

Analyse de données dans H1

Emmanuel Sauvan

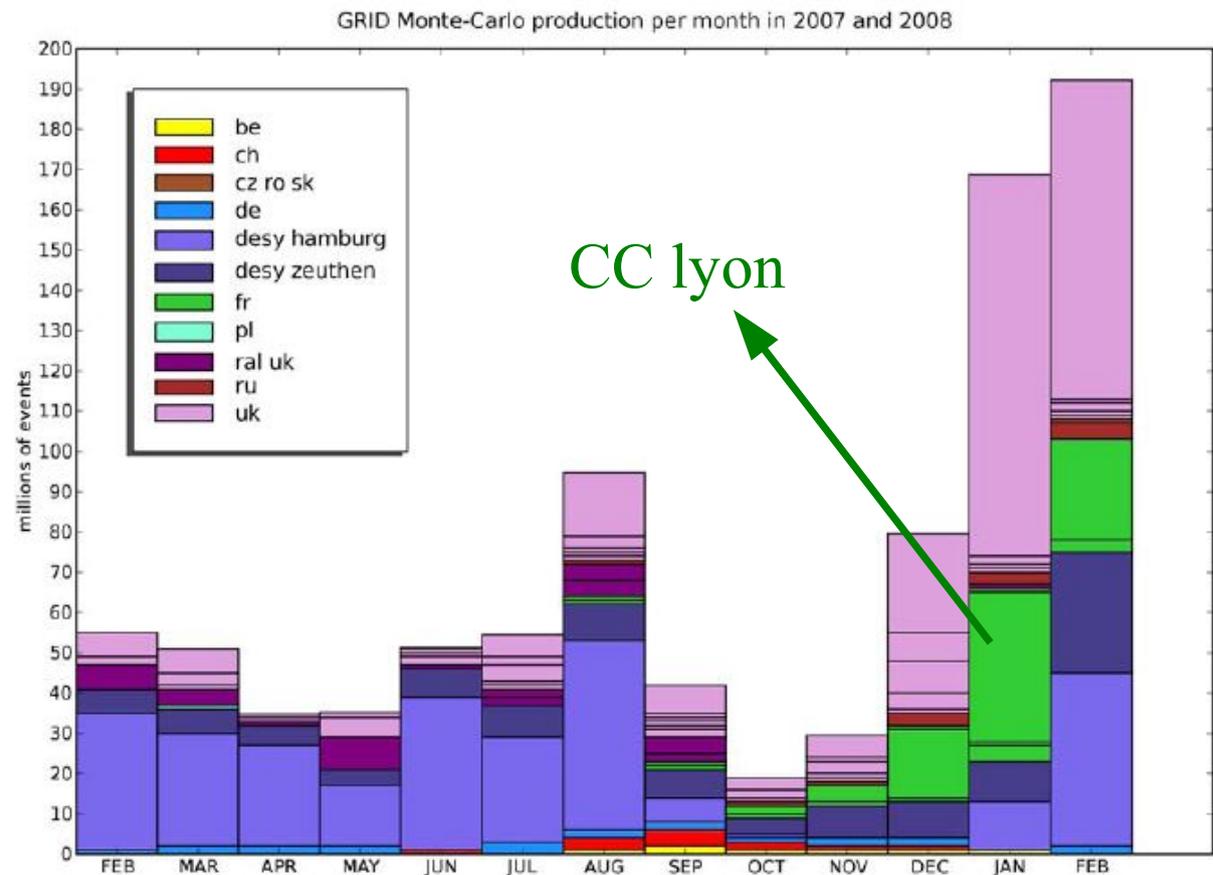
CPPM Marseille

Reconstruction des données et MC

- Reconstruction des données faite à DESY
 - ➔ Stockage sur bande à DESY
 - ➔ Présélection et copie au CC

