CMS Data Transfers

A.Sartirana (LLR, E.Polytechnique, Paris)

Intro overview

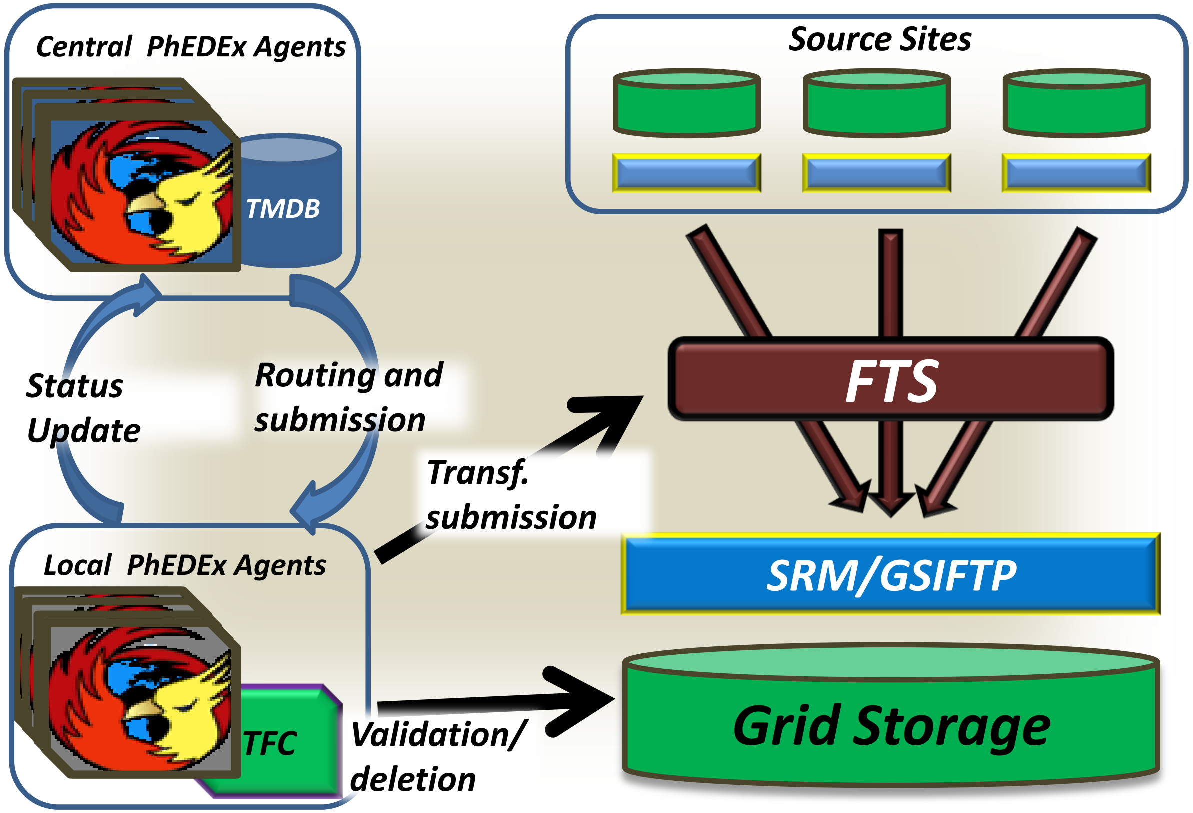
- The CMS ***Data Placement strategy evolved*** during years
  - computing model (2005) [*]: ***static***, hierarchical, ***local*** data privileged;
  - good ***reliability and performance of networks***: evolved (2008) into a "***full mesh***", more ***WAN dependent***;
  - ***infrastructures upgrades*** (LHCOne) and ***new tools*** (Xroot fed.): evolving into more ***dynamical*** and ***WAN-based*** data access;
- ***evolution*** possible ***thanks to*** CMS data mgmt tools
  - ***PhEDEX***: robust and ***flexible*** data placement system;
  - ***link commissioning*** system: ***monitoring*** on the real NW infrastructure ***performances***;
- in what follows…
  - brief ***intro*** to ***PhEDEx and Link Commissioning***;
  - ***evolution*** of the CMS data placement;
  - outlook to ***future evolution***.

[*] CMS C-TDR released (CERN-LHCC-2005-023)

Intro

CMS Data Transfer and Placement System

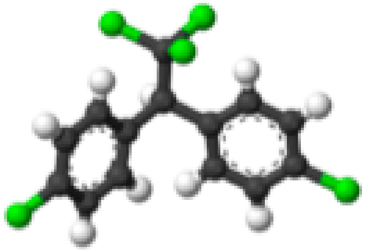
- central brain (CERN) and local agents at sites: ***routes data requested to a site*** from all available sources;
- extremely ***flexible***: can adapt to any data distribution model;
- ***performing***: able to saturate NW connections available between sites;
- ***reliable*** and robust.



