

CLOUD ECONOMICS

Avec le passage de l'architecture client/serveur au cloud computing, l'informatique connaît une profonde révolution dont l'importance et l'impact ne sont pas sans rappeler la transition entre les univers mainframe (sites centraux) et client/serveur. Les spéculations vont bon train et chacun se demande comment cette nouvelle ère évoluera au fil des ans, les responsables informatiques devant impérativement avoir une vision claire de l'évolution de l'informatique. Le meilleur moyen pour envisager l'avenir est de comprendre les critères économiques qui dictent la tendance sur le long terme. Dans ce document, nous aborderons l'économie du cloud en utilisant une modélisation précise. Ce cadre nous permettra ensuite d'anticiper le paysage informatique à long terme et de mieux le comprendre.

Si vous avez des commentaires et des questions à propos de ce document, envoyez-les à infoAzure@microsoft.com

1. INTRODUCTION

Lors de leur apparition, à la fin du XIXe siècle, les automobiles étaient considérées comme des « voitures sans cheval ». Naturellement, les premiers clients potentiels se montraient sceptiques car il leur paraissait inconcevable de remettre en cause le modèle connu depuis des siècles : le cheval et son attelage. Les toutes premières automobiles perpétuaient les formes des voitures hippomobiles sans le cheval. Les ingénieurs de l'époque n'avaient pas vu les nouvelles possibilités offertes (vitesse plus élevée et sécurité accrue). Ils continuaient même à prévoir l'emplacement du fouet dans les premiers modèles avant de se rendre compte que cet accessoire était devenu totalement inutile.



Figure 1 : Syndrome de la voiture sans cheval.

Au départ, personne n'a vraiment compris ce nouveau concept. Les banquiers prétendaient que « le cheval était une valeur sûre et que l'automobile n'était qu'une passade qui présentait simplement l'attrait de la nouveauté ». Même les pionniers de l'automobile ne mesuraient pas l'impact potentiel de leur invention dans le monde. Lorsque M. Daimler a tenté d'estimer les opportunités du marché automobile à long terme, il est arrivé à la conclusion qu'il ne serait pas possible de dépasser le million de voitures du fait de leur coût élevé et du manque de conducteurs compétents¹.

Dans les années 1920, le nombre de voitures avait déjà atteint 8 millions, et aujourd'hui nous avons dépassé les 600 millions de véhicules. Nous sommes très loin des prévisions de M. Daimler. Les premiers pionniers de l'automobile n'avaient anticipé ni la baisse des coûts ni la facilité de conduite des automobiles. Bien loin d'eux l'idée de voir les voitures prendre autant d'ampleur dans la vie quotidienne, outrepassant les premières contraintes et gagnant l'ensemble de la population.

Aujourd'hui, l'informatique connaît une révolution du même ordre avec l'avènement du cloud. Le cloud promet non seulement une baisse des coûts informatiques, mais aussi la rapidité, la simplicité, la flexibilité et l'efficacité.

Comme aux tout premiers jours de l'industrie automobile, il est bien difficile de savoir où ce concept nous mènera. **Ce livre blanc vise à définir un cadre qui aidera les responsables informatiques à planifier la transition vers le cloud²**. Nous avons adopté une démarche sur le long terme dans la mesure où les décisions et les investissements à évaluer concerneront vraisemblablement plusieurs décennies. Par conséquent, nous nous intéresserons uniquement à la dimension économique du cloud et n'aborderons ni les technologies spécifiques ni certains leviers particuliers comme les changements organisationnels dans l'entreprise. Les facteurs économiques donnent souvent les moyens d'appréhender clairement les transformations de cette nature.

Dans la deuxième section, nous présentons les facteurs économiques sous-jacents au cloud et soulignons plus particulièrement les différences qui séparent le cloud de l'architecture client/serveur. Dans la troisième section, nous étudions toutes les implications qui en découlent pour l'avenir de l'informatique. Nous abordons tous les aspects positifs du cloud, sans oublier de mentionner les obstacles qui existent encore aujourd'hui. Pour finir, la quatrième section nous permet d'envisager certains points importants à prendre en compte avant d'entamer un voyage vers le cloud.

¹ Source : Horseless Carriage Thinking, William Horton Consulting.

² Le terme cloud dans ce contexte réfère à l'architecture de cloud computing (informatique dématérialisée) et couvre aussi bien les cloud publics que privés.

2. LES ASPECTS ÉCONOMIQUES DU CLOUD

Les sciences économiques constituent un formidable moyen de suivre les transformations de l'industrie. Les discussions qui se tiennent aujourd'hui autour du cloud ont généralement pour thème les complexités techniques et les freins à l'adoption. Si nous reconnaissons que ces problèmes existent et ont toute leur importance, il convient aussi d'admettre que les aspects économiques ont toujours influencé les orientations et la rapidité des changements. Les défis technologiques ont toujours été résolus ou surmontés via une innovation rapide à laquelle nous nous sommes habitués (figure 2). À l'ère des mainframes, l'architecture client/serveur a été perçue, au départ, comme une technologie « gadget » et non comme une alternative parfaitement viable. Pourtant, au fil des ans, la technologie client/serveur s'est imposée dans l'entreprise (figure 3). De même, lorsque la technologie de virtualisation a fait son entrée sur le marché, les craintes liées à la compatibilité des applications et à la dépendance d'un seul fournisseur ont constitué de véritables freins à l'adoption. Pourtant, les économies de 20 à 30 % annoncées³ ont convaincu les directeurs informatiques les plus réticents à surmonter ces problèmes, et l'adoption s'est rapidement accélérée.

L'émergence des services de cloud computing bouleverse une nouvelle fois l'économie de l'informatique. Avec le cloud, il est possible de standardiser et de regrouper des ressources informatiques, puis d'automatiser de nombreuses tâches de maintenance, réalisées manuellement aujourd'hui. Les architectures de type cloud savent faire face à des charges fluctuantes, et permettent l'accès en libre-service et la facturation en fonction de l'utilisation.

Avec le cloud, les entreprises placent leur infrastructure informatique de base dans de vastes centres de données et profitent d'importantes économies d'échelle dans trois domaines :

- **Économies d'échelle côté fournisseur.** Les centres de données de grande envergure diminuent les coûts par serveur.
- **Groupement côté demandeur.** En regroupant les demandes de puissance informatique, il est possible de lisser globalement les variations et, par conséquent, d'optimiser les taux d'utilisation des serveurs.
- **Efficacité de la mutualisation** (locataires multiples). En cas d'adoption d'un modèle mutualisé, plus le nombre de locataires (entreprises clientes ou utilisateurs, par exemple) est grand, plus les coûts d'administration des applications et de serveur diminuent par locataire.

	Technologie	Rentabilité	Mode de commercialisation
Site central 	Stockage et calculs centralisés Clients légers	Optimisé pour l'efficacité en raison du coût élevé	Investissements initiaux élevés pour le matériel et les logiciels
Client / Serveur 	PC et serveurs pour du stockage et des traitements distribués	Optimisé pour l'agilité car coût faible	Licences perpétuelles pour applications et système d'exploitation
Cloud 	Grands centres de données, montée en charge, matériel ordinaire	Efficacité et agilité d'un niveau supérieur	Vous payez uniquement à l'utilisation

Figure 2 : Opportunité du cloud.

