Consultation sur les problématiques de recherche des Clouds

Membres du groupe de travail

- **INRIA:** Frédéric Desprez, Isabelle Chrisment
- **CNRS:** Michel Daydé, Denis Veynante, Christophe Berthon
- **CEA:** Christophe Calvin (DRF), Marc Duranton (DRT), Ahmed Jerraya (DRT)
- **IMT:** Adrien Lebre (IMT Atlantique), Gaël Thomas (Telecom SudParis/IP Paris)

Définition du terme « Cloud »

Dans les textes gouvernementaux, « *le cloud computing, en français informatique en nuage, correspond à l'accès à des services informatiques distants (serveurs, stockage, mise en réseau, applications) via internet à partir d'un fournisseur* », ce qui constitue la vision utilisateur du Cloud.

La recherche dans le domaine du Cloud s'articule autour de trois axes principaux. Le premier porte son attention sur le déploiement et la gestion de l'infrastructure autrement dit l'ensemble des éléments matériels et logiciels nécessaires à la mise en œuvre le Cloud, ce qui inclut : la puissance de calcul, le réseau, le stockage, ainsi que l'interface d'accès à aux ressources virtualisées des utilisateurs et aux services. C'est une vision orientée « fournisseurs ». Le second concerne les outils, méthodes et modèles de programmation pour gérer l'exécution des applications sur ces infrastructures et plus précisément le cycle de vie de ces dernières. C'est une vision orientée « DevOps ». Le troisième et dernier axe concerne les services de haut niveau proposés aux utilisateurs finaux. Il inclut par exemple les problèmes liés à la confidentialité et la protection de données. C'est la vision « usager ». A ces trois axes viennent s'ajouter des dimensions transverses comme la sécurité ou encore l'énergie.

Synthèse et recommandations

Nous assistons depuis plusieurs années maintenant à la généralisation irréversible d'une « cloudification » de l'ensemble des ressources des organisations quel que soit leur type.

Même si les travaux de recherche ciblant les infrastructures Cloud sont encore vivaces (énergie, protection/confidentialité des données, enclaves sécurisées, etc.), l'attention se porte aujourd'hui sur une nouvelle (r)évolution autour de la convergence entre les datacentres, le réseau et l'Internet des objets. Cette transition digitale nécessite le déploiement de nouvelles infrastructures informatiques plus décentralisées permettant de répondre aux besoins des nombreuses applications envisagées pour améliorer notre quotidien (e-santé, ville intelligente, véhicules autonomes, industrie du futur, etc.). Bien que nettement plus complexes en termes de gestion ou encore d'utilisation, la mise en œuvre de ces nouvelles infrastructures est une opportunité majeure pour la reprise en main d'un marché dominé par les GAFAM (« les *hyperscalers* ») mettant en péril la souveraineté de l'Europe. Soutenir le secteur académique et industriel dans cette transformation de l'Internet actuel vers un « nouveau modèle » plus décentralisé, et plus proche des préoccupations autour de la régulation des

données en France et en Europe, est donc un enjeu décisif pour notre pays et justifie de soutenir la recherche Cloud sous-jacente dans son ensemble.

La communauté de recherche autour du Cloud analysée dans cette note est principalement issue des organismes de recherche nationaux (CNRS, INRIA et CEA) et des principales écoles/universités. Dans le détail, on estime à 500 le nombre de Chercheurs et Enseignants-Chercheurs impliqués dans les recherches décrites ici (environ 50% du GDR RSD). Auxquels on peut ajouter les chercheurs sur des thèmes reliés (réseau, sécurité, calcul à haute performance, IA, Big Data, etc.) et qui adressent également des problématiques ayant un impact sur les Clouds d'aujourd'hui et de demain.

