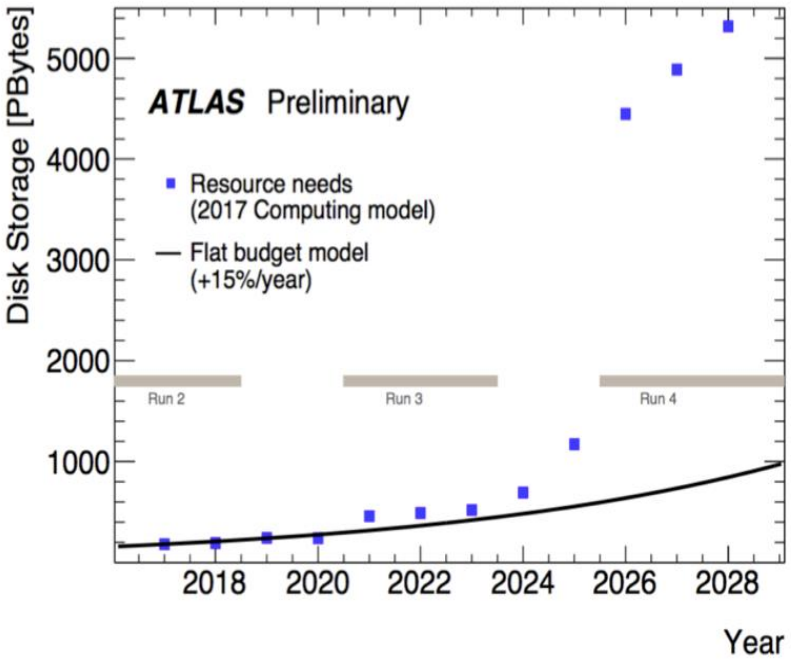
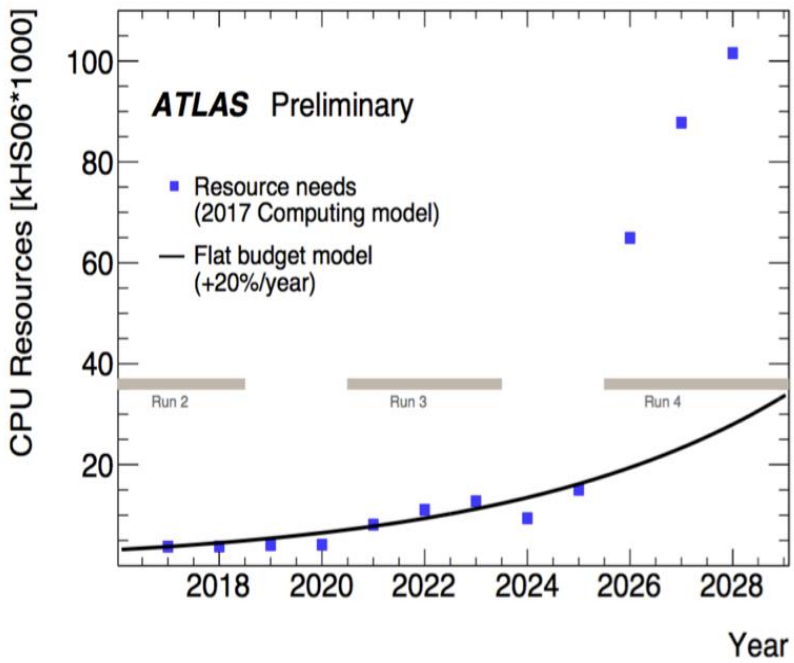
IDDLS: Italian Distributed Data Lake for Science

D.Cesini – CNAF

Motivation

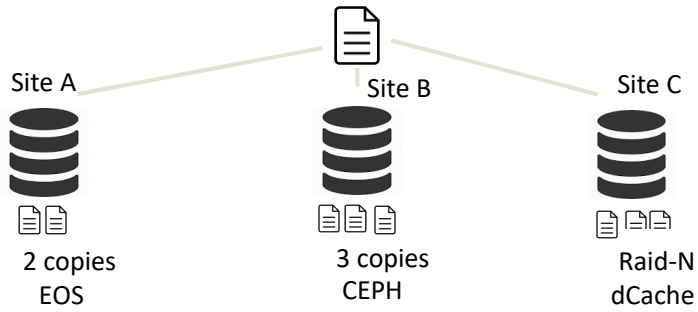
- HL-LHC computing needs are above the expected technology evolution (15%/yr) and funding (flat)
- We need to optimize hardware usage and operational costs
- High fraction of the infrastructure and operation costs is due to Storage





