



Une infrastructure pour la bioinformatique d'unité

Analyse autour de la maintenance, la viabilité, la charge financière et humaine

Sylvain Gaillard, IRHS, équipe bioinformatique

- Contexte
- Enjeux
- Infrastructure
- Charge financière
- Charge humaine
- Viabilité
- La vraie vie
- Freins / Leviers

IRHS :

- UMR ~250 personnes / 13 équipes
- INRA (BAP, SPE, EA), Agrocampus-Ouest, Université d'Angers
- Biologie des produits horticoles et de la production de semences
 - Génomique, métagénomique, épigénomique
 - Transcriptomique
 - Génétique
 - Phénotypage
 - Écophysiologie
 - Phytopathologie
- Membre de la SFR QUASAV

Héberge

- 3 plateaux techniques
 - Imagerie cellulaire (IMAC)
 - Analyse des acides nucléiques (ANAN)
 - Collection de microorganismes (COMIC)
- 1 plateforme de phénotypage
 - Phenotic

Héberge

- 3 plateaux techniques
 - Imagerie cellulaire (IMAC)
 - Analyse des acides nucléiques (ANAN)
 - Collection de microorganismes (COMIC)
- 1 plateforme de phénotypage
 - Phenotic

Génération assez rapide et volumineuse de données (accélération)

- Séquençage ADN/ARN
- Génotypage SNP/SSR
- Transcriptomique microarray/RNAseq
- Phénotypage champ/serre
- Imagerie microscopie/macrocopie/multispectrale/fluorescence

Héberge

- 3 plateaux techniques
 - Imagerie cellulaire (IMAC)
 - Analyse des acides nucléiques (ANAN)
 - Collection de microorganismes (COMIC)
- 1 plateforme de phénotypage
 - Phenotic

Génération assez rapide et volumineuse de données (accélération)

- Séquençage ADN/ARN
- Génotypage SNP/SSR
- Transcriptomique microarray/RNAseq
- Phénotypage champ/serre
- Imagerie microscopie/macroscopie/multispectrale/fluorescence

Traitement local des données

- Proximité acquisition / stockage / traitements
- Stockage / diffusion interne des données brutes / élaborées
 - Équipes
 - Plateaux
- Diffusion publique de données scientifiques élaborées (et brutes)
- Palier aux problèmes soulevés par systèmes EIC / DSI
 - 4 To pour l'unité avec quotas par équipes (samba / NAS centre)
 - 30 Go par espace share-point (accès difficile, trafic réseau)
- Intégration et centralisation données / méta-données / traitements
 - Bases de données
 - Outils développés par équipe bioinfo
 - Pipelines
- Protection / sécurisation des données

