

L 'informatique dans BaBar

- L 'environnement de l 'expérience
- Les choix
- ... et BaBar-France
- ... et le cc-in2p3

L 'environnement de BaBar

■ BaBar: expérience près de l '« Usine à B's » PEP-II de slac

luminosité nominale $3 \cdot 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ mai 1999 → ...

- très important flux de données

- taux trigger $100 \text{ Hz} \times 30$ semaines de prises de données/an
- $2 \cdot 10^8$ événements/an nominal (données réelles) ($15 \text{ M } B^0 \bar{B}^0$ /an)
- 350 TB/an à transférer au cc-in2p3 tous processings
- 100 TB/an à stocker dernier processing (data + MC)
- 1999 ~ 1/20 année nominale

- collaboration très large et dispersée

- USA-Canada-Europe 100 instituts 500 physiciens
- 6 pays européens (France, Italie, GB, Allemagne, Russie, Norvège)

Les choix de BaBar (1)

- Très important flux de données

→ stratégies de production des données

- données réelles : tout à slac
 - « prompt reconstruction » en ligne
 - actualisation en ligne des calibrations
 - filtrage en ligne
 - reprocessing minimal (à slac)
- données simulées : production répartie sur collaboration puis centralisées à slac
 - uniformisation ?
 - contrôle de qualité ?

Les choix de BaBar (2)

- Très important flux de données

→ stratégies de stockage des données

- **stockage dans une base de données**
 - utilisation d'une base de données OO Objectivity
 - structuration/éclatement des données (« clustering »)
 - accès hiérarchisé aux données par degré de processing (TAG, ..., RAW) ⊗ canal de physique
- **utilisation d'un système de gestion de stockage de masse hiérarchisé HPSS**
 - accès « transparent » aux données quelque soit le support
 - staging automatique et intelligent des cassettes
 - unique interface utilisateur = Objectivity

Les choix de BaBar (3)

■ Collaboration diverse et dispersée

→ standards

- choix d'un langage unique Orienté-Objet
 - C++ (compilateurs natifs)
 - mais ~pas de design OO !
- utilisation de logiciels commerciaux
 - RogueWave
 - Objectivity
 - Insure, ...
- minimum de plateformes à maintenir
 - AIX († juillet 98), HP-UX († février 99), OSF (?), **SUN** (+Linux ??)
 - pbs logiciels commerciaux ⊗ versions ⊗ plateformes ⊗
compilateurs (eg., Objectivity)

Les choix de BaBar (4)

■ Collaboration diverse et dispersée

mais données à slac →

- **stratégies de distribution des données**

- ▶ **structuration/éclatement des données (« clustering »)**

- transferts minimaux (→ GB, Italie, Univ. US...)

- ▶ **exception : cc-in2p3 « site miroir » de slac (MOU 1998) :**

- accès local à toutes les données de physique (tous degrés de processing RAW → TAG, tous canaux)

- ▶ → **utilisation *non-standard* d'Objectivity (import/export massifs et privés)**

Les choix de BaBar (5)

- Collaboration diverse et dispersée
mais repositories à slac →
 - **stratégies de distribution du code**
 - **CVS** (packages/versions, package coordinators, commits publics)
 - **SRT** (notion de releases et de « builds », outils et makefiles standard)
 - « releases » fréquentes (1/quinzaine)
 - ensemble *complet* des packages (sources+lib+dependances) +bin
→ 2.5 GB/release (Sun)
 - **mises à jour bases de données**
 - «condition database» (calibrations, alignements)
 - très important en 1999 (démarrage de l'expérience)
 - **stratégies de communications**
 - hypernews (centralisées à slac)
 - www (centralisé à slac)
 - → importance cruciale des réseaux

... et BaBar-France

■ 50 physiciens

- prog. centraux de physique (Geant4, B tagging, event tag,...) & analyse
- détecteurs (simulation/reconstruction/analyse)
 - _ DIRC, chambre à dérive
- « commissioning » machine (background)

■ Ingénieurs informaticiens → rôle majeur

- graphique (event display)
- distribution des données (Objectivity import/export)
- accès des données à distance
 - _ event servers / event browsers → display, debugging

... et le cc-in2p3 (1)

■ CPU

- analyse données réelles + MC 5 M h/an nom.
- production de MC 10% total BaBar 2.5M h/an nom.
- capacités actuelles :
 - abandon par la collaboration d 'AIX et HP-UX → SUN (OSF)
 - portage vers linux ~ an 2000 ?
 - → installation de nouvelles machines SUN
 - _ analyse 1999 → saturation des SUN existantes
 - _ MC → capacité actuelle ~0
- → développement de la ferme SUN ?

... et le cc-in2p3 (2)

■ Stockage des données

- 100 TB/an nominal 5TB en 1999
 - server Objectivity dédié
 - 15% (~TAG+AOD+ESD) sur disque
 - 85% (REC+RAW) HPSS

■ Réseaux → USA

- actuellement : 2Mbit/s
 - liaison via le CERN (« consortium »)
- avenir immédiat : 4Mbits/s
 - liaison à Internet2 (STARTAP), via le consortium
- ligne dédiée cc-in2p3-slac très haut débit ?

... et le cc-in2p3 (3)

■ Ressources humaines

- **installation et maintenance de matériels**
 - robot REDWOOD
 - nouvelle plateforme SUN
 - server Objectivity + disques
 - HPSS
- **installation et maintenance de logiciels**
 - gestion de licences commerciales
 - installations d 'OS conformes à slac (versions+patches)
 - installations et maintenance de librairies spécifiques
 - installations et maintenance de produits commerciaux (RogueWave, Objectivity, HPSS) versions conformes à slac

... conclusions ...

- PEP-II et BaBar ~prêts à démarrer (mai 1999)
 - « commissioning » des faisceaux
 - Luminosité atteinte (février 1999) ~ 1/6 nominal ($5 \cdot 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$)
 - Runs cosmiques
 - → tests du détecteur complet
- → et le software ?
- Environnement très contraignant
 - → choix nouveaux
 - stockage des données
 - langage OO
 - standardisation et utilisation de produits commerciaux

... conclusions ...

■ Choix ambitieux → questions

• évaluations préalables ?

» avantage réel des produits commerciaux ?

- eg., RogueWave / STL / CLHEP
- eg., Objectivity :
 - adaptation aux besoins (beaucoup de code BaBar) ?
 - difficultés de mise en œuvre (problèmes de « locks » et collisions, de capacité du server, de « clustering »,...)

» CPU .vs. C++ (optimisation ?)

• formation (C++) ?

- » → perte de savoir-faire et expérience du Fortran
- » inutilisation de méthodologie OO → duplications
- » concentration dans les mains d'experts (même l'analyse)

... conclusions

■ BaBar-France et cc-in2p3

- rôle particulier (privilegié) concrétisé par le MOU
→ cc-in2p3 site miroir de slac
- nécessité de ressources très importantes et nouvelles
 - capacités de stockage servers, disques et HPSS
 - CPU sur SUN
 - liaisons transatlantiques
- étroite interaction physiciens/ingénieurs/cc-in2p3
- → expertise, investissements à long terme