Intro link commissioning

- ***Infrastructure for commissioning*** (validate) ***links***
  - dedicated PhEDEx instance: ***constant testing;***
  - ***links have to be "enabled"*** to be available for "production" data transfer;
  - links ***should gain minimal performances*** to be enabled;

- ***Debugging Data Transfer (DDT) project***
  - created in 2007 to ***support link commissioning***;
  - experts to ***help and coordinate*** the sites administrators in ***debugging*** their links;

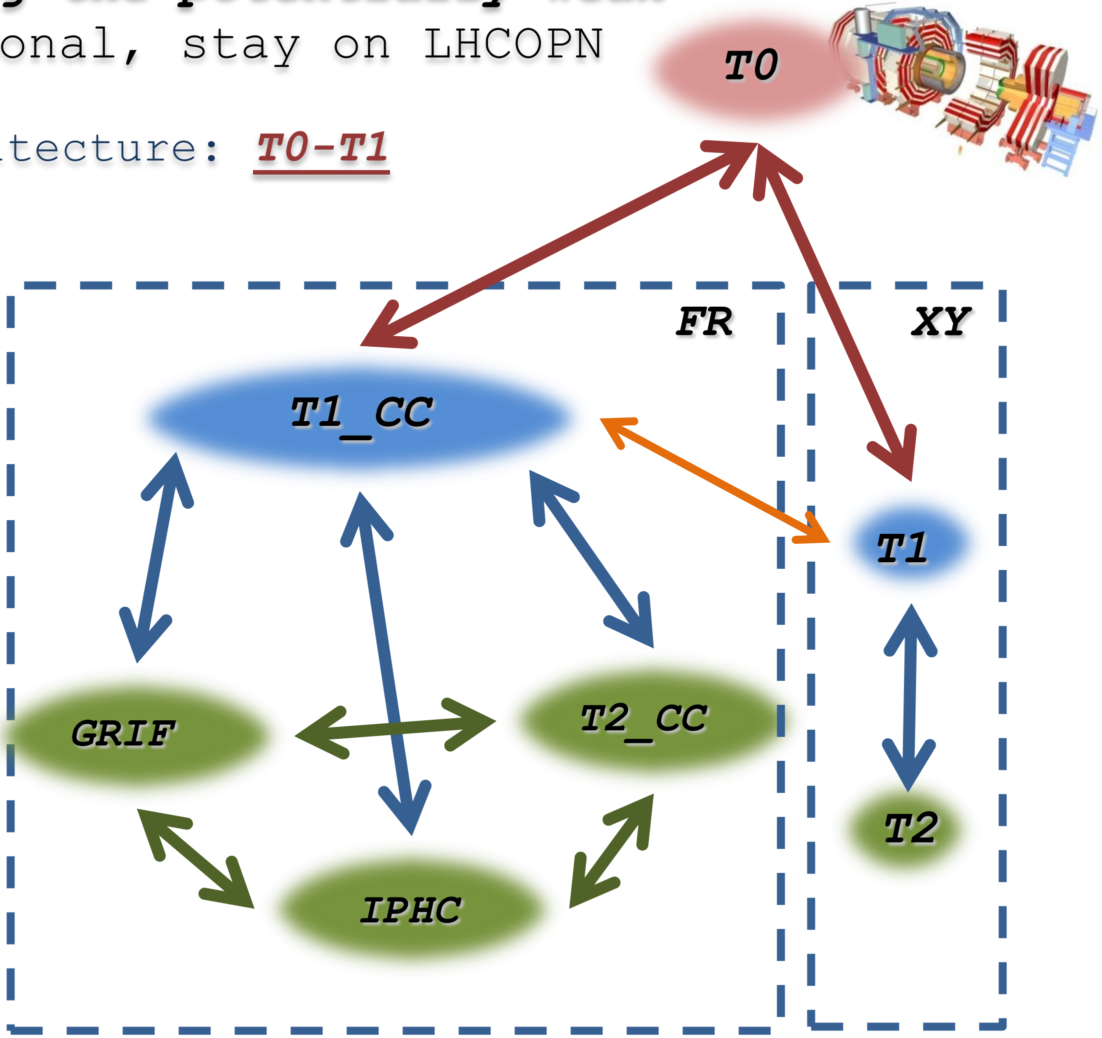


  - fundamental role in ***creating the actual backbone of*** CMS sites ***connections;***
  - ended in 2010: now maintained by the Data Transfer Team.

[*] "The CMS Data Transfer Test Environment in Preparation for LHC Data Taking", IEEE-2008
"Debugging Data Transfers in CMS" CHEP09
"Large scale commissioning and operational experience with T2-T2 data transfer links in CMS" CHEP10

