

# DataCloud: stato e prospettive

Davide Salomoni ( [davide@infn.it](mailto:davide@infn.it) ), per DataCloud  
C3SN, 4/7/2023

# Agenda

1. Stato di DataCloud
2. Punti di criticità
3. Proposta di cambiamento della governance di DataCloud



# Mandato di DataCloud – 1

- DataCloud si occupa di **diverse attività core del calcolo INFN**:
  - Sviluppo e gestione del modello architetturale di Data Lake
  - Sviluppo delle infrastrutture ISO-certified
  - Supporto alla gestione e all'operazione dei siti, agli esperimenti, ai progetti e agli utenti
  - Sviluppo di nuovi servizi
- Un **punto strutturale e fondamentale** delle attività di DataCloud è **l'integrazione tra l'infrastruttura tradizionale «a Tier» tipica di WLCG e il modello «cloud-native»**.
  - **Integrazione**: di risorse, metodi, persone, soluzioni.



# Mandato di DataCloud – 2

- L'ammmodernamento ed estensione dei Tier-2 con fondi PNRR comporta la **progressiva federazione di tutti i Tier-2 all'interno del datalake INFN** secondo il modello sviluppato da INFN Cloud e l'integrazione di tutti i servizi in un portafoglio Cloud-native. Le risorse nei Tier-2 non saranno più solo allocate per le cosiddette «pledged resources WLCG» ma serviranno anche a supportare le comunità coinvolte nei progetti PNRR.
- La **assunzione di personale a TD su fondi PNRR** legati alla «infrastruttura» avviene:
  - **Nei Tier-X**, in cui il personale TD lavorerà sia sul supporto locale (es. impiantistico, sistemistico), sia sulle attività distribuite definite in DataCloud.
  - **In alcune sezioni non sede di Tier-X**, in cui il personale TD lavorerà sulle attività distribuite definite in DataCloud.

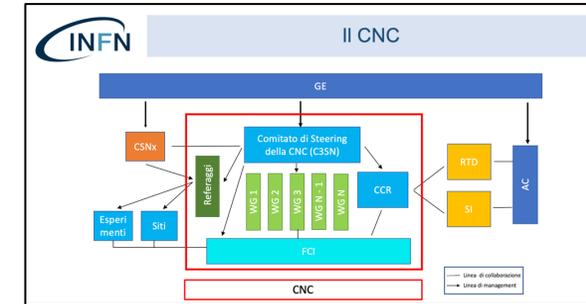


# Stato generale

- **Kick-off DataCloud: 17-18/10/2022.**
  - Da allora: 6 riunioni plenarie + incontro dedicato su Data Management + incontro dedicato con Aruba + incontro dedicato con GARR su TeRABIT.
  - Altre Riunioni:
    - PMB ogni 15 giorni.
    - Tutti i 7 WP sono ora attivi, con riunioni periodiche normalmente settimanali o bi-settimanali e report di tutti i WP alle plenarie mensili. Tutto è documentato su Teams e su Jira/Confluence.
- Dal kick-off, DataCloud è stata presentata o discusso in molti contesti:
  - IberGrid '22, SOSC '22, INFN4LS, meeting con ATOS su health, collaborazione con UCloud (DK), kick-off meeting TeRABIT, collaborazione con IFOM, DataCloud nel progetto DARE, CHEP '23, presentazione servizi/risorse a ICSC/Industrial Board, EGI Workshop, presentazione a plenaria Health Big Data, partecipazione come HBD-DataCloud a progetto flagship EU UNderstand CANcer (UNCAN).

