Federare lo storage distribuito nazionale

La prima esperienza in DataCloud verso il data-lake

Diego Ciangottini, Ahmad Aalkhansa, Alessandro Costantini, Alessandro Italiano, Andrea Rendina, Daniele Spiga, Enrico Vianello, Federica Fanzago, Lucia Morganti, Marco Verlato, Marica Antonacci, Massimo Biasotto, Massimo Sgaravatto, Stefano Stalio

Outline

- Motivazioni per un testbed
- Strategia
- Implementazione
- Benefici e casi d'uso
- Piani e conclusioni

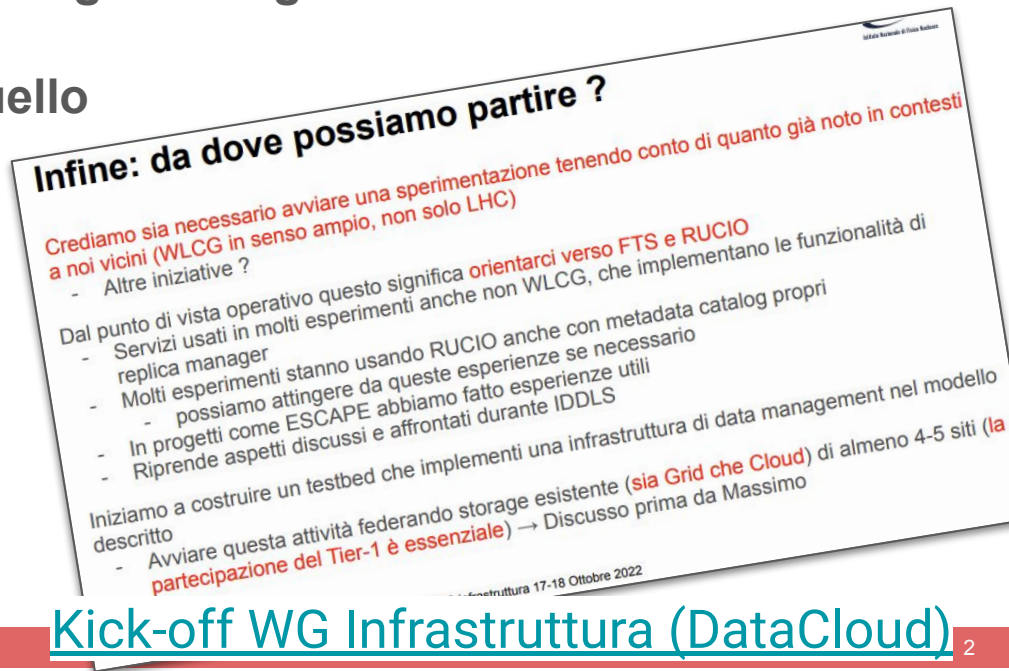
INFN

Istituto Nazionale di Fisica Nucleare

Recap: obbiettivo

Perché un **testbed per una infrastruttura nazionale sul modello data-lake?**

- **Federazione di storage con tecnologie eterogenee**
  - Sia “Grid” che “Cloud”
- **Astrazione livello “logico” da quello della gestione degli storage**
- **Interfacce** a vario livello
  - per utenti
    - E.g. gestione metadati
  - per admin (sia sito che VO)
    - Quote, gruppi etc



[Kick-off WG Infrastruttura (DataCloud)]

Motivazioni

- Valutare **una soluzione comune per utenti singoli e per piccole/medie collaborazioni**
  - Con unico caveat: accettare un set comune di policy
    - e.g. che tutti i nomi logici dei file siano visibili a tutti i partecipanti al data-lake
- Avere **un layer condiviso** da utilizzare come Service-as-a-Service
  - Gestione dei dati senza o con limitata expertise
    - Sfruttando esperienze di altri casi d'uso
  - Effort condiviso per la parte tecnica di "operazioni"
    - Mantenimento in produzione dei servizi
- Per **collaborazioni che hanno già un proprio setup** dedicato e comunque di interesse **valutare la possibilità di integrare l'utilizzo del data-lake nazionale per l'analisi dati**
  - Quindi come integrare due istanze gerarchiche di rucio ad esempio?