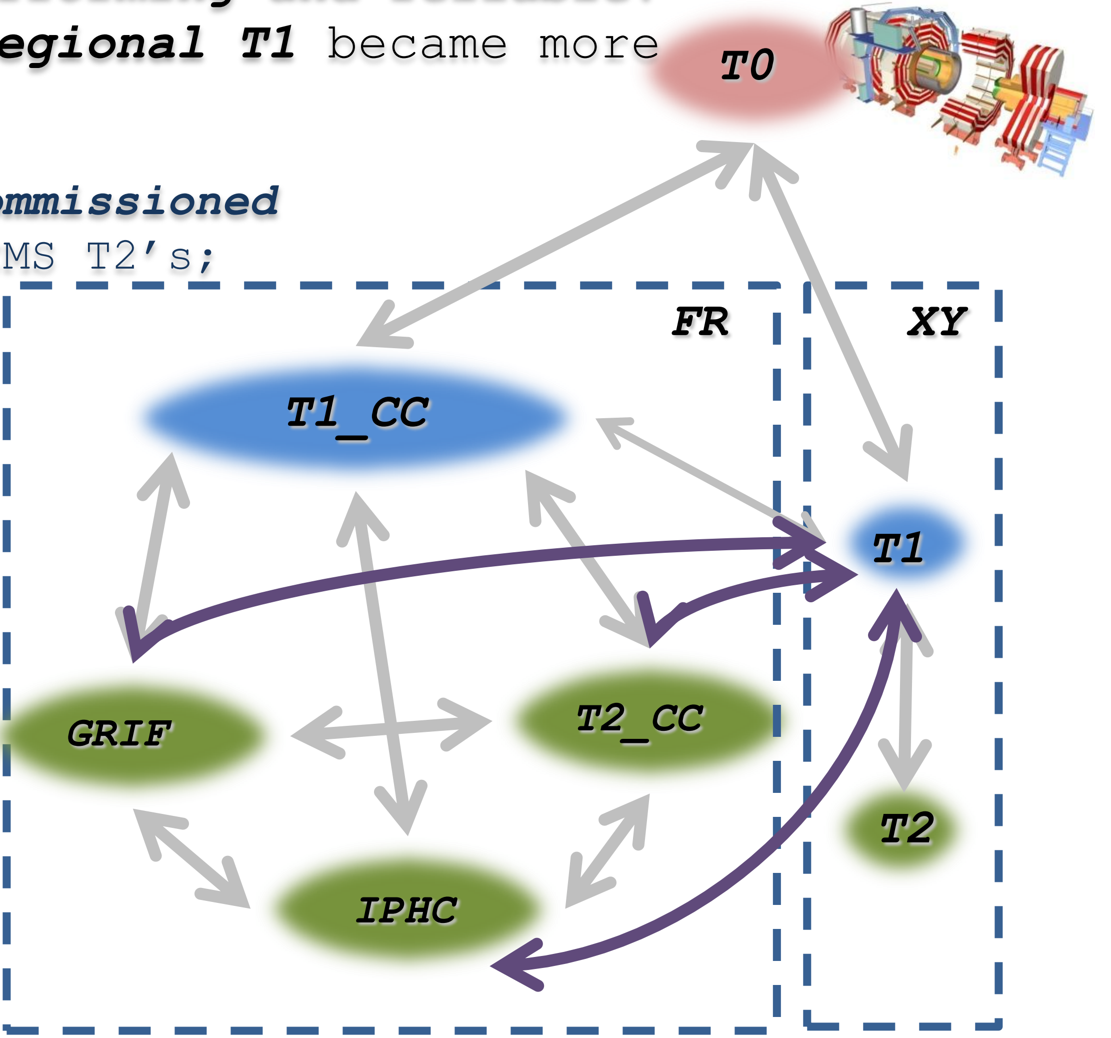
Data mgmt tiered arch

- ***Network*** considered ***among the potentially weak points***: keep local/regional, stay on LHCOPN
  - strict hierarchical architecture: ***T0-T1*** and ***T1-T2*** data flows;
  - good ***T1-T1*** connectivity for RE-RECO synch;
  - good ***T1-T2*** and ***T2-T2*** ***regional*** connectivity;
  - ***jobs access the data locally*** (i.e. job go where data are stored).