# Governance attuale di DataCloud

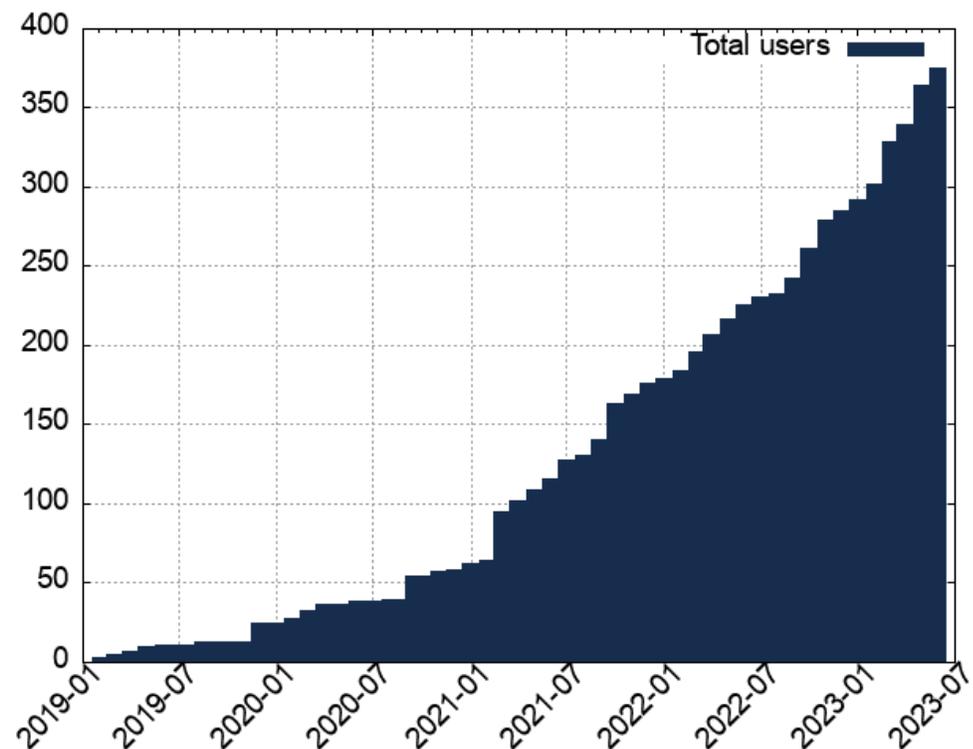
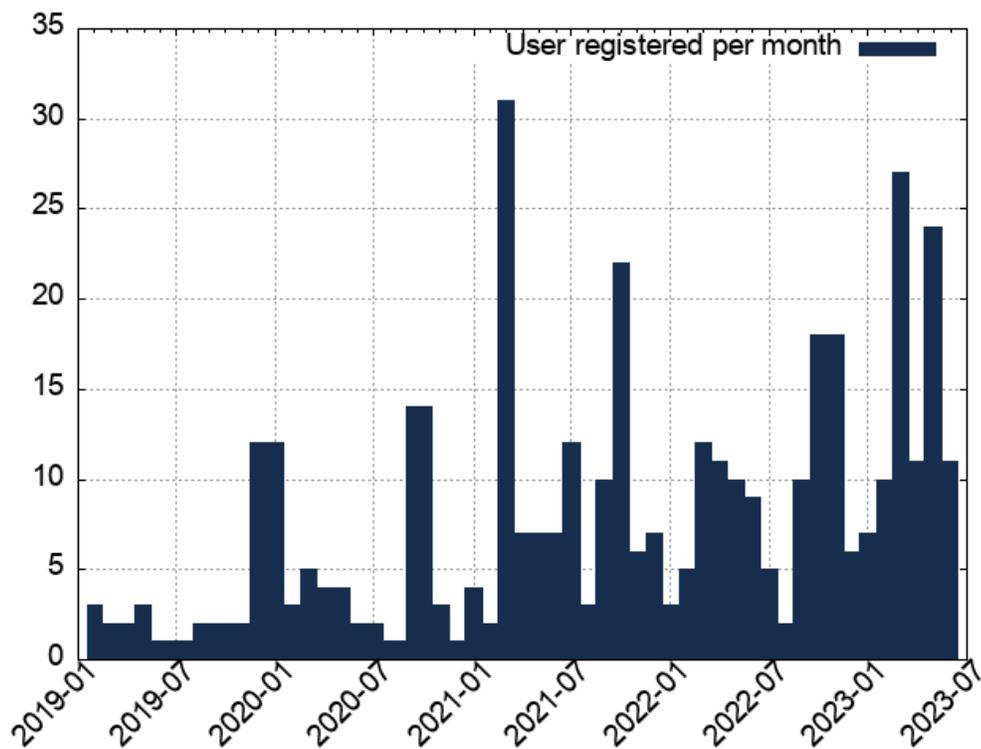
- DataCloud **riporta al C3SN** attraverso il proprio coordinatore (attualmente: Davide Salomoni).
- Internamente, sono definiti 7 “**Work Packages**” (WP).
- Ogni WP ha 2 responsabili (“WP Leader”). I WP Leader e il coordinatore del WG DataCloud compongono il “Project Management Board” (**PMB**) del WG. Il PMB si riunisce di norma ogni 15 giorni per discutere lo stato del WG e gestire eventuali problemi.
- Il WG si ritrova di norma ogni mese in una **riunione plenaria virtuale**, aperta a chiunque voglia partecipare, e almeno una volta all’anno in presenza.





# Supporto

- Ticket aperti: 9/1160 (attualmente ~45/mese).





# Guide, formazione

- **Enorme** lavoro svolto sulla **documentazione** da parte di tutti, in particolare WP2.
- Diverse guide pronte al rilascio:
  - MATLAB aaS
  - IAM aaS
  - HTCondor mini
- In scrittura:
  - Rclone per l'interazione con l'Object Storage di INFN-Cloud
  - Supporto a progetti di R&D che necessitano di allocazione speciale di risorse
- **Formazione:** webinar (da incrementare), corso utenti, corso amministratori; in programma: formazione neo-assunti (a ottobre), corsi «role-oriented» (sviluppatore, sistemista, ML/data analysis), corso interno su amministrazione OpenStack, formazione @ SOSC '23.

