

Accès aux données de



*Large Synoptic Survey Telescope*

Emmanuel Gangler - LPC

# LSST Data Management

- Flot de données important :
  - Image : 6 Gbyte/17 secondes
  - 15 TB/nuit
  - ~100 PB archives finales images
  - $40 \cdot 10^9$  objets ( table de 100-200 TB )
  - $5\,000 \cdot 10^9$  observations ( table de 1-3 PB )
  - Par nuit : alertes sur les transitoires ( $10^6$ )
  - Accès « ouvert » aux données : qui ? CC/NCSA ?
- Champ interdisciplinaire avec recherche informatique : BigData
- **2012** : mise en place consortium PetaSky
- **2013** : Test grande échelle au CC



Simulation 1  
CCD 4k x 4k

# Paradigme LSST :

Processing Cadence	Image Category (files)	Catalog Category (database)	Alert Category (database)
Nightly	Raw science image Calibrated science image Subtracted science image Noise image Sky image Data quality analysis	Source catalog (from difference images) Object catalog (from difference images) Orbit catalog Data quality analysis	Transient alert Moving object alert Data quality analysis
Data Release (Annual)	Stacked science image Template image Calibration image RGB JPEG Images Data quality analysis	Source catalog (from calibrated science images) Object catalog (optimally measured properties) Data quality analysis	Alert statistics & summaries Data quality analysis

- L2 : caractérisation → catalogues + stacked images
  - Delivré par projet
- L3 : analyse
  - Hors projet : ex. DESC
  - Quelle relation aux données L2 / reprocessing spécifique ?

# Tables des catalogues (extrait) :

Table	Taille	#enregistrements	#colonnes (arité)
Object	109 TB	38 B	470 ! = tendance à l'inflation !
Moving Object	5 GB	6 M	100
Source	3.6 PB	5 T	125
Forced Source	1.1 PB	32 T	7
Difference Image Source	71 TB	200 B	65
CCD Exposure	0.6 TB	17 B	45

- Besoins analyses DESC : (à raffiner sérieusement!)
- Supernovae :
  - Table source (+ forcées, différences ?)
  - Peu d'objets → extraction de données de travail
  - Réseau d'étoiles de calibration → quantité requise ?
  - Possible de fonctionner en full scan ?

# Tables des catalogues (extrait) :

Table	Taille	#enregistrements	#colonnes (arité)
Object	109 TB	38 B	470 ! = tendance à l'inflation !
Moving Object Source	5 GB	6 M	100
Forced Source	3.6 PB	5 T	125
Difference Image Source	1.1 PB	32 T	7
CCD Exposure	71 TB	200 B	65
	0.6 TB	17 B	45

- Besoins analyses DESC : (à raffiner sérieusement!)
  - Photo-z, BAO
    - Table objet suffit a priori, mais #galaxies élevé
    - Calcul photo-z sur un sous-ensemble d'arité faible → extraction, puis réinjection ? (colonne user)
    - Fonction de corrélation sur 3 attributs → extraction

# Positionnement :

- A-t-on besoin de maîtriser les produits du L2 ?
  - Pousser au maximum les algorithmes dans la chaîne de production !
  - En pratique, on utilisera des **sous-catalogues**
    - Possibilité d'interface avec DB via
      - tables utilisateur
      - Vues matérialisées (« cache » de nos sous-catalogues)
    - Note : un SGBD sert précisément à faire l'extraction des données utiles...
  - Le futur est incertain (CC as DAC?)
    - Mais les données seront au CC : pouvoir y accéder...
  - 2012-13 : a permis de mettre un pied dans le soft/les données LSST !

(... activité visible au CNRS ...)

# Les données LSST sont-elles « big » ?

“Big data” refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.

LDM-144

	SDSS	LSST 1 an	LSST 10 ans
Raw data	<b>14 TB</b>	6 PB	<b>60 PB</b>
Archive (bande)		19 PB	270 PB
Disque (DAC)		16 PB	90 PB
DB (baseline)	<b>7 TB</b>	0,5 PB	<b>5 PB</b>

- Remarques (1) :
  - Facteur ~6 entre DB/Raw entre SDSS et LSST !

# Les données LSST sont-elles « big » ?

“Big data” refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.