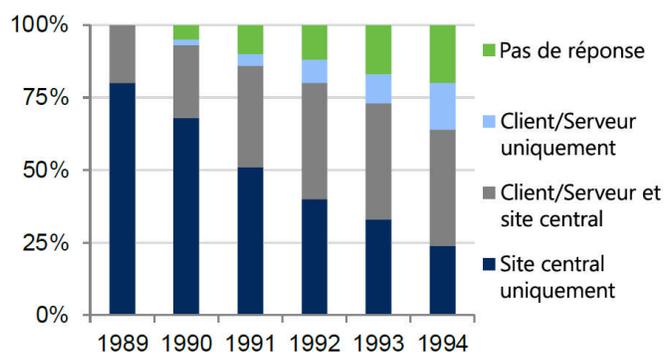


Figure 3 : Début de la transition vers la technologie client/serveur.

³ Source : « Dataquest Insight : Many Midsize Businesses Looking Toward 100% Server Virtualization ». Gartner, 8 mai 2009.

2.1 Économies d'échelle côté fournisseur

Le cloud computing reprend à son compte les aspects financiers les plus intéressants des mainframes et de l'architecture client/serveur. L'ère des mainframes se caractérisait par de considérables économies d'échelle du fait du coût d'entrée élevé des mainframes et de la difficulté à embaucher du personnel qualifié pour gérer les systèmes. Alors que la puissance de calcul (mesurée en MIPS - million d'instructions par seconde) augmentait, le coût déclinait rapidement dans un premier temps (figure 4). Mais, seuls les services informatiques de grande envergure disposaient des ressources nécessaires et enregistraient une demande suffisante pour justifier un tel investissement. Compte tenu du coût élevé, les entreprises faisaient passer la consommation des ressources matérielles avant la réactivité des utilisateurs. Les requêtes des utilisateurs étaient mises en file d'attente et le traitement n'intervenait que lorsque les ressources nécessaires étaient disponibles.

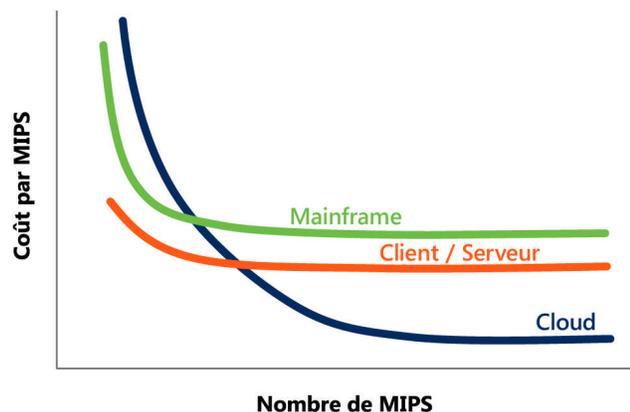


Figure 4 : Economies d'échelle.

Source : Microsoft.

Avec l'arrivée des mini-ordinateurs et, plus tard, de la technologie client/serveur, il a été possible de réduire le montant initial de l'achat, et d'exploiter et gérer les ressources beaucoup plus facilement. Cette modularité a fait considérablement baisser le ticket d'entrée aux services informatiques, d'où une augmentation radicale de l'agilité des utilisateurs. En revanche, nous assistions à un cruel dilemme en termes d'utilisation, menant à la situation suivante : les centres de données s'étendaient de manière tentaculaire avec des serveurs achetés à mesure des besoins, mais n'étaient utilisés qu'à hauteur de 5 à 10 % de leur capacité⁴.

Le cloud computing n'est en aucun cas un retour à l'ère des mainframes, comme certains le laissent entendre. Bien au contraire, il offre aux utilisateurs des économies d'échelle et une efficacité dépassant de loin celles des mainframes. La modularité et l'agilité sont également bien supérieures à ce qu'offrirait la technologie client/serveur. Fini les compromis !

Les économies d'échelle découlent des domaines suivants :

- **Coût de l'énergie.** Les besoins en énergie ne cessant d'augmenter, la facture d'électricité est devenue le principal composant du coût total de possession (TCO).⁵ Elle en représente aujourd'hui 15 à 20 %. L'indicateur d'efficacité énergétique (PUE - Power Usage Effectiveness)⁶ tend à baisser davantage dans les grands sites que dans les petits. Les opérateurs de petits centres de données doivent payer l'électricité au tarif local en vigueur, alors que les gros fournisseurs paient moins en implantant leurs centres de données dans des lieux où l'approvisionnement en électricité est moins coûteux et en signant des contrats d'achat en gros.⁷ De plus, l'étude montre qu'un opérateur qui gère plusieurs centres de données peut bénéficier de taux différents en fonction de la position géographique de chaque centre, ce qui allège encore les dépenses énergétiques.

⁴ Source : The Economics of Virtualization : Moving Toward an Application-Based Cost Model. IDC. Novembre 2009.

⁵ Hors main-d'œuvre. Les différentes études suggèrent que, en cas de faible efficacité des centres de données, l'énergie dépensée en trois ans pour l'alimentation, la climatisation, l'éclairage (infrastructure comprise) dépasse de loin les sommes dépensées en trois ans pour le matériel serveur.

⁶ L'indice PUE (Power Utilization Effectiveness) est calculé en divisant la dépense énergétique totale du centre de données par celle des serveurs qu'il héberge. Il mesure l'aptitude du centre de données à transformer l'électricité en puissance de calcul. La meilleure valeur théorique est de 1.0. Plus cette valeur augmente, plus la situation empire.

⁷ Source : U.S. Energy Information Administration (juillet 2010) et Microsoft. Le taux commercial moyen aux États-Unis est de 10,15 cents par kilowatt heure. Dans certaines localités, le prix du kilowatt heure peut descendre jusqu'à 2,2 cents.

- **Coûts de personnel de l'infrastructure.** Le cloud computing réduit considérablement les coûts de personnel à tous les niveaux en automatisant la plupart des tâches d'administration répétitives. Toutefois, les grands sites y parviennent mieux que les petits. Dans une entreprise traditionnelle, un seul administrateur peut s'occuper d'environ 140 serveurs.⁸ Dans un centre de données du cloud, le même administrateur a des milliers de serveurs sous sa responsabilité. Les informaticiens peuvent alors se consacrer à des activités à plus forte valeur ajoutée (création d'applications ou ajouts de fonctionnalités, par exemple) et répondre aux requêtes, toujours plus pressantes, des utilisateurs auxquelles le service informatique a affaire.
- **Sécurité et fiabilité.** Bien que souvent cité comme frein potentiel à l'adoption des clouds publics, le besoin accru de sécurité et de fiabilité donne lieu à des économies d'échelle. Pour atteindre des niveaux honorables, il faut généralement consentir de très lourds investissements. Les principaux fournisseurs commerciaux de cloud sont souvent mieux armés en la matière. Dotés d'une plus grande expertise qu'un simple service informatique d'entreprise, ils assurent une parfaite sécurité et fiabilité des systèmes du cloud.
- **Remises quantitatives.** Les opérateurs de grands centres de données bénéficient souvent d'importantes remises sur le matériel, de l'ordre de 30 %, par rapport aux acheteurs habituels. Cela est dû à la standardisation d'un nombre limité d'architectures matérielles et logicielles. À l'ère des mainframes, il n'était pas rare de voir coexister plus de 10 architectures différentes. Quant à une topologie client/serveur, elle pouvait regrouper près d'une douzaine de variantes UNIX, le système d'exploitation Windows Server et x86, ainsi que quelques plateformes RISC. Dans un environnement aussi hétérogène, il était difficile de compter sur des remises quantitatives importantes. Avec le cloud, l'homogénéité de l'infrastructure permet de réaliser des économies d'échelle.