Traitement (Dell R820) / Stockage (Dell MD3260) : 6 U

- 4 processeurs Xeon® E5-4620 8 cœurs
- 1,5 To RAM
- 160 To disques

Sauvegarde (Dell R420 + MD3260) : 5 U

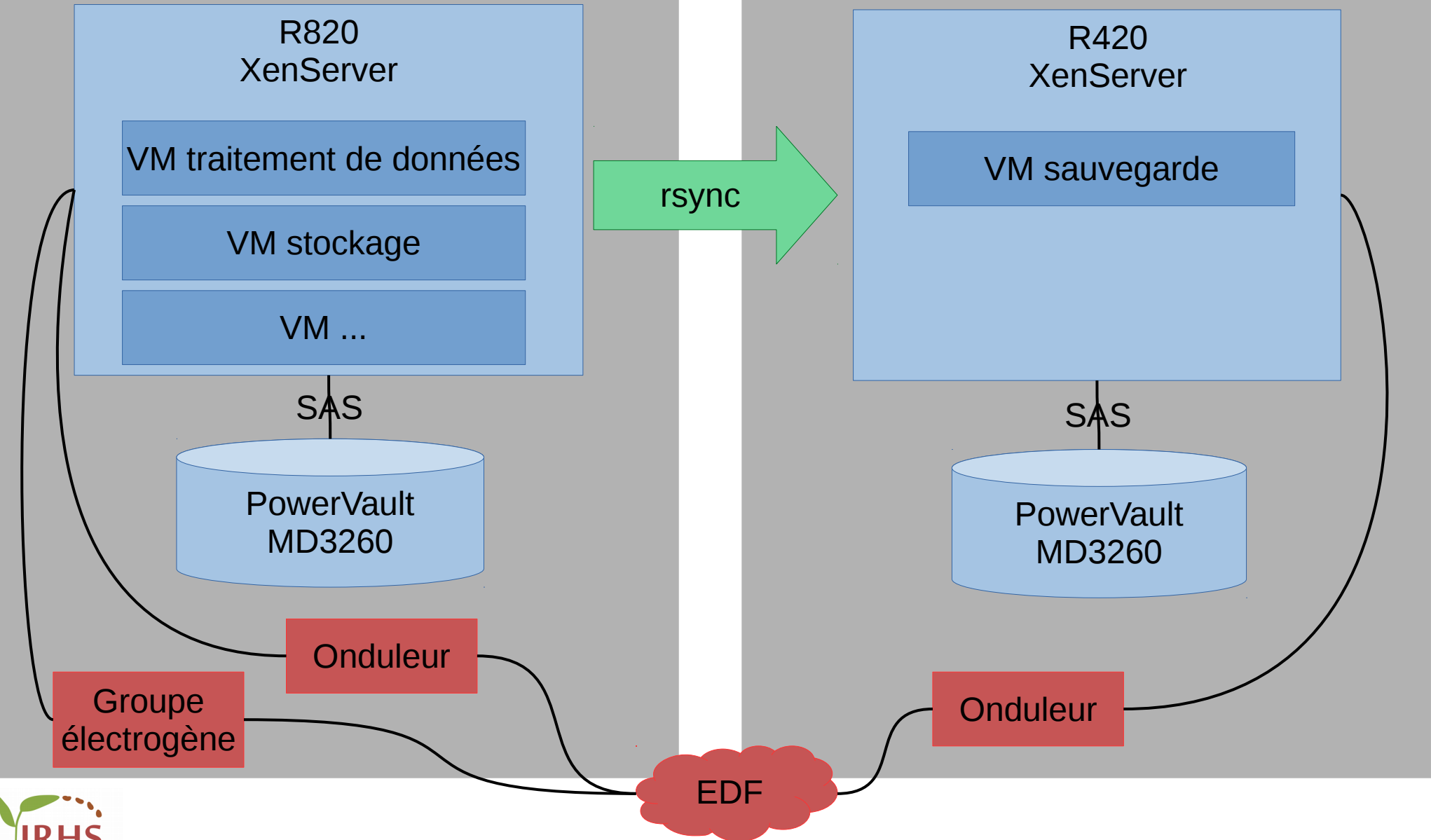
- 80 To disques

XenServer 7.1

Infrastructure

Salle serveur (bâtiment 1)

Salle serveur (bâtiment 2)



Infrastructure

Acquisition : novembre 2013

Mise en service : juillet 2014

Infrastructure

Acquisition : novembre 2013

Mise en service : juillet 2014

Supervision physique

- Machines à proximité
- Visite quotidienne

Acquisition : novembre 2013

Mise en service : juillet 2014

Supervision physique

- Machines à proximité
- Visite quotidienne

Maintenance matérielle

- Mise à jour firmware
- Intervention sur panne (2 disques durs)
- Changement de batteries onduleur

Acquisition : novembre 2013

Mise en service : juillet 2014

Supervision physique

- Machines à proximité
- Visite quotidienne

Maintenance matérielle

- Mise à jour firmware
- Intervention sur panne (2 disques durs)
- Changement de batteries onduleur

Maintenance logicielle

- Mise à jour hyperviseur et systèmes hébergés
- Gestion des comptes utilisateurs (100ne)
- Gardien/modérateur de l'espace

