



# sync & share su INFN-CC

stefano nicotri  
stefano stalio

INFN - Istituto Nazionale di Fisica Nucleare

Workshop di CCR: La Biodola, 3 - 7 giugno 2019

# intro

descrizione di attività svolte per realizzare soluzioni di storage

- di alto livello
- dal costo contenuto
- elastiche e scalabili
- facilmente dispiegabili
- amministrabili in maniera semplice
- utilizzabili da comunità vaste ed eterogenee
- con funzionalità avanzate e user friendly di condivisione e navigazione

# due tipi di servizi

## sync & share as a Service

- Nextcloud
- storage criptato
- setup semi-automatizzato

chi ne fa richiesta ottiene un sistema (di cui diventa amministratore) che può utilizzare per erogare a sua volta un servizio di personal storage (di piccole dimensioni) verso terzi

## large personal storage service

- ownCloud
- alta disponibilità
- replica geografica dei dati

sistema di personal storage robusto e affidabile in grado di offrire a comunità ampie ed eterogenee un servizio di sync & share per l'utilizzo quotidiano con un costo per TB relativamente basso

sync & share as a Service

# sync & share as a Service - componenti

- basato sul software open source Nextcloud
- storage criptato con metodo standard di cifratura LUKS
  - cifratura eseguita direttamente dall'amministratore dell'istanza
  - gestione semplice dei volumi criptati (<https://github.com/Laniakea-elixir-it/fast-luks>)
  - solo l'amministratore della piattaforma di storage ha accesso ai dati non criptati
  - gli amministratori della parte "hardware" vedono solo dati criptati
- setup semi-automatizzato
  - utilizzo di Docker + Docker Compose per il deployment del servizio
  - deployment: `cifratura + docker-compose up -d`
  - istruzioni su <https://github.com/stefanonicotri/nextcloud-luks>

# sync & share as a Service - architettura

istanza *proof of concept* (in collaborazione con E. Pasqualucci):

- singola VM su INFN-CC:
  - 4 CPU core
  - 8GB di RAM
  - disco SSD da 20GB per i dati del DB
  - disco HDD da 2TB per i dati degli utenti
  - docker container per il DB
  - docker container per l'application server

attualmente utilizzata come sistema di storage dall'amministrazione della Sezione di Roma dell'INFN

# sync & share as a Service - prossimi step

programmi per il futuro:

- rendere scalabile ed elastica l'architettura a microservizi
- completare l'automazione del dispiegamento, producendo una soluzione PaaS istanziabile via Ansible o TOSCA template, utilizzando gli outcome dei progetti R&D in corso

large personal storage service



# large personal storage service - caratteristiche

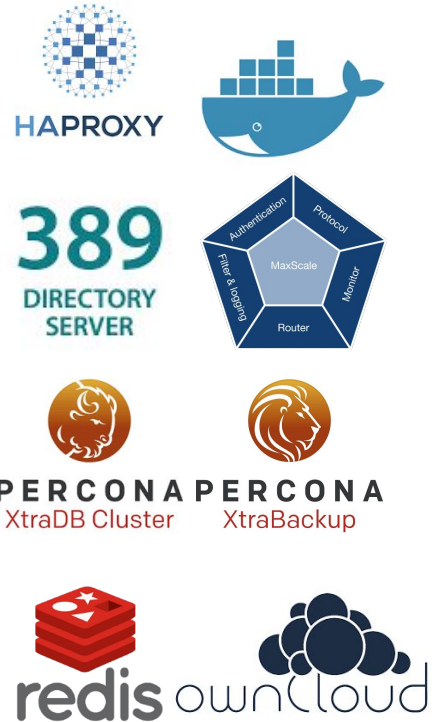
- ~ 10k utenti
- ~ 200M file
- spazio disco ~ 200 TB
- basso costo per TB
- realizzabile e dispiegabile su INFN-CC
- supporto ad autenticazione centralizzata (ad es. via AAI)

gruppo di lavoro: Stefano Stalio (LNGS) + Stefano Nicotri (INFN-BA) + chiunque sia interessato

