

Massimo Sgaravatto
INFN Padova
On behalf of the Cloud team

Stato installazione della Cloud

Personale

- Attivita` iniziata alla fine dello scorso anno
 - Kickoff meeting: 26/11/2014
- Personale coinvolto
 - Tecnici Universita`
 - M. Boccolini (CSIA)
 - P. E. Mazzon (DEI)
 - M. Menguzzato (DFA)
 - G. Paolucci (CSIA)
 - G. Sella (DiSC)
 - Personale INFN coinvolto nelle attivita` della Cloud INFN dell'Area Padovana
 - P. Andreetto, F. Costa, A. Crescente, A. Dorigo, S. Fantinel, F. Fanzago, M. Sgaravatto, S. Traldi, , M. Verlato, L. Zangrando
 - Tutti con una frazione molto limitata di tempo

Cloud INFN dell'Area Padovana

- Implementata unica infrastruttura Cloud tra la Sezione di Padova e i Laboratori Nazionali di Legnaro
 - Risorse distribuite nei due siti
- Utilizzato middleware OpenStack
 - Versione Havana
 - Migrazione ad Icehouse in preparazione
- Configurazione dei servizi in High Availability
- Sviluppo ed Integrazione di alcune nuove funzionalita`, non fornite da OpenStack
 - Integrazione con il sistema di autenticazione e autorizzazione dell'INFN, modulo per la registrazione degli utenti e dei progetti
- Altri sviluppi a piu` lungo termine in corso
 - Scheduler per gestione in modalita` fairshare delle risorse tra i diversi gruppi di utenti
- Dopo una fase di pre-produzione con funzionalita` ora avviata una fase di vera e propria produzione

