



Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

Spark au CC-IN2P3

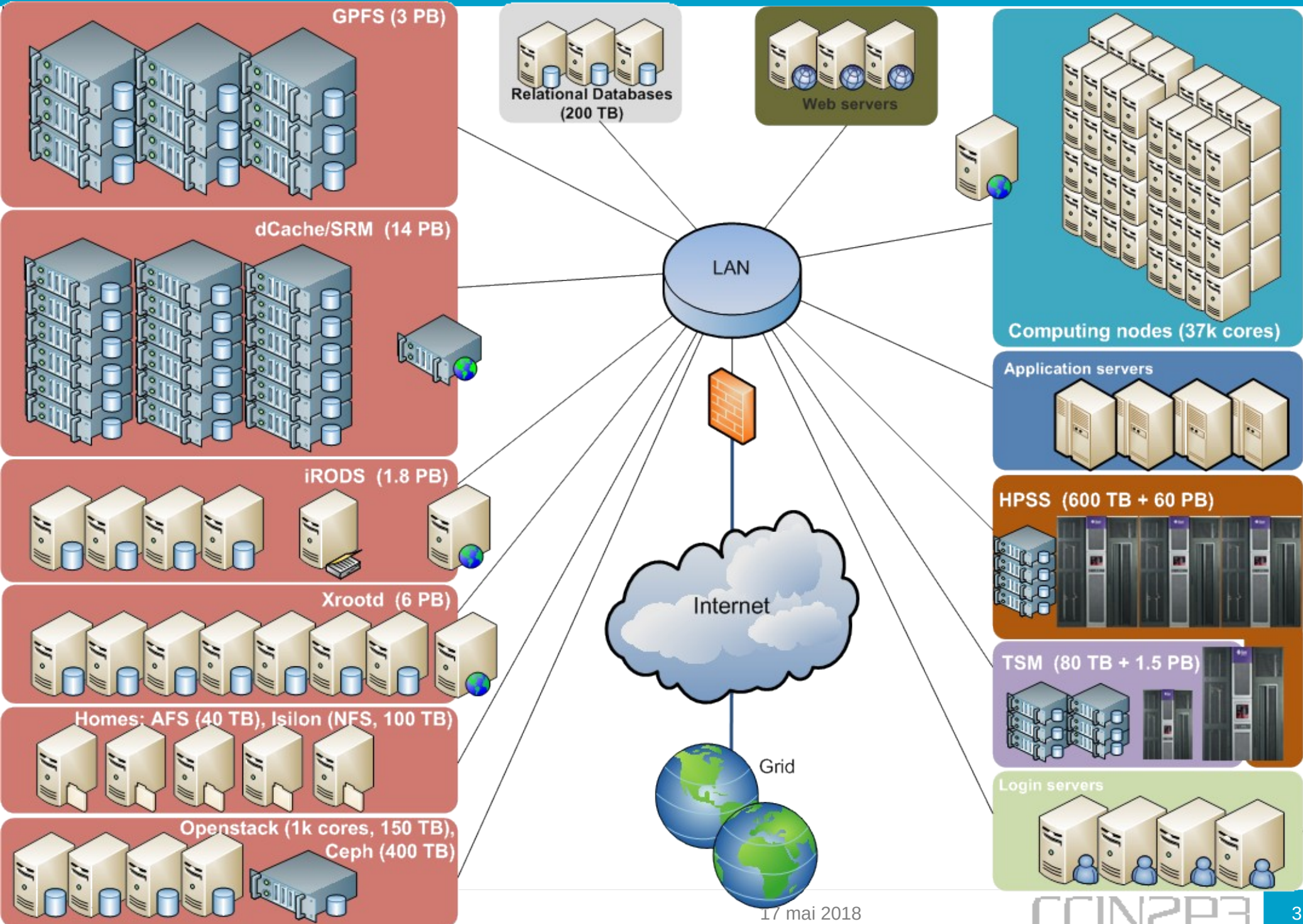
(Extrêmement rapide)

ANF UST4HPC, Fréjus 2018
Loïc Tortay



- *MapReduce*
- Analyse de données « en mémoire » sur n noeuds
- API Java, Scala, Python
- Intégration avec notebooks (Jupyter, Zeppelin, ...)
- Stockage distribué

CC-IN2P3



17 mai 2018

CCIN2P3

- Financement Région Rhône-Alpes (TIDRA)
- Besoins définis de façon (très) incomplète
- Faible retour des utilisateurs potentiels après premier(s) contact(s)
- Difficultés concrètes :
 - intégration environnement existant (accès aux données, batch, *shared nodes*)
 - ⇔ VMs OpenStack sur 20 hyperviseurs dédiés pour TIDRA (400 coeurs, 2.5 To RAM)
- Données dans Ceph (+HDFS plugin) ou/et GPFS (+HDFS plugin)