LDM-144

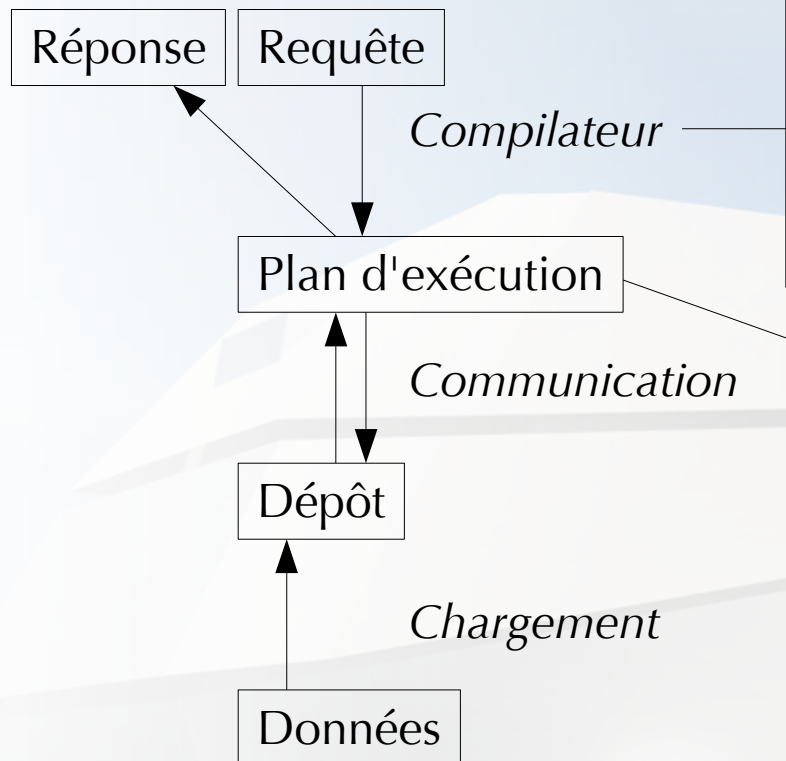
	SDSS	LSST 1 an	LSST 10 ans
Raw data	14 TB	6 PB	60 PB
Archive (bande)		19 PB	270 PB
Disque (DAC)		16 PB	90 PB
DB (baseline)	7 TB	0,5 PB	5 PB
Equivalent Moore 2014		12 TB	1.2 TB

- Remarques (2) : l'an 1 est le plus contraignant
  - 12 TB : → gros mais pas infaisable (>> SGBD classique → **Big**)
    - ~6H pour tout scanner à 600 Mo/s.
    - Nécessite de **distribuer** le système
  - La « loi » de Moore n'en est pas une ...



# Quelles implications ?

- SGBD « classique »



**Optimisation** heuristique

Réinterprétation de la sémantique

Choix des algorithmes

(ex . Utiliser index, hash table, full scan, ...)

