

Aide au choix d'un système de stockage distribué

ANF "Des données au bigdata: exploitez le stockage distribué !"
12-16 December 2016 - Complexe d'accueil de Gif-sur-Yvette

Plan de la présentation

- Différents types de besoins : profil des Entrées / Sorties
- Fonctionnalités et impacts sur les performances
- Conclusion

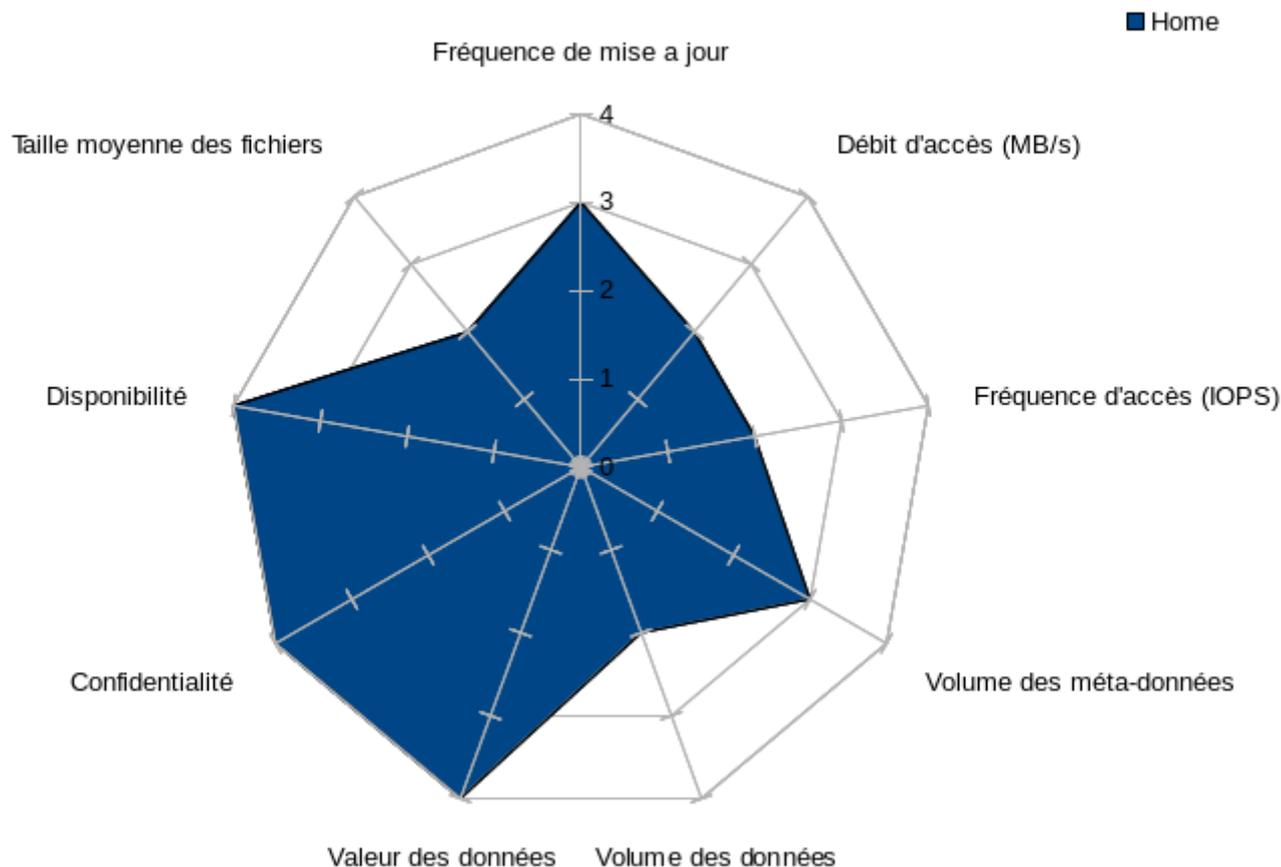
Différents types de besoins

Exemple de classification : HOME, Base de données, software area, scratch, acquisitions primaire (raw datas), analyses finales.

Il y en a beaucoup d'autres : HPC (checkpoint & restart), HTC (data mining...), streaming, distribution géographique des données, PRA...

Besoin de stockage type :

- HOME :
 - * disponibilité,
 - * données précieuses,
 - * confidentialité



Différents types de besoins

Exemple de classification : HOME, Base de données, software area, scratch, acquisitions primaire (raw datas), analyses finales