INFN Cloud

latest

Search docs

**GETTING STARTED**

- Getting Started
- How To: Request the "nomination to be system administrator"
- How To: Request the "nomination to be system administrator" (italian version)

**CENTRALISED SERVICES**

- Use the INFN Cloud Registry service
- Use the INFN Cloud object storage service
- Use the Notebooks as a Service solution

**GENERAL-PURPOSE GUIDES**

- Enable authenticated security scans
- Create a customized docker image for services
- Associate a FQDN to your VMs
- Access Sync&Share aaS cloud storage from a scientific environment
- Request to open ports on deployed VMs

**AAI SOLUTIONS**

- Deploy a dedicated INDIGO IAM (sys-admin nomination required)

**STORAGE SOLUTIONS**

- Deploy Sync&Share aaS (sys-admin nomination required)

**COMPUTING SOLUTIONS**

Read the Docs v: latest

Docs » Welcome to the INFN Cloud Use Cases Documentation

[View page source](#)

## Welcome to the INFN Cloud Use Cases Documentation

You'll find here useful information regarding the use-cases supported on the INFN Cloud infrastructure.

### Getting started

- [Getting Started](#)
- [How To: Request the "nomination to be system administrator"](#)
- [How To: Request the "nomination to be system administrator" \(italian version\)](#)

### Centralised services

- [Use the INFN Cloud Registry service](#)
- [Use the INFN Cloud object storage service](#)
- [Use the Notebooks as a Service solution](#)

### General-purpose guides

- [Enable authenticated security scans](#)
- [Create a customized docker image for services](#)
- [Associate a FQDN to your VMs](#)
- [Access Sync&Share aaS cloud storage from a scientific environment](#)
- [Request to open ports on deployed VMs](#)

### AAI solutions

- [Deploy a dedicated INDIGO IAM \(sys-admin nomination required\)](#)

### Storage solutions

- [Deploy Sync&Share aaS \(sys-admin nomination required\)](#)

### Computing solutions

- [Deploy Working Station for CYGNO Experiment \(sys-admin nomination required\)](#)
- [Instantiate docker containers using custom docker-compose files \(sys-admin nomination required\)](#)
- [Instantiate docker containers using docker run \(sys-admin nomination required\)](#)
- [Deploy Elasticsearch & Kibana \(sys-admin nomination required\)](#)
- [Deploy a HTCondor 8.9.9 cluster on Kubernetes \(sys-admin nomination required\)](#)

# Infrastruttura / risorse

- Manutenzioni, aggiornamento continuo middleware / PaaS. Molto lavoro su monitoring e su accounting (task complesso): problemi di supporto globali, anche per WLCG, su APEL / evoluzione di CaSO [es. supporto a GPU/FPGA] → **serve uno sforzo costante per seguire le evoluzioni.**
- Sono state definite le risorse “spare” che possiamo mettere a disposizione per l’INFN (extra “risorse referate”) e per progetti come ICSC – vedi prossima slide.
- **Definito calendario per incontri con i Tier-2:**
  - Fatti: LNL, Catania
  - 5/7 Roma, 17/7 Napoli, altri a settembre



# Cloud Services to ICSC



## Availability of INFN Cloud Services and Resources for ICSC

31 March 2023

Introduction .....	1
INFN Cloud Hardware Resources that can be made available to ICSC .....	1
INFN Cloud Services that can be made available to ICSC.....	1
Extension of the INFN Cloud Services .....	2
Use of OpenShift on INFN Cloud.....	2
Porting of INFN Cloud solutions to other Cloud providers .....	3
Future developments and more information .....	3

DRAFT

### INFN Cloud Hardware Resources that can be made available to ICSC

At the date of this document, INFN Cloud can provide the following resources to ICSC:

- 2000 vCPU
- 200 TB net disk space
- A marginal number of GPUs (NVIDIA V100 and A100)

### INFN Cloud Services that can be made available to ICSC

At the date of this document, INFN Cloud can provide the following services to ICSC:

- Unified federated authentication via OpenID-Connect (OIDC) tokens, X.509 digital certificates, username/password
- Cloud storage, offered via an S3 interface
- Virtual machines, with different sizes and operating systems (1/2/4 vCPU, Ubuntu/CentOS)
- Jupyter Notebook as a Service, connected to persistent storage space
- Private Container Image Registry
- Automated container deployment directly via Dockerfile or Docker-compose files

DRAFT

- Kubernetes as a Service, with a configurable number of k8s nodes
- Sync-and-share as a Service (“Dropbox-like”), with a choice of either the Owncloud or Nextcloud backends, possibly associated to an encrypted filesystem.
- HTCondor-based Batch System as a Service, with a configurable number of worker nodes
- Elasticsearch & Kibana as a Service
- Spark & Grafana & Jupyter as a Service, deployed on top of a Kubernetes cluster, with a configurable number of k8s nodes
- Automated security scans on the deployed services
- A PaaS-level interface to external Cloud infrastructures such as OpenStack, Amazon Web Services, Google Compute Cloud, Microsoft Azure



# WP3, T3.1 - Risorse HTC, Cloud e HPC



- Unico task effettivamente attivo in WP3 da Ottobre 2022
  - Preparazione Survey per ricognizione risorse
    - Da inviare in attesa di avere un quadro completo con le gare PNRR
  - Gare per evoluzione datacenter (PNRR e non)
    - Non solo WP3 (molto materiale da Claudio Grandi)
    - Personale WP3 direttamente coinvolto nella scrittura di capitolati e nelle gare
  - Presentazione Loano:  
[https://agenda.infn.it/event/34683/contributions/197347/attachments/105506/148357/WP3\\_Infra\\_CCR\\_Loano.pptx](https://agenda.infn.it/event/34683/contributions/197347/attachments/105506/148357/WP3_Infra_CCR_Loano.pptx)