Soutenir une recherche souveraine autour du Cloud peut se traduire par diverses actions emblématiques avec des appels à projets et actions spécifiques autour de thématiques bien identifiées :

- Favoriser l'essor de nouveaux modèles de programmation et de solutions distribuées pour le calcul et le stockage, liées aux domaines émergents construits sur le continuum Cloud IoT/CPS[1] (et repris sous le terme Edge Computing dans la suite du document). Cela inclut des aspects comme l'apprentissage fédéré, les micro-services géo-distribués, *Function as a Service*, etc. ;
- Développement de piles logicielles open-source pour la gestion et l'orchestration des infrastructures et des applications dans un contexte Edge. Cela inclut des problématiques liées aux aspects « système », à l'interopérabilité et également à des dimensions transverses telles que la sécurité, la protection des données personnelles ou encore l'énergie;
- Support et développement de solutions matérielles et de circuits intégrés supportant les dernières innovations techniques et logicielles (co-conception), afin de renforcer la souveraineté et la confiance ;
- Mise en place d'un nouveau type de programme visant à regrouper partenaires académiques et industriels autour du développement de piles open-source mentionnées préalablement avec un objectif de leadership pour la France ;
- Fédération et mise en visibilité de la recherche française sur les sujets liés au Cloud et à son évolution, par un moyen à définir dans le prolongement du mode de travail collectif qu'installera le PEPR, lien organisé avec l'industrie (par exemple par une structure trans-organismes ou un « campus », similaire à l'initiative en cybersécurité) ;
- Donner un support pour la participation de chercheurs français à des initiatives européennes et internationales ;
- Renforcement des formations (initiale, continue et professionnelle) autour du Cloud, avec là encore un rapprochement des grandes écoles/universités et de l'industrie ;
- Mécénat recherche/industrie pour favoriser le co-working entre chercheurs académiques et chercheurs de l'industrie, notamment afin de renforcer la participation à l'élaboration de standards internationaux et l'impact sociaux-économiques de la recherche Cloud ;
- Soutien à une infrastructure de recherche expérimentale permettant la validation des solutions logicielles en vue de son utilisation en production dans la recherche et l'industrie.

[1] Système cyber physique (*Cyber Physical System* en anglais).

Une nouvelle (r)évolution: un continuum des Clouds aux datacentres à la périphérie du réseau et à l'IoT

La transformation numérique est au cœur de notre société quelles que soient les activités envisagées. Elle change notre façon de penser le monde, notre façon de travailler, notre façon d'apprendre, notre façon de jouer, etc. Cette (r)évolution est rendue possible par le déploiement d'infrastructures de réseaux et de services (« *Everything as a Service* ») de plus en plus complexes, amplifiée par l'avènement des systèmes cyberphysiques (qui interagissent directement avec le monde physique) et des appareils IoT montrant les limitations du modèle de Cloud Computing « centralisé ». Les besoins émergents en termes de volumétrie de données, de réactivité ou encore de sûreté ont amené nos communautés à réfléchir à un nouveau type de paradigme où les ressources de calculs et de stockage ne sont plus uniquement consolidées de manière abondante dans quelques datacentres en nombre limité mais également réparties sur une multitude de sites proches des usagers finaux.

Connu aujourd'hui sous les termes « *Edge Computing* » (mais aussi avec des variantes sous les termes « swam », « Fog », « federed » , etc), **ce nouveau paradigme d'informatique à la demande consiste à maintenir un continuum entre les Clouds (hébergés dans des datacentres)** et les périphériques IoT grâce à une fédération massivement géo-distribuée de **petits datacentres placés en bordure du réseau**. Toutefois, la gestion de ces plateformes, l'orchestration des services proposés, intégrant des capteurs/actuateurs commandés, et leur utilisation dans des systèmes complexes de contrôle soulèvent de nombreux challenges scientifiques et techniques pour avoir des systèmes distribués à grande échelle **maintenables, efficaces, simples d'utilisation, sécurisés et ayant un impact énergétique minimal**.

De manière plus précise, si le *Edge Computing* comporte des caractéristiques similaires au modèle du Cloud (comme la mise à disposition de ressources de calcul, stockage et réseau à la demande) et repose sur les infrastructures de communication, sa complexité est considérablement accrue :

- les ressources sont distribuées au travers de plusieurs sites géographiques via des liens longues distances ;
- les ressources sont plus hétérogènes, en intégrant notamment du matériel spécialisé tel que des plateformes GPU / FPGA / NPU et autres accélérateurs pour des applications comme l'intelligence artificielle ou la réalité augmentée ;
- l'infrastructure dans sa globalité est plus dynamique : l'apparition/disparition d'une ressource/d'un service/d'un site devient la norme, certaines ressources sont mobiles, etc. ;
- Les sites sont dépourvus d'équipe technique en majorité et doivent donc être administrés à distance ;
- la localisation géographique devient un élément décisif pour les applications (placement, temps de réponse, consolidation des jeux de données) ;
- la fédération, ou orchestration facile des ressources à tous niveaux afin d'offrir les services/applications aux utilisateurs en tenant compte de leurs contraintes (sécurité des données, latence, coût, consommation, etc.) ;
- etc.