À l'avenir, beaucoup d'autres économies d'échelle seront probablement envisageables, mais nous manquons pour l'instant de visibilité à long terme. Les centres de données n'en sont qu'à leur début et nous voyons aujourd'hui sortir de terre des bâtiments d'une superficie de plusieurs dizaines d'hectares (figure 5). Compte tenu de la taille impressionnante de ces mégacentres de données, la R&D ne chômera pas pour optimiser leur exploitation et de les rendre encore plus attrayants aux clients. Les opérateurs de grands centres de données profiteront bien plus de ces avantages que les centres plus petits implantés à l'intérieur des entreprises.

Entreprise	Lieu	Coût (en millions \$)	Taille (en m ²)
Internet Villages Juil 2009	Annandale, Écosse	1 600	279 999
National Security Admin. Juil 2009	Camp Williams, Utah	2 000	92 900
Lockerbie Data Centers Déc 2009	Lockerbie, Écosse	1 500	Non applicable
Microsoft Sept 2009	Chicago, Illinois	500	65 000
I/O Data Centers Juin 2009	Phoenix, Arizona	N/A	50 000
Apple Mai 2009	Maiden, Caroline du Nord	1 000	46 500
Microsoft Juin 2010	Dublin, Irlande	500	N/A
U.S. Social Security Admin. Fév 2009	Baltimore, Maryland	400	N/A
Facebook Fév 2010	Princeville, Oregon	N/A	28 500
Next Generation Data Mars 2010	Cardiff, Pays de Galles	301	N/A

Figure 5 : Récents projets d'implantation de mégacentres de données.

Sources : Communiqués de presse.

2.2 Regroupement de la demande

Le coût global de l'informatique est déterminé non seulement par le coût de la capacité, mais aussi par le degré d'utilisation de cette capacité. Nous devons évaluer l'impact qu'aura l'agrégation des demandes sur

⁸ Source : James Hamilton, Microsoft Research, 2006.

les coûts des ressources réellement utilisées (processeur, réseau et stockage).⁹ Dans le centre de données non virtualisé, chaque application/charge de travail s'exécute généralement sur un serveur physique particulier.¹⁰ Autrement dit, le nombre de serveurs augmente proportionnellement en fonction des charges serveurs. Dans ce modèle, les serveurs ont été finalement très peu sollicités, aux alentours de 5 à 10%.¹¹ Avec la virtualisation, plusieurs applications peuvent s'exécuter sur un seul serveur physique à l'intérieur d'une instance optimisée du système d'exploitation. Il faut donc moins de serveurs physiques pour faire face au même nombre de charges, ce qui est un atout indéniable. En quoi cela affecte-t-il les économies d'échelle ? Si toutes les charges reposaient sur une utilisation constante, cela reviendrait à une simple réduction du nombre de serveurs physiques et n'influerait en rien sur les économies d'échelle. Mais, dans la réalité, les charges de travail varient énormément au fil des heures. Elles peuvent exercer une forte demande sur les ressources pendant une minute et ne pratiquement pas les solliciter la minute suivante. C'est pourquoi il convient d'optimiser l'utilisation des ressources en agrégeant des demandes diverses et variées.

Pour rédiger ce livre blanc, nous avons analysé l'origine des fluctuations d'utilisation, puis examiné les moyens de les éviter afin de réduire les coûts.

Cela nous a conduits à identifier cinq sources de variation :

- 1. Variation aléatoire.** Les habitudes d'accès des utilisateurs présentent un caractère aléatoire. Les utilisateurs, par exemple, consultent leurs courriers électroniques aux moments qui leur conviennent (figure 6). Pour respecter les accords de niveau de service, il faut prévoir des tampons de capacité susceptibles d'intervenir si jamais tout le monde entreprenait des tâches particulières en même temps. Si les serveurs sont organisés en groupes (pools), cette variation peut être atténuée.
- 2. Variation horaire.** Il existe des cycles dans le comportement des individus. Les utilisateurs ont tendance à accéder davantage aux services grand public dans la soirée et aux services professionnels dans la journée. La capacité doit tenir compte de ces pics journaliers sans être, pour autant, sous-utilisée le reste du temps. Pour remédier à cette variation, il suffit d'exécuter sur les mêmes serveurs la même charge de travail pour différents fuseaux horaires (figure 7), ou des charges complémentaires tout au long de la journée (accès aux services grand public et d'entreprise, par exemple).

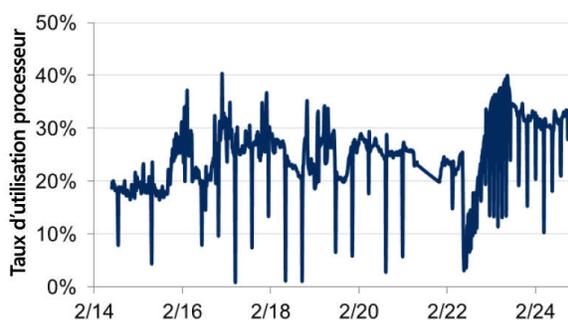


Figure 6 : Variation aléatoire (Exchange Server).
Source : Microsoft.

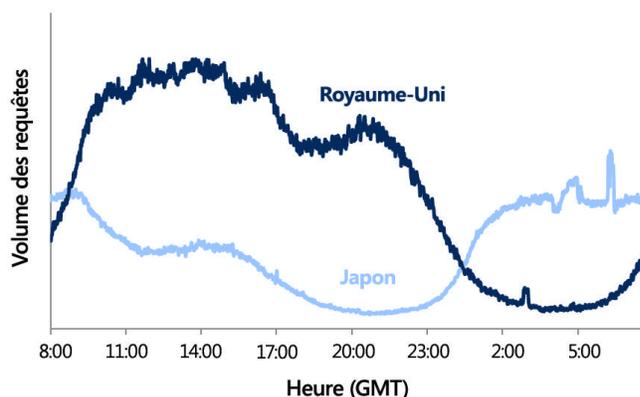


Figure 7 : Variation horaire dans le cas d'un service de recherche.
Source : Volume de recherche Bing sur une période de 24 heures.

⁹ Dans ce document, nous parlons globalement d'utilisation des « ressources ». Nous savons pertinemment qu'il existe d'importantes différences entre les ressources. Les unités de stockage, par exemple, connaissent peu de pics d'utilisation en comparaison avec les processeurs et les ressources E/S. L'impact dont nous parlons ici ne touchera que très peu le stockage.

