
DataCloud: prospettive e programma di lavoro

Claudio Grandi

Riorganizzazione di DataCloud

...veloce riepilogo

Contesto



Durante la riunione del C3SN di luglio e con successivo e-mail del 14/7, è stato dato mandato al gruppo di lavoro formato da:

Daniele Cesini, Giacinto Donvito, Claudio Grandi, Barbara Martelli, Daniele Spiga

Col mandato di redigere un documento contenente:

- una sezione con la descrizione degli obiettivi del progetto;
- la descrizione sintetica della struttura interna in Work Packages con i relativi deliverables e milestones associati;
- le connessioni con le altre realtà del calcolo INFN (es. siti, esperimenti, progetti ...)
- la struttura manageriale evidenziando anche la reporting line con il C3SN

Il documento prodotto è stato presentato al C3SN di ottobre e successivamente modificato secondo i suggerimenti e approvato dal management in dicembre

I compiti di DataCloud



DataCloud è uno dei WG del C3SN, col compito di **gestire ed evolvere l'infrastruttura per il calcolo scientifico dell'INFN**

DataCloud deve rispondere alle esigenze dei progetti delle CSN (**progetti interni**) e dei progetti derivanti da collaborazioni o call competitive nazionali ed internazionali (**progetti esterni**), che insistono sull'infrastruttura di calcolo scientifico dell'INFN

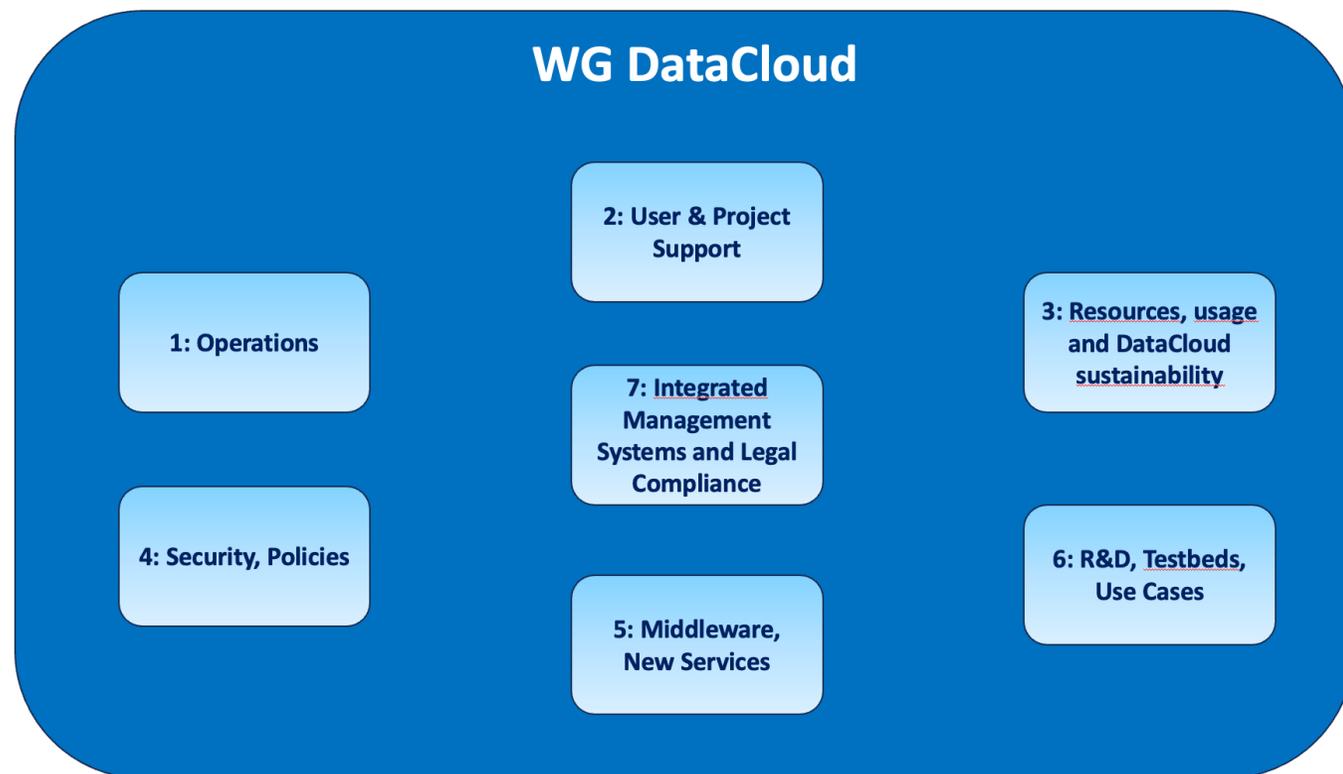
DataCloud ha un ruolo essenziale nella partecipazione ai progetti PNRR sul calcolo che coinvolgono l'INFN, in particolare il Centro Nazionale ICSC e TeRABIT, dove l'INFN ha ruolo di leadership