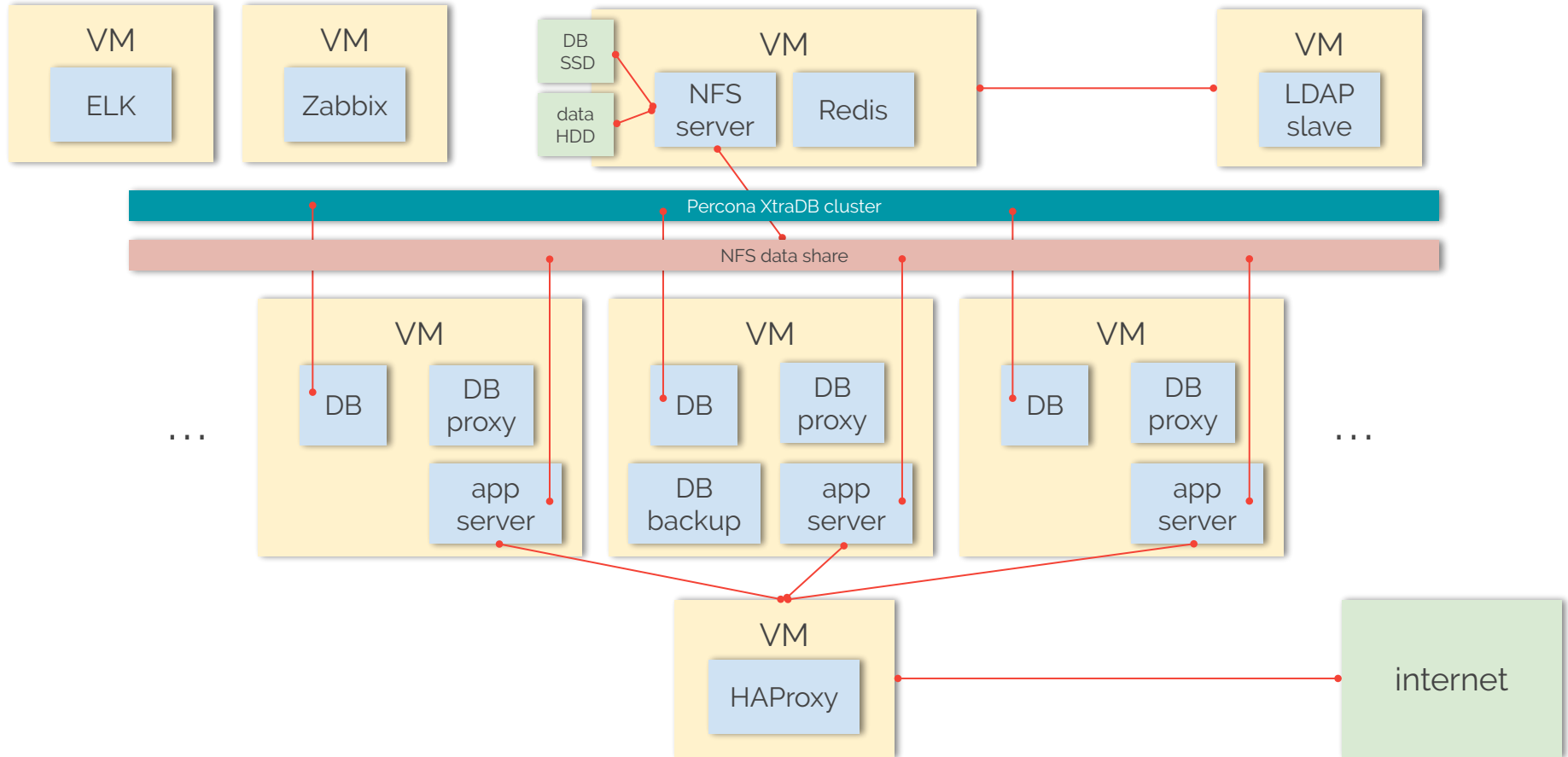
l'attività è cominciata "copiando" un sistema simile, SwitchDrive, realizzato da SWITCH (CH)

# large personal storage service - componenti

- setup su VM su INFN-CC (OpenStack)
- basato sul software open source ownCloud
- Setup semi-automatizzato: Docker + Docker Compose
- NFS
- HAProxy
- Redis
- 389 Directory Server
- Percona XtraDB cluster + Xtrabackup
- MariaDB MaxScale database proxy



# large personal storage service - architettura



# large personal storage service - storage

- CEPH Erasure Coded Pool → 4 + 2 (data + parity) (v. presentazione di Stefano Stalio)
  - risparmio di spazio disco senza rinunciare (?) a performance
- metadati di CEPH su SSD
- NFS server che esporta verso gli application server
  - share NFS per i dati (da dischi HDD)
- DB su SSD
- FSCache su SSD sui client NFS (application server)

# large personal storage service - autenticazione

- autenticazione via INFN-AAI (grazie al supporto del team AAI)
- slave LDAP dedicato (389 directory server) necessario per problemi di performance
  - mitigata lentezza nello startup, al primo caricamento dell'albero utenti, e nella condivisione dei file e cartelle