# WP3, T3.3 – Sostenibilità, Cost Models, Green Computing



## Da avviare

- Analisi dei costi delle gare degli ultimi anni
  - Evoluzione costi nel tempo per CPU, Storage, Tape, Energia Elettrica
- Cost Models
  - Flat budget vs altri sistemi (i.e. carbon footprint)
  - WP3 ha rappresentanti nel gruppo dedicato di WLCG
    - Poca attività fino a questo momento
- Green computing
  - Evoluzione della infrastruttura per rendere i datacenter più green
  - Capire con i siti quanto gli interventi che si stanno facendo in questa tornata di investimenti PNRR, miglioreranno il PUE e l'impatto ambientale (e quindi il TCO)



# Security

- OpenVAS operativo, in corso valutazione versioni a pagamento.
- Punto critico comune: Grafana istanziato da utenti (→ in corso valutazione di un Grafana aaS, mantenuto direttamente da DataCloud).
- NUCS: lavori SOC in corso, cf. presentazione di Gianluca Peco a Loano.
- Vedi dopo per criticità.



# Middleware

- Molto lavoro su **IAM**.
  - Aggiornamenti a ultime versioni su iam cloud, testing versioni development e StoRM (WebDAV, backend, puppet modules). Attesa per StoRM REST API di produzione.
- Molto lavoro su **orchestratore e dashboard**.
  - Diverse nuove feature introdotte, es. htcondor 9.x, Spark con IAM, NextCloud/Owncloud con IAM, IaaS-type features a livello PaaS.
  - Importanti attività in corso: es. elasticità k8s, ottimizzazione deployment, ottimizzazione apertura porte.
- **Criticità sulla manutenzione, aggiornamento, bug fixing di alcuni servizi PaaS**. Ci servirà a breve il supporto a MFA (vedi dopo).



# Data management & Software management



- Rispettato target di avere demo entro Loano – buon interesse.
- **Data Management:** Rucio + FTS
  - Interesse per integrazione di cataloghi esterni (alla DUNE).
- **Software Management:** integrazione trasparente per gli utenti di CVMFS, supporto a S3 buckets, in cantiere integrazione con git/gitlab.



# Follow-up Data Management

- Nella pianificazione delle attività, dobbiamo tenere in mente come pensiamo di esporre in produzione un sistema di DM agli utenti.
- Pensiamo che DataCloud debba supportare 2 modelli di deployment:
  1. **Sistema di Data Management di tipo «catch-all»**
    - Completamente gestito e supportato da DataCloud (come avviene ora per esempio per NaaS e servizio di object storage).
  2. **Sistema di Data Management ad uso di una singola comunità o esperimento**
    - DataCloud gestisce solo alcuni servizi (FTS, database). Gli altri (RUCIO, IAM, catalogo?) vengono gestiti dalla comunità.
- Stiamo definendo di conseguenza i task – anche per passare il DM in produzione verso WP1.



# Follow-up Software Management

- Attività di derivazione tesi magistrale di Giada Malatesta (LM Bioinformatics)
- Sono necessari fix, irrobustimenti, e nuove funzionalità:
  - A livello **infrastrutturale**
  - A livello **software DataCloud**
  - A livello di applicazione di **astrazione del sw CVMFS** (estendendo quanto fatto da Giada)
  - A livello di **“applicazioni avanzate”**
- È in fase di definizione di una lista di item (che si tradurranno in task) che dovranno essere assegnati alle persone
- Per il momento abbiamo discusso i primi 2 dei 4 bullet sopra.



# Ceph-rgw per object storage del backbone (migrazione da MinIO)



- Dimostrato in un testbed l'integrazione tra ceph-rgw e IAM.
- Dimostrato in un testbed l'integrazione tra ceph-rgw e OPA con relativa definizione di alcune policies di autorizzazione.
- Dimostrato in un testbed l'integrazione tra ceph-rgw e STS+rclone (per accesso "posix").
- Disponibile primo prototipo (anche se non con tutte le funzionalità necessarie) di una Web application per permettere l'accesso agli utenti all'object storage.
- Installato e configurato ceph-rgw nei 2 siti del backbone.
- Implementata tra i 2 siti la configurazione multi-site e fatti primi test.
- Altri test di funzionalità e performance in programma.