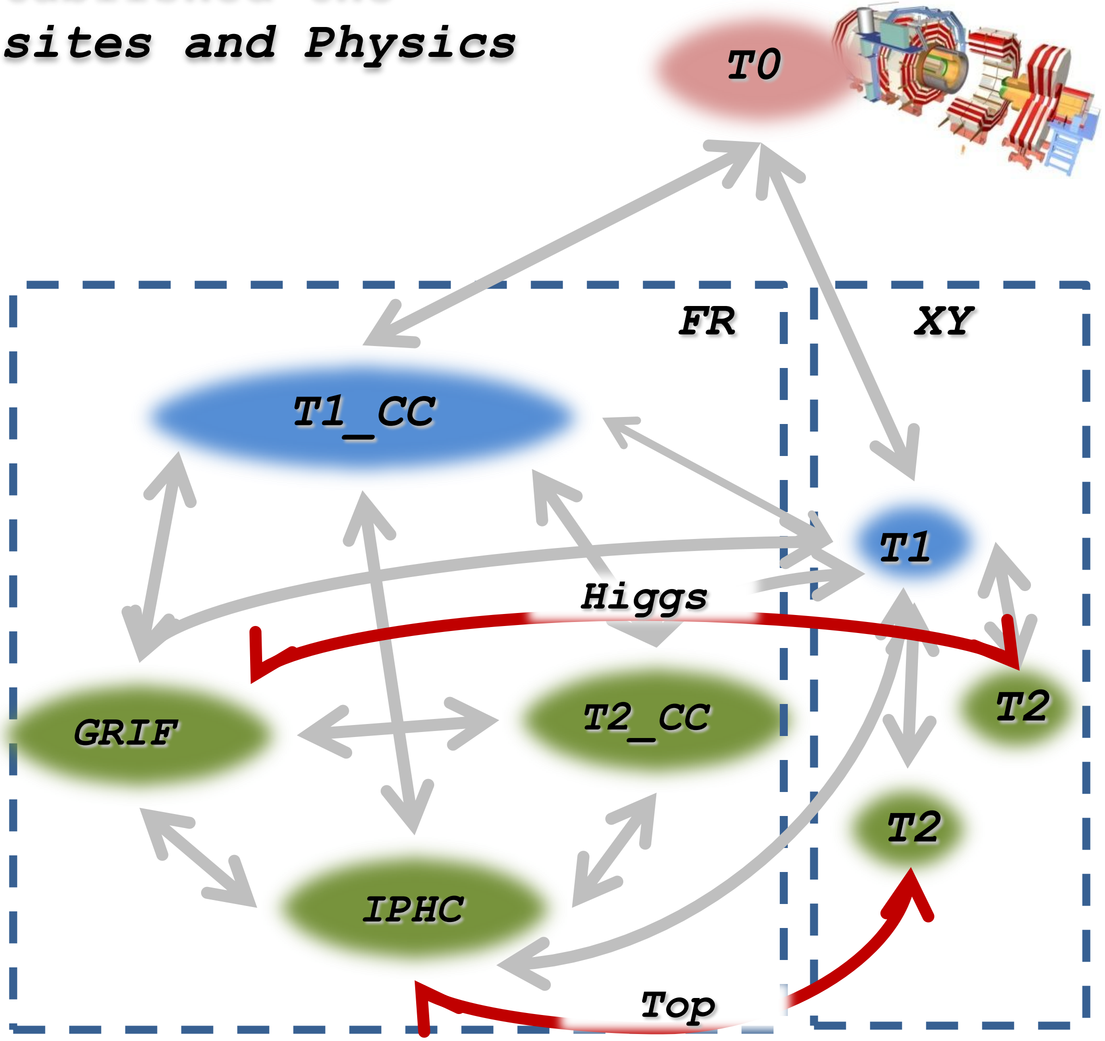
Data mgmt t1's full mesh

- ***Network*** showed to be ***performing and reliable***: ***T2 connections to non-regional T1*** became more and more important
  - having all ***T1-T2 links commissioned*** became a ***requirement*** to CMS T2's;
  - non regional ***T2-T1*** uplinks are more and more used as well;
  - required perfs: 20MB/s downlink, 5MB/s uplink;
  - currently ***most part of T2*** data ***import*** comes from ***non regional*** T1s.



Data mgmt t2's mesh

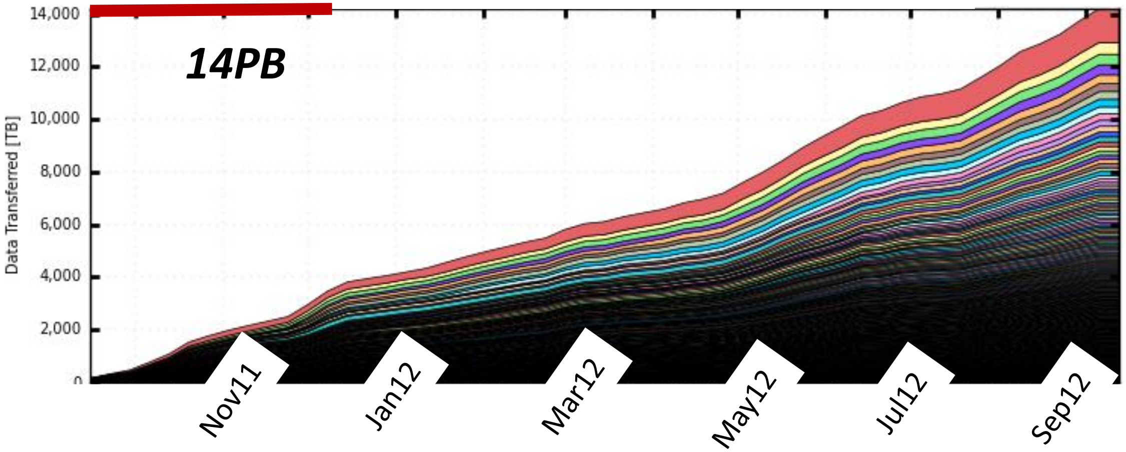
- With data taking CMS established the ***association between T2 sites and Physics Groups***
  - sites ***associated to the same Physics Groups*** started commissioning their ***links*** to better exchange data among themselves;
  - CMS computing turned this into a on ***official commissioning campaign;***
  - currently non-regional ***T2-T2*** links give important contribution.