¹⁰ De toute évidence, plusieurs applications peuvent s'exécuter sur un seul et même serveur, mais ce n'est pas pratique courante. Il est, en effet, très délicat de faire passer une application active d'un serveur à un autre sans déplacer le système d'exploitation. L'exécution simultanée de plusieurs applications sur une même instance de système d'exploitation risque de créer des goulets d'étranglement difficiles à juguler sans interrompre le service et nuire à la réactivité. La virtualisation permet de déplacer à volonté les applications et le système d'exploitation.

¹¹ Source : The Economics of Virtualization : Moving Toward an Application-Based Cost Model. IDC. Novembre 2009.

3. Variation sectorielle. Certaines fluctuations sont dues à des différences entre secteurs d'activité. Les magasins de vente au détail connaissent un pic d'activité à la période des fêtes alors que les cabinets comptables sont fortement sollicités avant le 15 avril (figure 8). Il existe plusieurs types de variations sectorielles. Certaines sont récurrentes et prévisibles (déclarations d'impôt ou jeux olympiques), d'autres le sont moins (événements ou communiqués de presse). Le résultat est le même : la capacité doit pouvoir faire face aux pointes en se ménageant même une marge d'erreur. Une large diversité entre de nombreux secteurs d'activité différents permet de réduire cette source de fluctuation.

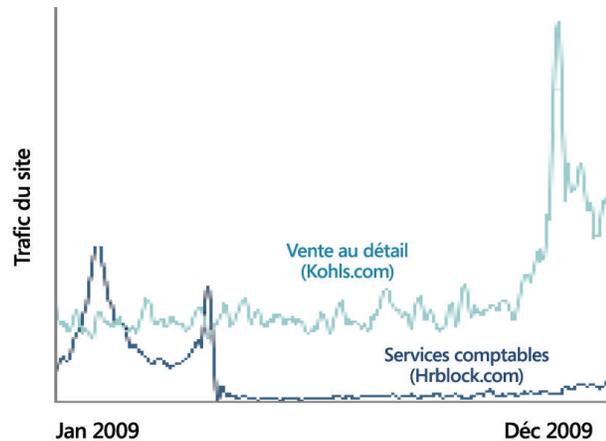


Figure 8 : Variation sectorielle.

Source : Alexa Internet.

4. Variation matérielle. Les ressources de calcul, de stockage et d'entrée/sortie (E/S) font généralement l'objet d'achats groupés : un serveur affiche une certaine puissance de calcul (processeur), une certaine capacité de stockage et un certain nombre d'E/S (réseau ou accès disque, par exemple). Certaines charges comme la recherche exercent une forte pression sur le processeur, mais peu sur le stockage ou les E/S. Pour la messagerie, c'est exactement l'inverse (figure 9). S'il est possible d'ajuster la capacité en achetant des serveurs optimisés pour la puissance de calcul ou le stockage, cela ne résout pas véritablement le problème puisque la flexibilité et les finances en souffriront. Cette variation conduira à une sous-utilisation des ressources à moins de regrouper des charges dont les profils en utilisation de ressources sont complémentaires.

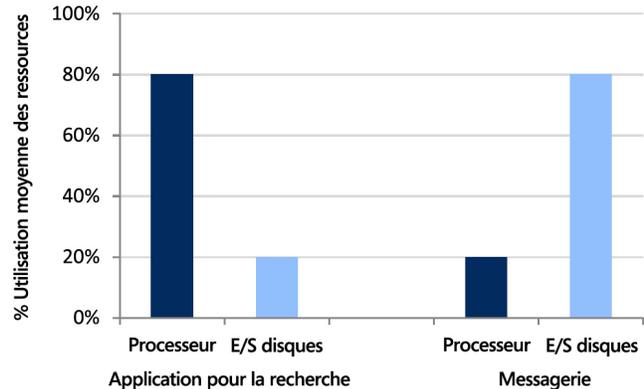


Figure 9 : Variation matérielle (exemple).

Source : Microsoft.

5. Variation prévisionnelle. Il est souvent difficile d'anticiper les besoins pour la puissance de calcul, et fastidieux d'ajouter de la puissance supplémentaire. Ces deux points noirs constituent une autre source de sous-utilisation (figure 10). Pour les startups, ce phénomène est souvent appelé l'effet TechCrunch. Les entreprises et les PME doivent toutes deux obtenir l'approbation de leurs investissements informatiques bien avant de savoir quelle sera la demande réelle. Les grands groupes privés n'échappent pas à la règle puisqu'ils doivent prévoir leurs achats six à douze mois à

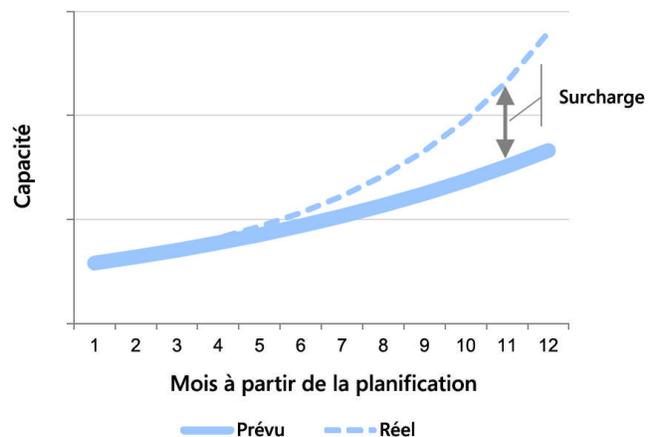


Figure 10 : Variation prévisionnelle.

Source : Microsoft.

l'avance (figure 10). En répartissant les charges sur plusieurs clients, les fournisseurs de cloud réduisent ce degré de variation dans la mesure où une demande plus forte que prévue pour certaines charges vient compenser une demande moins forte que prévue pour d'autres.

Le cloud présente un avantage économique indéniable : il offre la possibilité d'aplanir les variations découlant de ces facteurs pour mieux équilibrer l'utilisation des ressources. En regroupant les ressources, la variation s'atténue et les modèles d'utilisation sont lissés. Plus le pool de ressources est important, plus il est facile d'agréger les demandes, plus le taux d'utilisation est élevé et plus le service informatique est en mesure de répondre, à moindre coût, à la demande des utilisateurs.

Nous avons modélisé l'impact théorique de la **variation aléatoire** de la demande sur les taux d'utilisation des serveurs en augmentant le nombre de serveurs.¹² La figure 11 montre qu'un groupe théorique de 1 000 serveurs présenterait un taux d'utilisation de près de 90 % sans violer le SLA. Cela n'est vrai que dans la situation hypothétique où le caractère aléatoire est la seule source de variation et où les charges de travail peuvent passer instantanément d'un serveur physique à un autre sans interruption. Plus la durée de fonctionnement définie dans un SLA (accord de niveau de service) est élevée, plus il est facile de l'assurer à mesure que l'échelle augmente.