Hardware della Cloud INFN

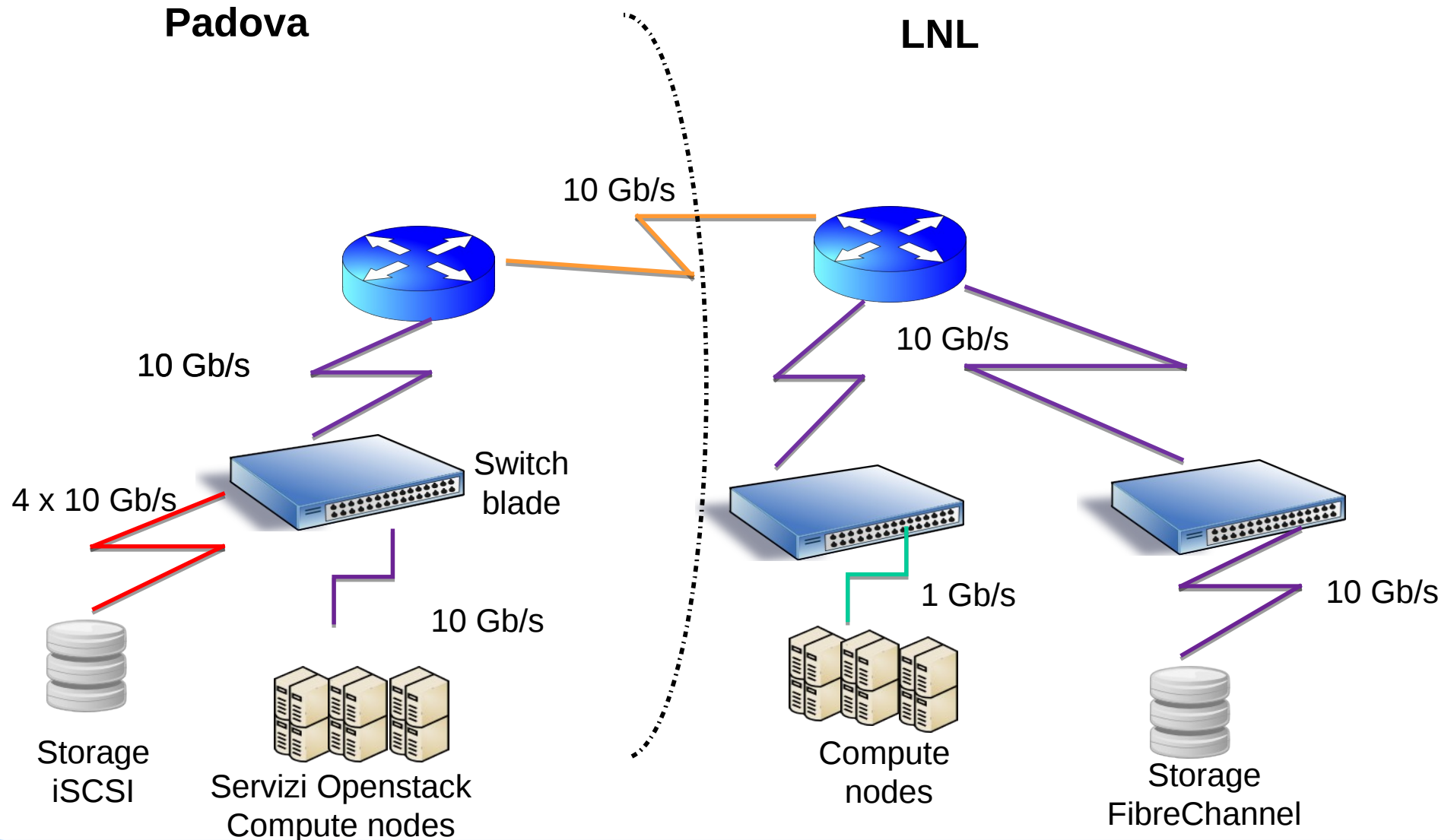
- Padova

- 4 lame DELL M620 ciascuna con 1 processore E5-2609, 32 GB RAM per servizi
- 5 lame DELL M620 ciascuna con 2 processori E5-2670v2, 96 GB RAM per calcolo
- 3 lame biprocessore Xeon E5-2650v3, 96 GB RAM per calcolo
- 1 server iSCSI DELL MD3620i, con 23 x 900 GB + MD1200 con 12 x 4 TB per storage di servizio e storage utente

- Legnaro

- 6 Fujitsu Primergy RX300S7 (2 processori E5-2650v2, 96 GB RAM) per calcolo
- Server storage DELL PowerVault MD3600 MD1200 (FibreChannel) per un totale di 24 x 2 TB per storage utente

Layout



Hardware della Cloud Universita`

- Server Blade
 - Enclosure PowerEdge M1000e
 - 2 switch moduli DELL Force 10 MXL 10/40 GbE DCB Switch for dual switch config (FI) 32 Port
 - 4 nodi PowerEdge M620 di gestione (ciascuno equipaggiato con: doppio processore Intel Xeon E5-2609 e 32GB di RAM)
 - 12 nodi PowerEdge M620 per calcolo (ciascuno equipaggiato con: doppio processore Intel Xeon E5-2670v2 e 160GB di RAM)
- 2 storage series Equallogic esterni per un totale di:
 - Cassetto con ~ 96TB di spazio Raw su 24 dischi NL-SAS a 7.200 giri.
 - Cassetto con ~ 18TB di spazio Raw su 17 dischi SAS a 10.000 giri + ~ 5.5TB di spazio Raw su 7 dischi SSD
- Installato in sala CED del DFA

CED-C

- Nella lettera di intenti firmata dai Direttori si parla di sviluppo, messa in opera e sperimentazione di un “Centro pilota di Elaborazione Dati Cloud a Padova – CED-C ” ad alte prestazioni a sostegno della ricerca dei partner coinvolti.
 - Nell’ottica di unire gli sforzi mettendo in comune le competenze esistenti e le esperienze già fatte
 - Per creare un centro di riferimento tecnologico e di competenze a livello regionale, non solo per il mondo della ricerca, ma anche per le imprese e per la Pubblica Amministrazione. Contatti in tal senso sono già stati avviati con la Regione Veneto.
- Possibile modello implementativo
 - Uso delle risorse Cloud dell'Università` per la prima implementazione del CED-C
 - Successiva integrazione delle risorse INFN, una volta verificato che tutto funziona