Some plots

overview

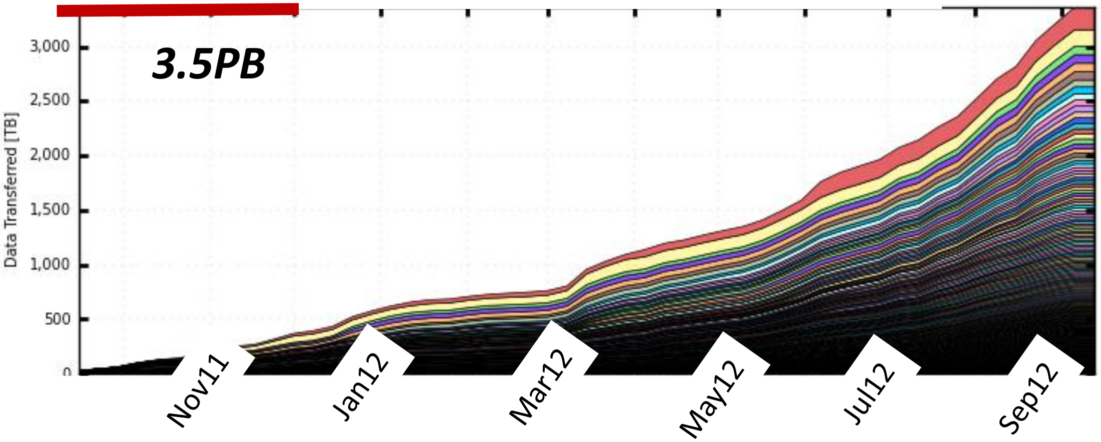
T1 to T2 transfers last year



- ***T1-T2*** transfers: ***14 PB***, in the last 12 months, over ***406 active links***;

- ***T2-T1*** transfers: ***3.5 PB***, in the last 12 months, over ***306 active links***.

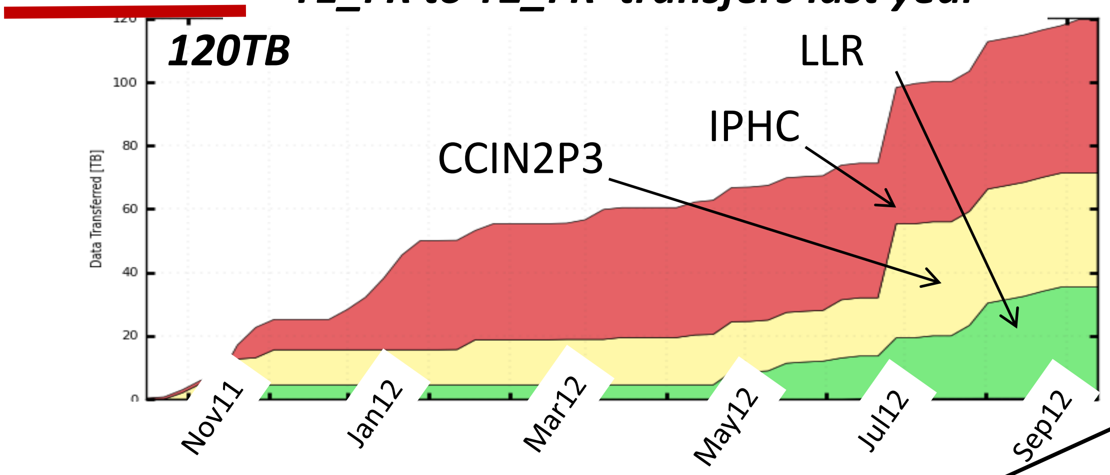
T2 to T1 transfers last year



[*] all PhEDEx plots in the following slides will plot ***effective*** (i.e. successful transfers) ***transferred volume*** in the ***last 12 months***.

Some plots regional

T1_FR to T2_FR transfers last year



120TB

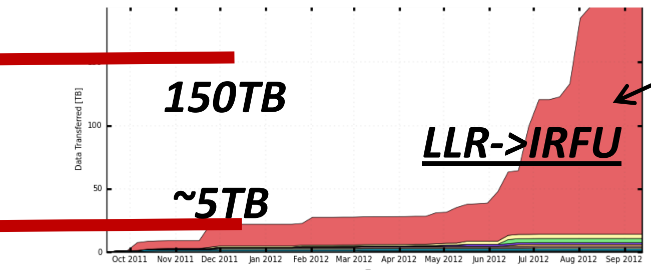
LLR

IPHC

CCIN2P3

- ❖ ***small*** contribution (see next slide);
- ❖ ***T2-T2*** mostly for ***non-regional multi-hop*** at GRIF.

