

Minute CDG 21/10/2022

Introduzione

Serie di riunioni di referaggio, attendiamo documentazione ufficiale dai referee, ma le tabelle con le risorse per il T1 sono abbastanza stabili.

Richieste di overpledge dagli LHC per 27 PB. No comunicazioni ufficiali in merito.

Richieste risorse dedicate su cloud – accesso interattivo. Non c'è un finanziamento dedicato al momento, comunque sono state finanziate con lo stesso prezzo delle risorse grid HTC. Il CNAF sta lavorando nell'ottica di unificazione risorse cloud e HTC.

Richieste rinnovi manutenzione accordati.

Stato delle gare:

- CPU in arrivo entro fine anno.
- Disco: aggiudicata gara da 14 PB, queste risorse verranno installate sul CNAF attuale. Consegna sperabilmente entro fine anno/inizio 2023 per consegna pledge ai no-LHC a gennaio.
- Tape siamo a pledge con i 130 PB grazie all'acquisto degli ultimi 14 PB.
- Rete: pubblicate 2 gare.

Rete: necessario down apparati al CINECA per configurazione HA. Nei dettagli farming.

LA: Per LHCb, I referee ci hanno fatto sapere che andranno avanti con la giunta per la proposta di overpledge, ma sarà più bassa per i ritardi di LHCB e perché gli acceleratori andranno il 20% in meno per il caro energia, quindi si pensa a un -9PB dell'overpledge

DS: siamo in attesa che ci dicano qualcosa, però lato CMS stessa storia, in attesa, ma ci si aspetta qualche modifica rispetto alle richieste lato tape.

Farm

Utilizzo risorse (vedi slide)

Nuova gara (vedi slide)

Risorse CINECA: sarà necessario down completo di tutta la parte CINECA per mettere in ridondanza la parte di rete.

Aggiornamenti effettuati per vulnerabilità varie. Terminata anche virtualizzazione di vecchie UI.

Aggiornamento da fare: vulnerabilità appena uscita. Necessario aggiornamento del kernel -> implica reboot di tutta la farm. Verrà fatto per passi alla volta. Data prevista: 8 novembre.

Apptainer: alcuni esperimenti reperiscono Apptainer tramite CVMFS. Siccome è spesso soggetta a vulnerabilità, se non è necessario avere il pacchetto installato allora sarebbe buono rimuovere il pacchetto e togliere una dipendenza.

Richiesta: fare indagine con US per capire se possibile per farming rimuovere Apptainer.

Configurazioni future: inasprimento delle misure di controllo sui job. Raccomandazione: non aprire servizi su porte alte.

SdP: Dopo l'HTCondor WorkShop di settimana scorsa, gli sviluppatori rimedieranno un problema che mandava in hold i job di ATLAS per una metrica che risulta troppo alta, ma non c'è enforcement di regole su questa modalità. Fino a pochi giorni fa, il sistema eliminava i job che saturavano lo swap della macchina, con una soglia a 35 MB (?) e refresh ogni 15 minuti dell'informazione. Questa soglia è obsoleta e inoltre non indica esattamente lo swap, ma la somma di altre metriche.

Questo succedeva anche per Alice, ma il refresh a 15 minuti era troppo lento.

Adesso il sistema è cambiato: monitorati i singoli job con un intervento più tempestivo.

Attivo da ieri sera alle 18. Questo modello si può perfezionare inserendo altre metriche.

LA: aggiornamento kernel provoca downtime su parte CINECA, ma anche su tutto?

AC: le macchine andranno riavviate tutte. Quelle del CNAF prima di tutte le altre.

Le vulnerabilità infatti sono su tutti i sistemi Linux.

LA: downtime?

AC: sì, ma soft. Tranne che per la parte CINECA, ma prima o poi ci voleva un downtime per intervento su apparato di rete.

DC: magari mettiamo un down at risk. Con intervento Cineca si perde metà della potenza, ma non sarà un vero e proprio down.

Storage

Disk storage in produzione (vedi tabella).

Prossimi acquisti: gara storage.

Stato Ticket: vedi slide di Lucia.

Problema thread saturati xrootd – CMS (vedi slide).

Stato tape: vedi slide di Enrico.

Repack: slide di Enrico. È previsto che resti spazio inutilizzato a fine anno per qualche esperimento?

Pensiamo di riprendere questa attività a marzo dell'anno prox quando avremo a disposizione nuovo nastro.

DS: tape – CMS sta lavorando a una seconda campagna. Ancora da capire se quello che togliamo non reca problema. Dopo forse riusciamo a capire impatto sul CNAF. Da 0 a qualche centinaio di TB.

Appena so qualcosa vi aggiornerò. Se avete bisogno di qualche info scrivetemi pure.

Sui ticket di Lucia:

1. Cancellazione manuale dei file – viene richiesta e va avanti. Non deve ricadere necessariamente sul supporto del CNAF, in realtà può farlo chiunque abbia il ruolo production in CMS.

Tra poco avrò una riunione interna per discutere di questo questo.

Baco lato sito? Per me invece FTS+Rucio, comunque problema lato esperimento, indago durante la riunione.

2. Rules approval. Suggesto Alessandro Pascolini per questo. Se noi approviamo le regole, i ticket non appaiono. Se con Ale ci aiutiamo non ci saranno più ticket aperti su queste cose.

VS: in realtà per noi sarebbe utile ricevere un avvertimento dall'esperimento nel caso in cui stiano procedendo con una campagna di movimenti verso o da te tape – potete comunicarlo anche a noi.

DS: Alessandro comunica con il CNAF. Per le recall penso vengano fatte da WLM, potrebbero sfuggirci, però sono abbastanza programmata. Possiamo aggiornarvi.