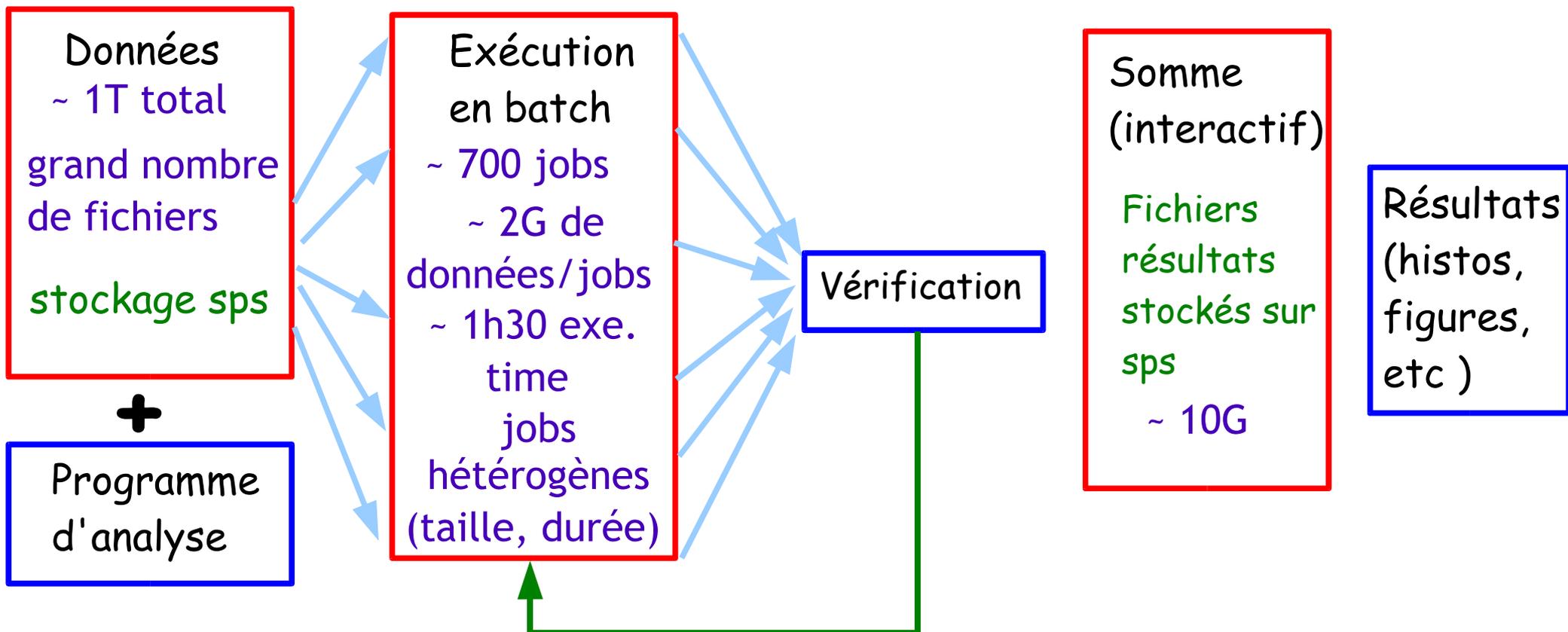
- MC = 10 x données

- ➔ Composante (très) importante pour l'analyse
- ➔ Production maintenant faite via la grille
- ➔ Contribution importante du CC depuis qques mois



Modèle d'analyse

- Analyse : après reconstruction et pré-sélection des données
 - ➔ ~6-8 % des données totales
 - ➔ Format avec information réduite ~5kO / evt (4-vecteurs des particules identifiées)
- Logiciels d'analyse : C++ et Root



Besoins spécifiques et limitations

↳ Besoins :

- « turn over » rapide
 - relancer l'analyse 1 à 3 fois par jour
- Besoin de tourner un grand nombre de petits jobs en parallèle
- Fiabilité
 - Taux échec des jobs doit être faible
- Utilisation non continue mais possibilité de pics

↳ Limitations actuelles :

- accès disque :
 - Nombre de connections simultanées aux mêmes données trop faible
 - Par ex.: sps/hera1 limité à ~80-100 jobs en parallèle
- Fiabilité

Expérience perso. de l'utilisation du CC

↘ Positif :

- Rapidité du CPU des machines récentes
- Bonne fiabilité depuis ~6 mois

↘ Négatif :

- Limitation du nombre de connections sps trop faible et rédhibitoire
- Manque de flexibilité dans la gestion des ressources
- Manque de réactivité face à des demandes spécifiques pour l'analyse par les expériences
 - Gestion des priorités de batch vs. demandes de CPU intégrées sur l'année adaptée aux grosses productions et pas pour l'analyse
 - Augmentation de l'utilisation de disques (sps) a été trop lente / aux autres centres (DESY)
 - Manque de dialogue avec les utilisateurs

CC vs. DESY

- DESY mieux adapté aux analyses depuis plusieurs années
 - Possibilité de grandes capacités disques (achetées par les groupes externes, ~6T pour H1-france ...)
 - Utilisation de dCache (disque, accès aux mêmes données par un grand nombre de jobs)
 - Gestion des priorités batch plus flexible
 - Actuellement: impossible de mener toutes les analyses H1-france au CC seul
- Fiabilité cyclique à DESY ou au CC
 - Nécessité de déployer les analyses sur 2 sites pour éviter de rester bloqué ...