# Architettura «generale» di DataCloud

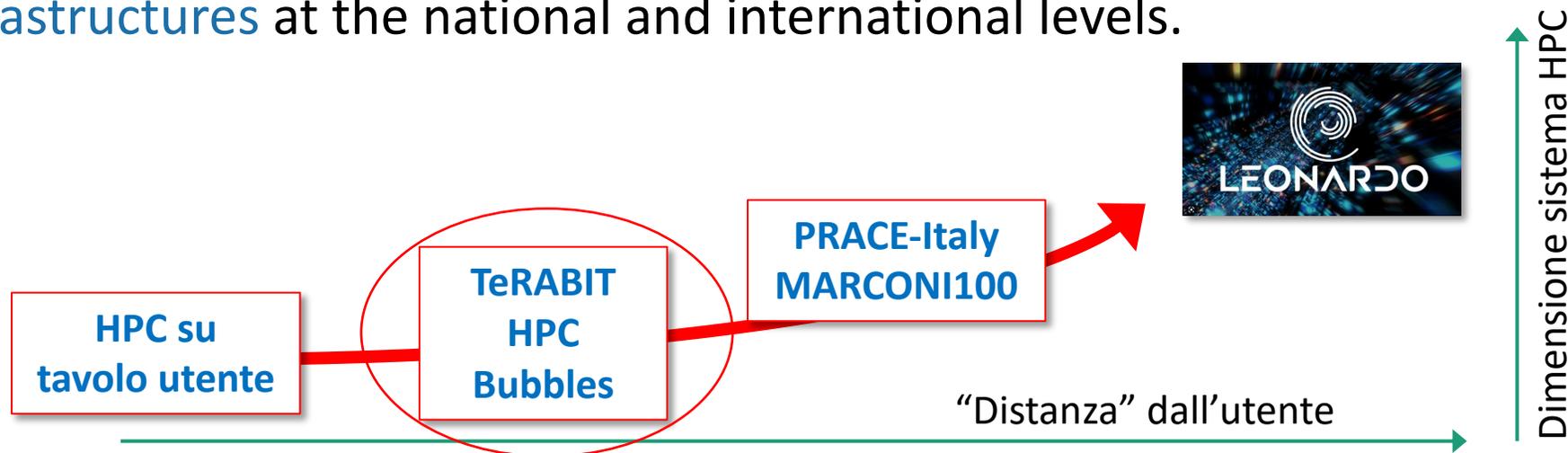


## Per riassumere:

- IAM media l'accesso "a tutto" (vedi anche dopo).
- La PaaS media l'accesso ai "servizi" ed espone un portafoglio espandibile.
- Il Data Management (Rucio, FTS, catalogo) gestisce le policy dei dati in modo indipendente dai back-end di storage (Tier-X, Cloud).
- CVMFS++ ottimizza la distribuzione di software e configurazioni.
- Le soluzioni di offloading (di cui non ho parlato, vedi InterTwin e TeRABIT) gestiscono la distribuzione dei workflow tra Tier-X, Cloud, HPC.

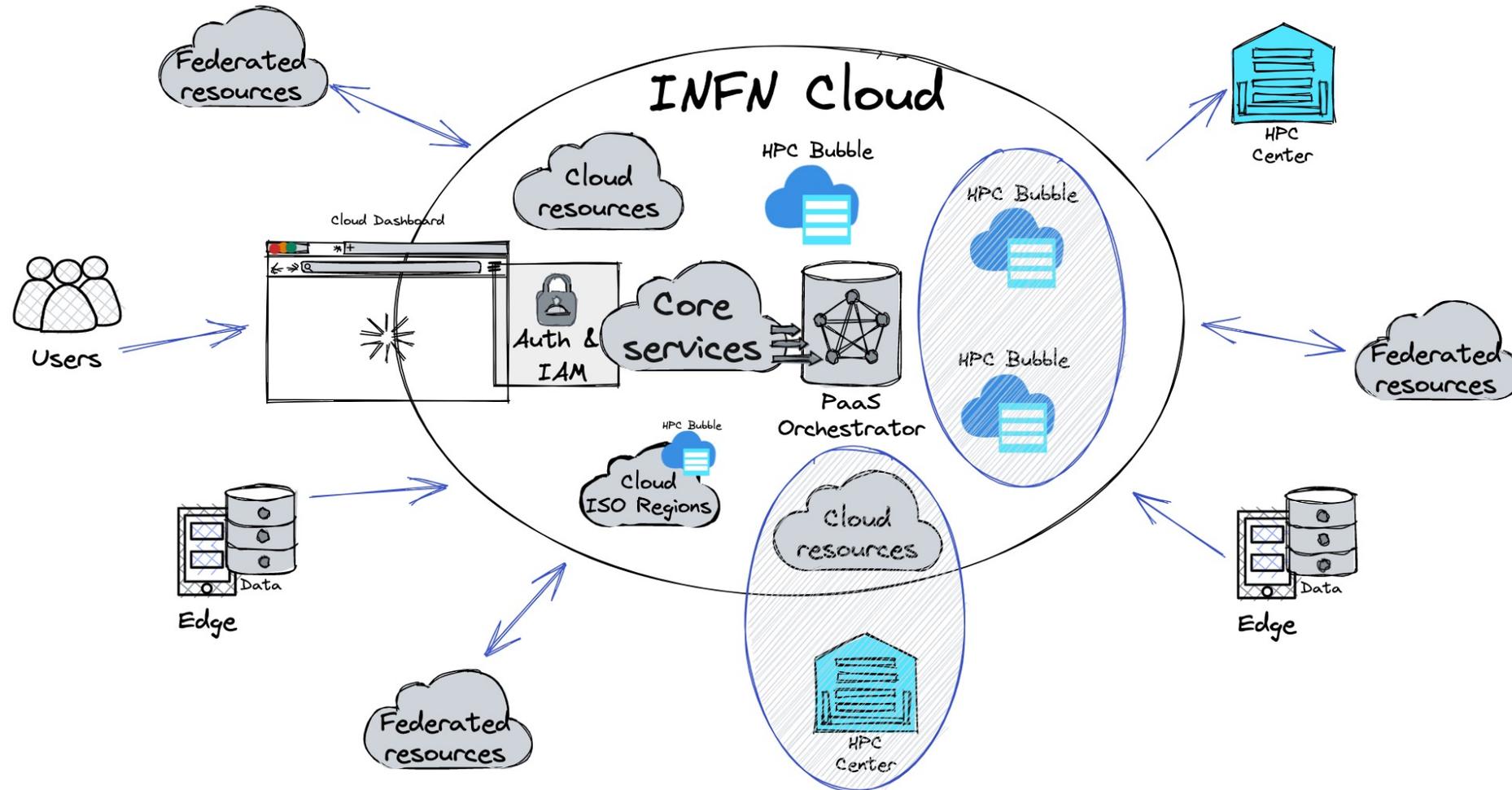
# An example: the “HPC Bubbles”