T2_FR to T2_FR

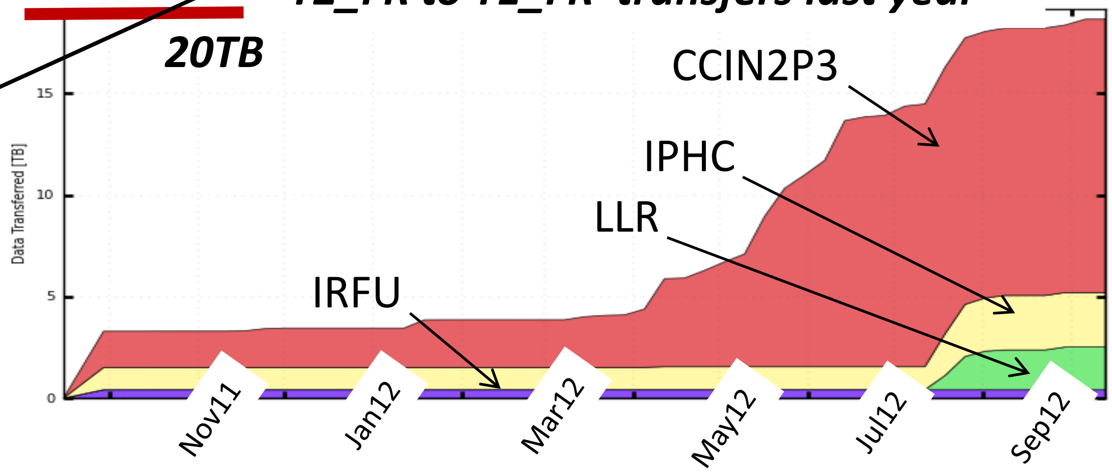


150TB

LLR->IRFU

~5TB

T2_FR to T1_FR transfers last year



20TB

CCIN2P3

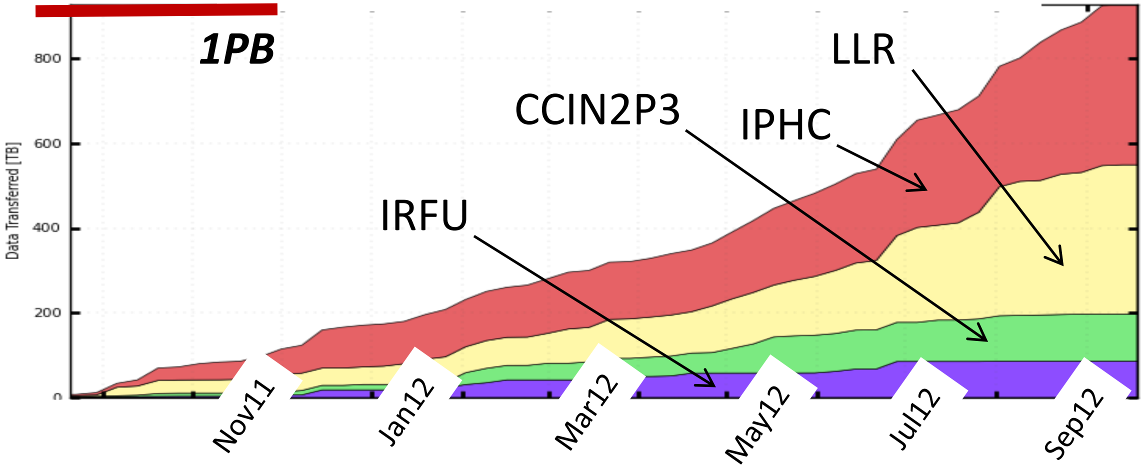
IPHC

LLR

IRFU

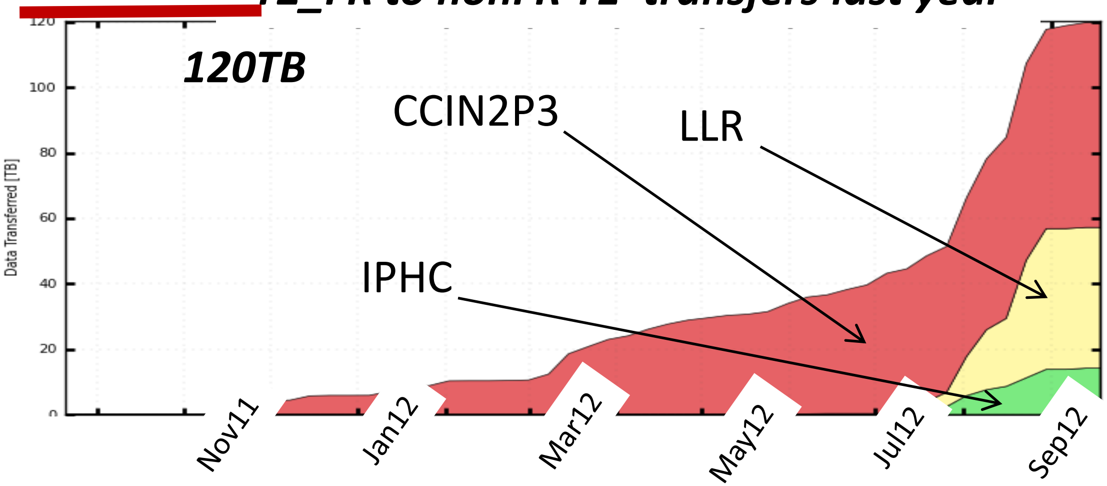
Some plots non-reg. T1's

nonFR T1 to T2_FR transfers last year



- ***89%** of the overall traffic **from T1's to T2_FR is non-reg;***
- ***85%** of the overall traffic **from T2_FR to T1's is non-reg;***

