

## Evolution des outils de traitement et de synchronisation des données et audit de la chaîne de calcul de l'IPSL

- *Support sur les outils élémentaires de traitement des données*
- *Synchronisation des fichiers d'entrée des modèles entre les centres de calcul utilisés par l'IPSL*
- *Audit et évaluation de la chaîne de calcul de l'IPSL*

## LOT 1 : ce qui a été fait

Lot 1 : Support sur les outils élémentaires de traitement des données

### Tâches – Etapes :

1. Prise en main de l'outil et des données sur le calculateur titane
2. Portage de l'outil sur Curie - passage en configuration réaliste
3. Validation/amélioration/tests de la robustesse
  - Gestion des pannes, reprises, gestion des erreurs
  - Commande d'information sur l'avancement (création de listes + traitement des données)
  - Outil de check données «avant» - «après» (fichiers netcdf)
4. Tests/amélioration des performances
  - Implémentation d'un parallélisme : soumission de plusieurs tâches en parallèles (utilisation d'un outil TGCC)
5. Portage IDRIS (version séquentielle de l'outil)
6. Documentation :  
[http://forge.ipsl.jussieu.fr/igcmg/attachment/wiki/BasculeCCRTTGCC/Prestation\\_CCRT\\_TGCC/Doc\\_outil\\_pack\\_IPSL\\_v2.pdf](http://forge.ipsl.jussieu.fr/igcmg/attachment/wiki/BasculeCCRTTGCC/Prestation_CCRT_TGCC/Doc_outil_pack_IPSL_v2.pdf)
7. Aide au démarrage de la migration (lien entre le TGCC et l'IPSL) (?)

## LOT 3 : Audit et évaluation de la chaîne de calcul de l'IPSL

Objectif : Evaluer la chaîne de production

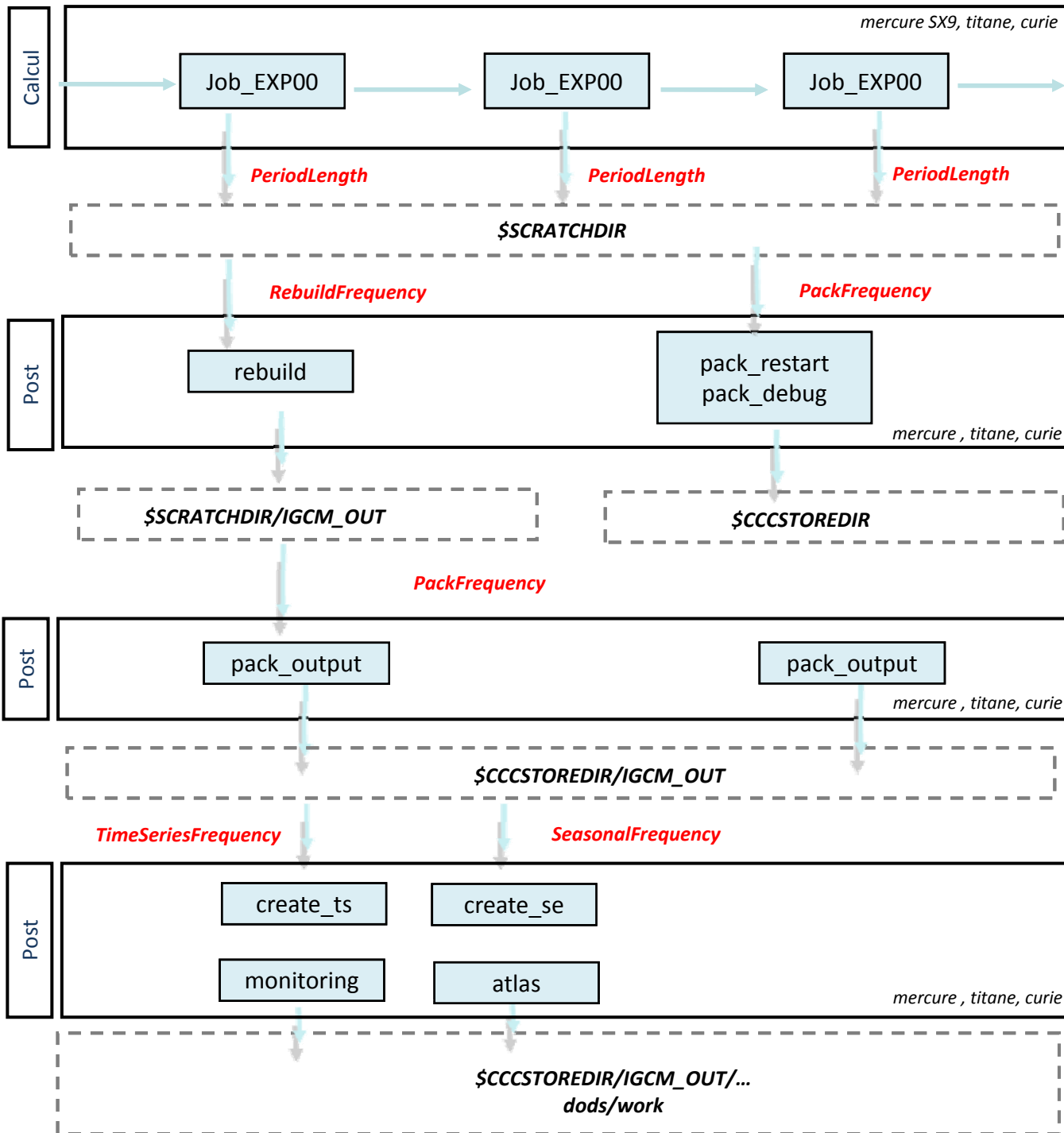
- sur les machines du CCRT/TGCC : mercure (calcul + post) et curie (calcul + post)
- sur les machines de l'IDRIS : vargas (calcul) et ulam (post).