Middleware Cloud

- Scelto di adottare OpenStack come middleware Cloud
 - Fondato da NASA e Rackspace nel 2010
 - Software Open Source
 - In forte e costante crescita in termini di funzionalità e sviluppatori
 - Forte supporto da parte dell'industria (RedHat, Intel, etc.)
 - Usato nella Cloud INFN Padovana e in molti altri ambienti (CERN, etc.)
 - Anche se non tutto è perfetto ...
 - Piuttosto complesso da installare, configurare, gestire
 - In veloce e profonda evoluzione (una release ogni 6 mesi, molti cambiamenti da una release a un'altra)
 - Ancora immaturo
- Scelto di fare il deployment iniziale considerando la release IceHouse di OpenStack
- Scelto di usare CentOS7 come OS delle macchine dell'infrastruttura Cloud
 - Nella Cloud INFN si usa attualmente SL6
 - CentOS7 scelto per rendere più agevoli successivi update (RHEL6 e derivati non più supportati dalla release Juno di OpenStack)

Deployment dei servizi

- Deployment dei servizi pensata in modo diversa wrt la Cloud INFN
 - Un po` per scelta
 - Alcune scelte fatte per la Cloud INFN si sono dimostrate non ottimali (v. storage condiviso con servizi)
 - Un po` per necessita`
 - Necessita` di ospitare tutti I servizi nelle lame della Blade
 - Mentre per la Cloud INFN si sono usati host esterni per certi servizi (database, HAProxy)
- Configurazione dei servizi in High Availability
 - Cluster Mysql HA usando Percona XtraDB MultiMaster formato da 3 host
 - Cluster HAProxy/Keepalived formato da 3 host per i servizi che richiedono load balancing
 - Servizi Openstack in HA su 2 nodi
- Uso delle lame
 - 2 lame di servizio per Storage
 - 2 lame di servizio per Controller + Network node in HA
 - 3 lame di calcolo per HAProxy + Compute node
 - 3 lame di calcolo per MySql + Compute Node
 - 6 lame di calcolo per Compute Node

Storage

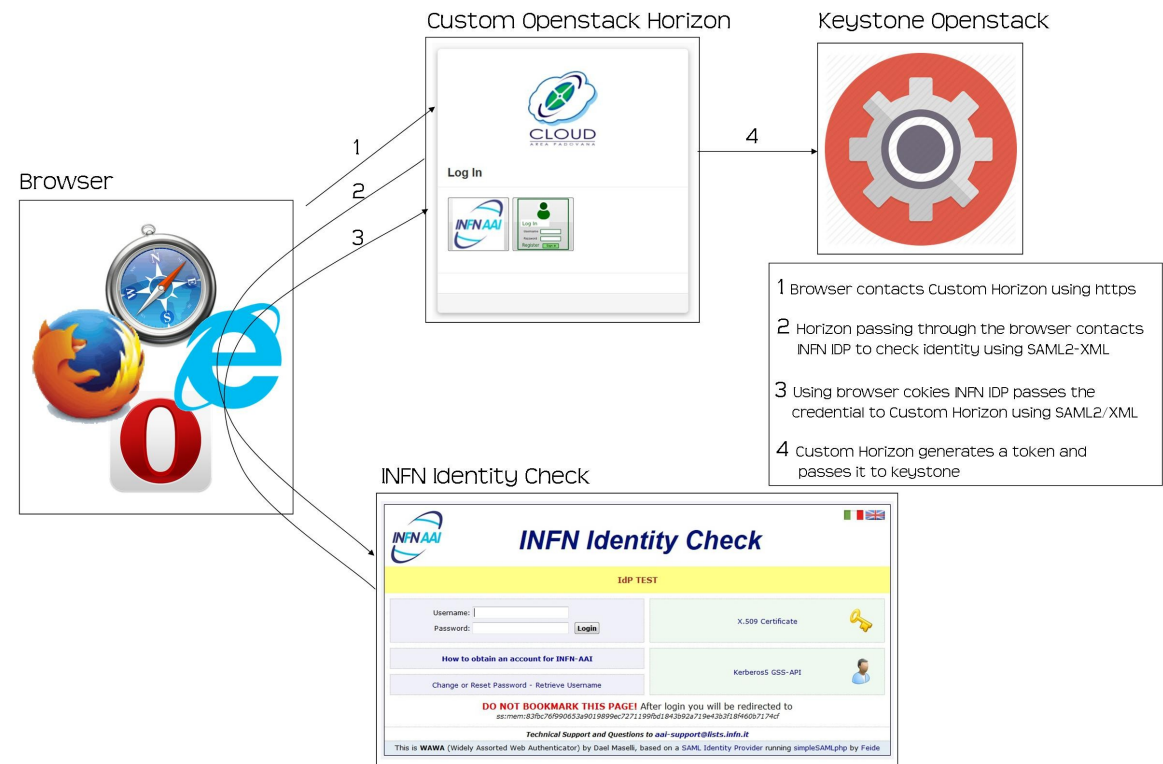
- Uso del cassetto con dischi veloci per le immagini (Glance) e per le istanze (Nova)
 - Uso di Gluster come backend
- Uso del cassetto con dischi piu` lenti per il servizio block storage Cinder
 - Possibilita` di creare dei volumi di storage “permanente”, ognuno dei quali puo` essere “agganciato” a una macchina virtuale
 - Senza un backend
 - “Interfacciamento” diretto con l'EqualLogic attraverso apposito driver