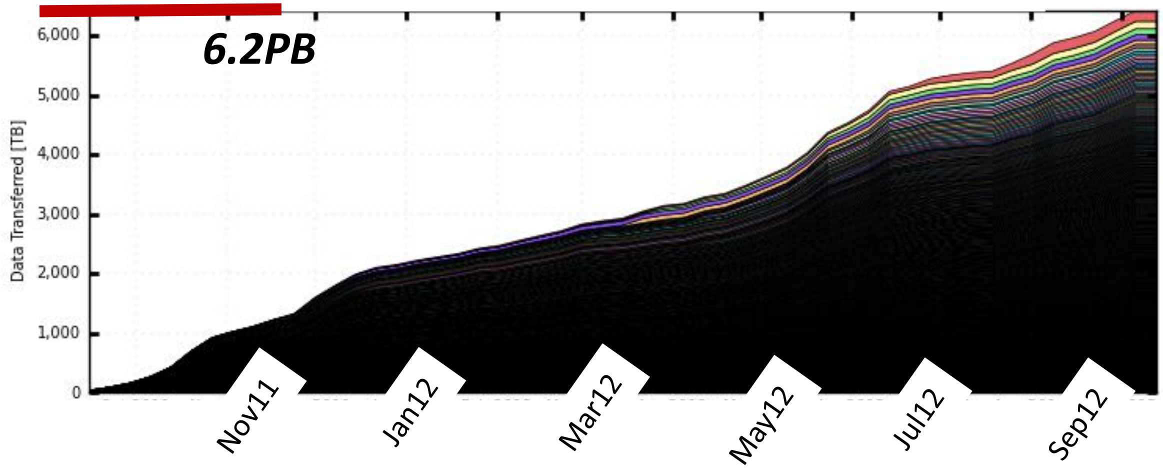
T2_FR to nonFR T1 transfers last year



- ***French T2's contribution** to global data movement is **~5%: in line with the expected ratio** of T2 CMS activity in France.*

Some plots non-reg. T2's

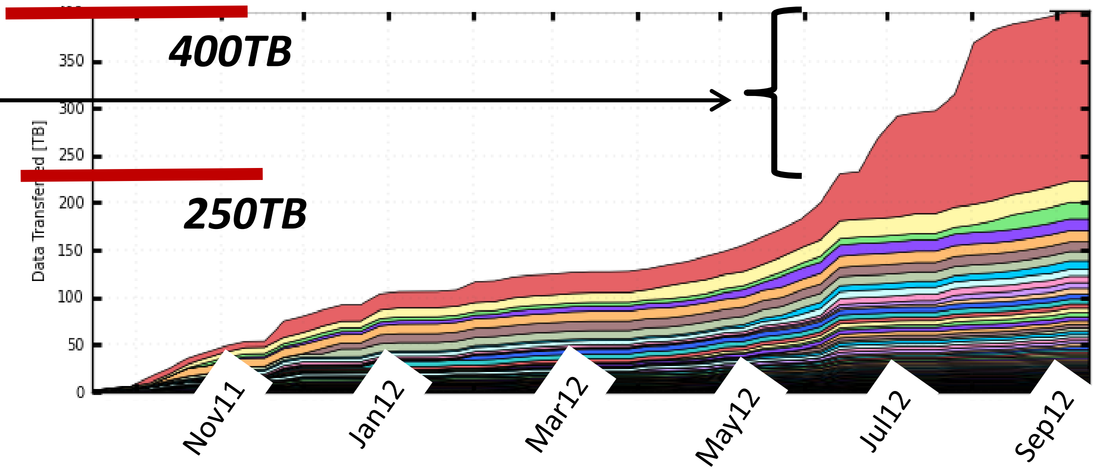
All T2 to all T2 transfers last year



- ❖ ***T2-T2*** transfers: ***~30% of transfers to T2*** sites;
- ❖ ***6.2PB*** in the last 12 months over ***1450 active links;***

All T2 to T2_FR transfers last year

- ❖ ***400TB, dominated by LLR-IRFU*** performing ***multi-hop*** transfers, ***actual volume*** is ***250TB*** (***20%*** of FR T2 ***imports***).