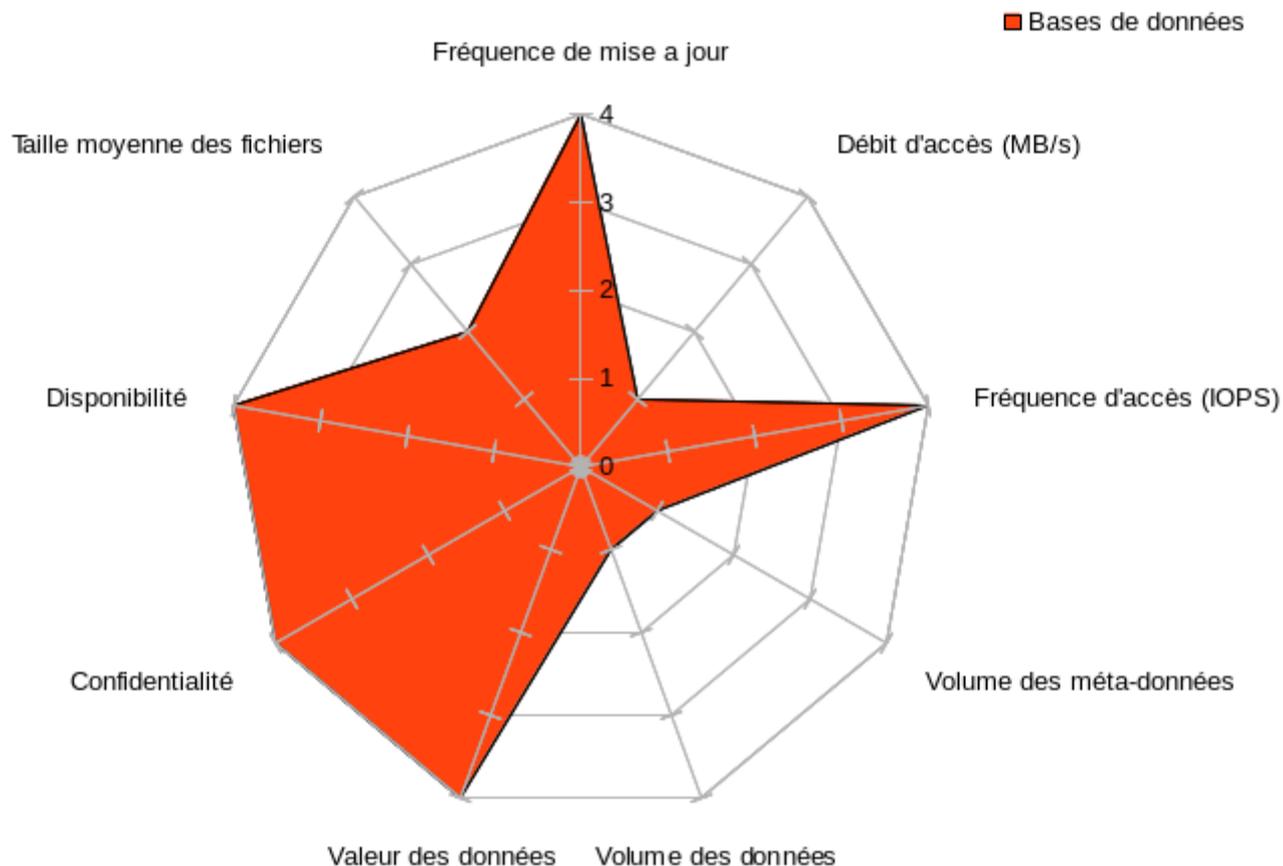
Besoin de stockage type :

- HOME :

- * disponibilité,
- * données précieuses,
- * confidentialité

- Bases de données :

- * accès et mises à jour fréquentes
- * disponibilité,
- * données précieuses,
- * confidentialité



Exemple illustratif non représentatif d'un cas général

Différents types de besoins

Exemple de classification : HOME, Base de données, software area, scratch, acquisitions primaire (raw datas), analyses finales

Besoin de stockage type :

- HOME :

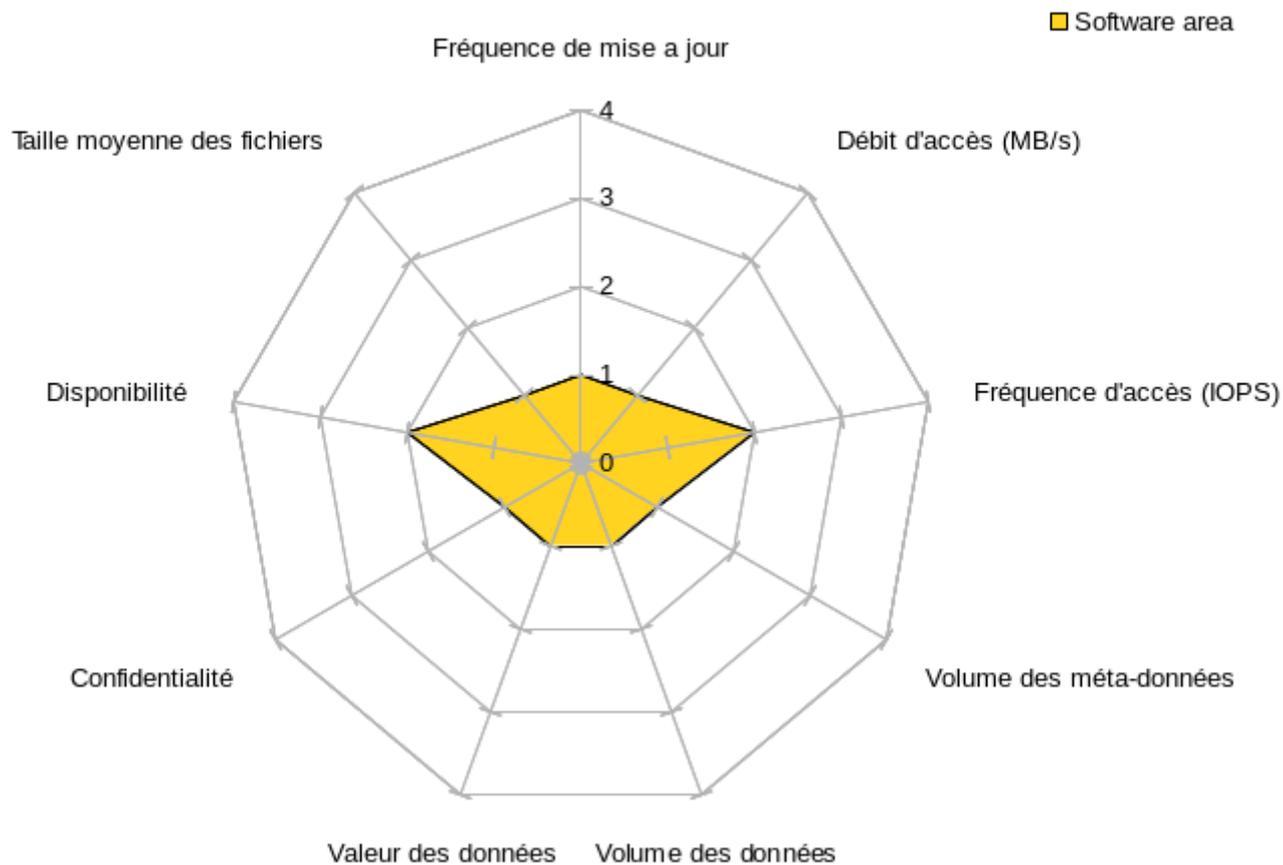
- * disponibilité,
- * données précieuses,
- * confidentialité