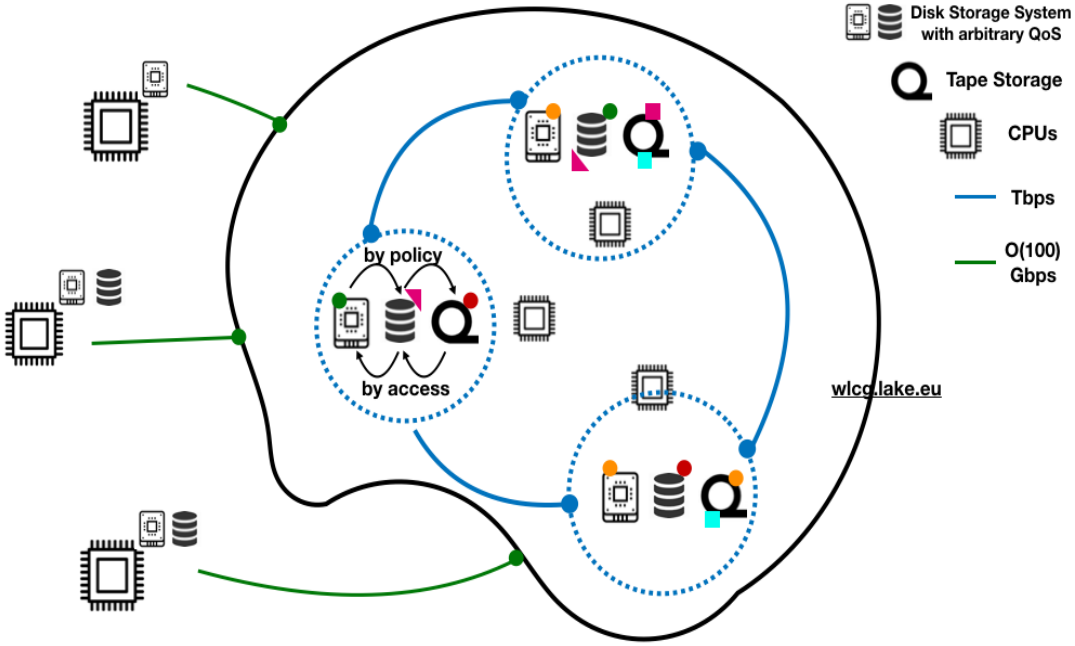
Some ideas on reducing Storage costs

- Reduce hardware cost: better exploiting the concept of QoS(Quality of Service)
  - Probably today we replicate more than we need
    - Reducing the number of copies



A stronger integration of sites could lead to a reduction of the number of copies