Charge financière

Acquisition :

- R820 : 40 625,00 €
- R420 : 1 504,00 €
- MD3260 : 15 990,00 €
- investissement initial : 74 109 €
- Tranche de 20 disques de 4 To : 3 920,00 €

Charge financière

Acquisition :

- R820 : 40 625,00 €
- R420 : 1 504,00 €
- MD3260 : 15 990,00 €
- investissement initial : 74 109 €
- Tranche de 20 disques de 4 To : 3 920,00 €

Maintenance :

0 € (toujours sous garantie)

Charge financière

Acquisition :

- R820 : 40 625,00 €
- R420 : 1 504,00 €
- MD3260 : 15 990,00 €
- investissement initial : 74 109 €
- Tranche de 20 disques de 4 To : 3 920,00 €

Maintenance :

0 € (toujours sous garantie)

Consommation électrique réelle : base 0,15 € / kWh

R820 + MD3260 (40 disques) : 1300 W

$1,3 * 24 * 365 * 0,15 = 1708,20 \text{ € / an}$

Coût inclus dans charges de structure (fluides bâtiments)

Charge financière

Acquisition :

- R820 : 40 625,00 €
- R420 : 1 504,00 €
- MD3260 : 15 990,00 €
- investissement initial : 74 109 €
- Tranche de 20 disques de 4 To : 3 920,00 €

Maintenance :

0 € (toujours sous garantie)

Consommation électrique réelle : base 0,15 € / kWh

R820 + MD3260 (40 disques) : 1300 W

$1,3 * 24 * 365 * 0,15 = 1708,20$ € / an

Coût inclus dans charges de structure (fluides bâtiments)

~~Cas théorique d'un déplacement en Data Center~~

~~Consommation nominale :~~

~~$2 * 1100 + 2 * 1755 = 5710$ W~~

~~Coût sur data center : 5710 € / an~~

~~Coût x 3,34 !~~

Correction après discussion

Consommation nominale (redondante) :

$1100 + 1755 = 2855$ W

Coût sur data center : 2855 € / an

Ce coût comprend la maintenance clim, onduleur, groupe électrogène qui n'est pas comptabilisé dans le coût d'hébergement local.

Charge humaine

Actuellement 0,4 ETP

- Bon fonctionnement des machines (intervention heures ouvrées, pas de certitude de fonctionnement 24/24 7/7)
- Maintenance système sur l'ensemble du parc VM et physique (serveurs)
- Gestion des comptes utilisateurs sur les différentes machines
- Installation d'outils
- Mise en place de répliquions rsync pour la sauvegarde

Charge humaine

Actuellement 0,4 ETP

- Bon fonctionnement des machines (intervention heures ouvrées, pas de certitude de fonctionnement 24/24 7/7)
- Maintenance système sur l'ensemble du parc VM et physique (serveurs)
- Gestion des comptes utilisateurs sur les différentes machines
- Installation d'outils
- Mise en place de répliquions rsync pour la sauvegarde

Manque de temps pour

- Déploiement d'outils de supervision (nécessité ?)
- Déploiement d'un gestionnaire de job sur les nœuds de calcul (critique)
- Gestion rationnelle du parc serveur (amélioré le cloisonnement, répartition de la charge)

Charge humaine

Actuellement 0,4 ETP

- Bon fonctionnement des machines (intervention heures ouvrées, pas de certitude de fonctionnement 24/24 7/7)
- Maintenance système sur l'ensemble du parc VM et physique (serveurs)
- Gestion des comptes utilisateurs sur les différentes machines
- Installation d'outils
- Mise en place de répliquions rsync pour la sauvegarde

Manque de temps pour

- Déploiement d'outils de supervision (nécessité ?)
- Déploiement d'un gestionnaire de job sur les nœuds de calcul (critique)
- Gestion rationnelle du parc serveur (amélioré le cloisonnement, répartition de la charge)

Recrutement en mobilité d'un AI sysadmin pour juillet 2018

Financement sur projets

- Planification d'investissement (achat / renouvellement)
- Pas de financement récurrent (incompatible avec abonnement)