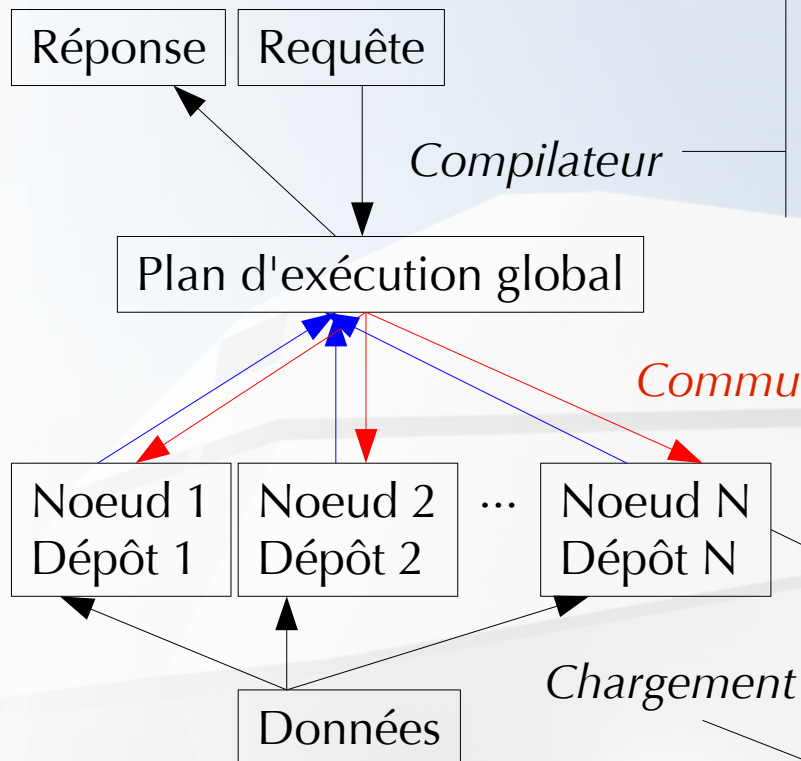
Modèle de coût : dominé par I/O

Non-adhérent au schéma de données

Effectue les calculs

# Quelles implications ?

- SGBD distribué



Modèle de coût : I/O + Communication

Adhérence au schéma de données ?

Problème d'intégration :

Schéma global complexe (overlaps!)

Réécriture de requêtes

- Spécifique ? Limite la grammaire d'entrée
- Générique ? Limite de l'informatique th.

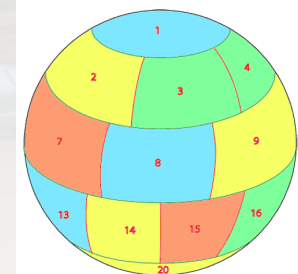
Synchronisation

Tolérance aux pannes !

Calcul au plus près du dépôt  
(map/reduce)

Choix du système du dépôt (MySQL, HDFS ...)

Choix du partitionnement



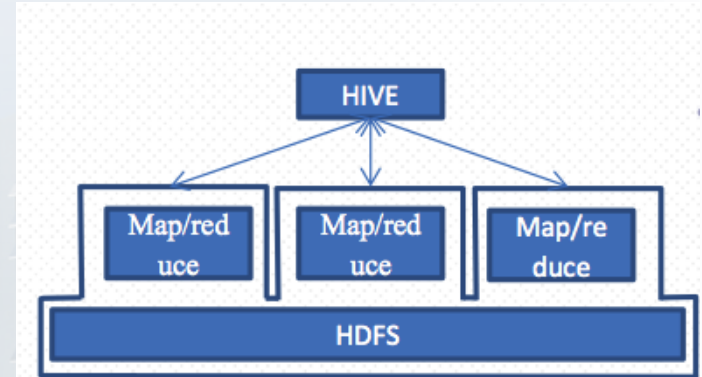
# Quelles solutions ?

- Clé en main : **Oracle** (*PetaSky*)
  - Utilisé à titre de comparaison
  - Avantages :
    - Gère le parallélisme
    - Offre le choix ligne ou colonne
    - Accès au plan avant exécution



# Quelles solutions ?

- Clé en main : **Hadoop** et cie...
  - Avantages :
    - Passe à l'échelle par construction
    - Map-Reduce
    - Bon en Full-scan
    - Bibliothèques de géométrie (*MASTODONS-GAIA*)
    - En plein essor (gestion des index, inclusion SGDB, ...)
  - Inconvénients :
    - Jointures inefficaces (et interdites si >2 tables)
    - Paramétrisation difficile

