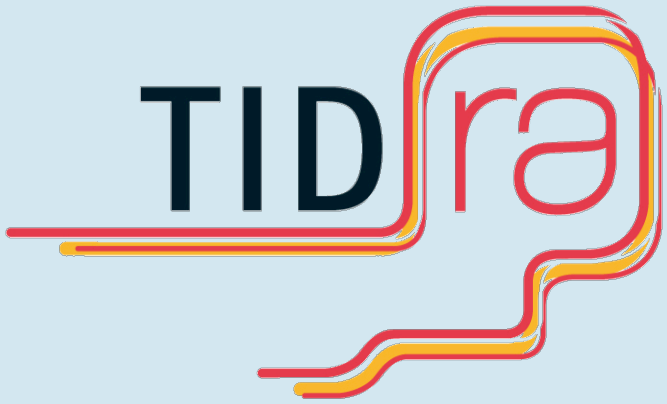
Journée de lancement TIDRA



De l'expression des besoins à la mise en production et aux résultats finaux

Yonny CARDENAS

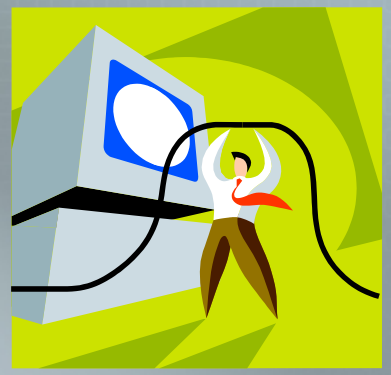
Traitement de données et Informatique Distribuée en Rhône-Alpes

Recherche et Innovation

- Étude ou simulation des événements ou phénomènes réels: naturels, sociaux, économiques, ...
- Développement des nouveaux produits, services.

Recherche et Innovation

- Modélisation informatique
  - Opérations complexes et répétitives par des ordinateurs
  - Objectif: extraire de l'information ou produire du savoir

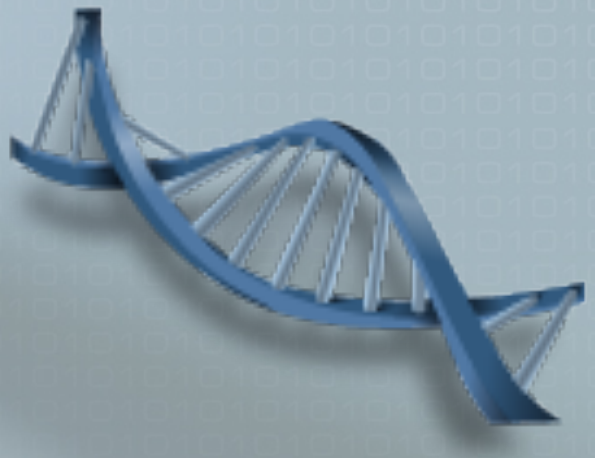


Traitement de Données

- Le traitement de données est une série de processus automatisés permettant d'extraire de l'information ou de produire du savoir à partir de données

Un exemple

- Description des gènes
  - Chaque cellule humaine contient 2 mètres d'ADN environ, correspondant dans les cellules diploïdes à 6,4 milliards de paires de nucléotides



Capacité de traitement

- La modélisation informatique demande régulièrement le traitement des volumes importants des données
  - Quantité grandissante
  - Qualité (richesse)

TIDRA

- TIDRA rassemble et mutualise des services informatiques pour les mettre à disposition de tous les scientifiques et industriels.
- Mission: Accélérer la recherche et l'innovation en Rhône-Alpes grâce à l'infrastructure informatique, le conseil et l'expertise

TIDRA

TIDRA

- Ressources sur les 5 sites (potentiellement) :
  - Calcul: 10000 Cœurs (*)
  - Stockage: 100 Téraoctets (*)
- Middleware (Job management):
  - gLite, LRMS (spécifiques)
- Stockage:
  - IRODS, GPFS, AFS, SRM
  - ORACLE, MYSQL, POSGRES

Démarche du conseil et de l'accompagnement

- Expression des besoins
- Évaluation des besoins
- Analyse et conception de la solution
- Mise en œuvre et production
- Résultats finaux

Démarche du conseil et de l'accompagnement

- Expression des besoins
  - Prise de contact
  - Réunions / entretiens personnels
  - Présentation globale du domaine et des activités (contexte)
  - Expression initiale des motivations et attentes (objectifs)