- As part of the TeRABIT PNRR project, we are implementing several highly innovative concepts:
  - Availability of **scalable HPC resources and services on INFN Cloud** through Cloud-native interfaces at the IaaS, PaaS and SaaS levels.
  - Strong **interaction between network, data and HPC/HTC resources**.
  - Communication and **federation between the HPC Bubbles and other HPC infrastructures** at the national and international levels.





# The *continuum* from Edge, to Cloud, to HPC





# Certificazioni, compliance

- **Attività chiave per diverse iniziative**, come ICSC/Spoke 8, TeRABIT, DARE, HBD-DataCloud.
- Definiti ambiti per analisi dei rischi.
- **Estensione dell'attuale certificato ISO 27001/27017/27018 da CNAF a AC, Bari, Catania**
  - Obiettivo: ottobre 2023.
  - ISO 9001 non per questa data ma previsto per il futuro.
- Svolti meeting a Catania il 29/6 e a Bari il 30/6 per la definizione delle procedure organizzative (vedi dopo).



# EPIC Cloud



Enhanced Privacy and Compliance Cloud – The INFN Cloud partition for personal and confidential data processing

- The GDPR states that Clinical and medical data (for instance, genomic) is personal data; i.e., it fits in the Art.9 special categories of personal data.
  - Genomic data is mostly impossible to be anonymized → GDPR shall always be applied
  - ISO/IEC 27001 is the main certification mechanism compliant with GDPR requirements (Art. 43, 58, 63)
- In order to comply with the requirements of health research projects INFN is involved in, we created **a region of the INFN Cloud infrastructure**, applied specific organizational and technical security measures, and certified it ISO/IEC 27001, 27017, 27018.
  - This is **EPIC Cloud**: a reference Cloud implementation for the treatment of sensitive data at INFN.

**From the Data Controller side, the fact that EPIC Cloud is ISO-certified is a way to demonstrate that processing is performed in accordance with the GDPR.**

# Punti di criticità

- DataCloud ha finora funzionato molto bene. L'impostazione generale pare sostenibile, l'architettura di riferimento è delineata sia verso l'interno sia verso l'esterno e le prospettive di sviluppo sono identificate.
- **Tuttavia ci sono alcuni punti che meritano un'attenzione particolare.**

# Integrazione delle nuove risorse hardware già arrivate in alcune sedi

- La federazione di nuove sedi sta procedendo, vedi ad esempio Catania. Questo in alcuni casi ci consentirà di incrementare in modo considerevole il pool di risorse disponibili.
  - Ad esempio, con la federazione di Catania in DataCloud, avremo a disposizione ulteriori ~15 NVIDIA A100, ~2000 vCPU e ~1PB di storage. Questo prima dell'installazione delle nuove risorse PNRR.

## MA

- Nelle prossime settimane occorre **concentrarsi sull'installazione delle risorse PNRR dei Tier-2.**
- Una **nota importante** è che lo storage acquisito per i Tier-2 è normalmente previsto per installazioni «tradizionali» (ovvero non è ad esempio ottimizzato per object storage alla CEPH). Questo lo gestiremo con i sistemi di data management citati sopra ma è un task che richiede forte collaborazione anche da parte dei Tier-2.

# Engagement delle persone

## Visione generale:

1. Abbiamo partecipato a **diversi bandi PNRR** legati al calcolo.
2. In ognuno di questi bandi abbiamo inserito **attività di interesse dell'Ente e del bando specifico**.
3. La integrazione di tutte le attività per tutti i progetti INFN relativi al calcolo **si realizza attraverso DataCloud**.