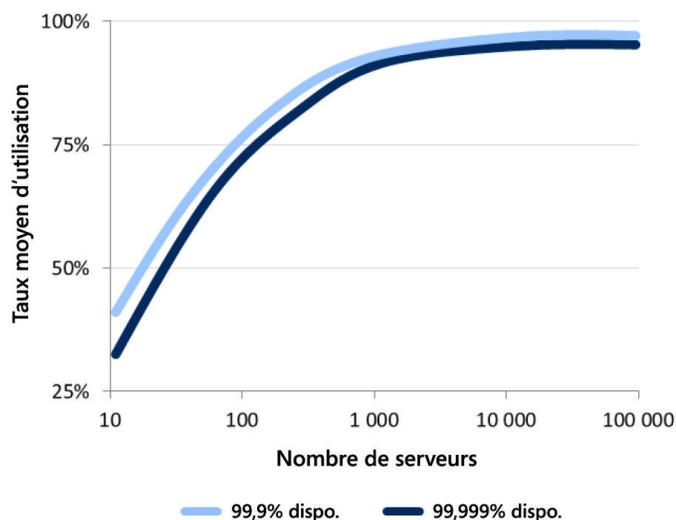


Figure 11 : Lissage d'une variation aléatoire.

Source : Microsoft.

Les clouds pourront réduire les **variations horaires** dans la mesure où les zones géographiques et les types de charge sont plus ou moins confondus. Dans une organisation moyenne, le pic d'utilisation informatique peut être deux fois supérieur à la moyenne journalière. Même dans les grandes entreprises internationales, la majorité des employés et des utilisateurs vit sous le même fuseau horaire, ce qui crée des cycles journaliers quasi-synchrones. Par ailleurs, les organisations n'ont généralement pas de schémas de charge capables de se compenser. Par exemple, les activités liées à la messagerie, au réseau et au traitement des transactions qui ont lieu pendant les heures ouvrées ne sont pas remplacées par un flux de travail aussi actif au milieu de la nuit. Le regroupement des entreprises et des charges permet d'aplanir ces pics et ces creux.

La **variation sectorielle** donne également naissance à des pics et à des creux au sein de chaque entreprise. Les systèmes du magasin de vente au détail, par exemple, tourneront quasiment tous à plein régime pendant la période des fêtes (serveurs Web, traitement des transactions, traitement des paiements, bases de données).¹³ La figure 12 explique cette variation sectorielle pour différentes industries et montre des pics de 1,5 à 10 fois supérieurs à la moyenne.

¹² Pour calculer les économies d'échelle découlant de la dissipation d'une variation aléatoire, nous avons créé un modèle Monte Carlo. Nous avons ainsi simulé des centres de données de tailles différentes traitant plusieurs charges aléatoires. Pour chaque simulation, les charges de travail (conçues de façon à se rapprocher de schémas d'utilisation Web hypothétiques) ont été ajoutées successivement jusqu'à ce que la disponibilité escomptée des ressources serveurs passe au-dessous d'un temps de fonctionnement donné de 99,9 % ou 99,99 %. Le nombre maximal de charges détermine le taux d'utilisation maximal auquel les serveurs du centre de données peuvent fonctionner sans nuire aux performances.

¹³ L'idéal serait d'analyser l'historique de l'utilisation des serveurs d'un grand nombre de clients pour approfondir ces modèles. Ces données sont toutefois difficiles à réunir et souvent peu exploitables. Nous nous sommes, par conséquent, servi du trafic Web pour mesurer la variation sectorielle.

Les services Microsoft comme Windows Live Hotmail et Bing tirent parti de la **diversification des ressources**. Pour optimiser les charges, ils mettent différents sous-services en corrélation avec différents profils de ressources (processeur ou stockage, par exemple). Il est difficile de quantifier ces avantages, c'est pourquoi nous n'avons pas retenu la variation matérielle dans notre modèle.

La **variation prévisionnelle** peut, dans une certaine mesure, être limitée par la standardisation matérielle et l'achat en flux tendu. D'après notre modélisation, le manque de visibilité sur la croissance des entreprises dotées de 1 000 serveurs entraîne un surdimensionnement de 30 à 40 % des serveurs par rapport à un service de cloud public. Pour les entreprises plus petites (startups Internet, par exemple), l'impact est bien plus grand.

Jusqu'à présent, nous sommes implicitement partis du principe que le degré de variation restera le même lors de la transition vers le cloud. En réalité, il est fort probable qu'il augmente et qu'il fasse, par contrecoup, réaliser d'autres économies d'échelle. Il y a deux raisons à cela :

- **Attente de performance plus élevée.** Aujourd'hui, les utilisateurs se sont habitués aux contraintes de ressources et ont appris à vivre avec. Prenons un exemple. Les utilisateurs planifient l'exécution de calculs complexes la nuit, évitent les itérations multiples de modèles ou renoncent à une optimisation fastidieuse et coûteuse de leur chaîne logistique. Le modèle commercial du cloud permet à un utilisateur de payer la même chose, qu'un seul système travaille pendant 1 000 heures ou que 1 000 systèmes travaillent pendant 1 heure. Aujourd'hui, l'utilisateur attendrait probablement 1 000 heures ou abandonnerait le projet. Avec le cloud, il peut faire tourner 1 000 systèmes sans coût supplémentaire, et donc accélérer le processus. L'impact sur la variation sera considérable. Pixar Animation Studios, par exemple, exécutent tous les processus de rendu de leurs animations numériques sous Windows Azure. Aujourd'hui, il faut environ 8 heures pour rendre chacune des images de leur film sur un seul processeur, autrement dit 272 ans au total pour l'ensemble du film. Comme ils le disent eux-mêmes, « Nous n'avons pas cette patience ». Avec Azure, ils mènent l'opération à leur rythme. Résultat : Azure connaît d'énormes pointes d'activité du fait de Pixar puisque les rendus sont créés à la demande.
- **Les traitements par lots se font en temps réel.** Bien des traitements, comme la disponibilité des stocks des distributeurs en ligne, qui avaient lieu par lots, se feront désormais en temps réel. Ainsi, les traitements à plusieurs phases qui se déroulaient jadis séquentiellement pourront désormais s'exécuter simultanément. Par exemple, un fabricant peut, tout en même temps, interroger ses stocks, vérifier son carnet de commandes et commander de nouveaux articles ou matériaux. Cela amplifiera inévitablement la variation de la consommation.

Notons que même les clouds publics, aussi grands soient-ils, ne pourront pas écartier toutes les variations. La variation des marchés restera probablement la même. Pour lisser davantage la demande, il convient de mettre en place une tarification sophistiquée. Comme pour l'électricité, par exemple, (figure 13), pourquoi ne pas inviter les clients à déplacer leur demande des périodes pleines vers les périodes creuses ? En outre, un tarif plus attractif les inciterait à consommer davantage du fait de l'élasticité du prix de la demande. La gestion de la demande augmentera notablement les avantages financiers du cloud.

Entreprise	Pic d'activité/ Trafic moyen
Services comptables	10x
Vente au détail	4x
Sports	2,5x
Voyages (avions, hôtels)	1,5x
Informations	1,5x - 2x

Figure 12 : Fluctuation sectorielle.

Source : Microsoft, Alexa Internet, Inc.