Critères d'évaluation

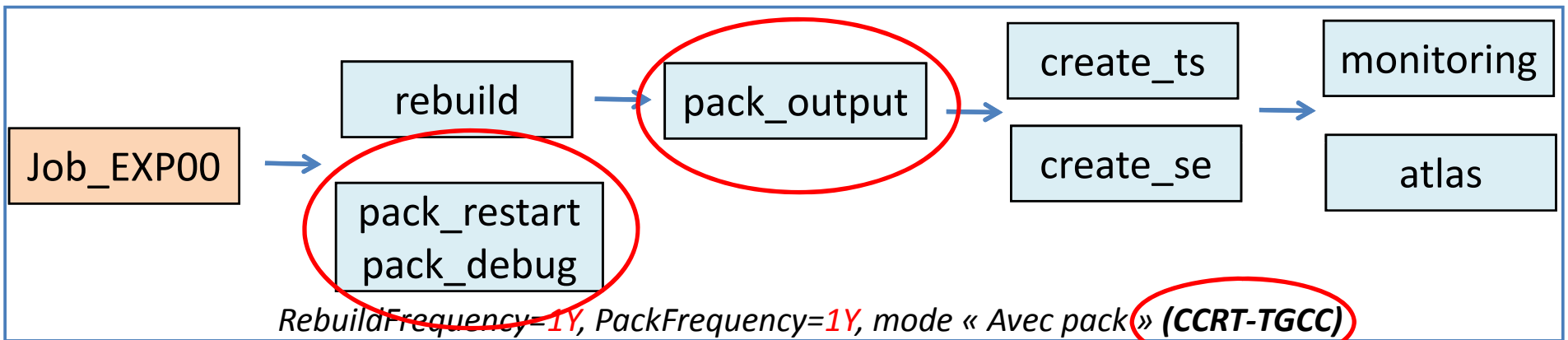
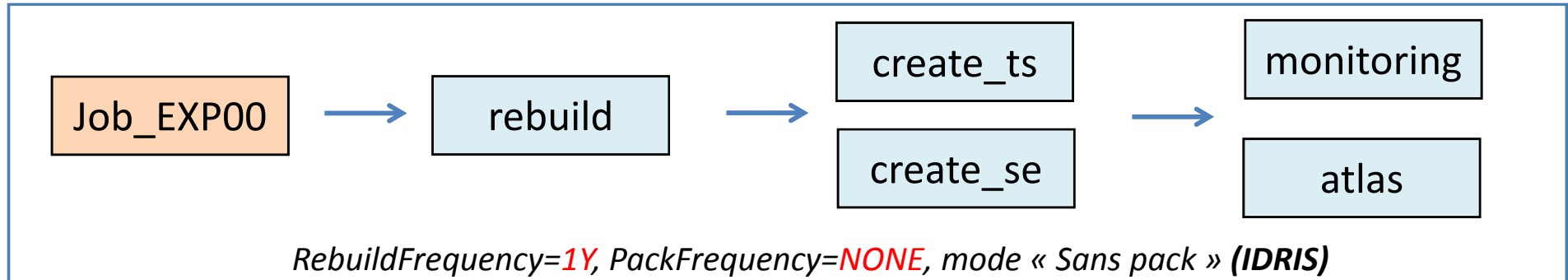
- fiabilité
- efficacité
- (facilité d'utilisation)