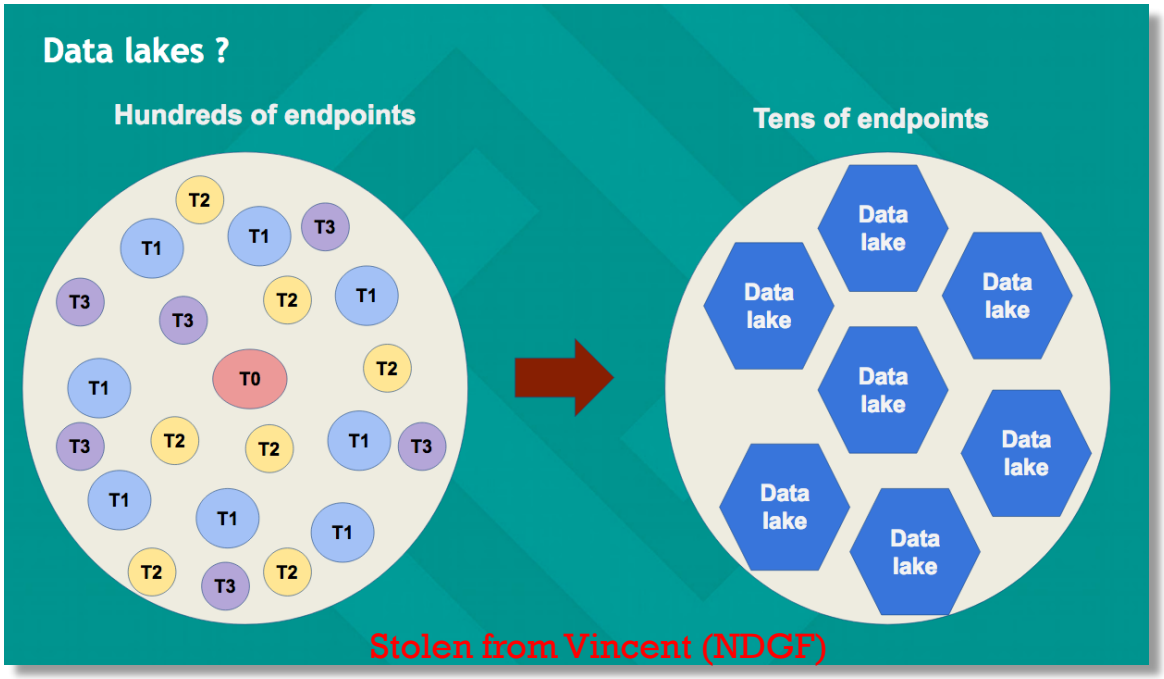
- Reduce Operational Cost: deploy fewer (larger) storage services maintaining high standards in availability and reliability
  - Create large storage repositories that "look like one, but it is composed of many" ➔ **the DataLake**
- Co-location of Storage and CPU will not be guaranteed anymore
  - Need technologies for quasi-transparent data access from remote locations
    - Smart Caching



On the definition of the *Data Lake*

- Very diverse understanding of the expression *data lake*.
  - Attempts to define it by 'name space' or 'region' or 'country' all failed.

