



Désassemblons le numérique

#Episode9 : Bientôt des supercalculateurs dans nos piscines ?

Bonjour à toutes et à tous et bienvenue dans ce nouvel épisode de *Désassemblons le numérique*. Aujourd'hui, nous accueillons Oliver Beaumont, directeur de recherche au sein de l'équipe-projet *Topal* du Centre Inria de l'université de Bordeaux, et Rémi Bouzin, Directeur scientifique de l'entreprise Qarnot avec qui nous allons aborder des questions de chaudières numériques. Bonjour Olivier, bonjour Rémi.

Olivier Beaumont (OB) – Rémi Bouzel (RB) : Bonjour.

Lors d'un article publié récemment sur inria.fr, nous avons pu voir que chez Inria, les scientifiques travaillent sur de nombreuses thématiques pour aller vers un avenir énergétique plus durable. Et c'est notamment dans ce contexte là qu'est née la collaboration entre Topal et Qarnot. Est-ce que vous pourriez nous expliquer un peu la genèse de cette collaboration ?

RB Alors je vais commencer avec un petit mot sur la collaboration avec Inria un peu plus générale. En fait, on était en discussion avec la direction générale de la direction scientifique de l'Inria, notamment Messieurs Gerbaud et Sainte-Marie, et on avait une volonté un peu commune de collaborer, mais on ne savait pas forcément sous quelle forme à l'époque. Donc il y avait plusieurs pistes dont celle d'un défi. Un défi, traditionnellement, c'est un projet qui est porté par plusieurs équipes Inria, qui est donc un peu transverse et qui crée du lien entre les équipes. Et puis il y a certains types de défis qui intègrent un industriel qui vient proposer ses problèmes et donc travailler avec le monde académique sur ces problèmes. C'est donc cette piste de collaboration qui a pris le mieux, je dirais. Je crois que Jacques Sainte-Marie a partagé un certain nombre d'équipes qu'il avait en tête pour cette potentielle collaboration et les équipes intéressées se sont manifestées. Et c'est comme ça qu'on a rencontré l'équipe Topal et qu'on a pu approfondir un sujet qui nous intéressait beaucoup côté Qarnot, parce que ça répondait à des problèmes qu'on n'était pas forcément capables de traiter avec nos compétences.

OB Du côté de Topal, la collaboration a commencé en 2022, au moment du montage du projet. Personnellement, ce n'était pas la première fois que j'entendais parler de Qarnot parce que j'avais déjà eu l'occasion d'assister à une soutenance de thèse qui était codirigée par Inria et Qarnot à l'époque. Mais entre les deux, le modèle et l'architecture cible de Qarnot avaient évolué de façon assez sensible. De notre côté, chez Topal, on s'intéresse à des problèmes d'allocation de ressources sur des calculs et d'optimisation. Donc la problématique générale,

c'est qu'on a un ensemble de calculs à réaliser, on a un ensemble de ressources et on essaye de maximiser l'occupation des processeurs ou de minimiser le temps pour réaliser cette tâche-là. Et notre plateforme cible, c'est plutôt un gros supercalculateur. Donc ce qu'on vise, c'est une application complexe, avec beaucoup de tâches, avec beaucoup de dépendances entre les tâches et à exécuter sur une architecture qui est elle-même complexe. Avec Qarnot, la problématique est un petit peu différente parce qu'on a plein de ressources à plus petite échelle et des tâches qui sont essentiellement indépendantes, mais en grande quantité. Et surtout, la grande différence, c'est que la problématique de minimisation du temps reste importante, mais la problématique de minimisation, enfin de contrôle de la consommation énergétique devient centrale. Et c'est ça qui nous intéresse, en particulier dans le défi.

Et du coup, c'était ça la problématique scientifique de départ ?

OB Oui. La problématique scientifique du défi avec Qarnot est plus large que celle-ci, mais du côté de Topal (et de l'équipe-projet Inria Storm), c'était sur ces questions-là qu'on voulait s'intéresser.

Chez Qarnot, au niveau de la problématique scientifique de départ, est-ce que c'était la même ?

