

Visio/Télé-conférence SA1-FR d'EGEE



Date : Mardi 24 juin 2008, 14h

Présidé par: Pierre Girard

Secrétaire: Pierre-Emmanuel Brinette

Fonctionnement : <http://vacs.in2p3.fr/rms/guide.pdf>

Prochaine visio. : le mardi 26 août 2008

Ordre du jour

Visio/Télé-conférence SA1-FR d'EGEE

Ordre du jour

Présences

Réunion

1) Approbation du CR de la dernière visioconférence (tous, 5')

<http://indico.in2p3.fr/conferenceDisplay.py?confId=916>

2) Points ROC / Sites / VOs (tous, 15')

a) Ticket GGUS

https://gus.fzk.de/ws/ticket_search.php?supportunit=ROC_France&status=open&radiotf=1&timeframe=no
Pensez à fermer les tickets quand ils sont résolus !

b) Point sur l'incident sécurité

3) Session spéciale "Stockage" (tous, 90')

a) Enquête sur les configurations du stockage sur les sites

b) Présentations

- Introduction à la solution Storm + Lustre (Pierre-François Honoré)
- Comparatif "HEPIX" de solutions de stockage (Loïc Tortay)

c) Discussions

- Doit-on traiter la redondance ?(LPC)
- Quelle est, ou y a-t-il une technologie de stockage plus adaptée à la grille ? (IRFU)
- Différence de prix SAN/NAS
- Niveau de disponibilité et d'intégrité des données
- Y a t'il des solutions de stockage plutôt adaptées à des type d'application ?
- Quelqu'un connaît-il les estimations des taux de transfert pour ALICE ? (SUBATECH)
- Le RAID est il indispensable pour le stockage d'un TIER 2 ?(LPSC)
- Y a-t-il des informations publiés concernant la fiabilité des fichiers ?(CGGVeritas)
- Remarques sur la commande « DPM Drain » (SUBATECH)
- Demande d'ajour d' un champ sur la taille des fichiers manipulés dans le questionnaire (CGGVeritas)

4) Divers (5')

Présences

VSites	Membres	Total
CC.IN2P3.FR	Loic Tortay	1
	Pierre Girard	1
	Pierre-Emmanuel Brinette	1
Total CC.IN2P3.FR		3
CGGVERITAS.COM	Jean-Bernard Favreau	1
Total CGGVERITAS.COM		1
CPPM.IN2P3.FR	Edith Knoops	1
Total CPPM.IN2P3.FR		1
IBCP.FR	Christelle Eloto	1
Total IBCP.FR		1
IPNL.IN2P3.FR	Denis Pugnere	1
	Guillaume Baulieu	1
	Yoan Giraud	1
Total IPNL.IN2P3.FR		3
IPSL / IPGP	David Weissenbach	1
Total IPSL / IPGP		1
IRFU/CEA.FR	Christine Leroy	1
	Pierre-François Honoré	1
	Pierrick Micout	1
Total IRFU/CEA.FR		3
LAPP.IN2P3.FR	Eric Fede	1
	Muriel Gougerot	1
Total LAPP.IN2P3.FR		2
LPNHE.IN2P3.FR	Liliana Martin	1
Total LPNHE.IN2P3.FR		1
SUBATECH.IN2P3.FR	Jean-Michel Barbet	1
	Jérôme Pinot	1
Total SUBATECH.IN2P3.FR		2
UREC	Alice de Bignicourt	1
Total UREC		1
Total		19

Réunion

1) Approbation du CR de la dernière visioconférence (tous, 5')

<http://indico.in2p3.fr/conferenceDisplay.py?confId=916>

Compte rendu de la dernière Visio non encore publié. Le document sera disponible sous peu.

2) Points ROC / Sites / VOs (tous, 15')

a) Ticket GGUS

https://gus.fzk.de/ws/ticket_search.php?supportunit=ROC_France&status=open&radiotf=1&timeframe=no

Pensez à fermer les tickets quand ils sont résolus !

b) Point sur l'incident sécurité

- Le CC a été contacté par le fonctionnaire Sécurité Défense sur CNRS qui a informé de la tenu d'une réunion le vendredi 27 au Secrétariat Général de la défense concernant la sécurité sur la grille.
- Guy Wormser ou Dominique Boutigny seront présents.
- Le CEA a contacté de son coté la Direction Centrale de la Sécurité.