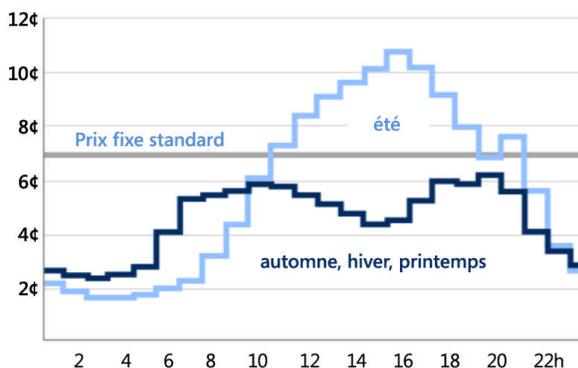


Figure 13 : Tarification variable de l'électricité.

Source : Ameren Illinois Utilities.

2.3 Efficacité de la mutualisation

Les économies d'échelle décrites au préalable pour l'approvisionnement et la demande peuvent être réalisées indépendamment de l'architecture applicative, qu'il s'agisse d'augmentation ou de diminution de charge classique, d'un seul locataire ou de mutualisation. Une autre source importante d'économies d'échelle ne peut être maîtrisée qu'en cas de mutualisation des applications. Autrement dit, il n'est plus question d'exécuter une instance d'application pour chaque client, comme en cas d'exécution sur site ou d'hébergement d'applications (instances dédiées de Microsoft Office 365, par exemple). En cas de mutualisation des applications, plusieurs clients utilisent simultanément une seule instance de l'application (cas du partage d'Office 365). Cette nouvelle approche s'accompagne de deux atouts financiers importants :

- Coût fixe d'administration des applications amorti sur un grand nombre de clients.** Dans une instance mono-licence, chaque client doit prévoir son propre budget pour l'administration des applications (autrement dit, pour le personnel chargé de l'administration des mises à jour et mises à niveau, ainsi que de la résolution des incidents). Nous avons étudié les données des clients ainsi que les offres Office 365-D et Office 365-S pour évaluer l'impact. Dans les instances dédiées, les mêmes activités (application de correctifs, par exemple) sont effectuées plusieurs fois - une fois par instance. Dans une instance mutualisée, comme Office 365-S, ce coût est partagé entre de nombreux clients, ce qui ramène les coûts d'administration des applications par client à quasiment rien. C'est là une économie non négligeable sur le coût total, surtout en présence d'applications complexes.
- Coût fixe d'exploitation des serveurs amorti sur un grand nombre de clients.** Pour chaque instance d'application, il y a une certaine partie de frais fixes pour les serveurs. La figure 14 présente l'exemple du service informatique de Microsoft. La variation au sein d'une journée n'apparaît pas frappante (16 % d'écart entre les périodes pleines et creuses) par rapport à la variation réelle des accès utilisateurs. Cela est dû aux coûts des applications et de l'exécution qui restent constants tout au long de la journée. En passant à la mutualisation avec une seule instance, ces frais généraux peuvent être amortis entre la totalité des clients. Nous avons analysé les données d'Office 365-D, d'Office 365-S et de Microsoft Live@edu pour évaluer ces frais fixes. Jusqu'à présent, il s'est avéré difficile, techniquement parlant, d'isoler ce phénomène des autres variations des données (nombre d'utilisateurs et utilisation du serveur) et des différences architecturales des applications. Par conséquent, nous n'en tirons aucun bénéfice dans notre modèle.

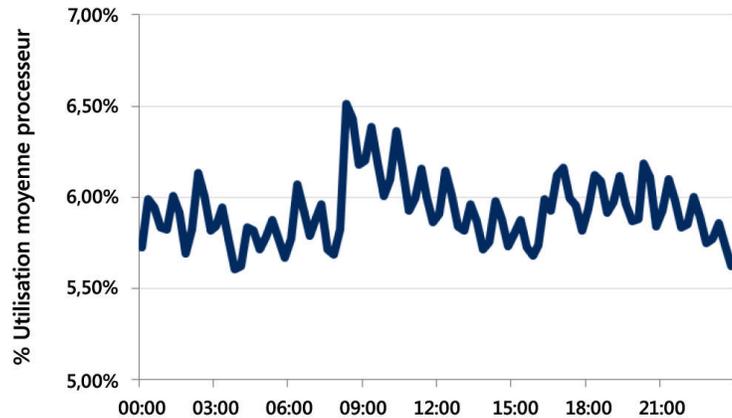


Figure 14 : Frais généraux d'exploitation.

Source : Microsoft.

Les applications peuvent être entièrement mutualisées en étant écrites dans ce but dès le départ ou le devenir partiellement en exploitant les services partagés par le biais de la plateforme cloud. Plus les clients auront recours aux services partagés, plus ils bénéficieront des économies d'échelle dues à la mutualisation des applications.

2.4 Impact global

Pour réaliser d'importantes économies d'échelle, il convient de jouer sur les différents tableaux : capacité des serveurs (amortissement des coûts sur davantage de serveurs), agrégation des charges de travail (réduction de la variation) et mutualisation des applications (amortissement des coûts sur plusieurs clients). Pour bien estimer l'ordre de grandeur, nous avons construit un modèle d'évaluation des coûts qui suit leur évolution sur le long terme.

La figure 15 présente les résultats pour une charge sollicitant 10 % d'un serveur traditionnel. Le modèle montre qu'un centre de données composé de 100 000 serveurs affiche une baisse de 80 % du coût total de possession (TCO) par rapport à un centre de données de 1 000 serveurs.

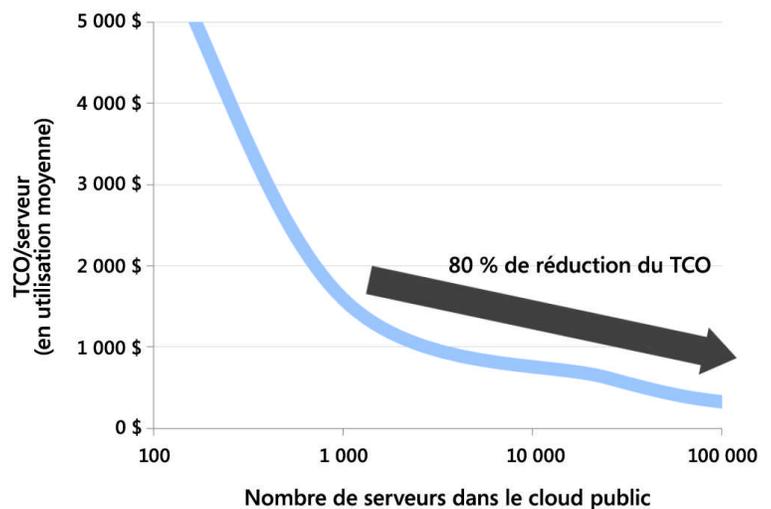


Figure 15 : Économies d'échelle dans le cloud.

Source : Microsoft.

Une question se pose maintenant : quel sera l'impact des aspects économiques du cloud dont nous venons de parler sur le budget informatique ? Du côté des données clients, nous savons à peu près comment se répartissent les coûts d'infrastructure, les coûts d'administration et de maintenance des applications et les coûts de développement de nouvelles applications (figure 16). Le cloud exerce une influence dans ces trois domaines. Les économies réalisées côté fournisseur et côté demandeur se répercutent essentiellement sur la partie infrastructure, qui représente plus de la moitié des dépenses. Les coûts de maintenance des applications couvrent la gestion des mises à jour et des correctifs, le support des utilisateurs et le versement des frais de licences aux fournisseurs. Ils entrent pour environ un tiers dans les dépenses totales et peuvent être réduits via la mutualisation.