RB Comme le disait Olivier, c'était un peu plus large que ça. Nous, ce qu'on voulait, c'était améliorer notre connaissance et notre capacité à diminuer l'empreinte carbone de nos services *cloud* tout en gardant une qualité de service égale. Et donc il y avait plusieurs volets là-dessus qui inclut évidemment ce projet avec Topal et Storm. Le premier axe c'était vraiment sur de l'évaluation et de la mesure. Pour l'évaluation, il faut se projeter. Par exemple, si on veut installer un nouveau site, comment est-ce qu'on peut évaluer avec des ordres de grandeurs et des paramètres la future empreinte carbone de ce site sur l'ensemble de sa durée de vie ? Ça, c'est un projet qu'on porte toujours avec une autre équipe Inria : Spirals. L'idée est vraiment de savoir quelles métriques on peut relever, quels capteurs on peut mettre... Typiquement, sur un ordinateur, on peut avoir plusieurs utilisateurs qui vont calculer en même temps. Comment on arrive à dissocier l'empreinte carbone de chacun ? Donc nous, notre objectif, c'est d'être le plus précis possible. Et ça, c'est un peu le socle indispensable pour pouvoir derrière optimiser. Donc je crois qu'il y a des citations qui sont attribuées à plein de gens différents mais qui disent en gros la même chose, qu'on ne peut pas optimiser ce qu'on n'a pas mesuré au préalable. Et c'est vraiment ça l'idée. Le deuxième axe qui réunissait quatre projets, dont celui avec Topal, c'était sur l'optimisation de l'empreinte environnementale du *cloud*. Mais effectivement, pas de n'importe quel *cloud* : celui d'un *cloud* un peu particulier qui est le format que je garde aujourd'hui, mais on espère qu'il y en aura beaucoup d'autres après. Et comme l'a dit Olivier, il y a deux grosses spécificités : d'un premier côté, c'est beaucoup plus distribué qu'un *cloud* normal qui va généralement être plutôt centralisé dans un ou quelques *data center* et d'un deuxième côté, il y a la contrainte de réutiliser la chaleur. Donc ça apporte une complexité supplémentaire et des pistes de recherche aussi supplémentaires.

Donc du coup, vous réutilisez la chaleur qui est produite. Comment est-ce que vous faites ça ? Comment vous la réutilisez cette chaleur ?

RB Effectivement, je peux commencer par là. Aujourd'hui, chez Qarnot, on installe ce qu'on appelle des chaudières numériques. Donc pour un plombier, de l'extérieur, ça ressemblera à une chaudière classique : il pourra l'utiliser à peu près comme une chaudière à gaz classique. Et pour un informaticien, s'il l'ouvre, il reconnaîtra un cluster de calcul. Nous, notre créneau, comme je l'ai dit, c'est d'installer nos serveurs là où on peut réutiliser la chaleur. C'est très difficilement faisable dans des *datacenters* parce qu'ils sont généralement isolés, parce qu'ils sont soumis à pas mal de contraintes. On va donc installer nos serveurs, ces chaudières numériques, dans des lieux où il y a des besoins en chaleur assez élevés. Typiquement, aujourd'hui, on va se concentrer sur des piscines d'un assez gros volume, sur des réseaux de chaleur ou sur des industries qui vont avoir des besoins en chaleur conséquents. Donc on va utiliser de l'eau relativement froide pour refroidir nos serveurs et en sortie, elle va être suffisamment chaude pour que l'on puisse utiliser cette énergie pour chauffer d'autres lieux.

Quelles étaient les attentes de chacun ?

OB Ce qui est intéressant pour nous, c'est qu'on ne cherche pas vraiment à minimiser l'énergie qui est consommée, mais l'énergie qui est gaspillée d'une certaine façon. Donc si une tâche produit de la chaleur en s'exécutant et que cette chaleur est réutilisée dans une chaudière pour chauffer une piscine ou dans un réseau de chaleur, d'une certaine façon, on fait d'une pierre deux coups, on ne gaspille pas. Du point de vue de l'optimisation, ça veut dire qu'on a des ressources qui ne gaspillent pas d'énergie. Et donc typiquement, on a travaillé sur des problèmes de ce type là et donc on a essayé de maximiser la quantité de calculs qui est faite dans les chaudières. Donc, ça, ça a abouti à un travail qui est commun entre Qarnot, Inria et l'Ademe.

RB Pour nous, à l'échelle de ce projet, ce qui était vraiment intéressant, c'est qu'on a des besoins forts, comme je l'expliquais, en termes d'ordonnancement. En revanche, on n'a pas du tout les compétences et l'expertise de l'équipe d'Olivier. Donc pour nous, c'était vraiment important et intéressant de pouvoir creuser ce sujet avec des gens qui ont l'expertise. Et puis peut être un peu plus largement, ça participe un peu à la création d'une communauté un peu transverse entre industriels et académiques qui s'intéressent à ces sujets, on va dire l'intersection du monde numérique et de l'empreinte environnementale. Donc j'ai l'impression que c'est une communauté qui est croissante et on est ravi de pouvoir en faire partie et d'y participer.