- **Looks like one thing, but is composed of many**

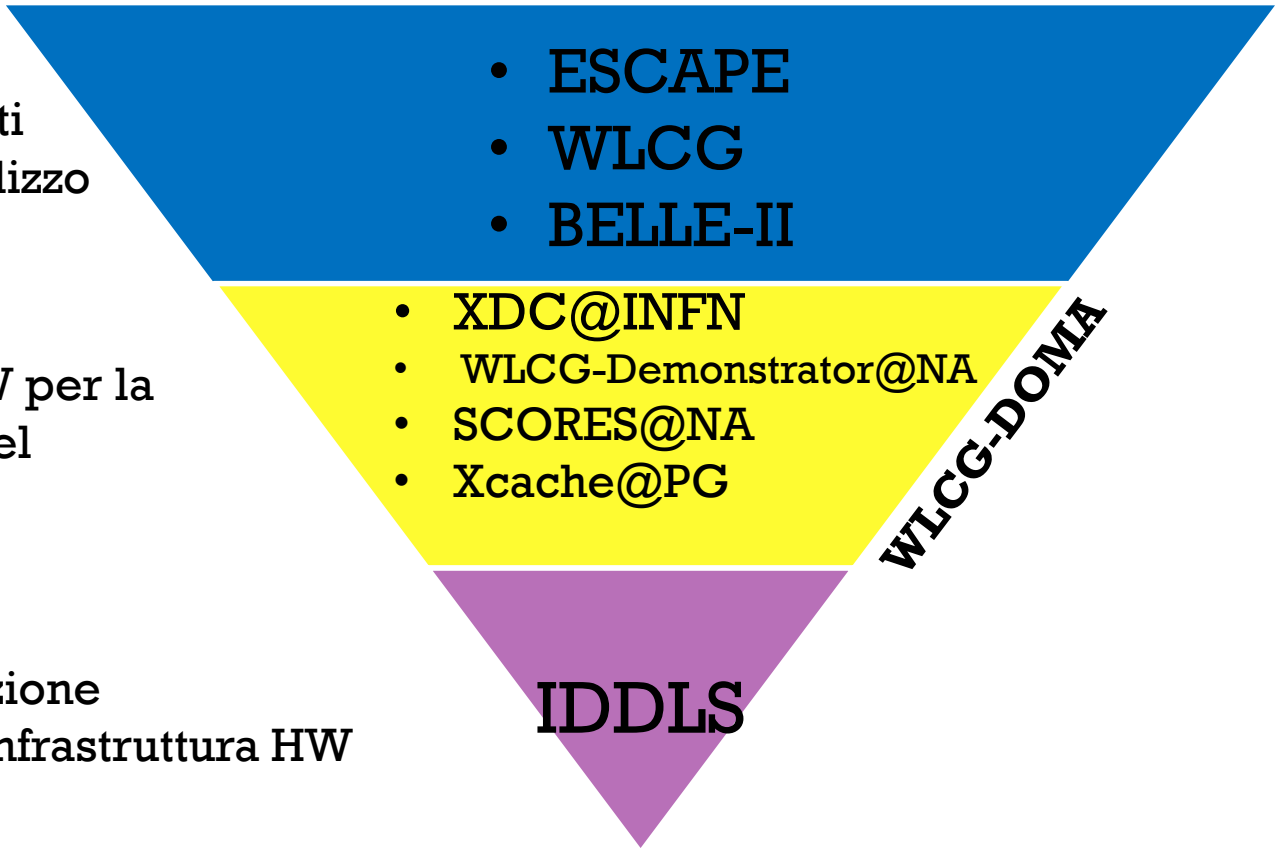


Sinergie in Europa

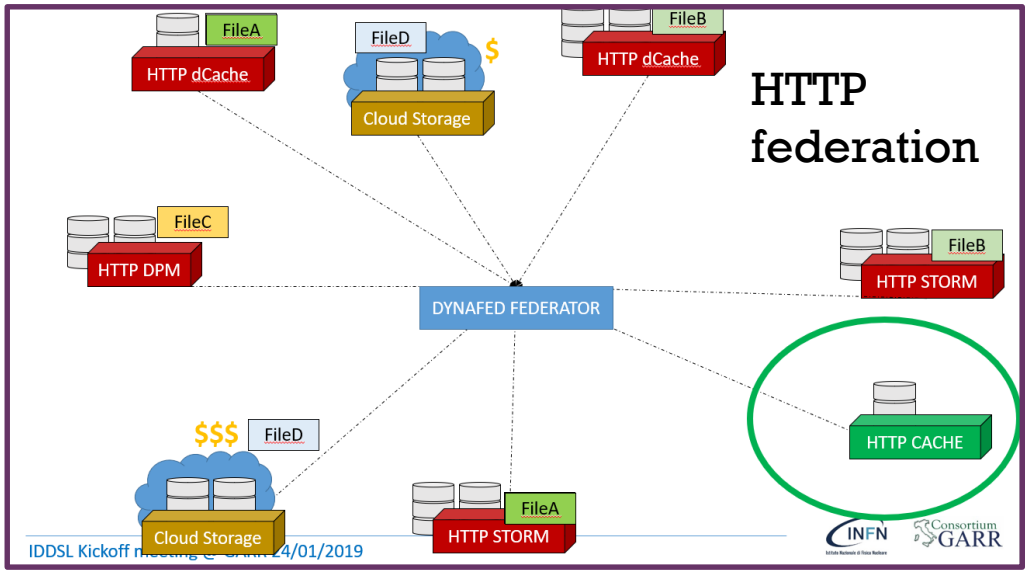
Comunità di utenti interessate all'utilizzo della tecnologia