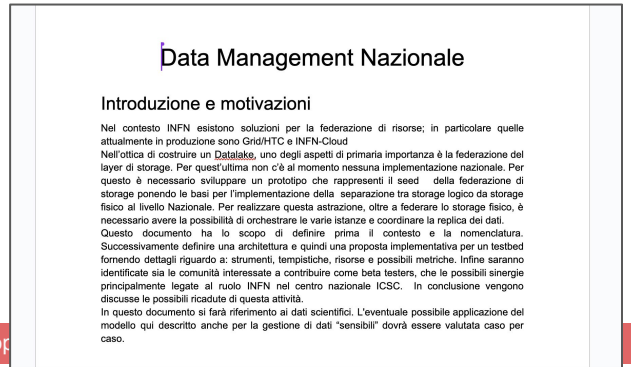
Attività WP6 DataCloud

L'esperienza in un contesto **data-lake è uno dei "pillar" dell'attività del WP6 di DataCloud.**

1	Software management	Gestione del software di un utente (o gruppo di utenti) in un ambiente distribuito
2	Data management	Gestione dei dati nel contesto "data lake"
3	Offloading di workload	Esplorazione di modelli per fare offloading di workload in un ambiente distribuito ed eterogeneo

Una serie di **passi coordinati per ottimizzare gli sforzi:**

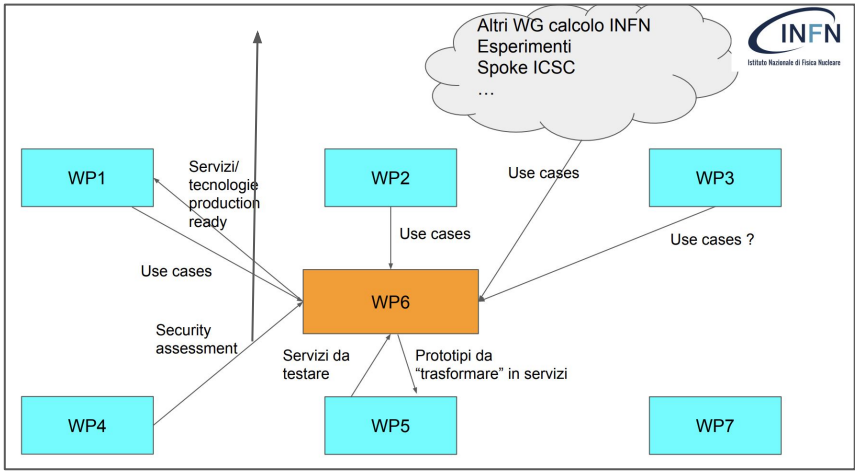
1. **"Scouting"** per soluzioni tecnologiche
2. **Stesura di un documento per:**
   a. Definire prima il contesto e la nomenclatura
   b. Definire gli obbiettivi per il testbed
   c. Definire i requisiti per siti e componenti
3. **Instanziazione e configurazione servizi**



Data Management Nazionale

Introduzione e motivazioni

Nel contesto INFN esistono soluzioni per la federazione di risorse; in particolare quelle attualmente in produzione sono Grid/HTC e INFN-Cloud
Nell'ottica di costruire un Datalake, uno degli aspetti di primaria importanza è la federazione del layer di storage. Per quest'ultima non c'è al momento nessuna implementazione nazionale. Per questo è necessario sviluppare un prototipo che rappresenti il seed della federazione di storage ponendo le basi per l'implementazione della separazione tra storage logico da storage fisico al livello Nazionale. Per realizzare questa astrazione, oltre a federare lo storage fisico, è necessario avere la possibilità di orchestrare le varie istanze e coordinare la replica dei dati.
Questo documento ha lo scopo di definire prima il contesto e la nomenclatura. Successivamente definire una architettura e quindi una proposta implementativa per un testbed fornendo dettagli riguardo a: strumenti, tempistiche, risorse e possibili metriche. Infine saranno identificate sia le comunità interessate a contribuire come beta testers, che le possibili sinergie principalmente legate al ruolo INFN nel centro nazionale ICSC. In conclusione vengono discusse le possibili ricadute di questa attività.
In questo documento si farà riferimento ai dati scientifici. L'eventuale possibile applicazione del modello qui descritto anche per la gestione di dati "sensibili" dovrà essere valutata caso per caso.



Strategia

Rucio+FTS è stato scelto come **punto di partenza**, per quello che conosciamo in **esperimenti LHC (de-facto standard)** ed ESCAPE ci ha insegnato essere **applicabile ad altre comunità.**

Il focus e' quindi sull'**integrazione di soluzioni che già conosciamo** facendo in modo da soddisfare le nostre esigenze specifiche.

E' fondamentale quindi avere un testbed che permetta di confrontarsi con gli utenti e verificare che queste esigenze siano soddisfatte.