LHCOne overview

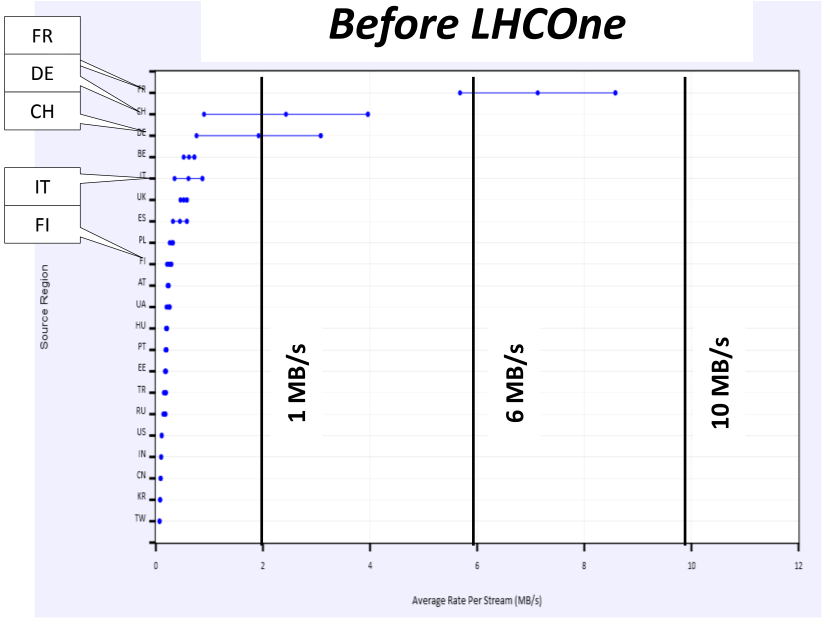
* **http://lhcone.net**

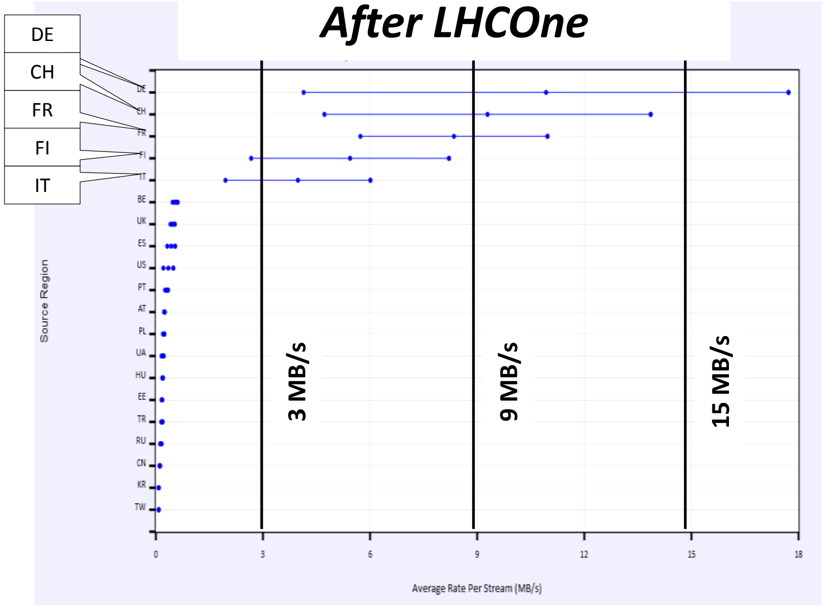
***L**HC* ***O**pen* ***N**etwork* ***E**nvironment*

"The objective of LHCONE is to provide a collection of access locations that are effectively ***entry points into a network that is private to the LHC T1/2/3 sites***. LHCONE is ***not intended to replace the LHCOPN but*** rather ***to complement it***." [*]

- Currently ***shared VLAN*** prototype;
- ***CMS*** has been ***much interested*** in the project since the beginning as a ***consistent part of CMS data placement*** is routed on Tn-Tm links (n,m>1);
- among CMS France sites (to my knowledge): ***GRIF, IPNL and CC*** are currently connected to LHCOne;
- to CMS: more than to improve overall performances it is important to fix critical points.

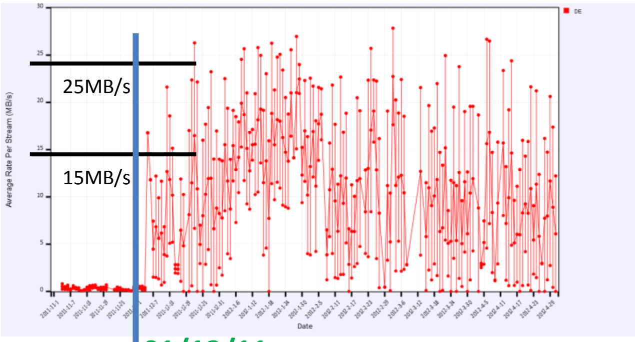
LHCOne example


Before LHCOne


After LHCOne

Average MB/s*stream in IRFU imports from different regions

Daily average MB/s*stream in IRFU imports from DE sites



01/12/11

- ❖ Import from ***some regions*** (DE,CH,FI,IT) ***significantly improved***

[*]quantity in plots: rate/stream (to get effective PhEDEx rate: multiply by nstream and by the number of parallel transfers)

Data mgmt the future

- ***CMS*** Data Management keeps ***evolving*** toward a more dynamical and distributed model
  - NW infrastructure: reliable + important improvements;
  - seek for ***more flexibility*** and ***less demanding operations***;

- ***Data Popularity*** and ***Site Cleaning*** services already in place;

https://cms-popularity.cern.ch/

- next step ***Dynamic Data Placement***
  - reduce pre-placed replicas and optimize storage usage;

- deploying ***Xroot federation*** for direct access over WAN
  - started at USCMS and now extending to all sites;
  - use cases: fallback of local access, re-brokerage of jobs, file caching & re-transfer of broken files.

[*]https://indico.cern.ch/getFile.py/access?subContId=4&contribId=30&resId=0&materialId=slides&confId=196073

Summing up…

- Over years CMS has developed its own ***Data Placement model***
  - relies on a ***reliable and performing NW infrastructure*** and on ***robust and flexible Data Management tools;***
  - ***Physics Groups*** can easily ***transfer and replicate their data at*** all supporting ***sites;***
  - still based on ***static data placing/deleting and local access;***
- ***LHCOne project perfectly suits the needs of CMS*** in terms of NW infrastructure;
- ***evolution*** toward a more ***flexible and dynamic model*** is foreseen
  - ***automatic cleaning and popularity gathering*** services are available;
  - ***dynamic data placement*** and ***direct WAN access*** via Xroot federation are in the plans.