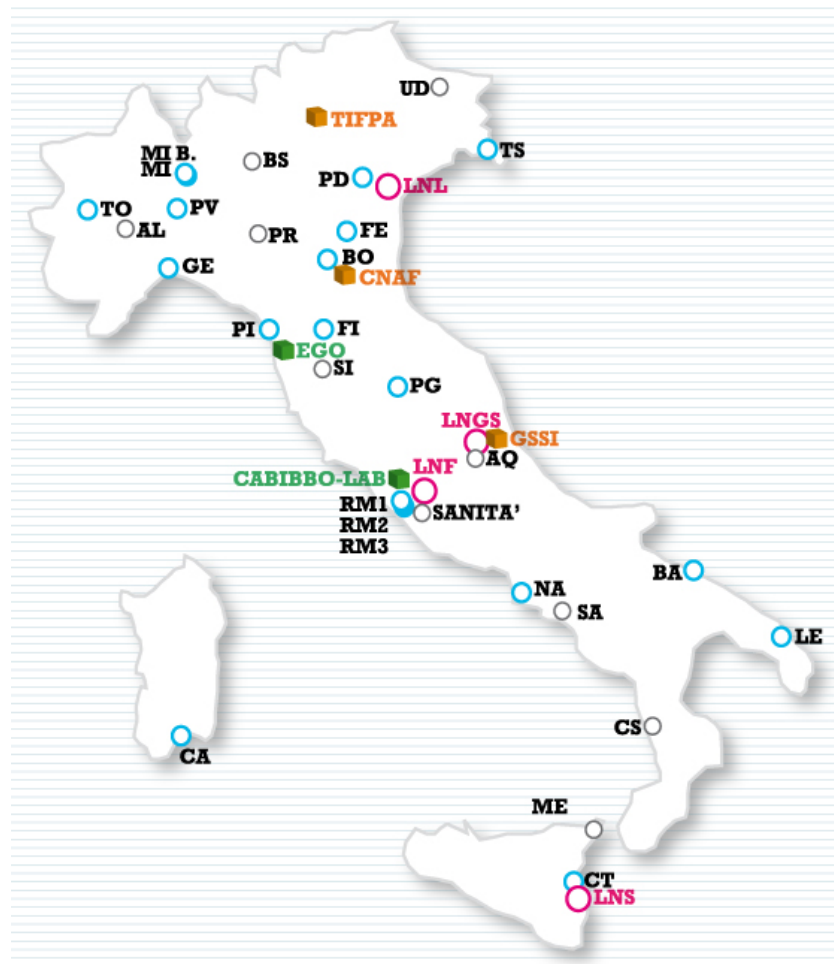
Grid e Cloud Computing nell'INFN

Davide Salomoni, INFN-CNAF
(davide.salomoni@cnafe.infn.it)

Workshop ICT INAF, 18/9/2014

Grid Computing nell'INFN

- Vedi slides di Dario per una introduzione al calcolo distribuito nell'INFN.
- In sintesi:
 - Dal punto di vista dell'**infrastruttura**, la maggior parte dei siti INFN sono connessi via Grid Computing tra di loro e alle infrastrutture mondiali (come WLCG, EGI) per calcolo scientifico di tipo HTC.
 - Dal punto di vista del **software**, l'INFN ha sviluppato e mantiene una importante suite di componenti per il Grid Computing.
 - Computing Element (CREAM), Accounting (DGAS), Storage Element (StoRM), Authentication/Authorization (VOMS/ARGUS), brokering (WMS).