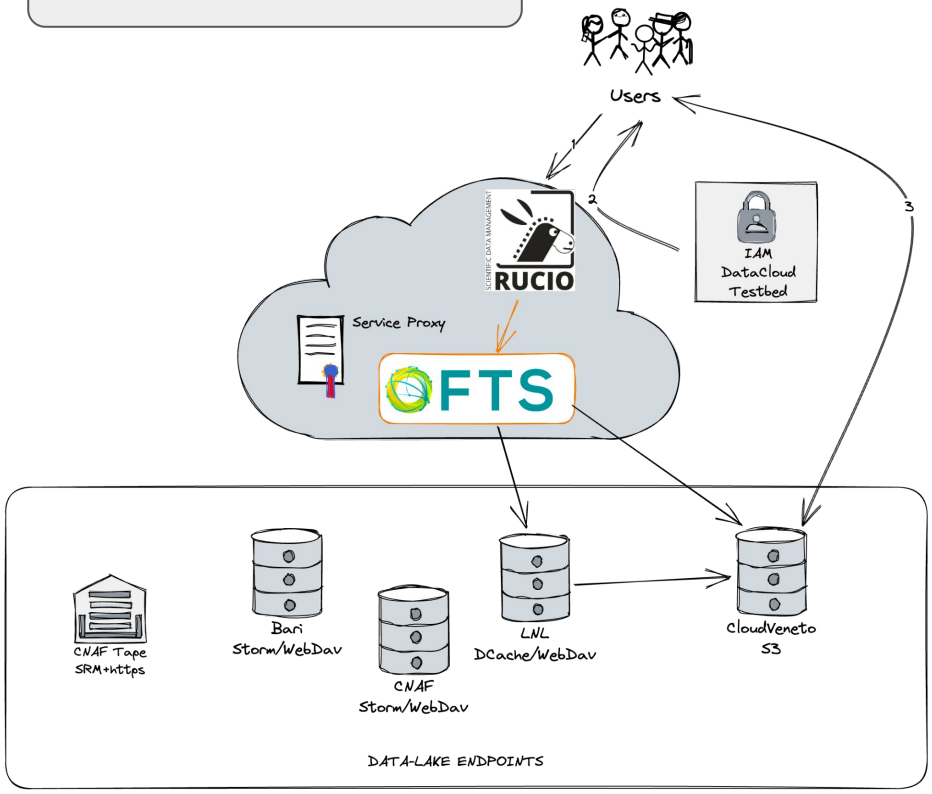
In sintesi:

Obiettivo	Avviare una sperimentazione al fine di implementare un sistema di data lake a livello di infrastruttura nazionale
Target	In particolare i piccoli esperimenti (WLCG è già autonomo) L'infrastruttura nazionale distribuita
Strumenti identificati	Possiamo partire dagli strumenti a noi noti (quelli di WLCG in particolare) e quindi a servizi quali RUCIO e FTS

Dove siamo

~6 mesi da inizio lavori!

- **Instanziato su risorse cloud i server IAM, FTS e RUCIO dedicati al testbed DataCloud**
  - FTS può essere mantenuto “centralmente” e servire istanze di RUCIO multiple
  - IAM per AuthZ fine gestita centralmente, vedi dopo
- **Federato 5 siti con storage eterogenei:**
  - Uno storage con protocollo S3 su ceph @CloudVeneto
  - Tre storage con protocollo WebDav
    - Due basati su STORM (CNAF, Bari)
    - Uno su dCache (LNL)
  - Un endpoint tape @CNAF
- **Automatizzato la registrazione e la gestione degli utenti via IAM**
  - AuthZ gestita centralmente, i siti autorizzano sulla base dei token rilasciati



Dopo un periodo di configurazione siamo ad una situazione dove **il sistema è in linea con quello che ci eravamo prefissati** (vedi dopo e demo Massimo)

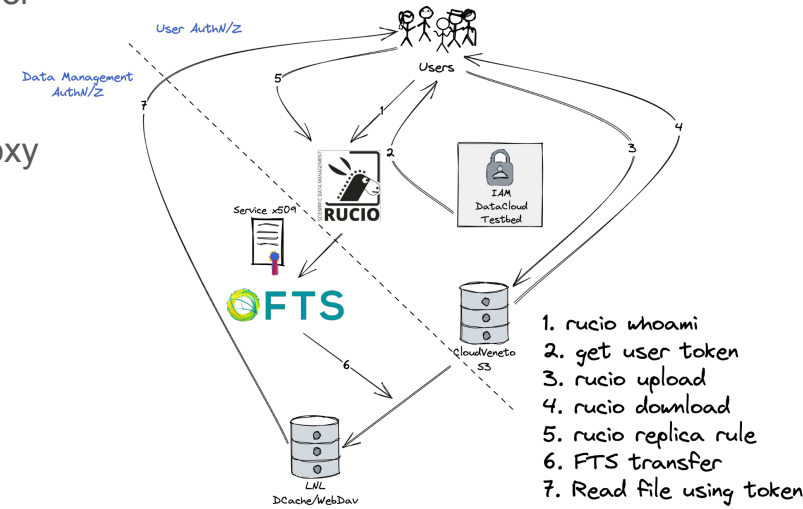
Modello AuthN/Z per primo testbed

Autorizzazione a grana fine:
- Utente ciangottini può scrivere solo su /user/ciangottini e leggere ovunque