DC: sarebbe utile sapere una stima del traffico attesa verso Google cloud o reti commerciali in generale.

LR: Non ricordo di questo traffico, ma posso indagare.

DC: sì, questo sulla rete potrebbe avere implicazioni non banali. Sarebbe buono avere delle stime di traffico su rete commerciali. Giusto per non avere sorprese con il GARR.

LR: traffico dal CNAF ai server google, da disco?

LM: sì.

LR: ok, chiedo.

LM: c'è un workaround.

SZ: sarebbe importante una stima in generale del traffico di tutti i tier italiani. Ci sono due peering per cui paghiamo a volume.

LA: non credo che LHCb riuscirà a scrivere tanto su tape entro fine anno.

Quindi buona parte non sarà utilizzata ed eventualmente utile per il repacking.

Cancellazioni già richieste dai referee, ma tempi lunghi: entro 12 mesi.

EF: grazie, sapere che eliminate qualcosa entro 12 mesi è già qualcosa.

Rete

SZ: Inaugurato il POP del garr T. Il CNAF è il primo a entrare e i primi link sono stati LHCOPN e ONE.

Nessun problema per la rete.

Down praticamente nullo su LHCONE/OPN e pochi secondi per il down su general internet.

Con un nuovo backbone, abbiamo fatto un paio di modifiche sul link di general internet che sono in LHCONE, da 20 a 40 Gps. Raggiunti picchi sporadici fino a 30 Gbps.

Sfruttata una finestra di lavoro GARR per raddoppiare la connettività da Milano a CERN per il nostro centro.

Connettività verso il CERN prima a 100Gbps, mentre passeremo a 2x100Gbps dal 18/11.

LHCb

No data taking (vedi slide).

La collaborazione è molto concentrata su commissioning.

Utilizzo delle risorse HPC: abbiamo iniziato a usare Marconi100 anche se non siamo ancora riusciti a integrarlo con DIRAC. In corso campagne di training su reti neurali.

DC: su m100 solo training, non c'è in ballo grosso trasferimento dati? Lucio: no. Dal punto di vista della rete tutto tranquillo.

Sdp: queste macchine con GPU100 hanno processore Power9, giusto? (sì)

Il compilatore Matlab non compila su Power9, quindi se uno ha software Matlab su Power9 non funziona.

Esiste una soluzione ibrida, ma se ti capita di sentire qualcuno che ha compilato Matlab avrei piacere di avere informazioni.

LA: so di alcuni lavori con Python. Lavorando a fondo si possono risolvere problemi di questo tipo.

DS: si può rivendere tutta la vostra esperienza. C'è un interesse generale.

LA: sono d'accordo.

ATLAS

Nessun aggiornamento paritocare.

Negli ultimi mesi situazione stazionario. Siamo a pledge.

L'occupazione del tape - siamo quasi a pledge. Possiamo aspettare un po' per il repack dati ATLAS.

Tra un mesetto si ferma presa dati, poi reprocessing. Probabilmente tra le due fasi è un buon momento.

Attività HPC su cineca un po' ferma. Ripresa a breve dell'attività.

CMS

Non molto da riportare. Utilizzo risorse (vedi slide): Non ci sono problemi lato CNAF, qualche problema lato CMS. Efficienza buona, a parte una perdita di efficienza dovuta a perdita di pressione per fallimenti dei SAM test che hanno mandato il sito in drain. I job fallivano e non arrivando hanno abbassato l'efficienza. Tutto comunque sotto controllo.

Test MTU: in sinergia con Carmelo, non ho niente da riportare.

LA: a cosa è dovuto il miglioramento delle performance?

DS: Legacy usa dei file di tipo mini AOD (o addirittura nano). Questi nano AOD vengono letti dai nano AOD tools che costituiscono il framework legacy. Mentre Root Data Frame (che è il nuovo tool utilizzato da DS) non utilizza i nano AOD tool, legge direttamente le root-tuple.

NO – LHC

DC:

Qubic: nel 2023 hanno cpu 300HS06, ricordiamocelo per quando dovremo discutere con loro.

Cuore: ci sono liberi 40 TB e trovate anche directory di back up. Già contattato l'esperimento.

DarkSide: ha già raggiunto il pledge, ma hanno un delta a gennaio che speriamo di assegnare.

Xenon: già scalato a Selvi. Non ci sono novità significative. Proviamo a scalare a commissione II.

Interagisco io, proviamo a sbloccare tutto.

Gamma: macchina fisica dedicata non c'è più. Cloudificata.

Eupraxia: verificare pledge 2023.

Herd

NM: FS ottimizzato per file piccoli. Molto poco utilizzato, ma prevediamo di utilizzarlo.

DC: se sono log, per quelli vecchi è consigliato compattare e zippare.

JUNO

GA: Niente da aggiungere. Grazie.

Virgo

LR: niente da aggiungere. Grazie.

SdP: dall'ultimo incontro PRIN con Bagnasco era sorta la richiesta di considerare di installare un submit node di Virgo. Ho avuto accesso alla documentazione per installarlo. Devo parlare con i contatti che ho preso per vedere il setup richiesto per capire se può stare al CNAF. È comunque simile a una ui.

La richiesta non è ufficiale, però SB la richiederà.

DC: hai fatto bene a tracciare la cosa anche qui. Non ci dovrebbero essere particolari problemi. Un po' modello VO box degli LHC.

LR: sono d'accordo. Da capire l'autenticazione, comunque grazie.

AOB

CG: vi invito a iscrivervi alla lista CNC forum per rimanere in contatto con tutta la comunità del calcolo INFN.

Prossimo CdG 18/11