# large personal storage service - alta disponibilità

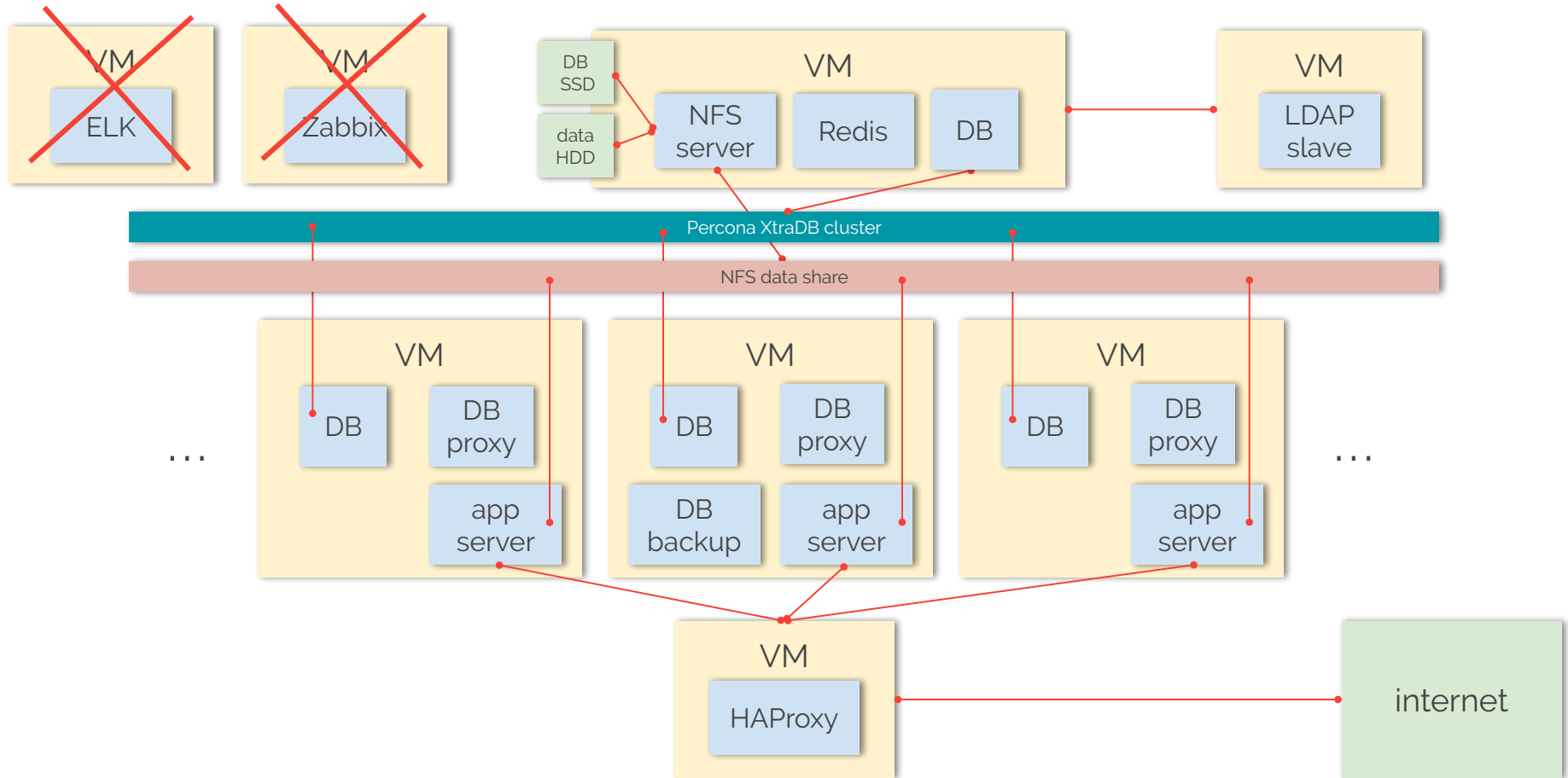
- Redis per il caching
  - supporto built-in in ownCloud
- dati replicati via RBD mirror
- DB master/slave
  - lettura dagli slave e scrittura sul master tramite DB Proxy MaxScale
- DB backup
  - Container Docker su una delle VM → backup via CRON
- applicazioni pronte a partire su sito a distanza geografica
  - sfruttando le diverse regioni di INFN-CC
- DNS-HA
- persistenza delle sessioni via HAProxy

# large personal storage service - test

- VM medium (2 CPU, 4 GB RAM, 20 GB disco, volume cinder da 20 TB, volume cinder highperf da 20 GB): server NFS, server MySQL
- VM large (4 CPU, 8 GB RAM, 20 GB disco, volume cinder highperf da 20 GB): application server, server MySQL, DB backup
- 2 VM large (4 CPU, 8 GB RAM, 20 GB disco, volume cinder highperf da 20 GB): application server, server MySQL
- VM small (1 CPU, 2GB RAM, 20 GB disco): HAProxy
- VM small (1 CPU, 2GB RAM, 20 GB disco): LDAP slave

tot: 16 CPU, 32 GB RAM >20 TB disco

# large personal storage service - test





# large personal storage service - test

- caricati, con 10 sync client circa 250 GB di dati su 200k file in 4 ore
- caricati, con 40 client a linea di comando, circa 280 GB di dati su 455k file in 3 ore (fino a 700Mb/s in upload)
- caricati, con 200 client a linea di comando, circa 120 GB di dati in 140k file, in tempi più lunghi anche a causa del sovraccarico lato client
- upload bandwidth proporzionale alla dimensione dei file
- necessario tuning sul numero di connessioni accettate dal database

# large personal storage service - sviluppi futuri

- monitoring
- gestione dei log: ELK
- cifratura
- miglioramento alta disponibilità → business continuity?
- management di un numero elevato di utenti (storage) → scalabilità
- aggiunta di un disco (+ automazione)
- Docker Swarm? INDIGO PaaS? Mesos/Kubernetes?

<https://baltig.infn.it/nicotri/nextcloud-cc>

grazie per l'attenzione