Malheureusement, les piles logicielles développées pour les infrastructures Cloud actuelles ne permettent pas de prendre en compte ces changements et ce pour la quasi-totalité des couches (aussi bien du point de vue opérateurs que de celui des utilisateurs). En effet, les systèmes existants ont été développés et conçus pour des infrastructures « centralisées » avec peu d'incertitudes sur les ressources et donc un grand contrôle par les fournisseurs de service. Par ailleurs, des dimensions

transverses comme la sécurité ou l'énergie doivent être également renforcées. Du point de vue de la sécurité, les mécanismes classiques doivent être adaptés pour prendre en compte l'augmentation de la surface d'attaques induite par le continuum Cloud-Edge-IoT (plus de données, plus de datacentres hétérogènes, plus d'appareils connectés, une mobilité des utilisateurs et des ressources, une hétérogénéité des réseaux d'accès, etc.). Du côté de l'énergie, les infrastructures Edge offrent plus d'opportunités et de levier d'actions (énergies renouvelables, réutilisation de l'énergie dissipée par le refroidissement, réduction des mouvements de données, etc.) que les grosses infrastructures centralisées et nécessitent du matériel qui doit offrir des nouvelles fonctionnalités et d'autres performances que celui utilisé par le Cloud centralisé.

La révolution actuelle du Cloud ralentit les GAFAM dans leur progression. Il faut en effet développer des solutions spécifiques pour les différents secteurs industriels pour réaliser la continuité des traitements de données de l'objet au Cloud tout en assurant le meilleur compromis possible entre les principales contraintes : temps-réel, performance et garantie de confiance (y compris sécurité et protection de la vie privée) sur l'ensemble de la chaîne de traitement ce qui crée une opportunité pour l'Europe et la France. Cette maîtrise est clé pour plusieurs marchés industriels stratégiques industriel : automobile, aéronautique, IT, etc. La maîtrise de la pile logicielle et du matériel (reposant sur un savoir-faire reconnu en systèmes embarqués à hautes performances) permettra à l'Europe et à la France non seulement de rattraper son retard face aux USA et à la Chine, mais aussi de prendre une place de leader dans les secteurs du Cloud du futur en considérant le continuum Cloud IoT et en ajoutant au-dessus de l'existant les éléments requis pour supporter ce « nouveau web ».

Dans la suite du document, nous avons choisi de diviser **les pistes de recherche** en cinq grandes catégories : **gestion du cycle de vie des applications, gestion des infrastructures, matériel, sécurité et énergie**. Deux dernières **actions transverses** concernent **la disponibilité requise** d**'infrastructures expérimentales** et les Clouds pour la recherche.

Volontairement, nous n'avons pas mentionné les orientations de recherche combinant réseaux logiciels (SDN), virtualisation des fonctions réseaux (NFV) et Clouds. Elles ont été décrites dans le rapport sur la stratégie d'accélération sur la 5G et les réseaux du futurs. Toutefois, nous tenons à préciser ici que ces sujets sont étroitement liés puisque la séparation historique entre les communautés des systèmes distribués et du réseau s'est fortement atténuée depuis les cinq dernières années.

Thématiques de recherche

Gestion du cycle de vie des applications (déploiement initial, configuration, reconfiguration, maintenance)

Les applications réparties sont encore majoritairement décrites en termes de processus ou de machines (virtuelles ou non). Ainsi, la gestion de leur cycle de vie se fait à un faible niveau d'abstraction. Cela conduit à un manque d'optimisation car la gestion est réalisée par les applications elles-mêmes qui souvent se contentent de solutions simples par manque d'expertise. Les mécanismes de gestion de l'élasticité dans les Clouds sont un exemple qui démontre comment un plus haut niveau d'abstraction (par exemple en fournissant des règles d'ajout et/ou de suppression de VM) simplifie la gestion de ressources tout en permettant de mieux optimiser leur usage.