## Dunque:

- Per quanto riguarda **l'hardware ed i servizi**: essi verranno integrati all'interno del portafoglio servizi di DataCloud.
- Per quanto riguarda **il personale**:
  - **Il personale assunto su un progetto PNRR** lavora al **100%** sulle attività di quel progetto, che dovrà dunque in qualche modo seguire (le modalità differiscono progetto per progetto). Allo stesso tempo, ognuno dei neo-assunti è previsto che partecipi attivamente alle attività di uno o più WP DataCloud. Questo per realizzare l'integrazione indicata sopra.
  - **Per ogni sede in cui viene assunta almeno una figura PNRR** identifichiamo un «**responsabile DataCloud**», con il compito di fare da collegamento con le attività locali e di progetto.
  - **Il personale staff** è incoraggiato a partecipare a uno o più WP DataCloud, indipendentemente dal fatto che partecipi o no a progetti PNRR (o ad altri progetti).

# Esempio: assegnazioni di personale *tecnologo* su TeRABIT

Sede	Nome / Cognome	WP DataCloud di riferimento
Bari ( <i>ne manca 1</i> )	Giovanni Savarese	WP5-P
Catania ( <i>ne manca 1</i> )	--	--
CNAF	Luca Giommi	WP5-P
CNAF	Ahmad Alkhansa	WP6
CNAF	Andrea Rendina	WP6
CNAF	Stefano Antonelli	WP4
CNAF	Roberta Miccoli	WP5-A
CNAF	Carmen Giugliano	WP5-A
CNAF	Federico Fornari	WP6

Sede	Nome / Cognome	WP DataCloud di riferimento
LNL	Davide Marcato	WP5-A
Milano Bicocca	Stefano Pozzi	WP4
Napoli ( <i>ne manca 1</i> )	Simona Stellacci	WP1
Padova	Antonino Troja	WP6
Perugia	Niki Tombolesi	WP6
Pisa	Alessio Fiori	WP1
Torino	Nicola Mosco	WP1
Trieste	Alex Lenni	WP4

Sono state allocate / assegnate anche posizioni di *tecnico*: Bari (Vincenzo Di Gravina) e Catania (Enrico Orlando). Manca il tecnico di Napoli. **Per i tecnici, il WP DataCloud di assegnazione di «default» è WP1, con impegno previsto anche in WP2.**

# Procedura di «engagement» (esempio di TeRABIT)



- La prima cosa da fare consiste nel parlare con:
  - Il responsabile locale DataCloud
  - Il referente TeRABIT di WP4: oggi Davide Salomoni
    - Il PI e l'Infrastructure Manager TeRABIT: Mauro Campanella, Silvia Calegari
  - Il responsabile nazionale DataCloud: oggi Davide Salomoni
  - I WP leader dei WP DataCloud corrispondenti.
    - WP1: Stefano Stalio, Diego Michelotto; WP2: Carmelo Pellegrino, Emidio Giorgio; WP3: Giacinto Donvito, Daniele Cesini; WP4: Vincenzo Ciaschini, Luca Carbone; WP5: Marica Antonacci, Enrico Vianello; WP6: Massimo Sgaravatto, Daniele Spiga; WP7: Barbara Martelli, Nadina Foggetti.
  - Le discussioni avvengono sul canale «INFN Cloud» (aperto a tutti) su Microsoft Teams.
  - **Tutti i WP hanno già avviato contatti con i neo-assunti PNRR.**

**A parte le attività dei WP, l'engagement è un compito normalmente cross-WP e anche cross-project, e vista la complessità delle interazioni richiede di essere seguito quasi full-time, almeno in questa fase.**

# Consolidamento dei servizi esistenti

- Mano a mano che le richieste crescono, è necessario **raffinare alcuni dei servizi esistenti**. **Questo è normale, ma è necessario dedicare tempo ed interagire con le comunità che utilizzano questi servizi.**
- Ad esempio:
  - La disponibilità di un servizio HTCondor as a Service (che viene realizzata in modo trasparente su cluster Kubernetes) è stata testata solo molto recentemente ad una scala significativa. Sono state evidenziate alcune criticità, risolte abbastanza rapidamente.
  - Ci sono tuttavia sicuramente altre ottimizzazioni da fare, anche in altri servizi.
  - Diversi servizi devono avere più ridondanza in termini di «ownership» e supporto. Questo a tendere lo faremo con il nuovo personale PNRR.

# Input dall'esterno verso DataCloud

- Finora l'input sulla richiesta di servizi è avvenuto essenzialmente da parte di comunità che già in qualche modo erano in contatto con DataCloud.
- **Sarebbe molto utile avere input più ampio almeno da User Forum e dal WG Computing Models.**
  - Il punto qui non sono tanto i computing models di WLCG, quanto quelli di altri esperimenti e comunità scientifiche – ad esempio, DUNE, SKA, CTA, i vari esperimenti in Cina, i requirement delle comunità mediche, i requirement delle comunità di cultural heritage.
  - Anche report da gruppi su Open Science, da IDEM o da gruppi di lavoro CERN/WLCG sarebbero molto utili. A volte questi arrivano perché le persone corrispondenti sono presenti in DataCloud ma a volte questo non succede.  
**Qui anche il C3SN potrebbe avere un ruolo più importante.**

# Relazione con i progetti

- Perché DataCloud possa essere un vero ombrello di integrazione progetti, **occorre una figura (o più di una) che mantenga un collegamento stretto tra gli obiettivi di DataCloud e quanto richiesto dai progetti.**
- Ad esempio: report scientifico periodico di ICSC/Spoke x, di DARE, di TeRABIT, etc. Chi segue (non nel senso di «scrive») i Deliverable e le Milestone di progetto? C'è consistenza tra questi e gli outcome di DataCloud?
- Questo non ha a che vedere con il mandato del WG Progetti.