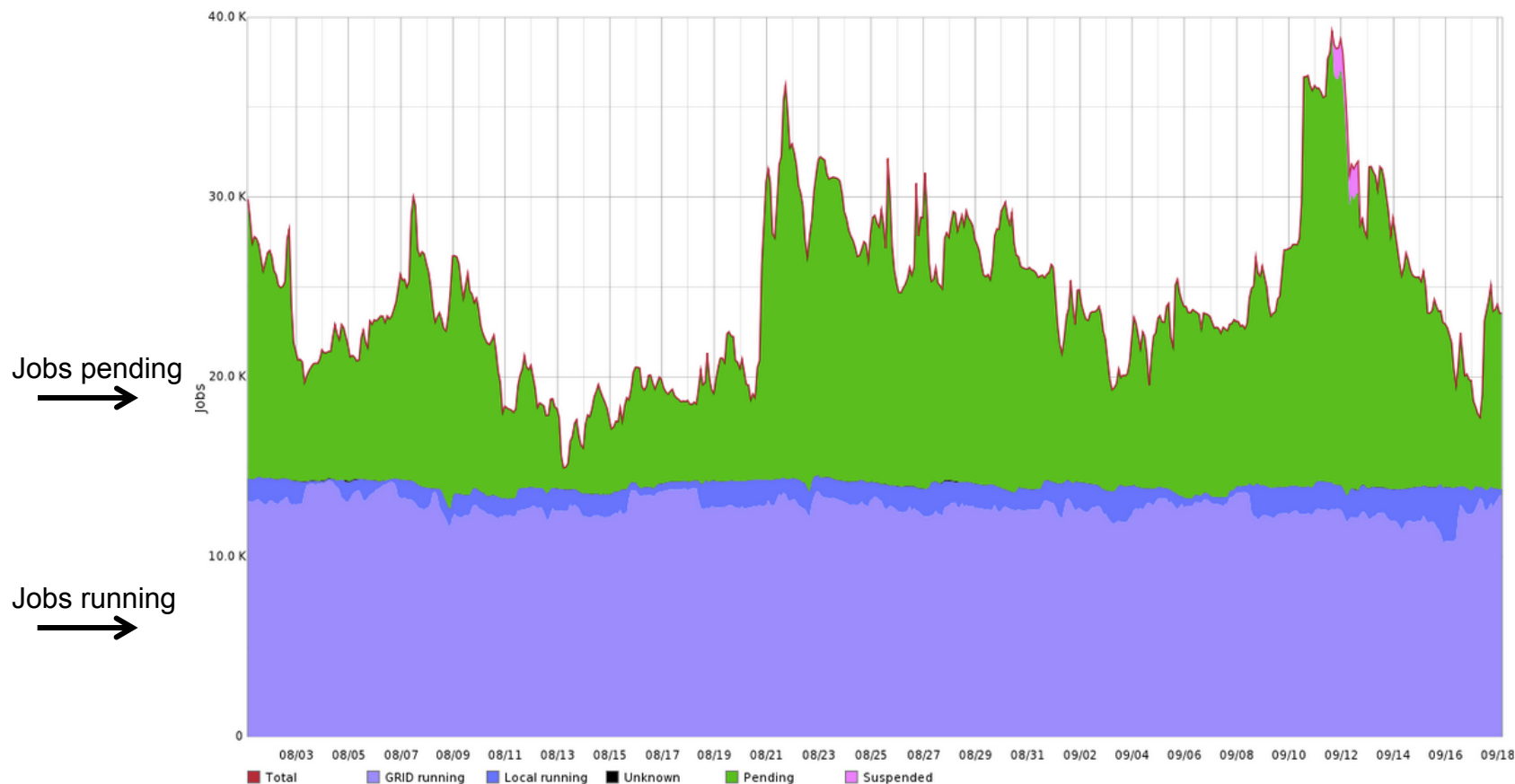


Esempio: il Tier-1 dell'INFN

- Tier-1: il Centro di calcolo nazionale dell'INFN, con sede presso il CNAF di Bologna.
 - **Sito multi-purpose:** 20 esperimenti scientifici supportati.
 - **Sharing di risorse locali:** un unico grande cluster di calcolo/storage, con condivisione delle risorse tipicamente via fairshare.
 - Attualmente circa 15.000 core di calcolo, 15 PB di spazio disco, 18 PB di spazio nastro.
 - Utilizzo in produzione dal 2009 di VM dinamicamente create per la gestione di custom environment (es. sistemi operativi legacy, richiesta di specifiche personalizzazioni di applicazioni/librerie/ etc).
 - **Sharing di risorse distribuite:** attraverso Grid Computing.