Le défi suivant consiste à abstraire la description de toute la structure de l'application afin de pouvoir optimiser globalement les ressources utilisées vis à vis d'objectifs multicritères (prix, deadline,

performance, énergie, etc.). Devant la complexité du choix, il apparaît ainsi important de découpler au maximum la description de la structure de l'application de l'infrastructure afin de pouvoir utiliser des services externes pour adapter l'application. Cela offre également un cadre pour aborder les défis liés à l'abstraction de la reconfiguration des applications afin d'adapter automatiquement l'usage des ressources. Cela demande de définir des modèles et des langages associés pour décrire les applications, leurs fonctions objectifs, les algorithmes de placement et d'ordonnancement supportant des critères au niveau système et applicatif, etc. L'encapsulation des applications et de leur environnement dans des containers légers et mobiles permet d'offrir une plus grande possibilité de migration, et permet de faciliter l'utilisation de la distribution des ressources. En plus d'être performants, de tels systèmes automatiques se doivent également d'être sûrs et frugaux, notamment pour gérer des défaillances.

Gestion des infrastructures

La gestion des infrastructures passe par trois axes de recherche majeurs autour de la virtualisation, du stockage et de l'administration.

Virtualisation. Pour faciliter l'administration et l'exploitation des fermes de serveurs, les opérateurs Clouds se tournent vers des architectures dites *hyper-convergées 2.0*. Une architecture hyper-convergée 2.0 consiste à concevoir un nœud de calcul comme un Lego constitué d'une myriade de composants matériels hétérogènes pouvant être ajoutés ou retirés dynamiquement. Gérer ces architectures arrive avec son lot de nouveaux challenges : il faut gérer la complexité, l'hétérogénéité, la dynamicité, le passage à l'échelle et la localité. Il faut aussi être capable d'agréger les composants matériels en des touts cohérents utilisables par les applications. La vision à terme est un système capable de virtualiser ce matériel complexe et dynamique en machines virtuelles dans lesquelles des logiciels patrimoniaux pourraient s'exécuter sans nécessiter de modifications. Mettre en œuvre cette vision nécessite de revisiter intégralement les couches systèmes du Cloud, et en particulier la couche de virtualisation.

Stockage. Bien que performants et permettant de stocker et d'analyser une grande quantité de données, les systèmes de stockage Cloud déployés en production reposent sur des travaux de recherche de plus de quinze ans. C'est à dire qu'ils n'ont pas été conçus pour appréhender les contraintes propres aux évolutions vers le Edge mais également les avancées technologiques des supports de stockage récents (e.g., mémoires non volatiles type NVRAM). Le déluge de données pressenti avec l'avènement de l'IoT et des CPS nécessitent de proposer des approches nouvelles où les données seront naturellement géo-distribuées (voir géo-répliquées) en utilisant les derniers supports de stockages. Ces évolutions obligent à repenser entièrement l'architecture traditionnelle car déplacer les données entre le tiers de stockage et le tiers de calcul est devenu trop coûteux, voire impossible pour certains domaines d'applications (e.g. médical). Aujourd'hui, il faut trouver des compromis entre déplacer les calculs vs les données. Les modèles et les structures pour stocker et manipuler les données collectées doivent également évolués (fichiers, BLOB, systèmes clé/valeur, base de données graphes géo-distribuée, etc.). Par exemple, quel est le système de stockage nécessaire permettant de faire de l'apprentissage fédéré au travers une multitude de périphériques IoT ?

Administration. Avec l'arrivée du Edge Computing, les plateformes Cloud vont exposer de nombreux nouveaux services obtenant des informations à partir de capteurs IoT et pourront contrôler différents actuateurs. Malheureusement, les techniques d'administration et d'orchestration actuels ne sont pas

préparés à cette évolution. L'orchestrateur va devoir être capable de passer à une échelle géo-distribuée, être plus intelligent, de s'adapter automatiquement aux besoins des utilisateurs, de gérer de nombreux capteurs et actuateurs en temps-réel et de grands volumes de données, d'offrir des interfaces permettant à des non-informaticiens d'administrer la plateforme, de garantir que des contraintes non fonctionnelles sont bien respectées (coût financier, temps de réponse, consommation énergétiques, etc...), de sécuriser de façon autonome le système ou d'assurer que les données privées sont protégées. Gérer un tel niveau de complexité nécessite d'aller bien au-delà des solutions actuelles, lesquelles reposent encore en grande partie sur une intervention humaine.