# Certificazione ISO

- Per l'espansione dello scopo della certificazione verso CNAF, Bari e Catania abbiamo bisogno di decisioni rapide.
  - Ri-certificazione di EPIC @ CNAF: **ottobre 2023**.
- La mia impressione è che qui stiamo procedendo lentamente. Mi preoccupa in particolare la situazione di Bari. Il rischio è di fallire la certificazione di EPIC.
- Attenzione a non sottostimare la necessità di competenze che ci consentano di interfacciarci con i vari comitati etici e DPO dei progetti ai quali partecipiamo.
  - Necessità di affrontare temi di compliance che richiedono competenze legali, unite ad una visione di ricerca.

# Security

- DataCloud si sta rapidamente espandendo verso un'infrastruttura che metterà a disposizione un numero significativo di risorse (CPU, GPU, storage, FPGA) a comunità che vanno molto al di là del solo INFN.
- **Questo richiede che investiamo rapidamente nella costituzione del SOC.** Probabilmente avremo anche bisogno di stabilire forti collegamenti con ACN e progetti PNRR collegati – vedi ad esempio talk di Luca Nicoletti a Loano.

# AAI e accesso alle risorse

- ICSC ha recentemente approvato la policy di accesso alle risorse. Questa va ora tradotta in termini implementativi consistenti con le policy INFN.
- Occorre finalizzare l'architettura, basata su INDIGO-IAM proxy, che consenta l'accesso a risorse INFN per utenti «esterni» (pubblici o privati).
  - Questo certamente richiederà l'implementazione dell'autenticazione a più fattori e la definizione di workflow appropriati.
- Questo non dipende solo da DataCloud e richiede una forte integrazione con CCR/Harmony, con il DPO e con tutti i progetti che si relazionano con DataCloud (es. ICSC, TeRABIT, DARE, HBD, etc.).
  - Il «RAC» è definito per ora solo per ICSC ma dovremo discutere un concetto analogo anche per altri progetti.

# Il coordinatore di DataCloud se ne va

- Questo lo sapete, ed è l'oggetto della prossima sezione di slides...
- È comunque una criticità.



# Proposta di nuova governance interna DataCloud da parte del PMB DataCloud



- Documento disponibile su <https://l.infn.it/datacloud-2023-07>.
- Si propone di mantenere inalterato il ruolo del Project Management Board (PMB) e di definire una nuova governance interna a DataCloud con:
  - Coordinatore di DataCloud.
  - Deputy Project Coordinator.
  - Chief Technology Officer (CTO).
- **Relazione con i progetti (PNRR e non):**
  - Istituzione di un **board dei PI di progetto** che si riunisca periodicamente insieme a coordinatore, deputy e CTO per tracciare avanzamento delle attività DataCloud vs. i vari progetti.
  - Creazione di una «**dashboard progetti**» con tutte le scadenze di es. Milestone e Deliverable, accessibile a tutti i PI di progetto.
  - Creazione di una «**hotline**» tra PI dei progetti e coordinatore, deputy e CTO DataCloud per richieste urgenti.

# Relazioni con strutture INFN

- In **ogni sede viene nominato un «referente DataCloud»** che ha il compito di coordinare le attività DataCloud a livello di sede.
- Si propone che coordinatore, deputy, CTO DataCloud e presidente C3SN si relazionino con **i Direttori delle sedi INFN** partecipanti a DataCloud per definire:
  - **Risorse umane e materiali** allocate a DataCloud.
  - **Obiettivi DataCloud** per ogni sede.
  - **Obiettivi Locali**, definiti da ciascun Direttore per la propria sede.
- Il coordinatore DataCloud **riporta al C3SN** gli obiettivi DataCloud, gli obiettivi locali e lo stato di avanzamento delle attività.
  - Il C3SN alloca le risorse al WG DataCloud e ne approva i piani di lavoro.
  - DataCloud riporta periodicamente lo stato di evoluzione del progetto al C3SN, facendo riferimento agli Obiettivi DataCloud e agli Obiettivi Locali.
  - Il C3SN formula raccomandazioni verso DataCloud, definisce le milestone di progetto e verifica periodicamente i risultati in termini di obiettivi raggiunti, rispetto dei tempi e scostamenti di budget.



# Proposta di nomi per la nuova governance di DataCloud



- Attualmente **hanno dato disponibilità a ricoprire i ruoli di coordinamento:**
  - Coordinatore di DataCloud: Barbara Martelli
  - Deputy del coordinatore di DataCloud: Daniele Cesini
  - CTO: Giacinto Donvito
- Da definire il board dei PI di progetto e il suo coordinatore.