Utilizzo delle risorse

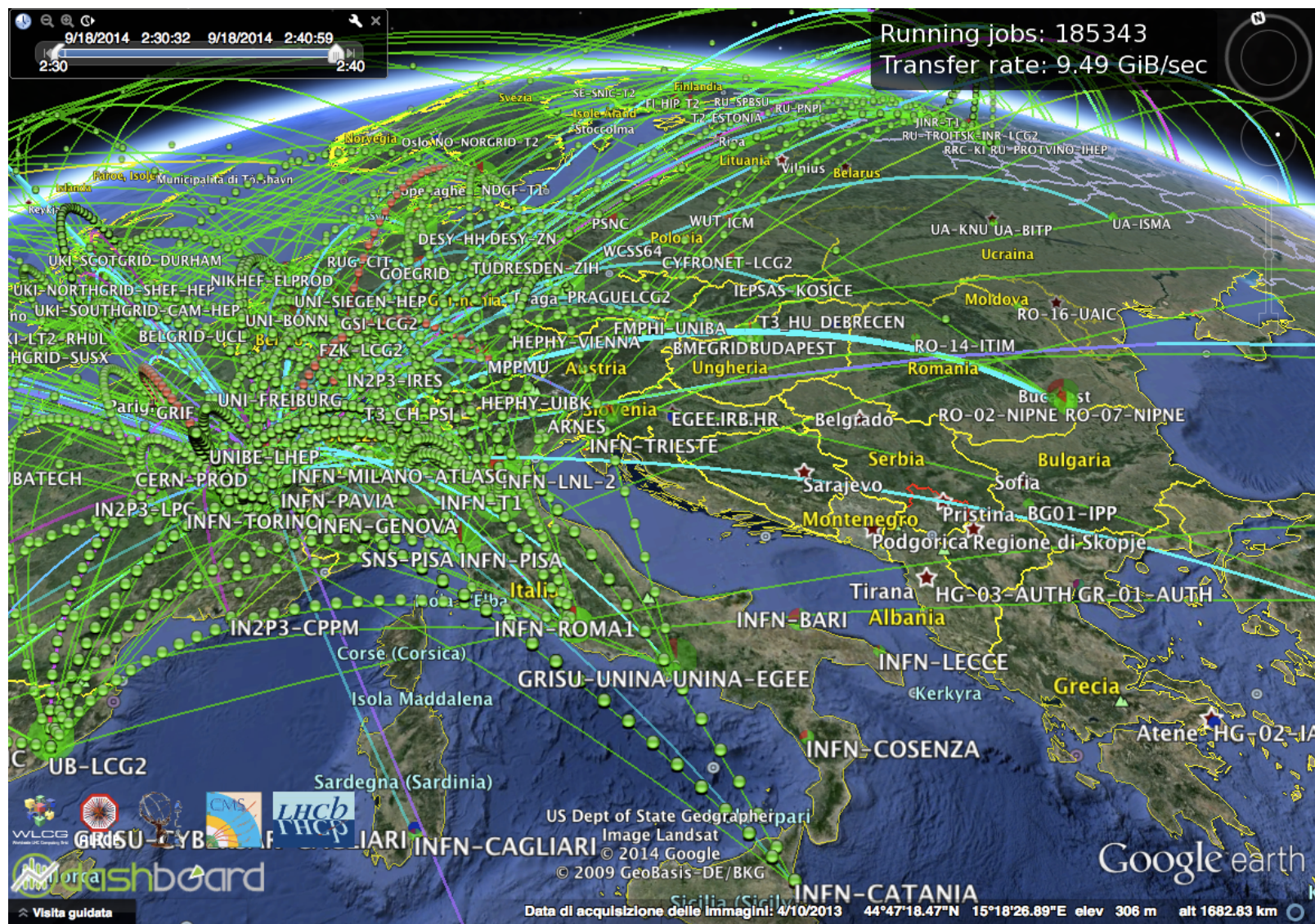
- Esempio d'uso delle risorse CPU al Tier-1 dal 1/8 al 18/9/2014
 - Notare che la grande maggioranza di job è di tipo Grid, e che tipicamente l'utilizzo del cluster è 100% 24/7 con 10k-20k job in coda, gestiti attraverso fairsharing delle risorse.