Matériel

La souveraineté européenne des Clouds passe aussi par une souveraineté sur le matériel et en premier sur les processeurs et les accélérateurs. La loi de Moore s'essouffle, l'efficacité énergétique est de plus en plus un challenge poussant à la créations d'accélérateurs spécifiques et à de nouvelles architectures (calcul en ou proche mémoire par exemple). La volonté économique d'accroître les performances persistera et nous prévoyons une augmentation du nombre d'architectures spécialisées à moyen terme afin d'exploiter les performances de la technologie CMOS et l'usage de nouvelles technologies spécifiques[2]. Les solutions de réduction d'énergie développées pour le Cloud classique ne s'appliquent pas forcément pour l'évolution vers le Cloud distribué ou à l'Edge, et l'industrie européenne et française ont des arguments à faire valoir par leur savoir-faire reconnu dans les systèmes embarqués qui doivent déjà satisfaire des contraintes de faible consommation, de faible latence, de sécurité qui seront clés pour les systèmes du continuum Cloud-Edge-IoT. Il est donc important de continuer les recherches dans ce domaine, y compris sur de nouveaux type d'accélérateurs, et de favoriser le transfert de ces résultats dans des prototypes puis dans des produits industriels.

D'autre-part, les innovations logicielles dans de multiples domaines (hyperviseurs, ségrégation des tâches, orchestration, chiffrement, ...) nécessitent des supports architecturaux pour être efficaces. La capacité de faire de la co-conception et de réaliser des architectures efficaces pour nos innovations sans dépendre de fabricants non-européens est une garantie de souveraineté et augmente les chances d'acceptation globale de nos solutions.

Enfin, il est important d'avoir des solutions matérielles (processeurs, accélérateurs, réseaux) de confiance, et pour cela il faut être capable de maîtriser - au moins[3] - la chaîne de conception du matériel : utiliser des méthodes renforçant des approches « correctes par construction » ou utilisant des preuves formelles, avoir accès à l'architecture fine afin de pouvoir l'auditer (par exemple, en la rendant en « open source » en se basant sur des approches du type « Risc V ») et vérifier l'absence de « cheval de Troie » [4]. Une mobilisation des industriels du domaine sur ces approches, et des collaborations entre recherche et industrie sont nécessaire afin de réaliser ces « architectures de confiance ».

[2] par exemple neuromorphiques ou optiques, comme l'accélérateur de la start-up LightOn intégré par OVH.

[3] La réalisation (fonderie) peut être faite hors Europe pour bénéficier des dernières avancées technologiques, tout en vérifiant que ce qui en sort est bien conforme à ce qui a été fourni.

[4] Voir la partie suivante « Sécurité » de ce document.

Sécurité

Le Cloud, comme tout système d'information demande la mise en place de mécanismes classiques pour assurer les services de confidentialité, d'intégrité et de disponibilité des données, des applications et services. De même, les principes de « *privacy* « et « *security by design* » doivent s'appliquer car la complexité accrue de l'architecture rendrait l'ajout ultérieur de fonctions dédiées à la sécurité quasi-impossible. Cependant, **les mécanismes de sécurité classiques doivent être adaptés aux caractéristiques du Cloud** pour faire face à des menaces spécifiques pour les différentes types de Clouds (SaaS, PaaS, IaaS) comme les failles des VM, des hyperviseurs et orchestrateurs, des technologies de réseaux virtuels (SDN, NFV), des interfaces de programmation ou d'accès :

