RINGOT, Patrice

De: RINGOT, Patrice
Envoyé: mardi 17 mars 2015 18:49
À: NIEDERLENDER, Claude; CARON, Etienne; PERRIN, Stanislas; PARENTIN, Jean-Joffrey
Cc: LUC, Martial; PONTICELLI, Sylvain; RINGOT, Patrice; SCHEFFER, Philippe; TURRI, Angel; VILLAUME, Michel
Objet: CR Réunion ISTEX/IPROD du 12/3/15

Voici donc la version finale du compte-rendu prenant en compte toutes les remarques.

Patrice

Présents IPROD : ML, SPo, PR, PS, AT, MV
Présents ISTEX : EC, CN, JJP, SPe

Durée de la réunion : 1h

Initialement la réunion devait porter sur une séance collective de monitoring ISTEX-IPROD.

Dans les faits, ces derniers jours, plusieurs opérations d'indexation ES puis réplication ont été lancées et suivies par les deux services.

Outils de suivi :

- Ligne de commande ou tty (vmstat,, htop)
- VSphere
- JConsole
- UI de la baie Compellent
- Autres outils utilisés ???

Cf : https://git.istex.fr/infra/istex-svp/issues/25
et https://git.istex.fr/infra/istex-svp/issues/28

Opérations également consignées dans le Wiki ISTEX :

http://wiki.istex.fr/elasticsearch/index/full-20150310

Cela nous a permis de procéder à des réajustements techniques :

- Un nœud isolé pour l'indexation (plutôt que de laisser le cluster ES assumer indexation et réplication + recherche en simultané)
- Augmentation de la HEAP size de la JVM d'indexation (1Go->8Go), ajout d'une possibilité de suivi JMX
- Augmentation en conséquence de la CPU et de la mémoire de la VM d'indexation
- Modification d'un paramètre d'ES permettant d'augmenter les débits de réplication

Le bilan, c'est que nous obtenons des temps d'indexation considérés comme satisfaisants (13,6M d'objets documentaires en 16h cumulées des différentes sessions optimisées de manières inégales), et une première mesure de référence relevée sur l'indexation des documents Elsevier (460 documents/s).

A été constaté un surdimensionnement CPU de la VM d'indexation (elle pourrait faire avec 2x ?????? moins).

ISTEX exprime ses interrogations quant aux performances ram et disque de l'infra par le biais de tests effectués à l'aide de la commande dd

- dd if=/dev/zero vers /run/shm (cpu vers ramdisk) – les temps sont moins bons que sur un pc de dev (cf un des items de https://git.istex.fr/infra/istex-svp/issues/20)
- dd if=/dev/zero vers /tmp (cpu vers disque) – 120 Mo/s, si 3 dd en // : 38Mo/s

IPROD répond qu'effectivement nous avons pu reproduire ce phénomène pour la partie ramdisk, mais que par ailleurs l'utilisation d'un logiciel adapté au test de perf mémoire (PMBW) montre d'excellents résultats sur l'infra. La conclusion d'IPROD, c'est que l'utilisation de la commande dd via une vm sur l'infra n'est pas représentative d'un potentiel de performance.

ISTEX va tester ses commandes dd auprès de Julien Marchall de l'UL (ils utilisent également une Compellent qu'ils ont paramétré pour l'utilisation de Zimbra). IPROD est intéressé par la connaissance technique de l'UL et prendra contact pour échanger avec eux.

IPROD dit qu'il faut agir prioritairement sur la mesure de l'efficacité de l'infra pour l'applicatif, vu du côté de l'utilisateur, avant d'examiner ces aspects « unitaires » qui ne sont pas forcément représentatifs des besoins techniques de l'applicatif. D'où la proposition d'utiliser Gatling pour obtenir des premières mesures communes « boite noire ». En commençant au plus près de l'API puis en déplaçant l'outil de mesure afin qu'il soit in fine positionné à un endroit équivalent à celui des utilisateurs (elasticsearch, api, nginx, reverse proxy, réseau universitaire), et que l'on puisse évaluer la part prise par chaque intermédiaire.

ISTEX dit que la communication sur l'API va engendrer une consommation croissante et une pression « production » qui peut interférer avec les tests.

IPROD propose de dédier un environnement figé de production de manière à ce que nous puissions collaborer au tuning en minimisant les couplages avec les utilisateurs réels.

ISTEX (WP) a lancé les batchs d'ingestion des 6 corpus qui mèneront au remplissage des 16 FS de 1,5To (cf : https://git.istex.fr/infra/istex-svp/issues/21). A l'issue de cette opération, une indexation ES complète aura de nouveau lieu.

IPROD propose d'assurer la répartition des 3 nœuds ES (tous utilisés pour servir l'API en l'état actuel) sur chacun des ESXi pour des raisons de disponibilité des index en cas d'incident sur l'ESXi qui les accueille actuellement. Restera à valider cette répartition par une étude de performance, notamment la co-localisation de la li et de l'index sur un même ESXi.

ISTEX a fait des essais concluants avec l'API Snapshot/Restore d'ES et va contacter AT/MV prochainement pour la mise en œuvre de la sauvegarde des données des nœuds ES. A noter qu'en l'état actuel de nos connaissances il n'est pas possible d'utiliser le snapshot d'un nœud X pour faire une restauration sur un nœud Y.

Le dojo Gatling (lundi 16 am) est repoussé à une date ultérieure (Patrice manque de temps pour la préparation – fin s+1 ou s+2).