Grid Computing (I)



Grid Computing (II)



Trends

- I trends (<http://www.google.com/trends/>) di *Grid Computing* vs. *Cloud Computing*

Esplora le tendenze

Ricerche più frequenti

Termini di ricerca ?

× Grid computing

× Cloud computir

+ Aggiungi termine

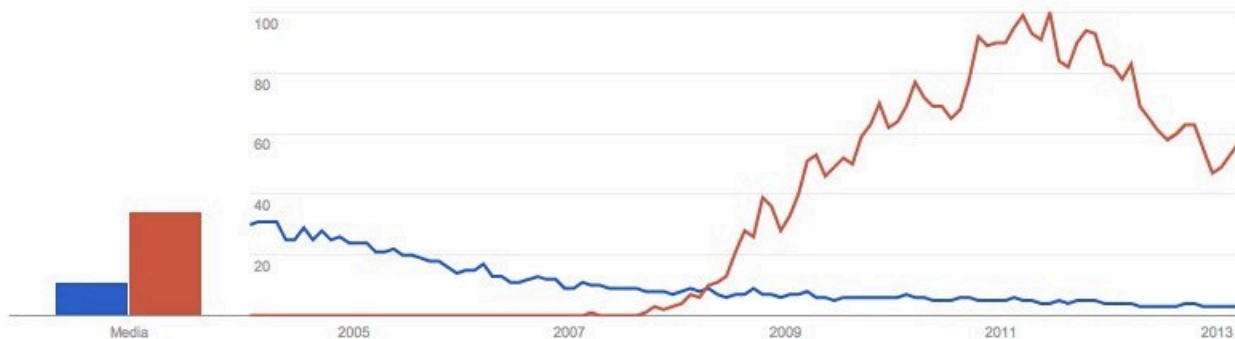
▶ Altri confronti

Limita a

Interesse nel tempo ?

Il numero 100 rappresenta il picco del volume di ricerca

☐ Intestazioni notizie ☐ Previsione ?



Perché questi trend?

- Il Grid Computing ha avuto un enorme successo in ambito scientifico HTC (cf. ad esempio “How grid computing helped CERN hunt the Higgs”, <http://goo.gl/dHvyGL>).
- Questo tuttavia ha funzionato (e funziona) soprattutto per **grandi collaborazioni scientifiche**.
 - Con una massa critica importante.
 - Naturalmente portate alla condivisione di risorse e tool.
 - Disponibili a “soportare” e/o adattare meccanismi di utilizzo delle risorse a volte non proprio user-friendly a livello sia sistemistico, sia di applicazioni.
- Il Grid Computing ha avuto **impatto sostanzialmente nullo al di fuori dell’ambito della ricerca scientifica**.
 - Il che pone problemi non semplici dal punto di vista della sostenibilità del software di Grid.

Cloud computing

- Non vi tedierò con le definizioni d'obbligo (cf. NIST, <http://goo.gl/eBGBk>), ma in sintesi il Cloud si occupa di

Fornitura di tecnologia di informazione e comunicazione (ICT) come servizio

- E si può riassumere in 5 caratteristiche principali:
 - **Self-service, on-demand**
 - **Accesso attraverso la rete**
 - **Pool di risorse**
 - **Elasticità**
 - **Pagamento a consumo**



Tuttavia...

- **“I come to bury Caesar, not to praise him.”** (JC, Act 3, scene II)
- Detto obbligatoriamente che oggi tutto è Cloud e Cloud è tutto, **che cosa manca** perché il Cloud diventi “mainstream” anche nel calcolo scientifico?
- **Diverse cose**, tra le quali:
 - Una allocazione sofisticata delle risorse per tenant multipli (vedi slide successiva, ma anche il plot di utilizzo del Tier-1 mostrato prima).
 - Meccanismi di autenticazione e soprattutto di policy di autorizzazione distribuiti.
 - Stabilità e standardizzazione del software (cf. i cicli di rilascio di OpenStack, per esempio).
 - Meccanismi di “federazione” tra resource providers, in particolare per cloud ibride (pubbliche + private).
 - Contorni legislativi chiari (cf. Terms of Contract di public Cloud providers, oppure le normative USA vs. EC vs. nazionali).