Le développement des nouvelles applications ne constitue que le dixième des dépenses¹⁴, même s'il a pour rôle de faire avancer l'innovation. En principe, les dirigeants informatiques sont prêts à investir dans ce créneau. Les avantages financiers apportés par le cloud leur permettront en libérant une partie du budget. Nous aborderons ce point plus en détail dans le prochain paragraphe, ainsi que dans la troisième section.

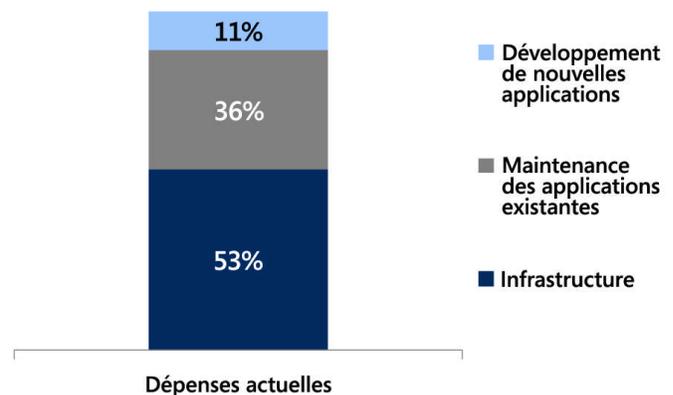


Figure 16 : Ventilation des dépenses informatiques.

Source : Microsoft.

¹⁴ Les coûts de développement de nouvelles applications ne comprennent que le coût de conception et d'écriture de l'application. En est complètement exclu le coût de l'hébergement sur une nouvelle infrastructure. Le cumul de ces coûts donne la répartition 80 % / 20 % habituelle.

2.5 Maîtrise des aspects économiques du cloud

Avec la technologie d'aujourd'hui, il n'est pas aussi évident que cela de tirer pleinement parti de ces avantages. Si, aux premiers temps de l'automobile, les ingénieurs ont dû revenir sur la conception des véhicules, les développeurs doivent aujourd'hui repenser la conception des applications. Les développeurs, voire les services informatiques plus sophistiqués, ont souvent des difficultés à mettre en place, à eux seuls, la mutualisation et l'agrégation de la demande. S'ils s'y prennent mal, ils risquent de voir les coûts de développement des applications augmenter (annihilant ainsi la part de budget libérée à cet effet) ou de ne profiter que d'une infime partie des économies décrites précédemment. L'approche à adopter varie selon que les applications sont des progiciels prêts à l'emploi (packagés) ou des solutions innovantes/personnalisées.

Applications prêtes à l'emploi :

Si la virtualisation des applications et leur migration sur les systèmes virtuels des clouds (Exchange virtualisé, par exemple) génèrent quelques économies, cette solution est loin d'être idéale et ne porte pas tous les fruits escomptés. Il y a deux raisons à cela. Premièrement, les applications conçues pour s'exécuter sur un seul serveur seront difficiles à faire évoluer sans programmation supplémentaire. Il faudra notamment prévoir l'équilibre de charge, le basculement automatique, la redondance et l'administration des ressources actives. Elles permettront difficilement d'agréger la demande et d'optimiser le taux d'utilisation des serveurs. Deuxièmement, les progiciels traditionnels ne sont pas écrits en vue de la mutualisation et le fait de les héberger dans le cloud ne change rien à cela. En cas de progiciels, la meilleure façon de profiter des avantages du cloud est d'opter pour des offres de SaaS (Logiciels + Services) comme Office 365, compatible avec une montée en charge et une mutualisation.

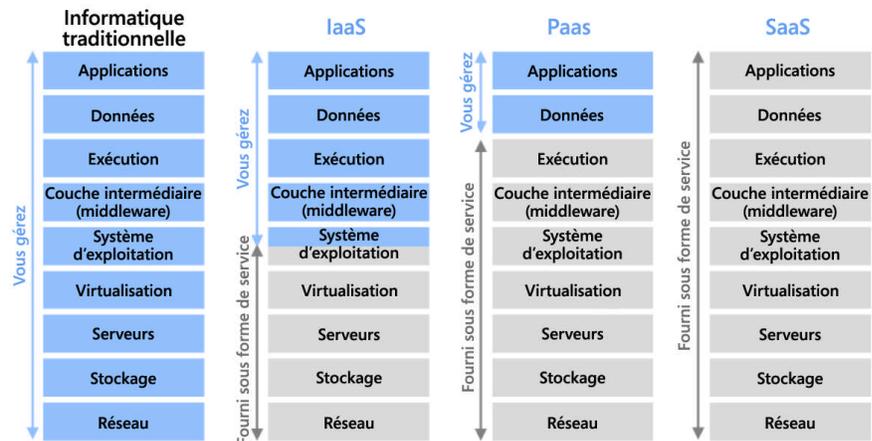


Figure 17 : Avantages du cloud.

Source : Microsoft.

Solutions innovantes et personnalisées : La mise en place d'une structure d'IaaS (Infrastructure-as-a-Service) permet de profiter de certains avantages financiers pour les applications existantes. Toutefois, la solution s'apparente un peu à une « voiture sans cheval », comme présenté dans l'introduction de ce document, en ce sens que la plateforme et les outils n'ont pas été conçus pour le cloud. Tout le potentiel du cloud computing ne peut se dégager qu'à condition d'investir dans une administration intelligente des ressources. Le gestionnaire des ressources doit comprendre l'état des ressources (réseau, stockage et puissance de calcul) et suivre l'activité des applications en cours d'exécution. Par conséquent, lors de l'écriture de nouvelles applications, mieux vaut privilégier le mode PaaS (Platform-as-a-Service) pour tirer le maximum du cloud. Une solution de PaaS propose des services partagés, une administration sophistiquée et des fonctions d'automatisation. Les développeurs ont alors tout loisir de se consacrer à la logique de l'application sans se préoccuper de sa future évolution.

Pour illustrer notre propos, prenons l'exemple d'une startup nommée Animoto qui a mis en place une solution d'IaaS (Infrastructure-as-a-Service) pour faciliter la montée en charge. En trois jours, elle a dopé sa capacité en ajoutant 3 500 serveurs pour desservir plus de 750 000 nouveaux utilisateurs. En examinant par la suite l'application, l'équipe d'Animoto découvrait pourtant qu'un fort pourcentage des ressources qu'elle payait n'était pas utilisé, souvent plus de 50 %, même dans un cloud supposé élastique. Ils ont repensé leur application et ont fini par abaisser les coûts d'exploitation de 20 %. Animoto est certes un témoignage de

réussite, mais ce n'est qu'en investissant dans une administration intelligente des ressources qu'elle a pu dégager tout le potentiel du cloud. Une solution de PaaS (Platform-as-a-Service) aurait apporté ces avantages directement sans autre tentative d'amélioration.

3. IMPLICATIONS

Dans cette section, nous aborderons l'incidence des aspects économiques du cloud évoqués plus haut. Nous examinerons la capacité des clouds privés à surmonter certaines barrières à l'adoption du cloud, et comparerons les coûts des clouds publics et privés.