Rete

- Servizio Cloud esposto
 - con nome “neutro” (non INFN, non Università, dominio TBD)
 - Per il momento usiamo cedc.csia.unipd.it
 - su una nuova rete “neutra” che abbiamo già chiesto e ottenuto
- Virtual Machine create nella Cloud saranno su rete privata
 - Raggiungibili attraverso un gate
- Se necessario sarà possibile assegnare IP pubblici a Virtual Machine (floating IP)
- Tre set di floating IP
 - Floating IP dell'Università
 - Floating IP dell'INFN
 - Floating IP della rete “neutra”

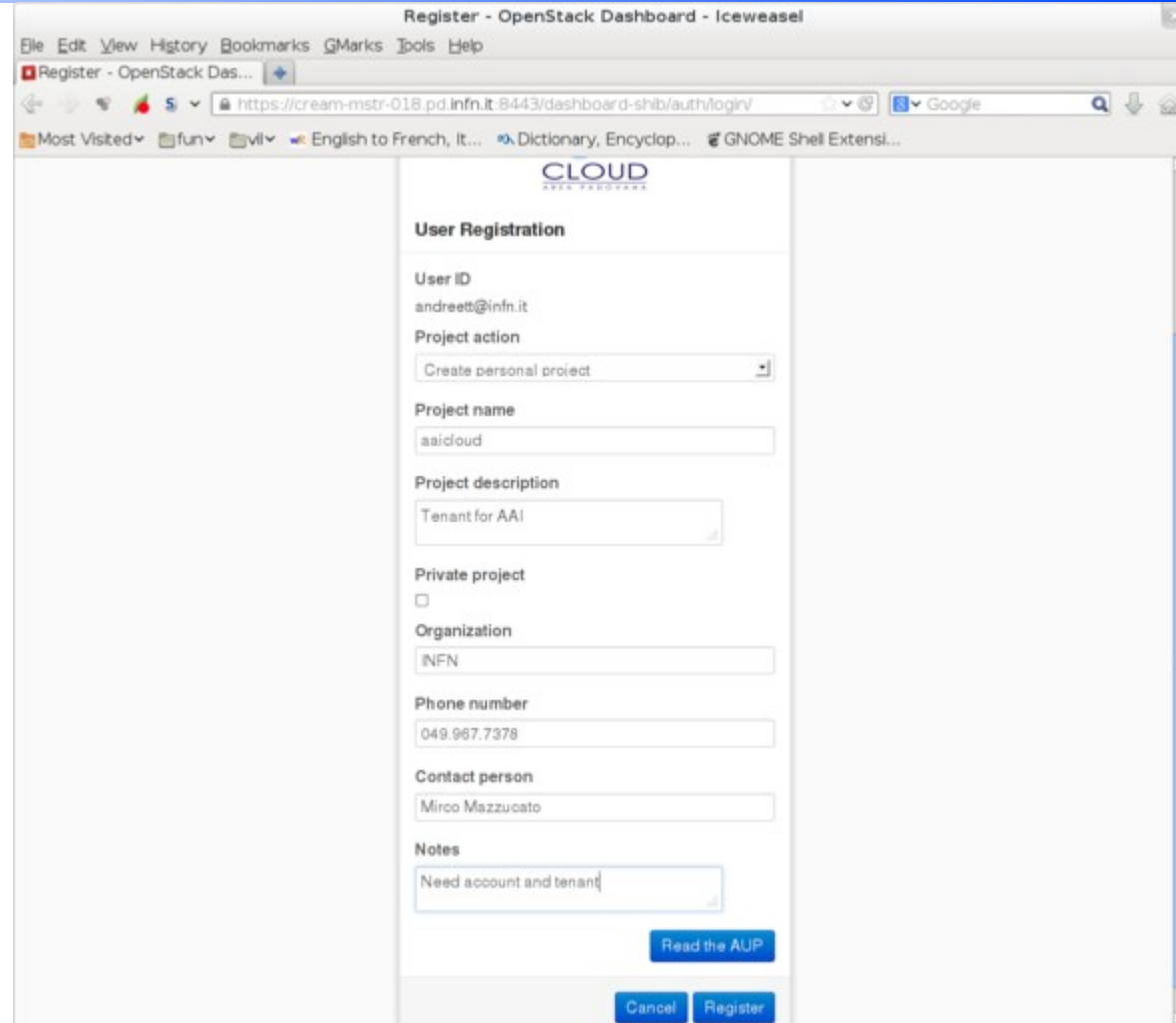
Autenticazione

- Possibilita` di autenticarsi alla Cloud sia attraverso username/password, sia attraverso un IdP
- Implementato per la cloud INFN per l'integrazione con l'IdP dell'INFN (INFN-AAI)
- Integrazione con l'IdP dell'Universita` in corso



Procedura di registrazione

- Nella Cloud INFN integrato un modulo sviluppato in casa per gestire la registrazione degli utenti e dei progetti
 - Un progetto rappresenta un gruppo/esperimento
- Definito un flusso di approvazione delle richieste di registrazione
 - Un utente che richiede la registrazione alla Cloud chiede anche la creazione di un nuovo progetto o l'affiliazione a uno o più progetti già esistenti
 - Project admin si fa carico di gestire le richieste di affiliazione al suo progetto
- Possibile utilizzo di questo modello anche per CED-C, almeno nella fase iniziale



The screenshot shows a web browser window titled "Register - OpenStack Dashboard - Iceweasel". The address bar shows the URL: "https://cream-mstr-018.pd.infn.it:8443/dashboard-shib/auth/login". The page content is a registration form for the Cloud Area Padova. The form fields are as follows:

- User ID:** andreett@infn.it
- Project action:** Create personal project (dropdown menu)
- Project name:** aacloud
- Project description:** Tenant for AAI
- Private project:**
- Organization:** INFN
- Phone number:** 049.967.7378
- Contact person:** Mirco Mazzucato
- Notes:** Need account and tenant

At the bottom of the form, there are three buttons: "Read the AUP", "Cancel", and "Register".