Il grande equivoco

- **La capacità delle risorse o dei resource providers non è infinita** (nonostante questo sia un assioma del Cloud).
 - Dunque, le risorse potrebbero essere non disponibili quando servono, o potrebbero non avere le caratteristiche che servono.



Nell'INFN, dunque...

- All'interno della Commissione Calcolo e Reti INFN, da un paio d'anni esiste un **Cloud Working Group**, con circa 80 iscritti, che lavora per la costruzione di una **Cloud nazionale INFN** principalmente su queste tematiche:
 - Ottimizzazione del calcolo distribuito all'interno di centri di calcolo esistenti, con **condivisione tra interfacce d'uso locali, Grid e Cloud**.
 - Meccanismi di **scheduling delle risorse nel Cloud computing**, in particolare con OpenStack.
 - OpenStack è la piattaforma Cloud di riferimento nell'INFN (esiste tuttavia una significativa esperienza anche con OpenNebula).
 - Sperimentazione **dell'utilizzo del Cloud computing principalmente per compiti di simulazione ed analisi** (oltre alla classica fornitura di risorse IaaS).
 - Sperimentazione di **condivisione di risorse tra diversi siti Cloud INFN**.
 - Questo implica la condivisione sia di informazioni di autenticazione (es. via Keystone in OpenStack) sia ad esempio di compute nodes o di storage (es. via Swift).
 - **Semplificazione di installazioni e di configurazioni**.
 - Ad esempio con creazione automatizzata di VM.

Progetti locali

- Attraverso e in aggiunta ai filoni generali del cloud-wg, diverse sezioni INFN hanno maturato **esperienze di installazione ed uso** di framework Cloud.



Cloud dell'area Padovana



- **Obiettivi**
 - Uso piu` efficiente delle risorse
 - Gestione sistemistica meno onerosa
 - Evitare la proliferazione di cluster dei diversi gruppi/esperimenti
- → **Implementazione di una Cloud tra la Sezione di Padova e i Laboratori Nazionali di Legnaro**
 - Per sfruttare le infrastrutture, il manpower, le competenze in entrambi i siti
 - C'e` gia` l'ottima esperienza del Tier-2 in tal senso
- **Integrata in una Cloud INFN Nazionale**
- **Sinergie con altri progetti Cloud (v. Progetto Cloud dell'Universita` di Padova)**

Workshop CCR, Maggio 2014
Massimo Sgaravatto
2/28

Trasferimento tecnologico

- E' un **punto fondamentale** e che fa parte del mandato dell'Ente.
- Abbiamo tenuto e teniamo **corsi di base e avanzati su Cloud computing** presso INFN (formazione interna), presso università e presso enti pubblici.
 - Ad esempio attraverso il progetto di trasferimento tecnologico “Marche Cloud” (cf. <http://goo.gl/3HTGwE>) con la Regione Marche. Analoghi progetti sono in fase di studio con altre Pubbliche Amministrazioni.

Progetti nazionali

- **PRISMA** (Piattaforme cloud Interoperabili per Smart-government), progetto di ricerca industriale finanziato dal PON Ricerca e Competitività, cf. <http://goo.gl/9Q2p6c>



- **OCP** (Open City Platform), Smart Cities and Communities and Social Innovation – Bando MIUR, cf. <http://goo.gl/56mD1I>



OCP, numeri (soft)

Durata: 30 mesi (1 gennaio 2014 – 30 giugno 2016)

Personale Coinvolto: oltre 100 unità

Mesi uomo: oltre 2000

Obiettivi Realizzativi: 11

Deliverable: 63 (tra ricerche e prototipi)

Budget Progetto: 11.949.448,89 euro

Totale Agevolazioni: 10.688.786,90 euro

Cofinanziamento: 1.260.662,00 euro

Oltre 9.000.000 di cittadini potenzialmente coinvolti dalle sperimentazioni regionali e comunali