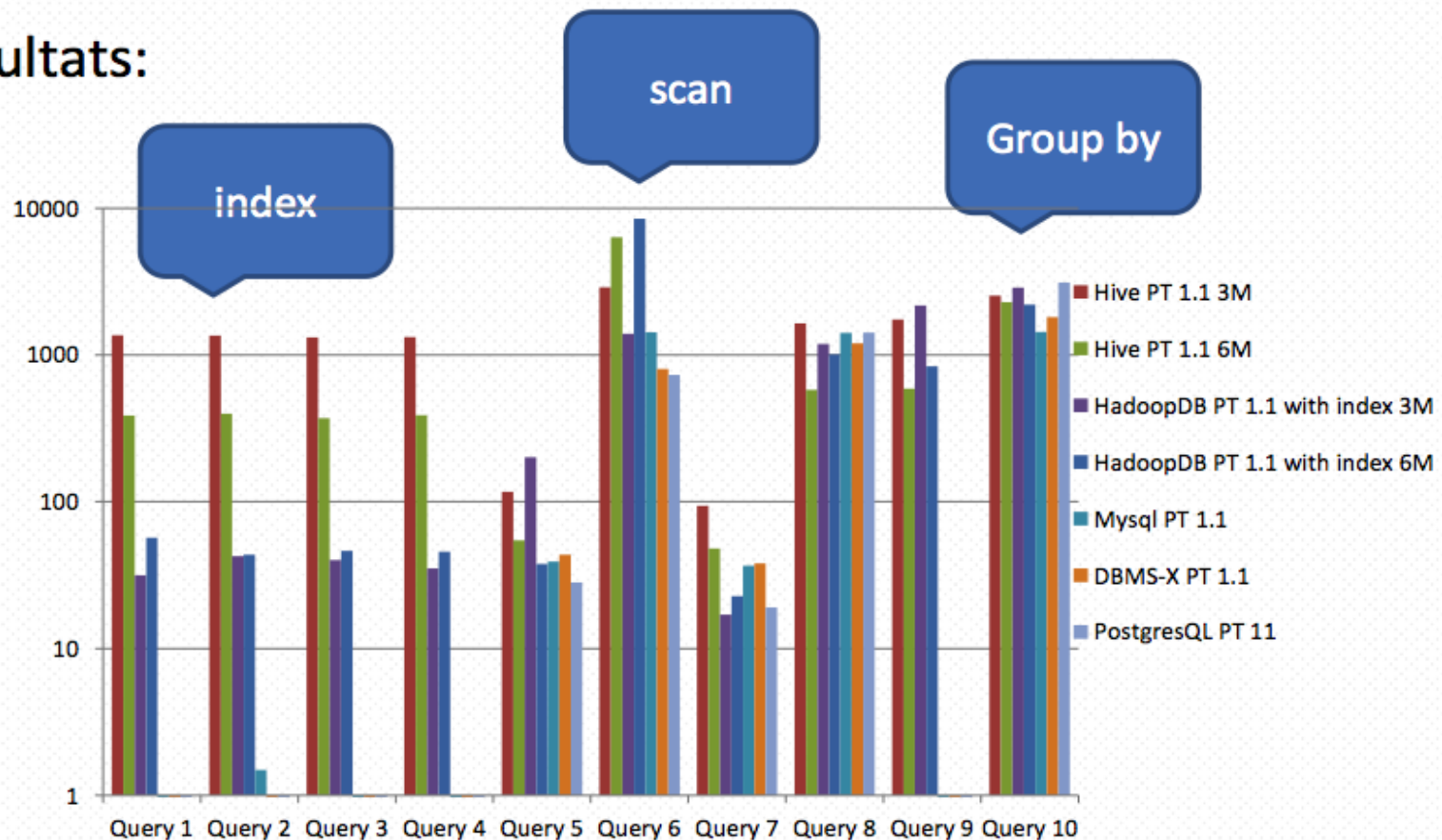


# Quelles solutions ?

- Clé en main : **Hadoop** et cie...

Amin Mesmoudi (LIRIS)

- Resultats:



- Loin derrière systèmes classiques en benchmark
- Limitation théorique ?

# Quelles solutions ?

- Cousu main : **Qserv** (SLAC) Cf. Fabrice Jammes
  - Avantages : c'est adapté à notre problème
  - Inconvénient : il faut (re)coder des composants complexes
  - Compilateur :
    - D. Wang (SLAC) : approche adhérente
    - E. Medernach (LPC) + F. Toumani (LIMOS) : approche générique
  - Communication
    - Hack Xrootd
  - Loader (D. Smith remplacé)
  - ... Et il faut que ça fonctionne ... (F. Jammes)

# Conclusion

- Le Use Case pour la science DESC demande encore à être précisé
- Le travail sur l'aspect **Big Data**
  - Est attractif pour la communauté informatique
    - Manpower gratuit, met en avant le LSST en Fr !
  - Met en avant la communauté Fr dans LSST
  - Enjeu local, national et européen
- Autres aspects
  - la **fouille de données** (en cours. C. Roucelle)
  - Les problèmes de visualisation (cf. G. Barrand)