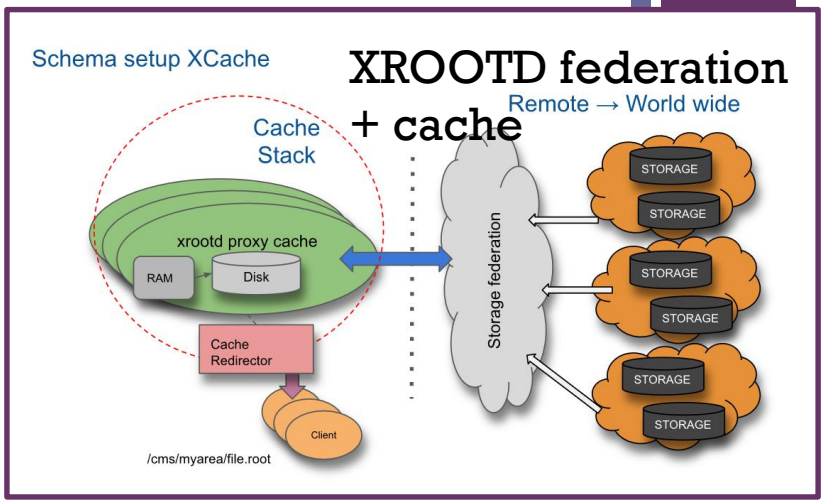
Sviluppo SW per la creazione del DataLake

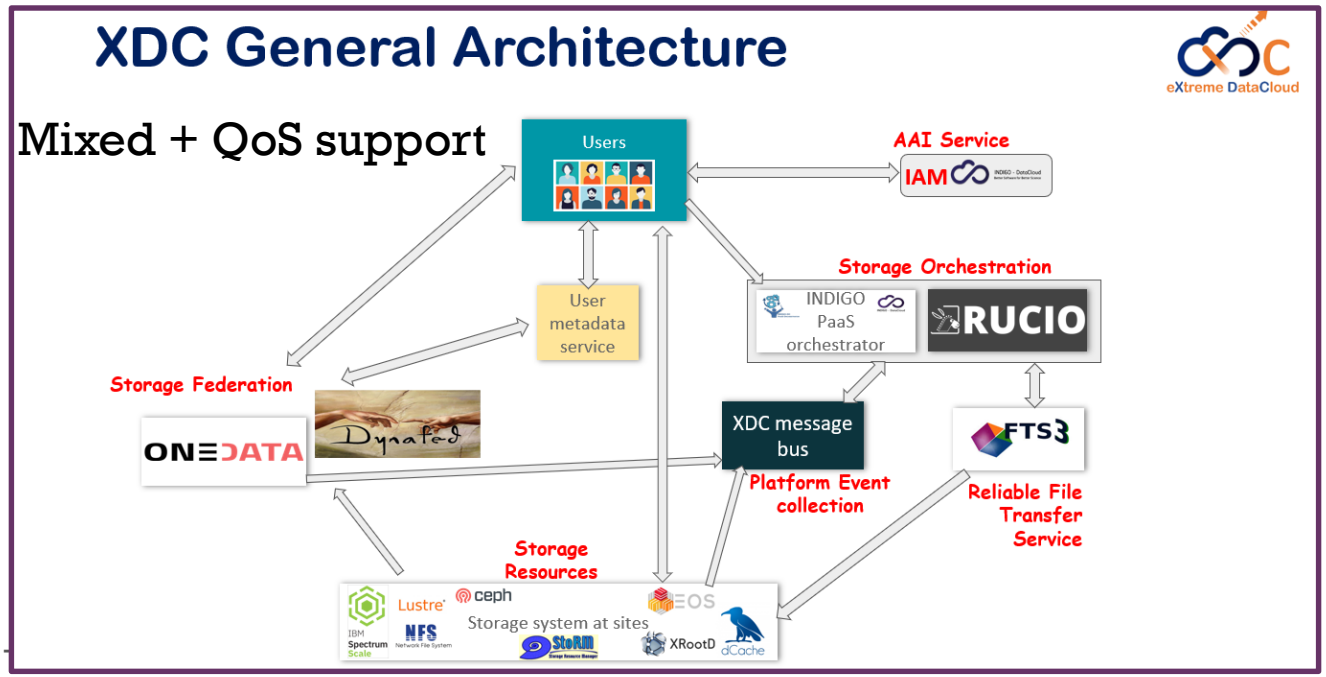
Creazione dell'infrastruttura HW



Lake Architecture(s)