ATI Marche



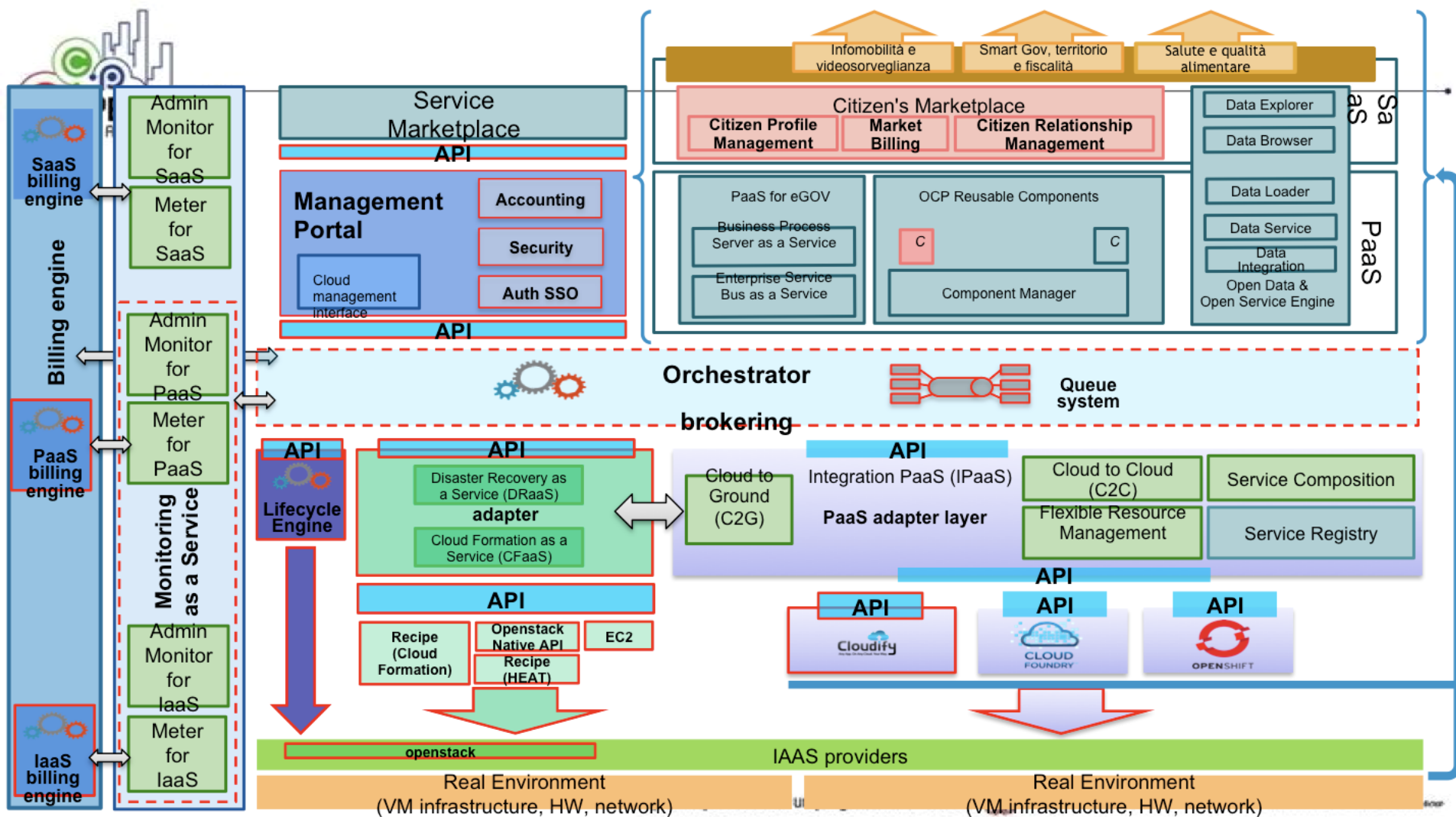
ATI Toscana/E.R.