Ce qu'on attend : un document/rapport

- décrivant et analysant, suivant ces critères-là, la chaîne de calcul de l'IPSL.
- destiné à une diffusion interne IPSL et externe (centres de calcul)
- base de dialogue entre l'IPSL et les centres de calcul.



# Simulations avec libIGCM : IDRIS et TGCC



# LOT 3 : Audit et évaluation de la chaîne de calcul de l'IPSL

## Rapport (idée de plan)

1. **Description** : description globale de la chaîne de calcul de l'IPSL. Pas de détail mais introduction des analyses à venir.
2. **Analyse** : analyse de la chaîne de calcul.
  - critère de fiabilité : déterminer la capacité de la chaîne, sur une machine particulière, à permettre de réaliser une simulation sans avoir besoin d'intervention humaine. Estimation des causes de plantage et temps humain passé à résoudre.
  - critère d'efficacité : évaluer l'adéquation chaîne de calcul/centre de calcul, mettre en avant les mauvais points en termes de performances.
    - 1<sup>er</sup> niveau : enchaînement des jobs, temps d'attente, temps d'exécution
    - 2<sup>ème</sup> niveau : job : part des tâches au sein d'un job (copie fichiers, execution de la commande, choix filesystems, performances filesystems (débit théorique vs débit réel))
3. **Conclusions** :
  - adéquation chaîne de calcul/centre de calcul ET centre de calcul/chaîne de calcul : conseils d'amélioration.
  - idées/conseils pour instrumenter la chaîne afin d'affiner et faciliter les analyses ultérieures de fiabilité.
  - mise en évidence du temps humain/temps machine.

## LOT 3 : Audit et évaluation de la chaîne de calcul de l'IPSL

### Tâches à réaliser

1. prise en main de la chaîne de calcul et formation à son utilisation (fait)
2. lancement et analyse d'une simulation type de 10 ans sur les calculateurs Curie (TGCC) et Vargas-Ulam (IDRIS) (fait)
3. choix de l'information utile à extraire d'une simulation (en cours)
4. développement de scripts d'extraction de cette information (en cours) :
  - temps de calcul, temps de restitution, temps d'attente d'un job,...) à partir des scripts de sortie d'une simulation donnée.
  - part/performances des tâches au sein d'un job
5. analyse de la chaîne de calcul suivant les critères d'étude choisis.
6. écriture du rapport

### Données à disposition

- simulations "type" de l'IPSL ayant déjà tourné sur les machines du CCRT/TGCC et l'IDRIS.
  - simulation avec calcul et post-traitement sur Curie (TGCC)
  - simulation avec calcul sur Vargas(IDRIS) et post-traitement sur Ulam(IDRIS)
  - simulation avec calcul et post-traitement sur Mercure (CCRT)
- simulations à lancer avec « nouveaux scripts » pour profiling au sein d'un job