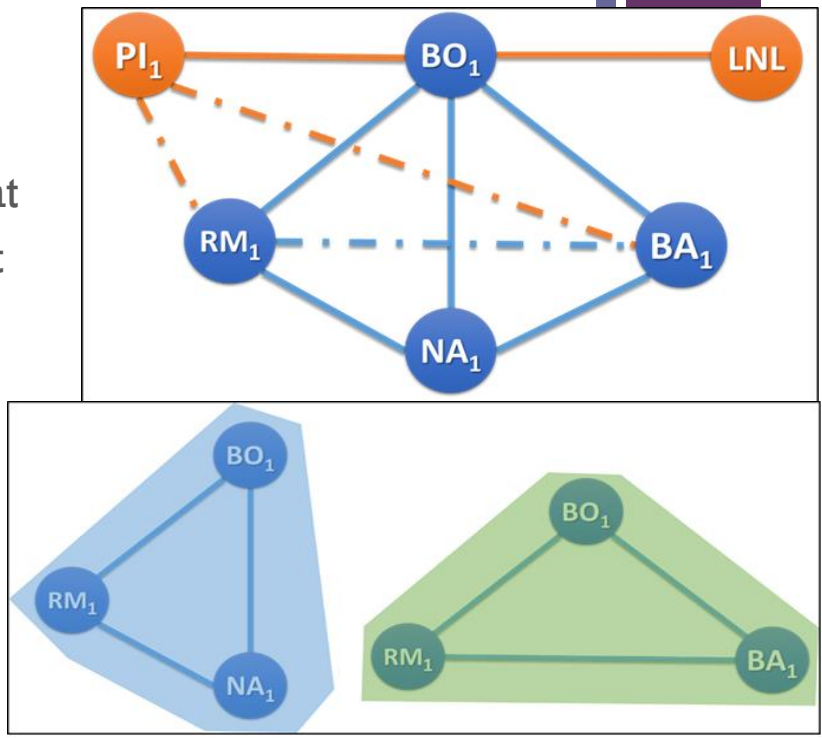






The project

- INFN-GARR collaboration to realize a prototype of an Italian DataLake exploiting:
  - Last generation networking technologies provided by GARR
    - DCI (Data Center Interconnection) equipment
    - SDN (Software Defined Network) deployment
  - Software for creating **scalable storage federations** provided by INFN
    - eXtreme-DataCloud project
    - SCoRES project (INFN-NA)
  - Real life use cases for testing
    - CMS
    - ATLAS
    - BELLE-II
    - Possibly involving LNGS experiments (XENON) and VIRGO