OCP, obiettivi (*medium*)

- Gestione federata di piattaforme cloud eterogenee
- Integrazione di componenti PaaS per automatizzare l'attivazione su Cloud di nuovi servizi per la PA in particolare Paas for eGov
- Monitoraggio integrato e supporto a sistemi di billing
- Progettazione e reingegnerizzazione delle applicazioni Cloud
- Disaster recovery as a Service
- Open data e Open Service e integrazione in modelli di business
- Gestione federata dell'identità, autorizzazioni e dei relativi rapporti di trust

OCP, architettura (*hard*)



Progetti Cloud in Horizon 2020

- Dal 2013 abbiamo cominciato a seguire attivamente le **call H2020**, in particolare il programma 2014-2015.
- Per quanto riguarda lo specifico ambito di Cloud computing, abbiamo finora sottomesso due proposte:
 - **FONDACloud** (Federated Environment for Data Analysis in the Cloud), call ICT-7 (aprile 2014).
 - **INDIGO-DataCloud** (INtegrating Distributed data Infrastructures for Global ExpLOitation), call EINFRA-1 (settembre 2014), **proposta coordinata da INFN**.

INDIGO-DataCloud

- 26 partner europei (tra i quali INAF), comprendenti enti di ricerca, e-infrastructures, collaborazioni scientifiche, industrie.
- Budget richiesto: circa 11M€ su 30 mesi.
- Indirizza due argomenti della call EINFRA-1:
 - **Topic 4:** large scale virtualization of data/compute centre resources.
 - **Topic 5:** development and adoption of a standards-based computing platform (with open software stack).

INDIGO Work Programme

- **6 Work Packages**

- WP1 (NA): administrative and financial management, project quality assurance and global oversight of activities
- WP2 (NA): Support to Research Communities
- WP3 (SA): Software Management and Pilot Services
- WP4 (JRA): Resource Virtualization
- WP5 (JRA): PaaS Platform Development
- WP6 (JRA): Science Gateways, Workflows and Toolkits

INDIGO Development Tasks

- **WP4 (local virtualization):**
 - T4.1 – Cloud Computing Virtualization
 - T4.2 – Cloud Storage Virtualization
 - T4.3 – Network Virtualization
- **WP5 (PaaS development):**
 - T5.1 – PaaS architecture and implementation
 - T5.2 – Security and Authorization
 - T5.3 – High-level geographical application/service deployment
 - T5.4 – Unified Data Access
- **WP6 (Science Gateways, Workflows and Toolkits):**
 - T6.1 – Libraries and Toolkits
 - T6.2 – Science Gateways and User Interfaces
 - T6.3 – Support for big data driven workflows for e-Science