- Bases de données :

- * accès et mises à jour fréquentes
- * disponibilité,
- * données précieuses,
- * confidentialité

- Software aréa :

- * pas de grandes contraintes



Exemple illustratif non représentatif d'un cas général

Différents types de besoins

Exemple de classification : HOME, Base de données, software area, scratch, acquisitions primaire (raw datas), analyses finales

Besoin de stockage type :

- HOME :

- * disponibilité,
- * données précieuses,
- * confidentialité

- Bases de données :

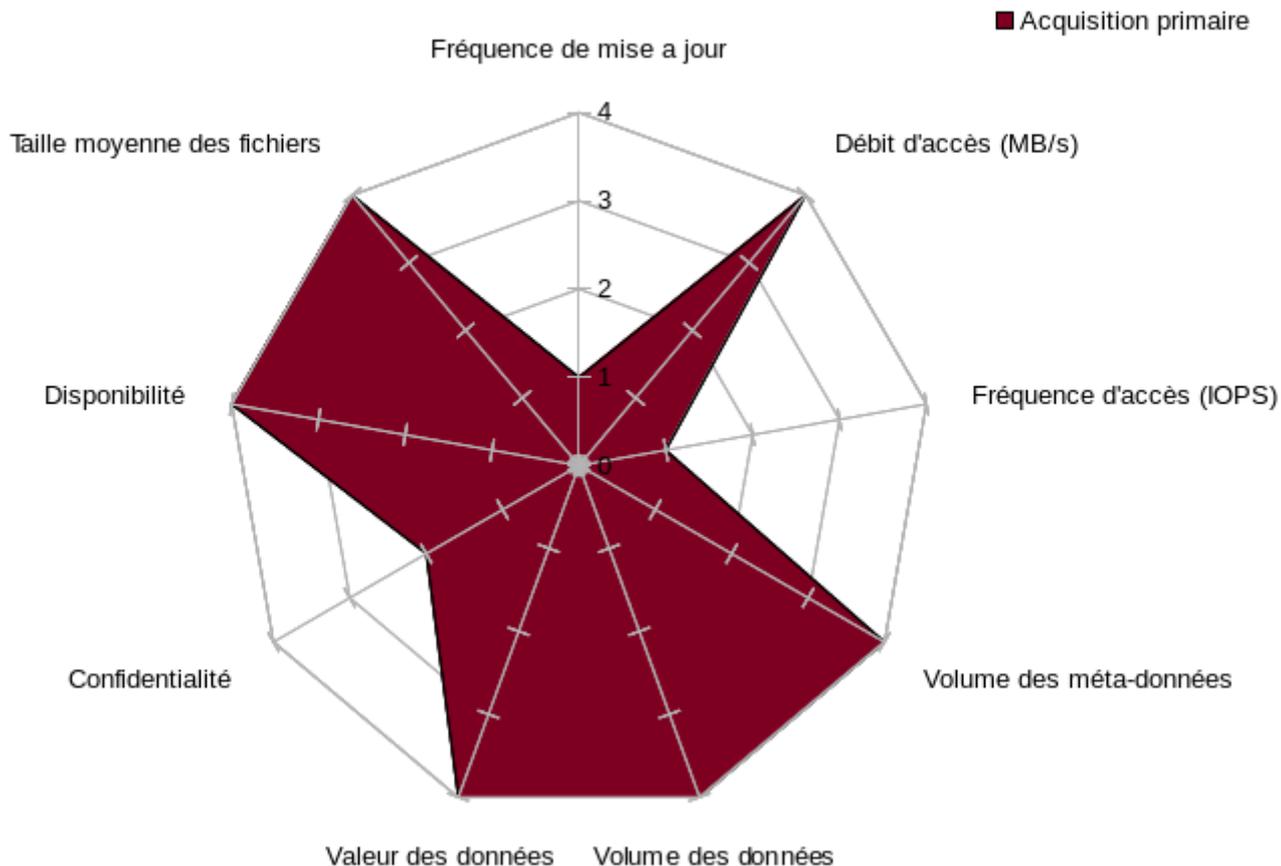
- * accès et mises à jour fréquentes
- * disponibilité,
- * données précieuses,
- * confidentialité

- Software aréa :

- * pas de grandes contraintes

- Acquisition primaire :

- * contraintes sur les volumes
- * disponibilité,
- * contraintes sur les débits



Exemple illustratif non représentatif d'un cas général

Différents types de besoins

Exemple de classification : HOME, Base de données, software area, scratch, acquisitions primaire (raw datas), analyses finales