Possible topologies of the GARR Network with DCI and SDN for the DataLake

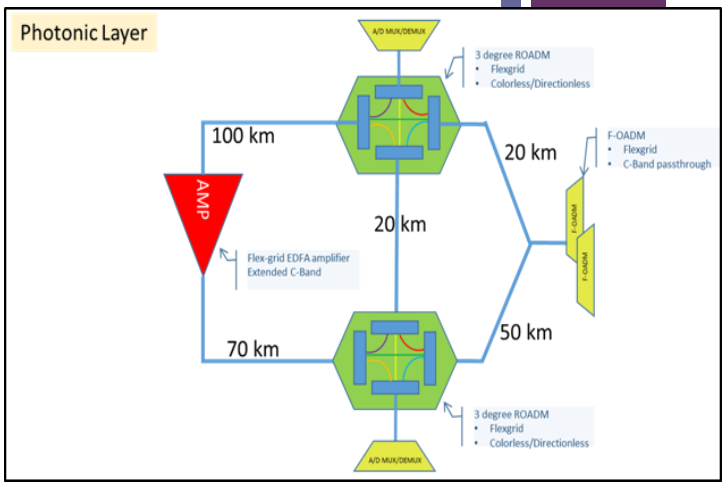
Timeline

- **3 years project**

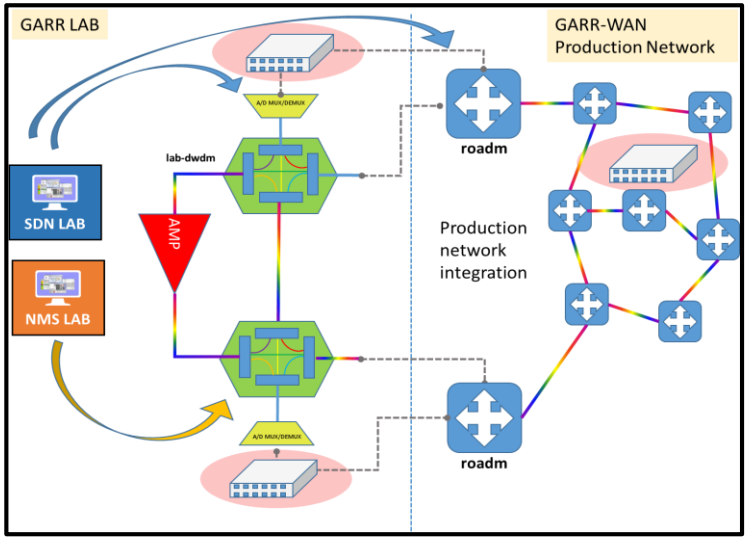
- First year
  - Technology scouting for DCI equipment to be deployed by GARR
  - Application (INFN) requirements analysis
  - Network equipment acquisition (INFN and GARR) and Lab testing
  - Deployment on mixed Lab+WAN environment of the networking equipment
  - Creation of the DataLake on sites connected with standard networking and first prototype using DCI

- Second year
  - Testing of the mixed (Lab+WAN) configuration
  - Final creation of the DataLake on the 3 INFN sites with DCI systems
  - Performance evaluation and comparison
  - Possible acquisition of new equipment with increased performance

- Third year
  - Deployment only on WAN of the networking equipment
  - Optimization of the DataLake
  - Performance evaluation
  - Final consideration



Lab deployment at GARR for testing



Mixed Lab+WAN deployment

Milestone 2019 e stato

- **30/06/2019: Scelta degli apparati di networking per la creazione del Datalake**
  - Apparati GARR Identificati
    - Infinera Groove G30 ?
  - Richiesto e ottenuto in CSN5 sblocco fondi SJ per acquisto switch con porte a 100GB
    - Arista 7050 o 7020
      - Capitolato in fase di definizione
- **10/09/2019: refereggio CSN5**
- **31/12/2019: Deployment degli apparati** di rete in una configurazione Lab+WAN – primi portotipi DataLake su apparati DCI e standard
- **31/12/2019: Primi run per la valutazione delle performance sui prototipi**

Budget Assegnato

Tasca 40k al cnaf
Per 2/3 switch 100G
5k per NIC

Sez. & Suf.	MISS				CON				ALTRICONS				TRA				SEM				PUB				MAN				INV				LIC-SW				APP				SPSERVIZI				TOTALE			
		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.		Sj	Dot.	Ant.
BA	1.5																																											1.5			
	0.5																																											0.5			0
CNAF	3.0																												65.0															68			
	1.5																												5.0	40.0														6.5	40.0		0
LNF	1.5																																											1.5			
	0.0																																											0			0
LNL	1.5																																											1.5			
	0.5																																											0.5			0
NA	1.5	0.5																																										1.5	0.5		
	0.5	0.0																																										0.5			0
PI	1.5																																											1.5			
	0.5																																											0.5			0
RM1	1.5																																											1.5			
	0.0																																											0			0
TOTALE	12	0.5																											65															77	0.5		
				12.5				0				0				0				0				0				0				65				0				0				0				77.5
	3.5		0	0																							5	40	0	0													8.5	40.0	0.0	0.0
				3.5				0.0				0.0				0.0				0.0				0.0				0.0				45.0				0.0				0.0				0.0				48.5