Attività`

- Definita l'architettura e le scelte implementative (v. slide precedenti)
- Installato e configurato l'hardware terminata
- Installato OS nelle lame
 - Usando Foreman
- Integrate le macchine in una infrastruttura di monitor basata su Ganglia e Nagios
 - Configurazione via Puppet
- Installato MySql in High Availability
- Installato HAProxy

Attività` (cont.ed)

- Configurazione dei servizi OpenStack in corso
- Fatto
 - Configurazione dei servizi sui controller node in High Availability
 - Keystone (Autenticazione, Autorizzazione)
 - Glance (Image service)
 - Nova (Compute)
 - Cinder API (Block storage)
 - Horizon (dashboard)
- Next steps
 - Configurazione neutron (rete), configurando i 3 set di floating IP
 - Configurazione dei servizi in SSL
 - Configurazione di un compute node
 - Configurazione di cinder volume
 - Configurazione di un nodo che ospita haproxy anche come compute node
 - Configurazione di un nodo che ospita mysql anche come compute node
 - Configurazione di tutti i compute node rimanenti
 - Integrazione con IdP (UniPD e INFN) per l'autenticazione
 - Modulo per registrazione utenti e progetti

Schedula

- Prevediamo di riuscire a mettere a disposizione qualcosa di testabile, anche se con funzionalità basilari per fine Giugno, salvo grosse sorprese
- Abbiamo l'esperienza del deployment della Cloud INFN ma ci sono molte differenze (e ci sono diversi aspetti su cui non abbiamo ancora esperienza)
 - SL6 vs CentOS7
 - Havana vs IceHouse
 - Deployment dei servizi fatto in modo diverso
 - Storage system diverso (v. es. Cinder volume)
 - Configurazione rete (v. es. 3 set di floating IP)
 -

Come funzionera` ?

- L'utente chiede la creazione di una macchina virtuale (VM)
 - Specificando l'immagine da usare (SL6.6, Ubuntu, etc.) e il "flavor" (quanti cores ? quanta RAM ? quanto disco ?)
- La macchina virtuale viene istanziata su un compute node
 - Lo schedulatore dell'infrastruttura Openstack sceglie qual e' il compute node piu' adatto
- Quando l'utente ha terminato di usare la VM, la cancella, e le relative risorse vengono rilasciate
- Possibilita` di "agganciare" dei volumi di storage permanente alle VM
- Uso "elastico" delle risorse e in modalita` self-provisioning (l'utente si "arrangia" a istanziarsi la VM)

Allocazione delle risorse

- Risorse partizionate tra diversi progetti
 - Corrispondono in genere a esperimenti/gruppi di ricerca/...
 - Un utente puo` far parte di piu` progetti
- Esaurita la quota di risorse allocata a quel progetto, non c'e` possibilita` di richiederne altre, anche se ce ne sono di non utilizzate in altri progetti
 - La richiesta viene semplicemente "respinta"
- Sviluppi in corso c/o INFN Padova per superare questi limiti
 - Possibilita` di accodare le richieste che non possono essere immediatamente soddisfatte
 - Uso fairshare delle risorse senza necessita` di partizionarle
- Prototipo esistente ma tempi lunghi per avere questa funzionalita` production ready e integrata in OpenStack

Modus operandi

- Meeting settimanale (mercoledì pomeriggio) dove vengono pianificate e discusse le varie attività
 - Prima parte del meeting riservato al setup della Cloud CED-C
 - Seconda parte riservata alla gestione della Cloud INFN
 - Agenda e minute in:
<https://agenda.infn.it/categoryDisplay.py?categId=658>
- Attività svolte da singole persone o gruppi
- Attività tracciate in Jira

Conclusioni

- Implementazione in corso
- Si stanno sfruttando le esperienze nella implementazione della Cloud INFN, ma non si tratta semplicemente di replicare quell'installazione
 - Perche` gli use case sono diversi
 - Perche` il deployment verra` fatto in modo diverso
 - Perche` la tecnologia nel frattempo e` evoluta: v. versione OS, versione OpenStack, etc.
- Salvo imprevisti, prevista una prima implementazione con funzionalita` basilari testabile ~ a fine Giugno

THE END