Besoin de stockage type :

- HOME :

- * disponibilité,
- * données précieuses,
- * confidentialité

- Bases de données :

- * accès et mises à jour fréquentes
- * disponibilité,
- * données précieuses,
- * confidentialité

- Software aréa :

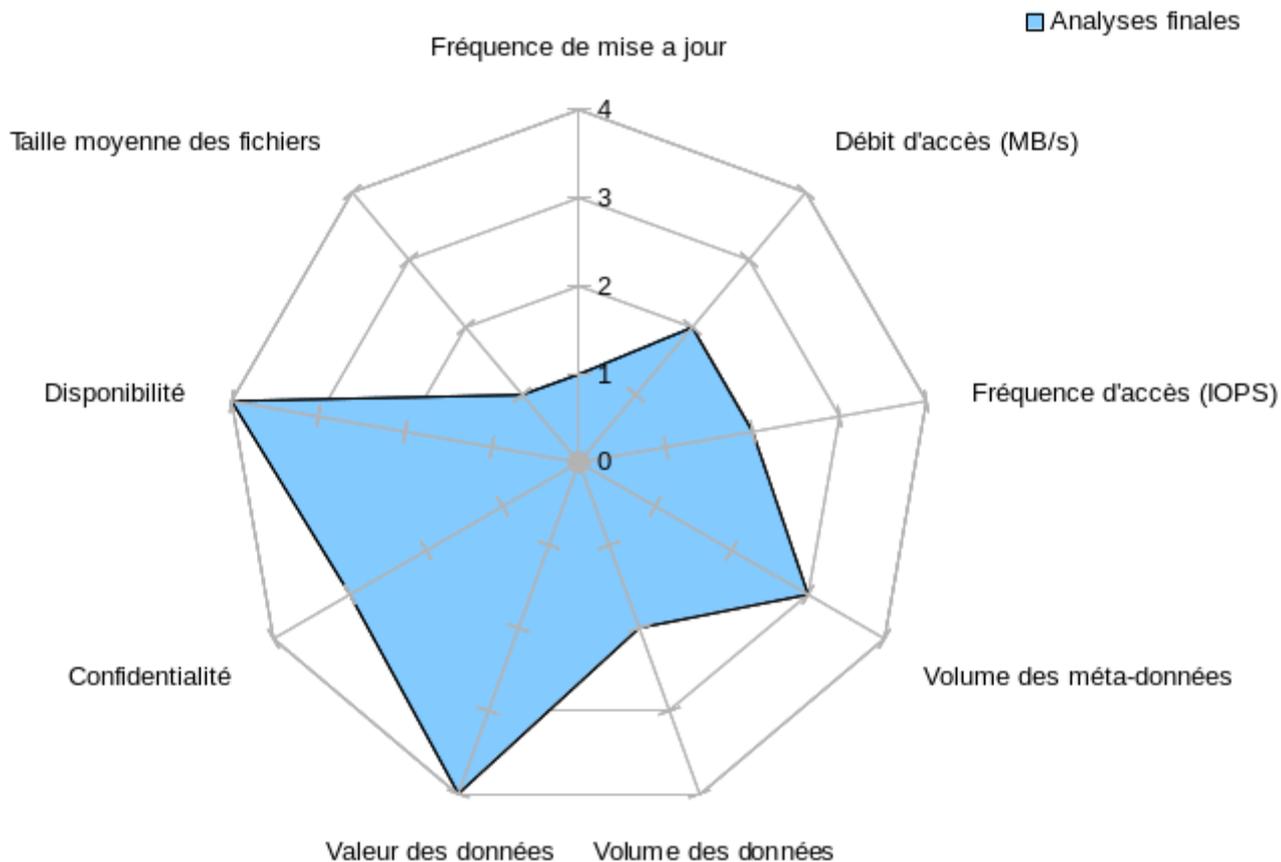
- * pas de grandes contraintes

- Acquisition primaire :

- * contraintes sur les volumes
- * disponibilité,
- * contraintes sur les débits

- Analyses finales

- * données précieuses,
- * disponibilité,



Exemple illustratif non représentatif d'un cas général

Différents types de besoins

Exemple de classification : HOME, Base de données, software area, scratch, acquisitions primaire (raw datas), analyses finales

Besoin de stockage type :

- HOME :

- * disponibilité,
- * données précieuses,
- * confidentialité

- Bases de données :

- * accès et mises à jour fréquentes
- * disponibilité,
- * données précieuses,
- * confidentialité

- Software aréa :

- * pas de grandes contraintes

- Acquisition primaire :