Lavori iniziati con IAM per gestire questo tipo di scope policy

L'obiettivo è provare a fornire questa soluzione a tutti gli utenti → **serve authZ a grana fine.**

- L'autenticazione è fattorizzata in **due livelli logici**
  - **Utente:** interagisce con il server RUCIO via IAM Token per tutte le operazioni
    - Mappato automaticamente ad un account RUCIO
  - **Data management:** è RUCIO che agisce per conto dell'utente con una identità di "servizio". E.g. un x509 proxy
    - Delegato a FTS per effettuare i trasferimenti
- **"Strawman" di autorizzazione:**
  - Lettura permessa a tutti gli utenti appartenenti al gruppo data-lake users - via IAM token
  - Scrittura permessa a:
    - X509 proxy di servizio
    - IAM token rilasciato da RUCIO client
      - Necessario per rucio upload



Federare uno storage: la nostra esperienza

I requisiti per un sito sono minimi! (Uno degli obbiettivi che volevamo raggiungere.)

- Una singola area del FS dedicato a Rucio (e.g. /data/rucio)
- Token authZ -> autorizzare sulla base del token IAM presentato
  - Dipendentemente dal protocollo utilizzato, ci sono già configurazioni usabili e testate durante il setup dai siti partecipanti

NO agenti da installare!

Valori aggiunti:

- **Quote:** Rucio e responsabile per la gestione fine delle quote interne al namespace
  - Il sito deve solo indicare quanto spazio riserva alla istanza di Rucio
    - Può, se vuole, anche avere permessi di editing sugli account che desidera attraverso RUCIO
- **AuthZ:** attraverso l'authZ via token le "decisioni" sono prese/gestite a livello di IAM policy
  - No mapfile custom
  - Rimane la possibilità di offrire funzionalità/protocolli aggiuntive se si desidera

Esperienza utente -> demo Massimo domani

L'obiettivo è **rendere intuitivo** per un utente:

- **Caricare e leggere dati nei vari endpoint senza preoccuparsi dei protocolli utilizzati**
  - Semplice come *rucio upload/download MY_FILE MY_STORAGE*
  - *Non devo più preoccuparmi di scp, gridftp, rsync etc..*
- Organizzare i **dati in un catalogo**
  - “Attaccare” metadati:
    - Cluster: definire dataset o collezioni di dati affini
    - Metadati generici: json o key/value per dataset o per file
      - Query e policy basate su queste informazioni

Possiamo dire che **siamo al punto in cui è possibile “sfidare” il testbed su questi punti**.

Oggi l’idea è solo lasciare i **punti salienti di quello che è possibile fare e di quello che potrebbe esserlo in futuro.**

Gestione dei dati

Gestire in maniera dichiarativa policy per replica dei dati:

- *Voglio questo file o questo set di dati replicato N volte. In topologie complicate a piacere*
  - Almeno N copie nei siti [A,B,C] ed M copie nei siti [C,D,E]
  - Ogni nuovo file che arriva e che appartiene a questo dataset deve soddisfare questa lista di regole
- *QoS:*
  - Posso indicare un attributo del sito come requisito per la replica. Ad esempio la QoS
    - Voglio sempre una copia su tape e una su disco per questi dati
- *Gestione ciclo di vita*
  - Voglio che i miei dati rimangano in un buffer su disco per 15gg, mentre la loro copia su tape sia permanente