OB Un autre point qui nous intéresse beaucoup chez Topal, c'est la problématique du non-renouvellement des ressources de calcul. C'est à dire comment est-ce qu'on peut proposer des outils logiciels qui permettent d'augmenter la durée de vie des calculs et qui évite aux utilisateurs d'être obligés d'acheter de nouveaux GPU parce qu'ils n'arrivent plus à faire tourner le dernier modèle d'IA ou la dernière application. Et donc pour ça, ce qui est important du point de vue de l'impact carbone, c'est donc d'éviter le renouvellement de ces ressources, parce qu'il faut extraire toutes les ressources naturelles qui sont nécessaires à l'extraction des matériaux. Il faut recycler en fin de vie et tout ça, ça a un gros impact. Et ce qui est compliqué, c'est que quand on met en œuvre ces solutions pour essayer d'augmenter la durée de vie des machines, souvent ce qu'on est obligé de faire, c'est de rajouter des calculs d'une certaine façon ou de rajouter des transferts de données. Donc du coup, on augmente la consommation énergétique. L'autre point, c'est que les constructeurs sont assez efficaces pour fabriquer des machines qui

sont de plus en plus efficaces en termes de nombre de calculs qu'on peut faire par jour. Et donc utiliser de vieilles ressources, ça veut aussi dire utiliser des ressources qui sont moins efficaces énergétiquement. Et donc il y a une sorte de compromis qui est un peu compliqué et qui est difficile à quantifier, avec d'un côté un impact carbone positif au niveau de la fabrication des ressources et de l'autre un impact qui est plutôt négatif au niveau de la consommation de l'énergie. Et pour nous, ce que nous offre Qarnot, c'est un cadre dans lequel finalement l'énergie n'est plus si importante que ça. Parce que cette énergie, elle n'est pas gaspillée, elle est réutilisée pour faire quelque chose d'utile. Et donc du coup, ça offre un cadre qui est plus favorable au développement de ces techniques auxquelles on croit beaucoup, qui visent à augmenter la durée de vie des ressources.

Ça ouvre de belles perspectives. Une dernière petite question est ce que vous pensez que dans cinq ans, toutes les piscines municipales par exemple, pourraient être chauffées grâce aux supercalculateurs ?

RB J'avais fait l'exercice de regarder un peu les ordres de grandeur. Je crois que pour les piscines en France, on est à environ 500 mégawatts qui doit correspondre à un réacteur nucléaire. Sauf qu'ici elles sont traditionnellement chauffées au gaz et pas à l'électricité. Et si on regarde un peu la consommation électrique de tous les *datacenters* en France, on arrive à peu près au même chiffre. Donc on va dire que, en termes d'ordre de grandeur, que ça marcherait. Après, je ne pense pas personnellement qu'on verrait dans cinq ans toutes les piscines équipées de systèmes comme ceux de Qarnot. D'une part parce que souvent on s'attaque aux cas les plus faciles, c'est à dire là où il y a les plus grosses quantités d'énergie qui sont demandées, et ça va être de manière privilégiée plutôt les réseaux de chaleur. Donc, je crois qu'il y en a une cinquantaine en France. C'est peut-être un peu optimiste, mais c'est possible et envisageable qu'une bonne partie de ces réseaux de chaleur aient un système comme celui de Qarnot dans les années à venir.

OB Pour compléter cette idée de réutiliser la chaleur produite, elle est super naturelle et elle paraît extrêmement prometteuse. Quand on fait des gros calculs en entrée, on a de l'électricité, en sortie, on a les résultats des calculs et on a de la chaleur. Et pour l'instant, que ça soit dans les supercalculateurs ou dans les *datacenters*, cette chaleur, elle a toujours été vue comme une source de problèmes et pas comme une ressource qu'on pouvait exploiter, ce qui paraît un peu un peu stupide. En même temps, ce n'est pas vraiment ni de la stupidité, ni de la bêtise. C'est difficile à faire, parce que Rémy l'a dit au début, c'est une architecture qui est super distribuée. La conception même des chaudières de Qarnot, techniquement est difficile donc. Mais j'ai l'impression que Qarnot a maintenant les bons outils et le bon modèle. Et donc du coup, on peut arriver à faire des choses, des choses sérieuses. Ce qui est important, c'est de faire des gros calculs. C'est quelque chose qui est important, y compris pour l'environnement. Toutes les simulations du GIEC nécessitent l'utilisation massive de supercalculateurs. Le but, c'est de limiter l'impact du numérique et pas de renoncer à faire ses calculs. Et l'impact du numérique vient de la fabrication des ressources. On a vu que les constructeurs faisaient un bon boulot pour arriver à augmenter la quantité de calcul qu'on peut faire par unité d'énergie. Et avec le modèle de Qarnot et ce qu'on propose, on peut essayer à la fois d'augmenter la durée de vie des machines et de réutiliser l'énergie qui est produite au moment où on fait les calculs. Et donc du coup, potentiellement, on a la possibilité d'avoir un impact important, même si toutes les

piscines ne seront probablement pas équipées de Qarnot. Mais en tout cas, ça paraît être une bonne idée pour essayer d'atténuer l'impact du numérique dans ce cadre-là !

Merci beaucoup Olivier et Rémi pour ces échanges passionnants autour des chaudières numériques ! J'espère que nous aurons l'occasion de re-échanger ensemble dans un nouvel épisode de Désassemblons le numérique ! A très bientôt.