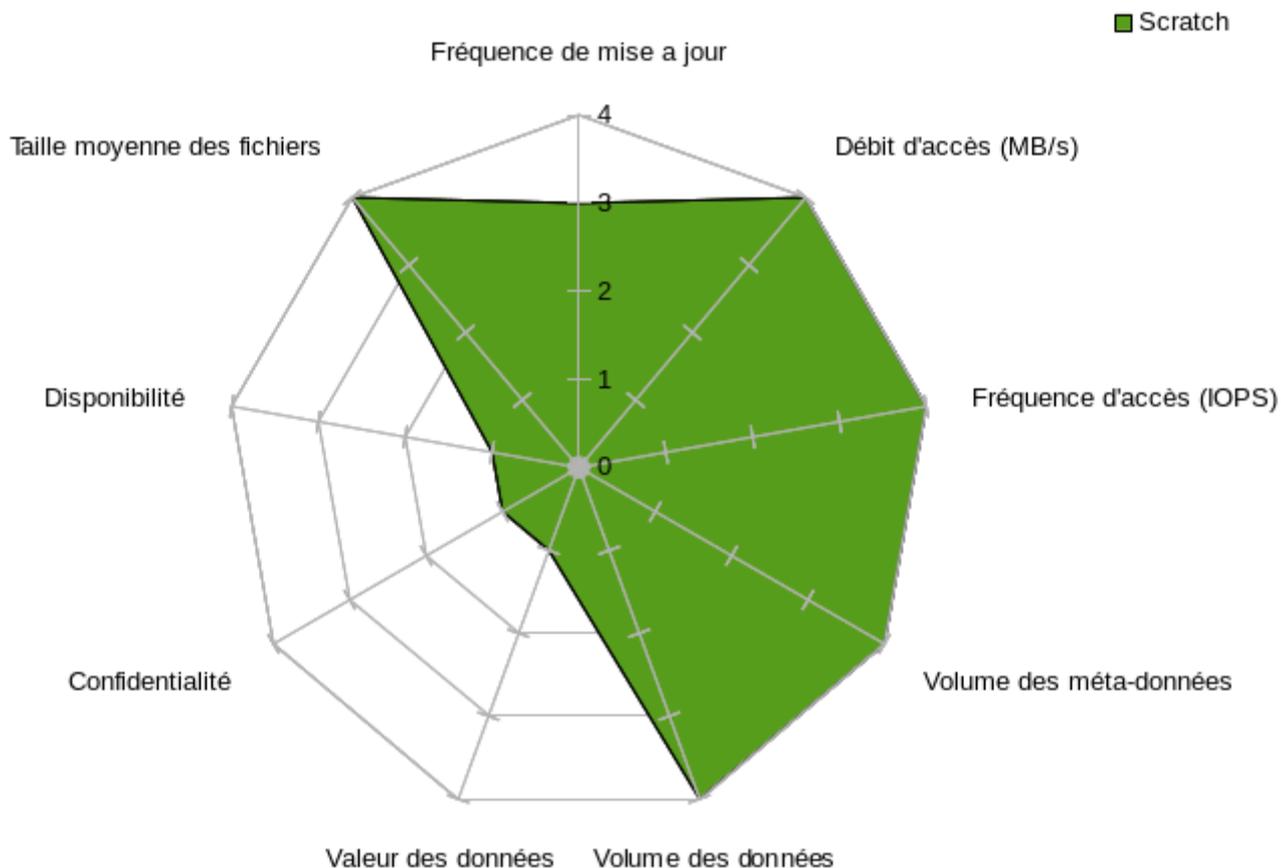
- * contraintes sur les volumes
- * disponibilité,
- * contraintes sur les débits

- Analyses finales

- * données précieuses,
- * disponibilité,

- scratch :

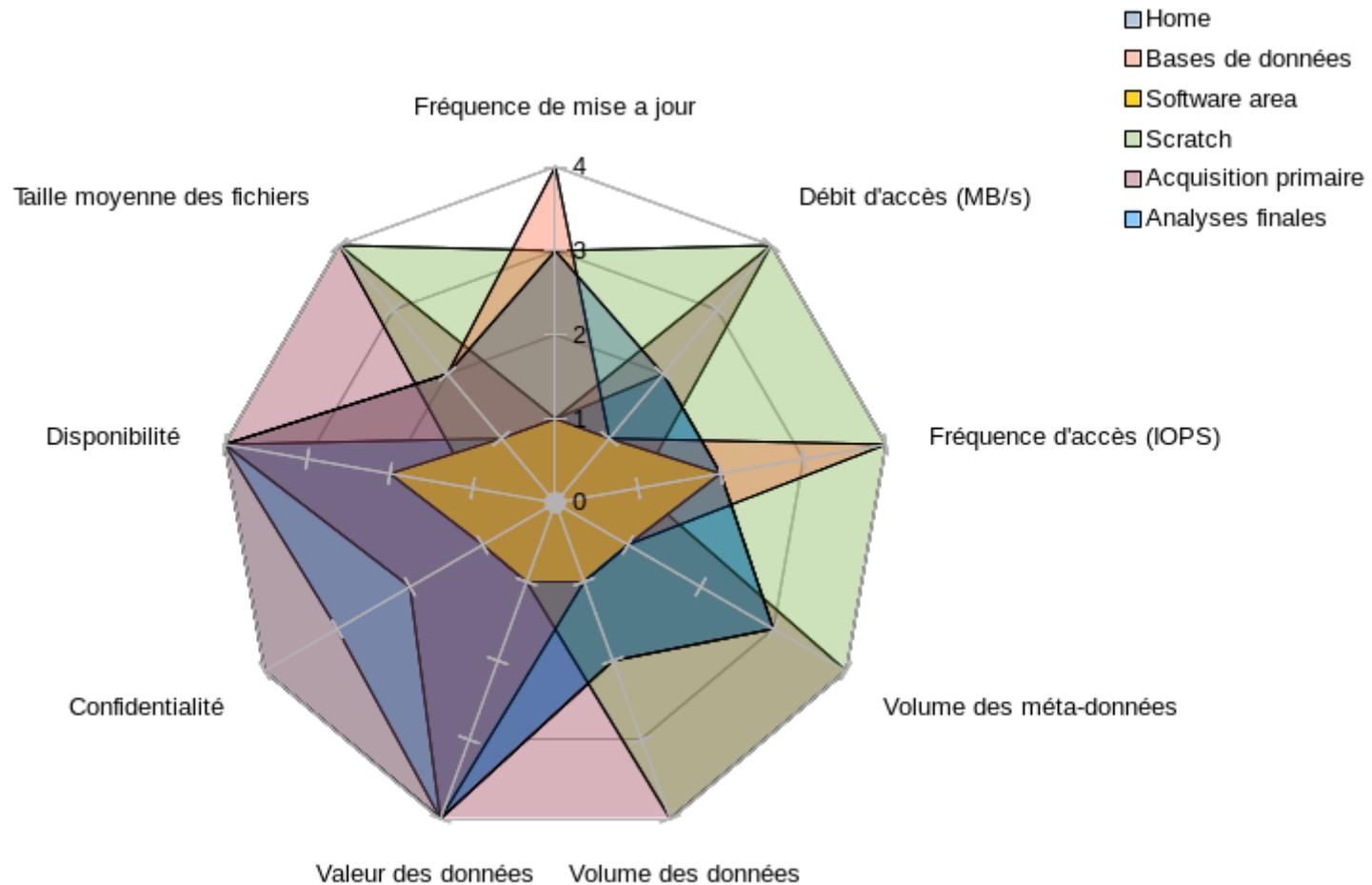
- * données volatiles
- * contraintes sur les débits
- * contraintes sur les volumes