INDIGO Global Architecture

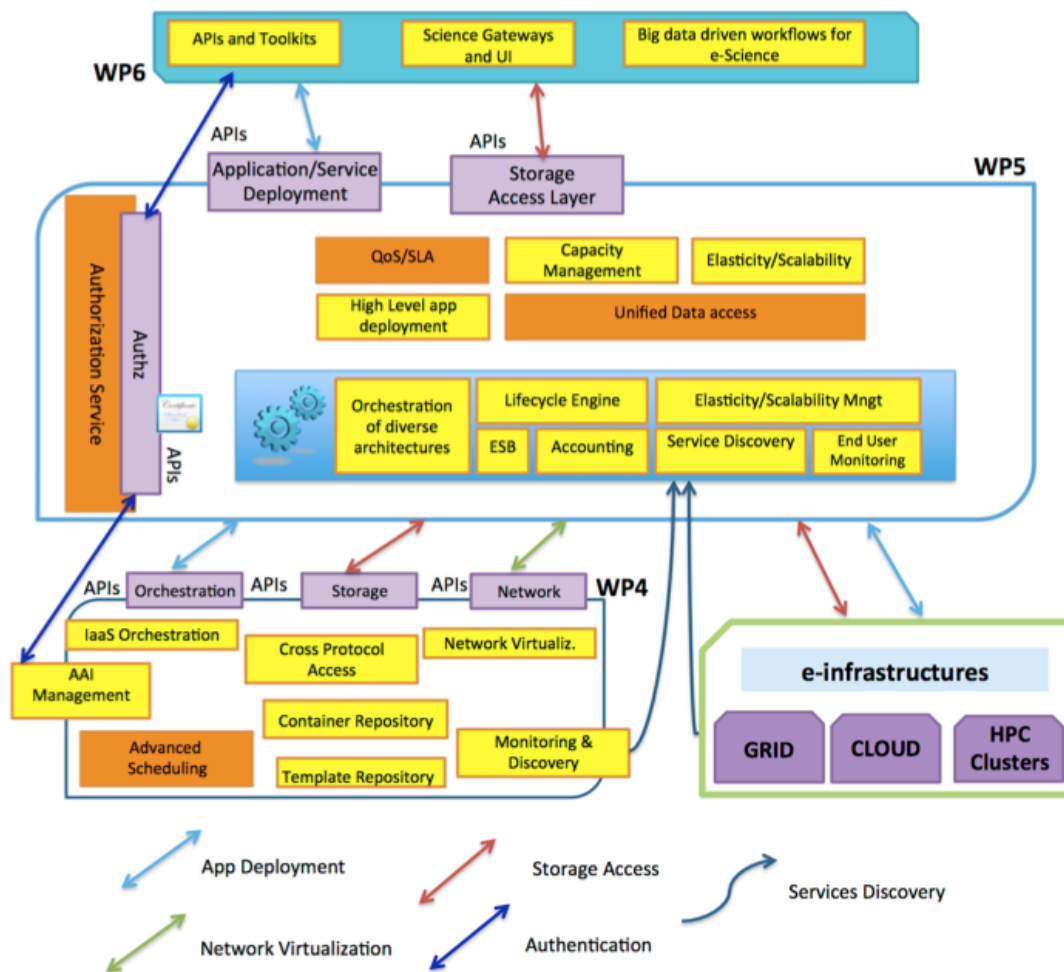


Figure 2: INDIGO global architecture. Color codes: Yellow: implementation based on already available solution to be improved/changed; Orange: newly implemented services.

Nel prossimo futuro...

- **Le prossime calls H2020 sono dietro l'angolo!** (ad esempio, cf. EINFRA-9-2015, “e-Infrastructures for virtual research environments”, <http://goo.gl/SlwtU2>), scadenza 14/1/2015.
 - E la scrittura delle call 2016-2017 (“consolidation calls” secondo la EC) è in corso.
- **Consolidamento** del lavoro che sta avvenendo nel **Cloud Working Group della CCR INFN**.
 - Il 2015 vogliamo che sia per noi l'anno di una generale messa in produzione di alcuni degli sviluppi sperimentali sui quali il gruppo sta lavorando.
 - Con focus sulle necessità dell'Ente sia per analisi e supporto alle esigenze di calcolo degli esperimenti, sia per l'ottimizzazione dell'utilizzo delle risorse distribuite.
- Attraverso progetti esterni come PRISMA, OCP, i progetti di trasferimento tecnologico e i potenziali progetti H2020, vogliamo **continuare la costruzione di know-how sul calcolo distribuito e il reperimento di fondi** a beneficio dell'Ente, della comunità scientifica italiana e della società in generale.
 - Focalizzando l'interesse non tanto sulla parte IaaS, quanto su PaaS e SaaS.

Grazie! (e a proposito del futuro...)

- “Il cavallo è una cosa che resterà, l'automobile è solo una novità, una moda.” (Presidente della Michigan Savings Bank, 1903)



- “La televisione non durerà, perché la gente si stancherà presto di passare ogni notte a fissare una scatola di legno compensato.” (Darryl Zanuck, 20th Century Fox, 1946)



- “Entro il 2005 o giù di lì, diventerà chiaro che l'impatto di Internet sull'economia non è stato maggiore di quello del fax”. (Paul Krugman, economista del New York Times e premio Nobel per l'economia, 1998)



- “Se avessi chiesto alla gente che cosa voleva, avrebbe detto che voleva dei cavalli più veloci.” (Henry Ford)