- Adaptation des approches classiques de prévention à ces menaces spécifiques avec le chiffrement de bout en bout et la sécurité des applications au niveau langage ;
- Adaptation des politiques de sécurité à un environnement plus complexe. De véritables challenges se posent pour assurer la cohérence globale, pour traiter les incohérences/incompatibilités entre tenants ou entre administrateurs/gestionnaires. Une fois, ces politiques définies, une couche d'orchestration doit être capable de les interpréter et d'interagir avec les éléments composant le Cloud afin de les mettre en œuvre ;
- Adaptation des approches classiques de la supervision elle-même avec la prise en compte de la plasticité et de la grande dynamique du système global tout en considérant les besoins spécifiques des différents tenants. Il faut être à même de gérer dynamiquement les ressources allouées à la supervision et de reconfigurer automatiquement les sondes. De nouvelles méthodes devront être proposées pour monitorer à la fois les échanges de données pour la plupart chiffrés à différents niveaux de la pile logicielle et les comportements des « tenants » et/ou des administrateurs du cloud ;
- Adaptation aux contraintes induites par la réglementation des différentes localisations des machines constituant le Cloud. A défaut de rapatrier les données dans une localisation de confiance, il faut être capable de traiter ces données sous forme chiffrée comme le permettent les mécanismes de chiffrement homomorphe et fonctionnel. Du travail de recherche est encore nécessaire pour atteindre l'efficacité requise en termes de temps de traitement, de volume des chiffrés et de gestion des clés.

Les **caractéristiques des Cloud peuvent aussi être utilisées en elles-mêmes pour assurer la sécurité**. Ainsi l'intégrité des données peut résulter de leur réplication en plusieurs localisations. De plus, la protection des données personnelles (confidentialité) peut résulter de leur localisation en certains endroits, voire chez les personnes elles-mêmes (*personal Cloud*) qui retrouvent ainsi la souveraineté sur leurs données. Dès lors, des travaux d'algorithmique distribuée sont nécessaires pour permettre l'exploitation de ces données. Il faut aussi noter que la sécurité des nœuds du *personal Cloud* est à assurer et nécessite d'être automatisée pour être le plus transparent possible pour les utilisateurs.

Renforcer la protection des données passe aussi par l'innovation dans les systèmes, les langages et la matériel. Depuis quelques années, les fabricants de processeurs proposent de nouveaux jeux d'instructions permettant de chiffrer à la volée nos données (e.g., Intel SGX ou ARM Trustzone ou ARM CCA - Confidential Compute Architecture). Ces jeux d'instructions garantissent que les données

privées ne peuvent être accédées que par les codes déployés par le propriétaire des données[5], ce qui évite que des données privées ne soient exfiltrées par des pirates ou des opérateurs Clouds (et comme les détails fins des processeurs ne sont pas connus, rien ne garantit l'absence de « *back doors* » ou de mécanismes pouvant être exploités avec de mauvaises intentions[6], d'où un intérêt certain pour des systèmes qui peuvent être totalement audités, par exemple grâce à de l'open source ou faits par des constructeurs français/européens). Malheureusement, les applications n'utilisent quasiment pas ces jeux d'instructions car les couches logicielles du Cloud et les langages de programmation sont incapables d'exposer de façon simple ces abstractions. Il convient donc de revisiter les langages de programmation afin d'offrir des abstractions simples pouvant être facilement utilisées par les développeurs d'applications Clouds.

Énergie

La consommation énergétique et les impacts environnementaux du numérique continuent de croître. La virtualisation des infrastructures et des services ne freinent pas cet accroissement malgré l'exploration et la mise en œuvre de solutions améliorant l'efficacité énergétique. Les datacentres et l'informatique virtualisée reposent encore fortement sur des approches surdimensionnées permettant d'obtenir une grande qualité de service et des usages hétérogènes. Par ailleurs, l'arrivée d'infrastructures Edge apporte de nouvelles problématiques énergétiques et implique de nouvelles architectures matérielles plus efficaces énergétiquement. Il est urgent d'aborder de nouveaux défis au sein du Cloud pour orienter les futurs conceptions et déploiements dans une démarche de sobriété numérique :

- Analyse et gestion énergétique de bout en bout d'infrastructures hiérarchiques à grande échelle Cloud/Edge/Fog sur les aspects traitements, réseaux et stockage ;
- Monitoring et profiling des ressources virtuelles (containers logiciels, machines virtuelles) sur des infrastructures hétérogènes (CPU, GPU, ...) ;
- Compromis entre efficacité énergétique et autres métriques de performances dans des infrastructures virtualisées, placement de tâches entre des infrastructures de Cloud, Fog et Edge en fonction des besoins et des sources d'énergie ;
- Éco-conception d'applications et de services numériques : IA frugale, blockchain, streaming ... exposition des caractéristiques énergétiques des différents services afin d'offrir le choix aux orchestrateurs ;
- Prise en compte des effets rebonds lors d'amélioration de l'efficacité des techniques de Cloud.