Organizzazione in WP



L'organizzazione in Work Package di DataCloud definita nel 2022, è stata considerata ancora valida

Nel documento definiti collaborazioni con WG del C3SN (WG Tecnologie, WG Computing Model)



Davide Salomoni, "WG Infrastruttura: mandato, struttura generale e operativa" Kick-off WG Infrastruttura (DataCloud), 17-18 ottobre 2023 <https://agenda.infn.it/event/32364/contributions/178083/attachments/98155/135704/WG%20DataCloud%2020221017%20finale.pdf>

Progetti interni ed esterni



I **progetti interni** (AKA esperimenti) hanno durata in generale lunga, hanno canali consolidati di comunicazione con chi gestisce l'infrastruttura e meccanismi di referaggio ben precisi. In alcuni casi (ad es. esperimenti LHC) utilizzano servizi che, indipendentemente dall'evoluzione dell'infrastruttura, devono essere mantenuti (Grid)

I **progetti esterni** hanno durata in generale breve e natura molto variabile, con esigenze diverse; devono essere gestiti in modo dinamico

Interazione coi progetti



I progetti interni mantengono la possibilità di utilizzare i canali di comunicazione e supporto con DataCloud in essere ed è garantito il funzionamento dell'attuale infrastruttura Grid; estensione per supporto per accessi di tipo Cloud

È essenziale che DataCloud sia coinvolto nelle decisioni riguardo alla partecipazione dell'INFN a progetti che hanno impatto sull'infrastruttura di calcolo

- Per valutare se le richieste del progetto sono compatibili con le risorse disponibili e con i piani di evoluzione dell'infrastruttura

- Per valutare se l'effort messo a disposizione dal progetto per svolgere le attività previste è sufficiente ed è possibile integrarlo in DataCloud

- Per concordare a priori le modalità di gestione delle risorse e del personale che entra a tutti gli effetti a far parte di DataCloud

DataCloud porta il contributo tecnico alla discussione nel WG Progetti e nel C3SN

Creazione di una [Dashboard progetti](#) in cui tracciare le caratteristiche e le scadenze

Struttura di governo 1/2



Project Management Board (PMB)

Composto dai coordinatori dei WP, gestiscono le attività tecniche di DataCloud, definisce ed evolve l'architettura di Datacloud

Coordinatore

Coordina le attività, presiede il PMB, rappresenta DataCloud verso il C3SN, i WG del C3SN, le strutture INFN e gli stakeholders di DataCloud, sovrintende al bilancio

Mandati legati al C3SN, nomina dal C3SN con l'accordo del management

Struttura di governo 2/2



Deputy

Coadiuvava il coordinatore in tutti gli aspetti della sua attività

Chief Technology Officer (CTO)

Coadiuvava il coordinatore nell'analisi delle tecnologie da usare, propone strategie, contribuisce alla definizione dell'architettura e alla sua evoluzione, mantiene i contatti con i responsabili locali

Program Officer

Mantiene la dashboard dei progetti, assiste il coordinatore e il PMB nella definizione del programma di lavoro di DataCloud

Mandati legati a quello del coordinatore, nomina da C3SN con l'accordo del coordinatore

Implementazione della riorganizzazione

Spunti per la discussione

PMB e WP manager



WP1 Operations:

Stalio, Michelotto

WP2: User & Project support:

Pellegrino, Giorgio

WP3: Risorse, Data Lake, Sostenibilità:

Cesini, Donvito

WP4: Security, Policies:

Ciaschini, Carbone

WP5: Middleware, Nuovi Servizi:

Antonacci, Vianello

WP6: R&D, Testbed, Use Cases:

Spiga, Sgaravatto

WP7: Sistemi integrati di gestione e Legal Compliance:

Martelli, Faggetti

Gruppo di coordinamento



A parte il coordinatore, tutti i membri del coordinamento sono già membri del PMB:

Coordinatore: CG

Deputy: Martelli

CTO: Donvito

PO: Spiga

Inoltre, Cesini sarà deputy di CG nel coordinamento di ICSC Spoke 0 e sarà comunque parte del coordinamento DataCloud

Commenti generali



Bisogna imparare a lavorare in un modo nuovo, con un gruppo di coordinamento invece che con un singolo coordinatore

Le attività del PMB devono ripartire – da troppo ormai manca il coordinamento

Necessario coinvolgere il nuovo personale – l'impressione è che oggi non stiamo sfruttando appieno le potenzialità