Exemple illustratif non représentatif d'un cas général

Outil idéal ?

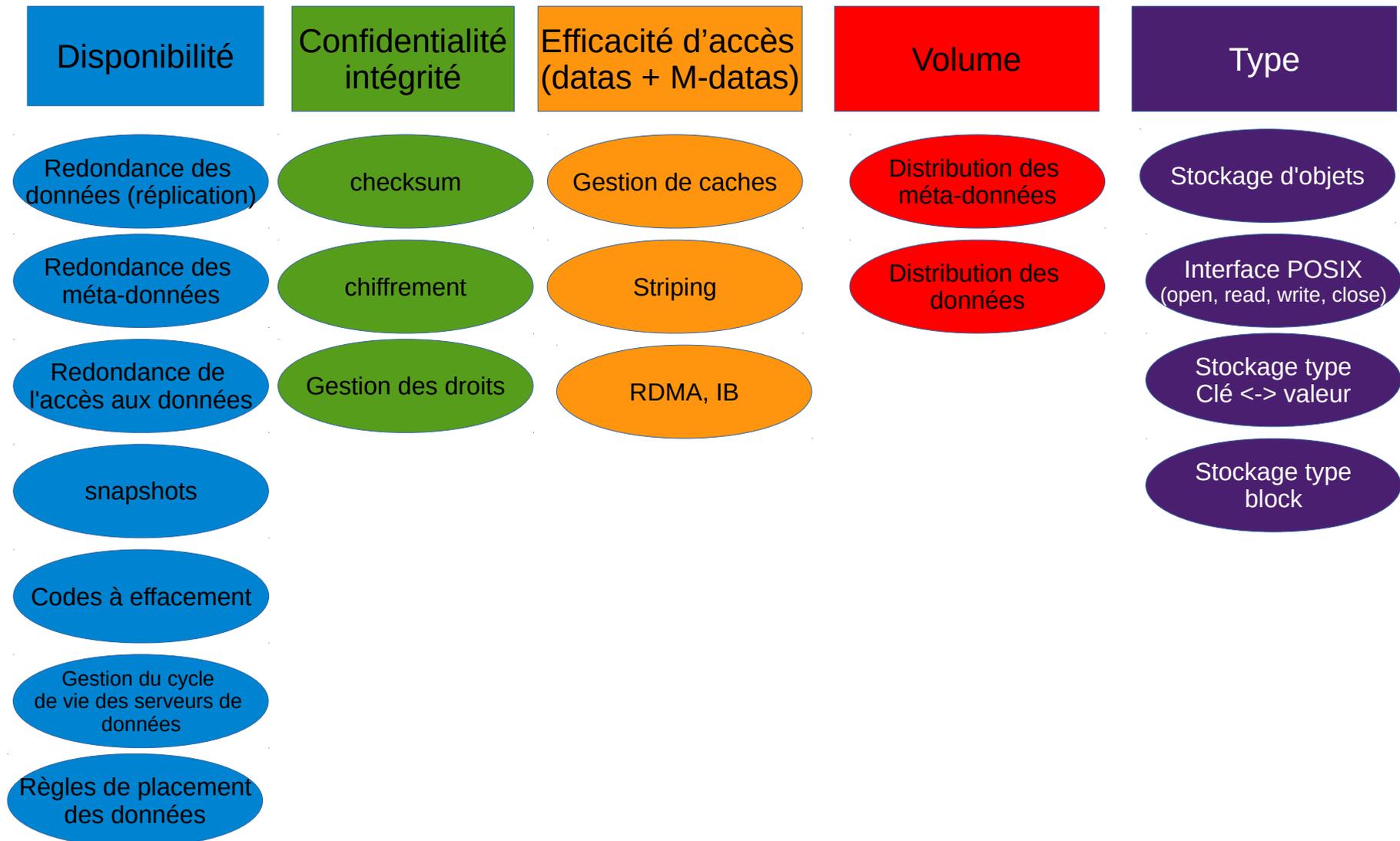
Existe-il un système qui couvre tous ces besoins ?