+200k inkind dal GARR per apparati

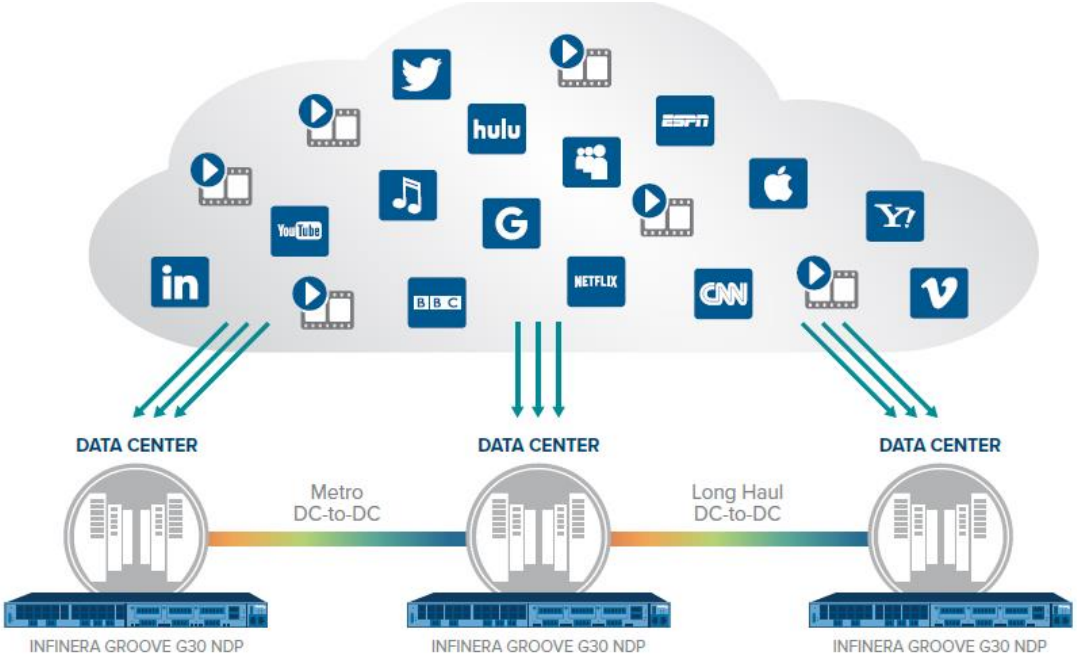


Figure 1: Powering High Performance, Cost-efficient Data Center Connectivity

THE PURPOSE-BUILT INFINERA GROOVE G30 NETWORK DISAGGREGATION PLATFORM

The Infinera Groove G30 Network Disaggregation Platform (NDP) is an innovative 1RU modular open transport solution for cloud and data center networks that can be equipped as a muxponder terminal solution and as an Open Line System (OLS) optical layer solution. Purpose-built for interconnectivity applications, the disaggregated Groove G30 delivers industry-leading density, flexibility, and low power consumption.

Infinera Groove G30 Open Line System (OLS)

Infinera Groove G30 Muxponder (MUX)

BACKUP

Gruppi di lavoro

- WP1 – Management
  - Coordinamento, rapporti CSN5 e referee, organizzazione meeting
  - Progress report periodici
  - Procedure acquisti
- WP2 - Studio, definizione e implementazione dei link ad alta velocità
  - Scouting tecnologico delle soluzioni Data Centre Interconnect (DCI)
  - Identificazione dei requisiti degli esperimenti INFN
  - Integrazioni delle componenti HW e SW delle tecnologie DCI
  - Sperimentazione mista laboratorio e infrastruttura di rete geografica
  - Condivisione dello spettro in ambiente protetto;
  - Modelli di provisioning
  - Modelli di gestione e controllo
  - Sperimentazione su infrastruttura geografica su 3 siti

Gruppi di lavoro

- WP3 – Creazione del DataLake
  - Definizione dello stato dell'arte delle tecnologie esistenti
  - Implementazione del DataLake con tecnologie basate su protocollo HTTD/XROOTD con e senza sistemi di caching
  - Implementazione del DataLake con tecnologie differenti (eventuali)
- WP4 – Testing del DataLake
  - Definizione della testsuite del progetto basata sul software degli esperimenti rappresentati, almeno CMS, ATLAS e BELLEII
  - Esecuzione della testsuite sul DataLake sfruttando sia i siti interconnessi con tecnologie di tipo DCI che di tipo legacy
  - Interazione con sedi INFN o legate all'ente produttrici di dati (i.e. LNGS, CASCINA) che possano essere interessate a testare le soluzioni dell progetto
  - Valutazione delle performance ottenute