Démarche du conseil et de l'accompagnement

- Expression des besoins
  - Inventaire des besoins et contraintes
  - Première évaluation des besoins
  - Définition des actions à suivre (notions du plan de travail)
    - Techniques (tests, vérifications,...)
    - Administratives
    - Réactivité (actions en parallèle)

Démarche du conseil et de l'accompagnement

- Évaluation des besoins
  - Focalisée sur des aspects techniques
    - Traitement de données (stockage, accès, protection, …)
    - Calcul intensif
    - Adaptation du logiciel (portage)
    - Optimisations
  - Révision approfondie
  - Spécification précise (délimitation)
  - Établir les bénéfices à atteindre

Démarche du conseil et de l'accompagnement

- Analyse et conception de la solution
  - Analyse de fiabilité
  - Évaluation des alternatives techniques
  - Estimation des ressources
  - Identification des techniques et procédures à utiliser
  - Conception de la solution (système)

Démarche du conseil et de l'accompagnement

- Mise en œuvre et production
  - Réingénierie de l'application (adaptation)
  - Prototype
  - Tests de montée en puissance progressive
  - Validation et optimisation
  - Adaptation aux outils et services distribués (environnement)
  - Plan de mise en production
  - Production

Démarche du conseil et de l'accompagnement

- Résultats finaux
  - Post-production
  - Extraction des résultats
  - Validation
  - Gestion des données produites
  - Publication ou exploitation

TIDRA

- Bilan 2010

Utilisateurs

- 6 groupes (unités ou laboratoires) en production régulière
- 4 domaines différents
  - Astronomie (1)
  - Aéronautique et Transports (1)
  - Santé (2)
  - Sciences de la Vie (2)
- 2 nouveaux groupes
  - Médecine vétérinaire (1) {avant fin 2010}
  - Biologie humaine (1) {premier trimestre 2011}
- Description sur www.tidra.org

Gestion des Données

- 5 groupes (unités ou laboratoires)
- 10 projets actifs de recherche
  - Nationaux et internationaux
- **17,4** To en ligne pour **8,2** millions de fichiers
  - 80 % domaine de la santé
    (imagerie médicale)
  - 14 % domaine de la biologie
    (séquence des gènes)
  - Gestion de métadonnées, anonymat, accès sécurisé, sauvegarde,…

Gestion des Données

- Deux premières mondiales
  - Premier projet en production
  - Première intégration en production avec la grille de calcul (gLite)
- www.irods.org

Calcul – Utilisation CPU

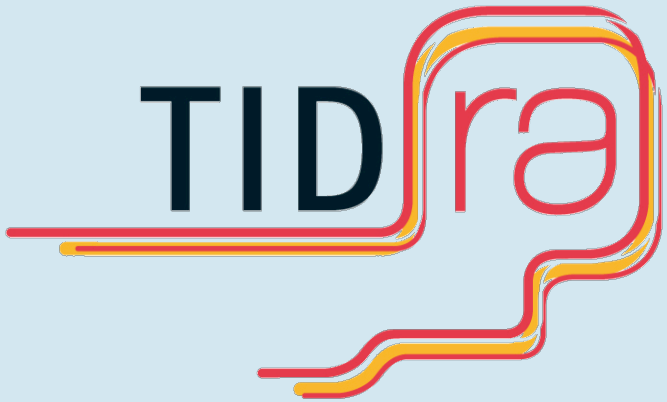
- 6 groupes (unités ou laboratoires)
- **13,5** millions d'heures HS06
- **0,58** millions de jobs

  - 74 % domaine de la biologie
  - (biochimie intracellulaire, phylogénie,...)
  - 19,4 % domaine de l'aéronautique et transports
  - (écoulements tournants)
  - 15,8% en grille VO (vo.rhone-alpes.idgrilles.fr)
  - 191 ans d'un ordinateur Intel Xeon 2.50 GHz

Résumé

- Adressé aux projets scientifiques et industriels en Rhône-Alpes
  - Recherche et innovation
- Démarche du conseil et de l'accompagnement
  - Expression des besoins
  - Résultats finaux
- Service de production
  - Approche pragmatique
- Gestion intégrale de données

Merci de votre attention



De l'expression des besoins à la mise en production et aux résultats finaux

Yonny CARDENAS

Traitement de données et Informatique Distribuée en Rhône-Alpes