[5] En supposant que le « superviseur » du système n'y ai pas accès non plus. La solution ultime serait de faire les opérations dans le domaine crypté (*Full Homomorphic Encryption*), mais cela demande actuellement du travail de recherche pour rentrer les solutions FHE moins gourmandes en calcul.

[6] comme le « Intel Managment Engine » qui a accès à toutes les ressources et mémoires d'une machine, même en l'absence d'OS, voir https://en.wikipedia.org/wiki/Intel_Management_Engine ou https://www.howtogeek.com/334013/intel-management-engine-explained-the-tiny-computer-inside-your-cpu/

Infrastructures

Infrastructure pour la recherche sur les Clouds

Le modèle économique qui sous-tend les Clouds publics repose sur la fourniture de services abstraits : calcul, stockages, réseaux, environnements d'exécution, applications. La manière dont ces services sont fournis est tenue secrète, avec dans le meilleur des cas des accords de niveau de service (SLA) pour formaliser la relation entre le consommateur de services et le fournisseur. L'observabilité et le contrôle, nécessaires pour comprendre les performances des applications ou des algorithmes, sont donc très limités dans ces Clouds publics. Par ailleurs, ces infrastructures ne permettent pas des expérimentations de bout en bout en environnement contrôlé, depuis les capteurs/objets connectés par des réseaux sans fils jusqu'au datacentres de grandes tailles et en passant par les micro-datacentres de type Edge.

Il est donc indispensable, pour soutenir la découverte et l'innovation, d'investir dans des plateformes permettant d'adopter des approches fondées sur l'expérimentation, accordant une grande importance à l'observabilité et au contrôle et passant à l'échelle, pour offrir aux universitaires et aux entreprises un large éventail de configurations possibles, allant des datacentres classiques hautement instrumentés et contrôlables à des plateformes hautement innovantes et distribuées. En termes d'observabilité, elles devraient fournir aux expérimentateurs des métriques sur tous les niveaux de l'infrastructure : réseau filaires et sans fils, éléments de stockage et de traitement, hyperviseurs, etc. C'est le but de l'infrastructure SILECS[7] retenue dans la feuille de route 2021 des infrastructures de recherche.

Cloud pour la recherche

L'émergence des Clouds scientifiques basés sur des techniques de virtualisation permet d'envisager de créer des infrastructures informatiques beaucoup plus souples et accessibles à l'ensemble des communautés scientifiques qui répondent à un certain nombre de besoins (e.g. analyse sur les données, Intelligence Artificielle). On assiste donc à une urbanisation des infrastructures de calcul et de données, interfaçant calcul haute performance, Grilles, Cloud et les grands datacentres des Instituts s'insérant dans le contexte national et européen. Toutes les sciences sont maintenant concernées par ces plateformes à haute disponibilité (mathématiques, physique, sciences de l'univers, sciences humaines et sociales, biologie et bio-informatiques, ...). Une infrastructure européenne comme GAIA-X[8] pourrait offrir aux chercheurs et industriels des services pour leurs recherches de manière souveraine.

Les avancées de recherches présentées dans les sections précédentes et les logiciels développés et validés sur une plateforme expérimentale comme SILECS pourront directement bénéficier à ces utilisateurs finaux. Par ailleurs, étant donné les problèmes d'accès à des instruments répartis sur le territoire, les futures infrastructures de type Edge seront en première ligne pour offrir aux scientifiques d'autres disciplines les meilleures plateformes pour leurs recherches. Réciproquement, ces mêmes utilisateurs finaux, notamment au travers de leurs usages avancés, vont apporter de nouvelles problématiques que la communauté en recherche Cloud pourra alors aborder. En d'autres termes, les deux communautés ont à gagner à travailler ensemble.

[7] https://www.silecs.net/

[8] https://data-infrastructure.eu/