Exemple illustratif non représentatif d'un cas général

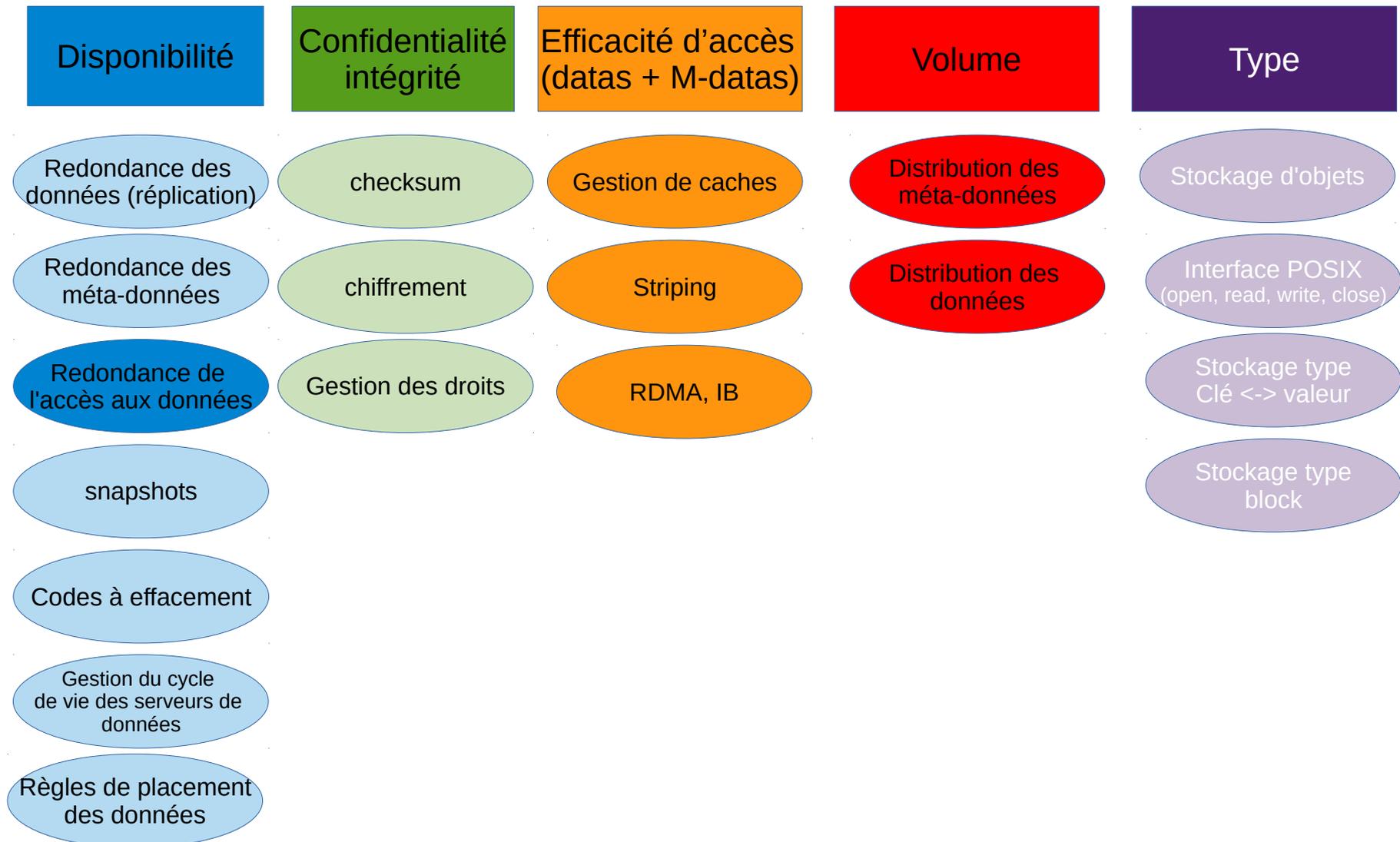
Différents types de fonctionnalités

Fonctionnalités des systèmes de fichiers/stockage distribués



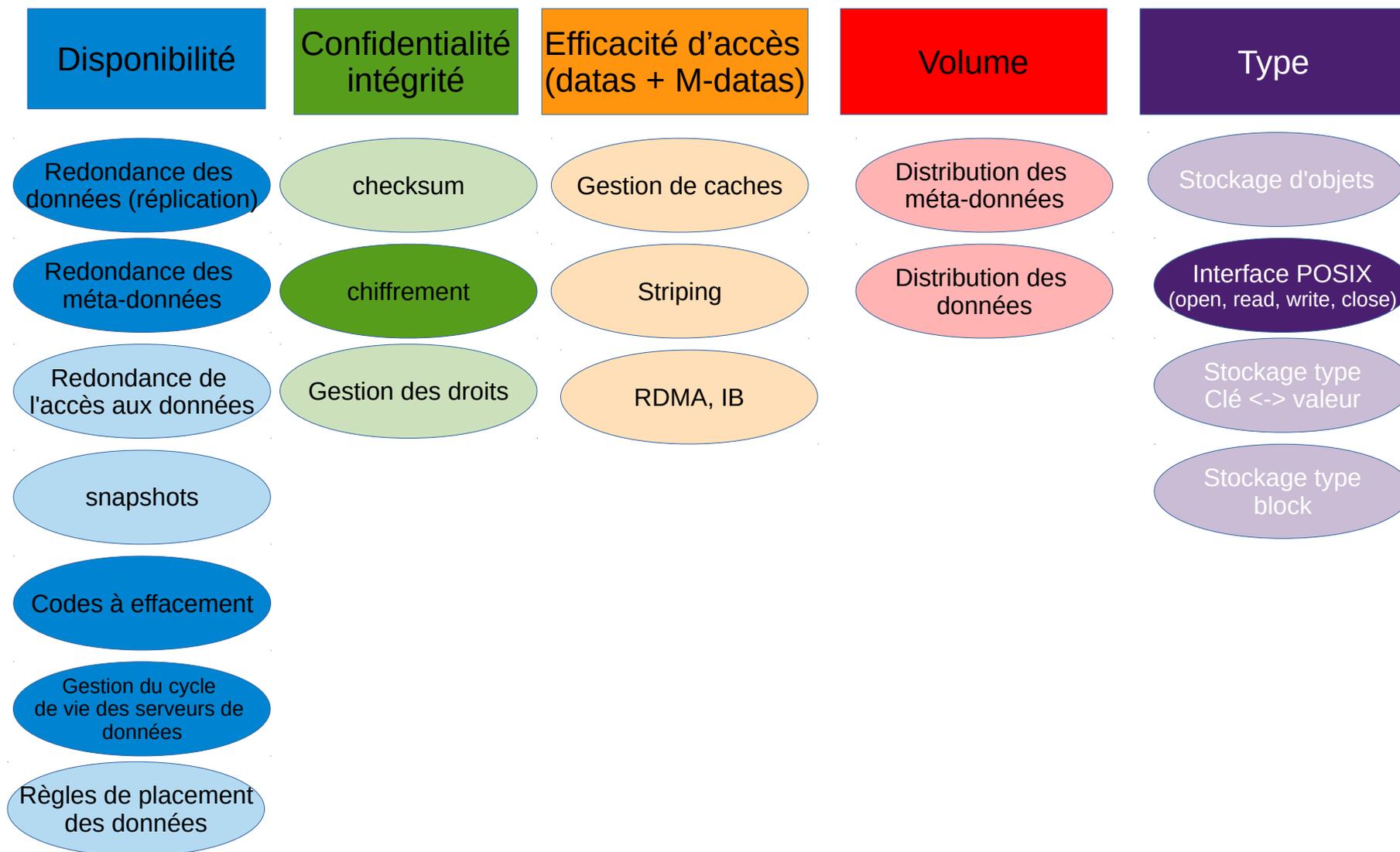
Quelles fonctionnalités ?

Quelles fonctionnalités ont un impact **positif** sur les performances ?



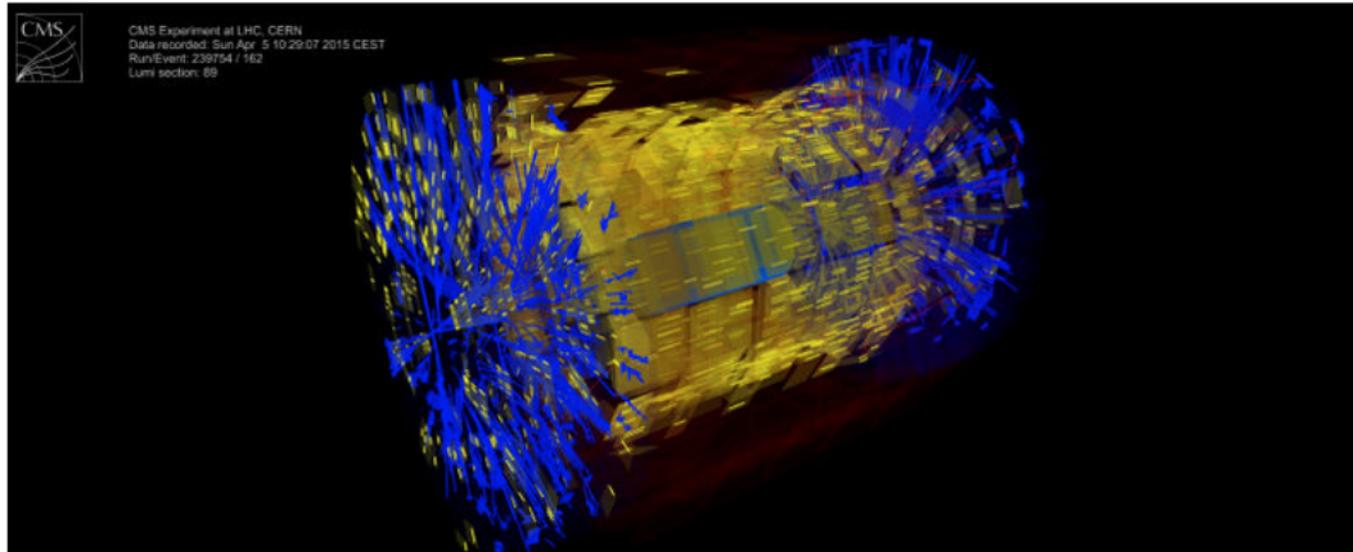
Quelles fonctionnalités ?

Quelles fonctionnalités ont un impact **négatif** sur les performances ?



Ce que je pense, tout le monde souhaite éviter :

Conclusion



Event display of one of the first particle splashes seen in CMS during Run2

... only a few minutes before one of the OSS servers crashed...

... and the failover mechanism failed ...

27 / 29

Online data handling with Lustre at the CMS Experiment
Lavinia Darlea, on behalf of CMS DAQ Group, September 17, 2015
2015 European Open File System conference

Conclusion

- Critères généraux :
 - Fonctionnalités,
 - Rapport coût / TB,
 - Rapport performance / coût,
 - Scalabilité de la solution,
 - Robustesse,
 - Complexité versus facilité de maintenance dans le temps,
 - Coût humain d'exploitation de la solution.
- Quantifier le service rendu :
 - En fonction du besoin
 - En fonction des fonctionnalités réellement utilisées
- Ne pas négliger le coût de sortie de la solution