Gestione metadati

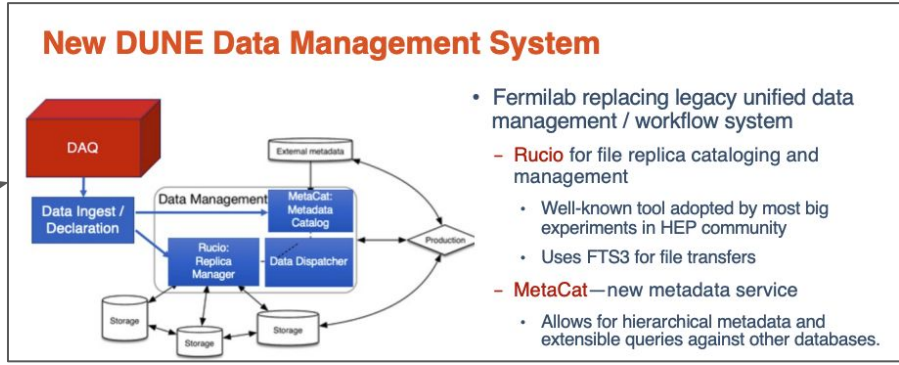
Sinergia

Gestire i metadati è una **necessità trasversale a gran parte degli use case**.

- Nella implementazione “vanilla” del testbed **è possibile associare a file e dataset una serie di metadati nella forma di chiave/valore**
  - *Dammi la lista di file che soddisfano questa tipologia di query: solo i dati 2022 del run 201231 per il canale X*
  - *Poi creami un dataset rucio con questi ultimi*

L’utilizzo di **motori per metadati esterni** può essere utile in determinati casi d’uso.

-> integrazione di un **modello a plugin** in corso in Rucio. Ad oggi **DUNE** **è uno degli esempi più production-ready**



Casi d'uso e sinergie

Vogliamo valutare **questo testbed anche nell'ottica di varie sinergie:**

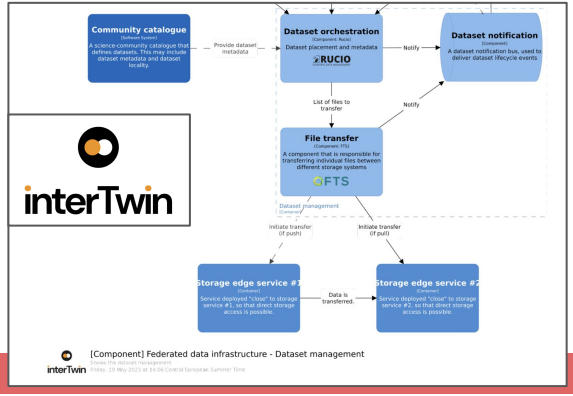
- Possibile soluzione **cross Spoke ICSC**
  - Penso **a 2 e 3** per conoscenza, dove per verificare/validare il modello è fondamentale avere velocemente un testbed su cui lavorare
  - Ma anche una **estensione ad altri casi che NON hanno ancora esperienza di queste tecnologie** (spoke > 3)
- Sinergia con **attività legate alla analisi interattiva**
- In progetto **interTwin il concetto di Data-lake è abilitante per la gestione dei dati** in un ambiente eterogeneo come quello dei provider coinvolti (talk precedente di Daniele).

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

HEP Analysis Facilities

"Analysis facility" could be any type of managed computing / storage resources shared between multiple users used for end-user analysis

Fast network connection	Orchestration, Provisioning & Configuration Management	Multitenant VO support	Distributed disk storage system	Batch



Test con utenti e con comunità

Siamo al punto in cui **il testbed sarà aperto a breve a beta-tester.** Se interessati non esitate a contattarci.
Per quanto riguarda le comunità ci sono contatti già stabiliti con attività preliminari in corso:

Cygno

- Il mio daq scrive in un buffer “locale”
- Rucio automaticamente lo prende e lo importa nel data-lake
- Da lì può partire il workflow di processamento
  - Analisi interattiva e output registrati indietro nel lake
  - Dump su tape sulla base di metadata

Dampe

- Necessità di tenere in sync i contenuti di due centri
  - CNAF e centro cinese
- Caso d’uso sensibile a policy restrittive di siti e ad automazione repliche

Integrazioni e piani

Dopo 6 mesi di attività WP6, abbiamo un primo testbed su cui far affacciare i primi volontari per ogni use case.

L'attività è **aperta a tutti i siti che sono interessati a registrarsi e a partecipare al data-lake!**

Ci sono chiaramente un buon numero di passi e iterazioni davanti a noi (ma anche già diversi dietro), per poter avanzare verso uno stato di pre-produzione:

- Apertura a primi early adopters
- Primo giro di feedback e definizione policy:
  - Quote per account, per sito
  - Capacita' per account (chi può scrivere/creare regole e dove)
- Integrazione altri pattern di data discovery e accesso
  - Analisi interattiva: integrazione con JupyterLab
  - Accesso programmatico: lancio N job su una coda e trovo gli N file di input attraverso una query RUCIO
- Pianificare evoluzione servizi verso un setup di pre-produzione
  - DB, ridondanza dei servizi etc..