Financement sur projets

- Planification d'investissement (achat / renouvellement)
- Pas de financement récurrent (incompatible avec abonnement)

Parc de petite taille et à proximité

- Supervision relativement simple (aidé par le faible nombre de pannes)

Financement sur projets

- Planification d'investissement (achat / renouvellement)
- Pas de financement récurrent (incompatible avec abonnement)

Parc de petite taille et à proximité

- Supervision relativement simple (aidé par le faible nombre de pannes)

Stockage évolutif

- 50 % de la capacité en disques
- Faible coût par tranche de 40 disques (finançable sur projets de recherche)

Financement sur projets

- Planification d'investissement (achat / renouvellement)
- Pas de financement récurrent (incompatible avec abonnement)

Parc de petite taille et à proximité

- Supervision relativement simple (aidé par le faible nombre de pannes)

Stockage évolutif

- 50 % de la capacité en disques
- Faible coût par tranche de 40 disques (finançable sur projets de recherche)

Pression de la DSI pour

- Fermer salles serveurs de centre
- Placer les machines en Data Center

Financement sur projets

- Planification d'investissement (achat / renouvellement)
- Pas de financement récurrent (incompatible avec abonnement)

Parc de petite taille et à proximité

- Supervision relativement simple (aidé par le faible nombre de pannes)

Stockage évolutif

- 50 % de la capacité en disques
- Faible coût par tranche de 40 disques (finançable sur projets de recherche)

Pression de la DSI pour

- Fermer salles serveurs de centre
- Placer les machines en Data Center

Proximité traitement / stockage

Salle serveur = salle brassage

Data Center Université d'Angers (tutelle, membre SFR)

La vraie vie

Installation :

- Faible coût (surtout de type abonnement) **VMWare**
 - Facile à maintenir / paramétrer **Xen**
 - Environnement de virtualisation stable
 - Capable d'exploiter 1,5 To RAM **VMWare VirtualBox**
 - Capable de gérer des machines virtuelles avec 500+ Go RAM
 - Open Source
- juillet 2014 : **XenServer 6.2**

La vraie vie

Installation :

- Faible coût (surtout de type abonnement) **VMWare**
 - Facile à maintenir / paramétrer **Xen**
 - Environnement de virtualisation stable
 - Capable d'exploiter 1,5 To RAM **VMWare VirtualBox**
 - Capable de gérer des machines virtuelles avec 500+ Go RAM
 - Open Source
- juillet 2014 : **XenServer 6.2**

Paramétrage :

- Prise en compte du multipath pour les baies sous XenServer
- Suppression de la limite de RAM de 1 To
- Augmentation de l'espace disque du Dom0

La vraie vie

Installation :

- Faible coût (surtout de type abonnement) **VMWare**
 - Facile à maintenir / paramétrer **Xen**
 - Environnement de virtualisation stable
 - Capable d'exploiter 1,5 To RAM **VMWare VirtualBox**
 - Capable de gérer des machines virtuelles avec 500+ Go RAM
 - Open Source
- juillet 2014 : **XenServer 6.2**

Paramétrage :

- Prise en compte du multipath pour les baies sous XenServer
- Suppression de la limite de RAM de 1 To
- Augmentation de l'espace disque du Dom0

Conclusions :

- 9 mois de tests, installations et paramétrages
- 3 ans 1/2 sans problème majeur
- Indisponibilité : 1/2 journée par an (mises à jours de l'hyperviseur)
- Coût « invisible » comparable à un congélateur -80°C

Freins :

- ETP (maintenance, administration)
- Infrastructure :
 - locaux
 - électricité
- Pression DSI
- Pérennisation / renouvellement (financement sur projets)

Leviers :

- Proximité des sources de données
- Proximité du traitement des données
- Financement non récurrent (pas de coût type abonnement) compatible avec le fonctionnement en mode projet
- Maîtrise complète de l'infrastructure