

Working Group 6 – Methodology
Completing challenging experiments on Grid'5000
Ou: *Faire de la meilleure science sur Grid'5000*

Lucas Nussbaum (ALGORILLE) – Olivier Richard (MESCAL)

Deux axes de travail

- ▶ Méthodologie de l'expérimentation
- ▶ Outils pour l'expérimentation

Méthodologie de l'expérimentation

Travail sur la description des expériences et des campagnes :

- ▶ Scénarios, plans d'expériences
- ▶ Conditions expérimentales (souhaitées, rencontrées)
- ▶ Métriques
- ▶ Cahier de laboratoire

Méthodologie de l'expérimentation

Travail sur la description des expériences et des campagnes :

- ▶ Scénarios, plans d'expériences
- ▶ Conditions expérimentales (souhaitées, rencontrées)
- ▶ Métriques
- ▶ Cahier de laboratoire

En collaboration avec la communauté **SimGrid** :

- ▶ Besoins similaires
- ▶ Cadre mieux défini

Méthodologie de l'expérimentation

Travail sur la description des expériences et des campagnes :

- ▶ Scénarios, plans d'expériences
- ▶ Conditions expérimentales (souhaitées, rencontrées)
- ▶ Métriques
- ▶ Cahier de laboratoire

En collaboration avec la communauté **SimGrid** :

- ▶ Besoins similaires
- ▶ Cadre mieux défini

Objectifs à long terme :

- ▶ Un ou des formalismes de description d'expérience assez riche pour servir d'entrée à des outils de contrôle d'expérience
- ▶ Diffusion des bonnes pratiques au sein de la communauté

Outils pour l'expérimentation

Objectif : faire de meilleures expériences

- ▶ Augmenter la confiance et la maîtrise
contrôler les causes d'erreurs ou de biais expérimentaux
- ▶ Augmenter la reproductibilité
- ▶ Se placer dans les conditions expérimentales souhaitées
- ▶ Être plus efficace (expériences à grande échelle)

Outils pour l'expérimentation

Objectif : faire de meilleures expériences

- ▶ Augmenter la confiance et la maîtrise
contrôler les causes d'erreurs ou de biais expérimentaux
- ▶ Augmenter la reproductibilité
- ▶ Se placer dans les conditions expérimentales souhaitées
- ▶ Être plus efficace (expériences à grande échelle)

Besoin d'outils pour :

- ▶ Contrôler efficacement les ressources
 - ▶ Lanceurs parallèles
 - ▶ Gestion des données (diffusion, collecte, ...)
 - ▶ Instrumentation et monitoring
(entre autres métrologie réseau – WG Metrology)
- ▶ Se placer dans les conditions expérimentales souhaitées
 - ▶ Etalonnage
 - ▶ Émulation
 - ▶ Injection de charge et de fautes

Orchestration d'expériences : approches

- ▶ Scripts ad-hoc
- ▶ Fichier de description / configuration hiérarchisé [Plush]
- ▶ Visual Programming / Workflows [LabVIEW, Taverna, Kepler/Ptolemy II, Lily, TUNe]
- ▶ **Langage de haut niveau (DSL) pour l'orchestration d'expérience**
 - ▶ Abstractions : gestion de groupes de nœuds et de configurations
 - ▶ Interaction avec des outils spécialisés existants ou à développer
lanceur parallèle TakTuk, émulateur : Wrekavoc

Première étape : collecte de *use cases* (bibliothèque d'expériences)

Orchestration d'expériences : approches

- ▶ Scripts ad-hoc
- ▶ Fichier de description / configuration hiérarchisé [Plush]
- ▶ Visual Programming / Workflows [LabVIEW, Taverna, Kepler/Ptolemy II, Lily, TUNe]
- ▶ **Langage de haut niveau (DSL) pour l'orchestration d'expérience**
 - ▶ Abstractions : gestion de groupes de nœuds et de configurations
 - ▶ Interaction avec des outils spécialisés existants ou à développer
lanceur parallèle TakTuk, émulateur : Wrekavoc

Première étape : collecte de *use cases* (bibliothèque d'expériences)

Problématiques secondaires :

- ▶ Constructibilité de l'environnement d'expérience
(Capacité à reconstruire les composants de l'expérience de zéro)
- ▶ Archivage des expériences
- ▶ Partage d'expérience et réseaux sociaux [<http://myexperiment.org>]
- ▶ Valorisation dans le processus de relecture (école G5K 2011 ?)

Vers un défi

Reproduire l'expérience de [1] :

Déployer 10 000 000 machines virtuelles sur Grid'5000

- ▶ De manière efficace, reproductible et réutilisable
- ▶ *Stress* de la plateforme : déploiement massif, support de la virtualisation
- ▶ *Stress* des outils : exécution de commandes, diffusion de données

Déjà démarré (projets d'étudiants)

Mais il y a du travail pour un doctorant ou un post-doc !

[1] Ronald G. Minnich, Don W. Rudish. **Ten Million and One Penguins, or, Lessons Learned from booting millions of virtual machines on HPC systems.** In HPCVirt'10 : 4th Workshop on System-level Virtualization for High Performance Computing, 2010.