3) Session spéciale "Stockage" (tous, 90')

a) Enquête sur les configurations du stockage sur les sites

Le questionnaire est accessible à l'adresse :

<http://indico.in2p3.fr/confDisplayEvaluation.py/display?confId=947>

Les résultats de ce questionnaire sont accessibles à l'adresse :

<http://grid.in2p3.fr/html/Public/SA1FR/Survey/results.htm>

<http://grid.in2p3.fr/html/Public/SA1FR/Survey/sites.html>

(NB : La synthèse des réponses au questionnaire sera refaite pour offrir une vue plus synthétique)

Les différents documents relatifs aux infrastructures de stockage sont disponibles sur l'agenda :

<http://indico.in2p3.fr/conferenceDisplay.py?confId=947>

b) Présentations

- *Introduction à la solution Storm + Lustre (Pierre-François Honoré)*

<http://indico.in2p3.fr/materialDisplay.py?contribId=5&sessionId=2&materialId=slides&confId=947>

- *Comparatif "HEPIX" de solutions de stockage (Loïc Tortay)*
<http://indico.in2p3.fr/materialDisplay.py?contribId=6&sessionId=2&materialId=slides&confId=947>

c) Discussions

- *Doit-on traiter la redondance ?(LPC)*
 - ➔ *Problème de redondance/disponibilité, faut-il que les SE soient disponibles à tout prix ?*
Cas LCG :
 - *Le MOU des T2 est de 95 % pour la disponibilité des données, par contre la sécurité des données est plus faible. Théoriquement les données reçus sur les Tier 2 peuvent être perdus et redemandés*
 - *T3 : Les données générées par les sites Tier 3 doivent être sécurisées*
 - ➔ *Le site GRIF a-t-il travaillé à mettre en place une certaine redondance entre ses sous-sites ?*
A priori, non.
 - ➔ *L'Aspect redondance n'est-il pas directement géré par les VOs ?*
 - *CMS aurait commencé à étudier la possibilité de répliquer les données entre les sites Tier 2.*
 - ➔ *Le middleware dispose de fonctionnalités de réplication très limitées.*
 - *Cas d ESR : Chaque application se charge de répartir ses données sur des SE et un ensemble de script a été développé pour récupérer les informations de façon intelligente. Il n'existe pas de notion de topologie géographique dans le MW.*
 - *Les VO se retrouvent à gérer les faiblesses du middleware*
- *Quelle est, ou y a-t-il une technologie de stockage plus adaptée à la grille ? (IRFU)*
 - ➔ *Faut-il mieux un boîtier en Direct Attachement (NAS) plutôt qu'une infrastructure haute disponibilité unifiée (SAN) ?*
 - ➔ *LAPP : La solution SAN + GPFS est très efficace et satisfaisante.*
 - + Performances.
 - + Sécurité des données.
 - + Facilité d'administration.
 - - Prix
 - ➔ *Subatech : Solution aussi basée sur une baie SAN et GPFS.(Cf. présentation attachée à l'agenda)*
- *Différence de prix SAN/NAS*
 - ➔ *Pour des achats prévisionnels de 10To/an (avec contrat de maintenance), Il faut prendre en compte le problème d'extension du contrat de maintenance pour l'évolution de la baie de stockage*

- *Boitier NAS : l'intérêt est de faire 1 seul achat matériel (+ maintenance) en fonction du besoin du moment. Et lors de besoins supplémentaires, on ne se préoccupe pas de l'existant, on rachète une solution clef-en-main, en fonction des offres du marché à ce moment-là.*
- *Niveau de disponibilité et d'intégrité des données*
 - *NAS : Les fichiers sont stockés avec comme référence l'adresse du serveur NAS.*
 - *DPM impose l'enregistrement du serveur ce qui plutôt incompatible avec des fichiers distribués/recopiés sur plusieurs boîtes*
 - *Xrootd+SRB permet d'améliorer la redondance par rapport à DPM*
 - *GPFS et lustre sont des solutions efficaces qui imposent une infrastructure performante et fiable pour en exploiter pleinement les possibilités. Avec Lustre, le NAS peut être utilisé, mais le SAN est recommandé.*
- *Y a t'il des solutions de stockage plutôt adaptées à des type d'application ?*
 - *ESR : On essaye d'avoir des solutions de stockage par application. Supportant les réplicas. Pas de modèle de propagation de données, la réplication est faite artisanalement.*
 - *CGG/EGEODE : Pour la production NON grille :*
 - *Fichiers de petits volumes important pour les utilisateurs → baie de stockage Raid 5*
 - *Traitement de gros volumes de données sur des boîtes NAS redondés pour la disponibilité et disposant d'une agrégation réseau. La redondance des serveurs de disques est assurée Heartbeat*
 - *Besoin de calcul pour la production : 40 To/semaine*
 - *Certaines données on une faible durée de vie , uniquement une campagne de calcul (de 1 à 3 jours)*
 - *Coté grille, une petite baie de stockage RAID 5*
 - *Subatech : s'interroge sur la scalabilité de DPM dans le cas des jobs d'analyse.*
 - *A noter, qu'hormis le CC (en partie à cause d'HPSS derrière Dcache), aucun site n'a pour l'instant déployé de configuration spéciale pour une VO.*
 - *LPC : Problème biomed a Clermont pour l'accès simultané à un grand nombre de fichiers sur DPM.*
 - *Les utilisateurs on été contactés directement pour leur expliquer le problème induits par leur mode d'utilisation.*
 - *Est-ce du ressort du site ou de la VO ?*
 - *A partir du moment où ça impacte la production du site, ce dernier est impliqué.*
 - *LAPP : GPFS est intéressant pour ce type d'usage, car il permet d'avoir des taches de fond qui permettent par exemple de déplacer des fichiers automatiquement sur des critères de taille, d'accès, ... Par contre ce 'fine tuning' peut être délicat a mettre en place. Ce n'est pas fait au LAPP d'ailleurs.*