Non vedo criticità per quanto riguarda il supporto ai progetti interni, ma è essenziale vigilare perché gli impegni su quelli esterni non tolgano risorse

Il supporto ai progetti esterni, in particolare quelli del PNRR, richiederanno uno sforzo notevole e, almeno nelle fasi iniziali, impegneranno molto il WG

Relazioni coi siti



Coinvolgimento dei siti nelle attività di DataCloud anche per il supporto ai progetti esterni e alle attività per i PoC

Coinvolgere i direttori quando necessario

Supporto per le applicazioni Cloud

Più intense relazioni fra siti e WP2 per il supporto agli utenti

Migliore e più uniforme gestione della sicurezza informatica

Estensione a tutti i siti delle buone pratiche identificate per i siti certificati, quando opportuno

Relazioni con i progetti interni



Rafforzamento delle relazioni con i referee del calcolo e col WG tecnologie per quanto riguarda gli aspetti di programmazione e sostenibilità

Maggiore interscambio con gli attuali forum di comunicazione (CdG Tier-1, Forum, ...) e di supporto

Speriamo che il prossimo coordinatore del Forum del Calcolo INFN sia meglio di quello uscente!

Più organica raccolta dei requisiti e comunicazione delle opportunità su DataCloud, in collaborazione anche con il WG Computing Model

Relazioni con i progetti esterni



Con l'introduzione del PO e della dashboard progetti, si conta di riempire alcuni gap fra le attività di DataCloud e quelle dei progetti esterni

Maggiore integrazione, per evitare che i progetti esterni siano percepiti come «corpi estranei»

Non solo supporto ai progetti ma anche aiuto dai progetti per implementare funzionalità utili per tutti

L'impressione è che in passato ci si sia basati soprattutto sulle relazioni interpersonali fra partecipanti ai progetti e a DataCloud. Serve maggiore copertura

Migliore programmazione, in collaborazione col WG Progetti, per evitare sovraccarichi

ICSC e TeRABIT (ma non solo)



Dobbiamo creare l'infrastruttura Cloud nazionale

Essenziale concordare con CINECA e gli altri attori gli strumenti per interoperare

Infrastruttura di Autenticazione Autorizzazione

Trasferimento e accesso ai dati

Accesso alle risorse di calcolo (in particolare tramite slurm per HPC)

Monitoring e accounting

Supporto utenti (ticketing)

Abbiamo già iniziato a discutere su come costruire l'infrastruttura AAI

Procedendo allineati con le attività nazionali e internazionali (e.g. eduGAIN)

Siamo in ritardo rispetto alle scadenze dei progetti!

Scadenze di «alto livello»



Nel documento erano state individuate alcune scadenze di «alto livello» derivanti dai principali progetti PNRR

Deliverable/Milestone	Data	DataCloud WP (Resp.)	DataCloud WP (Coll.)
Analisi dei requisiti dei progetti attivi	ASAP	PMB	WP2, WP3
Descrizione dell'architettura dell'infrastruttura	ASAP	PMB	WP1, WP3
Identificazione del middleware stack e del catalogo iniziale di servizi	ASAP	PMB	WP1, WP5, WP7
Deployment di un PoC di accesso federato, anche da utenti non INFN, alle risorse di siti INFN, incluse risorse HPC	31/12/23	WP6	WP1, WP3, WP5
Inclusione di componenti ad alta sicurezza nel PoC	30/06/24	WP7	WP1, WP4, WP5, WP6
Inclusione di risorse non INFN nel PoC	30/09/24	WP6	WP1, WP3, WP5
Infrastruttura federata italiana in produzione	30/06/25	WP1	
Piattaforma con componenti ad alta sicurezza in produzione	31/08/25	WP1	WP7

*Completata
l'impostazione
In discussione
con CINECA
e GARR
In discussione
con CINECA e GARR*

Nel brevissimo termine



Dobbiamo completare l'acquisizione delle risorse

Gare CPU, storage, router: già avviate le indagini in collaborazione con il WG Tecnologie

In attesa di meccanismi più sofisticati, dobbiamo dare accesso alle risorse disponibili oggi agli utenti esterni all'INFN

Definire le risorse fisiche

Definire la metodologia (accesso Grid, Cloud, ...)

Definire le policy (registrazione utenti, nomine ad amministratori, ...)

Abilitare il supporto utenti «allargato»

Dobbiamo definire al più presto nel PMB la lista delle attività e delle priorità

Dobbiamo organizzare una sorta di *all-hands* col nuovo personale

Il gruppo ristretto del C3SN



Deve essere inteso come un aiuto a DataCloud

per l'identificazione delle priorità

per portare all'attenzione del C3SN le esigenze

per affrontare le situazioni critiche