- Quelqu'un connaît-il les estimations des taux de transfert pour ALICE ? (SUBATECH)
 - ➔ Réponse de Yannick Patois. :
 - Valeurs théoriques : 2,6 Mo/s /cœur
 - Valeurs observées : 600 ko/s max, 100 ko/s moyenne
 - ➔ CPPM : Les performances observées sont meilleures pour 10x20 jobs lancés à la suite plutôt que 200 jobs simultanés
 - ➔ Il pourrait être intéressant que le réseau puisse être déclaré comme une ressource au niveau du système de batch.
 - ➔ Pour les jobs Atlas/CMS : Les tests effectués ont montré qu'il faut un débit de 200/Mb(ytes ?)s par To traité (par WN) !

- Le RAID est-il indispensable pour le stockage d'un TIER 2 ? (LPSC)
 - ➔ Raison : sur une baie DELL, le RAID réduit le stockage utile de 1,4 To
 - ➔ Attention. Il ne faut pas négliger l'effort humain. En effet, l'humain est souvent une ressource critique des sites. Il faut donc vérifier, si la surcharge de travail engendrée par une défaillance de disque n'aurait pas un coût supérieur au gain obtenu en supprimant le RAID.
 - ➔ LPC : Théoriquement les Tier 2 peuvent perdre leurs données, celle si étant dupliqué. Dans la pratique, ce n'est pas toujours vrai (pb rencontré lors de migration vers un SE DPM)

- Y a-t-il des informations publiées concernant la fiabilité des fichiers ? (CGGVeritas)
 - ➔ Par exemple, peut-on savoir si les fichiers sont stockés sur du disque RAID ?
 - ➔ Il y a des informations dans le Glue schema concernant la durée de vie, si l'espace est permanent, durable, etc.
 - ➔ Mais ce n'est pas réellement des informations sur la fiabilité du stockage.
 - ➔ Il y a aussi les « Space tokens » qui permettent de dire si le fichier ou le pool doit être sur 'TAPE' ou 'DISK', ou les deux.

- Remarques sur la commande « DPM Drain » (SUBATECH)
 - ➔ DPM considère que si l'attribut « permanent » n'est pas placé sur un fichier, il n'est pas nécessaire de conserver lors du déplacement de File System. Attention, les fichiers sont purement et simplement supprimés.
 - ➔ Le déploiement de cette configuration (pool disque « volatile ») est dû à l'utilisation d'XROOTD et de RFIO qui nécessite que le pool autorise la création de fichier de ce type.

→ Information à vérifier :

- Lorsque le pool est plein, le Garbage Collector (DPM Drain) supprime les fichiers volatiles ?
 - Les fichiers XROOTD et RFIO sont-ils susceptible d'être supprimé par DPM Drain ?
-
- Demande d'ajour d' un champ sur la taille des fichiers manipulés dans le questionnaire (CGGVeritas)
 - L'infrastructure peut dépendre de ce type d'information.
 - PB de taille de fichier plus facile à gérer pour des sites qui ne supporte qu'une seule VO.

4) Divers (5')

Le questionnaire reste ouvert jusqu'à la prochaine visioconférence du 8 juillet.
Prochaine visio : le mardi 8 juillet 2008.