3.1 Possibilités et obstacles

Les facteurs économiques décrits dans la section précédente influenceront considérablement sur l'informatique. Pour de nombreux responsables informatiques, le problème réside aujourd'hui dans le fait que 80 % du budget est consacré à la maintenance de l'infrastructure et des services existants. Il ne reste donc que peu de ressources pour l'innovation ou les requêtes des utilisateurs. Le cloud computing permet de libérer un grand nombre de ressources et de les affecter à l'innovation. Les demandes de technologies généralistes telles que l'informatique ont toujours connu une forte élasticité de prix (figure 18). C'est ainsi que de nombreux projets informatiques, aux coûts autrefois prohibitifs, deviennent viables du fait des économies du cloud. Cependant, la baisse du coût total de possession n'est que l'un des leviers favorisant l'innovation informatique :

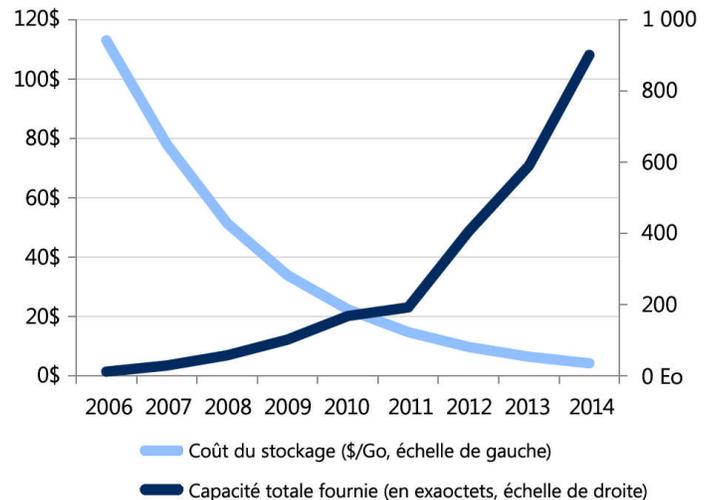


Figure 18 : Élasticité du prix du stockage.

Source : Coughlin Associates.

1. **Cette élasticité change la donne.** Comme nous l'avons vu précédemment, il revient quasiment au même, dans le cloud, de louer un système pendant 1 000 heures que de louer 1 000 systèmes pendant une heure. Les utilisateurs et les agences réalisent ainsi rapidement des tâches complexes, jusque-là inaccessibles pour des raisons de coût ou de temps. Le fait de pouvoir augmenter ou réduire instantanément la consommation des ressources donne lieu à une nouvelle classe d'expérimentation et d'esprit d'entreprise.
2. **L'élimination des dépenses d'investissement** réduit notablement la prise de risque et se traduira très certainement par une hausse des expérimentations. S'ensuivra une baisse des coûts liés au démarrage d'une opération, ou aux pannes et arrêts intempestifs. Lorsqu'une application n'a plus besoin de certaines ressources, celles-ci seront simplement retirées sans frais ni perte.
3. **Libre-service.** En activant les serveurs via un simple portail Web, plutôt que par une chaîne complexe d'approvisionnement et d'approbations, il est possible de réduire les frictions dans le modèle de consommation et d'accélérer la mise en place et l'intégration de nouveaux services. Un tel système permet par ailleurs de réaliser les projets plus vite et à moindre risque, tout en diminuant la charge administrative.
4. **Simplification.** La complexité constitue depuis longtemps l'un des principaux freins à l'innovation informatique. Du point de vue de l'utilisateur, le modèle SaaS (logiciels sous forme de services) engendre l'apparition de nouveaux logiciels conviviaux. Pour les développeurs, les solutions de PaaS (Platform-as-a-Service) simplifient grandement la création de nouvelles applications, de même que les voitures ont modernisé les transports.

Ces facteurs contribuent à accroître considérablement la valeur ajoutée par l'outil informatique. Le phénomène d'élasticité permet, par exemple, des applications de tarification en temps réel, de traitement d'événements complexes, d'optimisation de la logistique ou de simulation Monte Carlo, dont les charges de

travail exercent une demande quasi-infinie sur les ressources informatiques. Il est ainsi possible d'améliorer l'expérience et d'effectuer des analyses décisionnelles ou des calculs intensifs à grande échelle.

De nombreuses études font néanmoins ressortir une appréhension certaine vis-à-vis du cloud computing. Comme le montre la figure 19, cette appréhension porte notamment sur la sécurité, la confidentialité, la maturité et la conformité. De nombreux directeurs informatiques s'inquiètent également de la compatibilité des équipements existants : il n'est pas toujours simple de déplacer des applications dans le cloud.

- Sécurité et confidentialité.** Les directeurs informatiques doivent être en mesure de rendre compte de la sécurité et de la confidentialité des données à la direction générale. Les données et processus importants financièrement et stratégiquement sont souvent protégés par des exigences de sécurité complexes. Les anciens systèmes ont généralement été personnalisés pour respecter ces objectifs, et le passage à une architecture de cloud peut s'avérer délicat. De plus, la connaissance des mécanismes de sécurité standards du cloud est encore limitée, et les directeurs informatiques se sentent toujours plus à l'aise avec les systèmes actuels.
- Maturité et performance.** Le cloud exige des directeurs informatiques qu'ils comptent sur d'autres personnes pour la fourniture de services fiables et hautement disponibles. À la différence des pannes sur site, les interruptions dans le cloud ne passent pas inaperçues et risquent d'accroître les craintes.
- Conformité et souveraineté des données.** Les entreprises font l'objet d'audits et de divers contrôles, à la fois internes et externes (fisc, AMF, par exemple). Dans de nombreux pays, elles sont soumises à des obligations en matière de souveraineté des données, qui limitent sérieusement les possibilités d'hébergement des services de données. Les directeurs informatiques souhaitent identifier les clouds compatibles avec ces systèmes et les mesures à prendre pour qu'ils le soient.

Tandis que la majeure partie de ces questions trouvera aujourd'hui une réponse dans le cloud, certaines subsistent et incitent les responsables informatiques à envisager les clouds privés pour bénéficier des avantages du cloud et résoudre ces problèmes. Examinons ce concept plus en détail et évaluons le pour et le contre.

3.3 Clouds privés

Microsoft fait la différence entre clouds privés et publics, selon que les ressources informatiques sont partagées par plusieurs entreprises (cloud public) ou dédiées à une seule et même entreprise (cloud privé). Cette distinction est illustrée à la figure 20. Comparés aux centres de données virtualisés traditionnels, les clouds privés et publics bénéficient d'une administration automatisée (suppression des tâches répétitives) et d'un matériel homogène (réduction des coûts et plus grande souplesse). Du fait du partage à grande échelle caractéristique des clouds publics, l'une des principales différences entre clouds publics et privés réside dans l'échelle et la portée de la centralisation des demandes.



Figure 19 : Craintes vis-à-vis du cloud public.

Source : Sondage Gartner auprès des directeurs informatiques.

- **Les centres de données virtualisés traditionnels** permettent généralement la mise en commun des ressources à l'intérieur de limites organisationnelles.

Autrement dit, le service informatique d'une entreprise virtualise ses charges de travail tandis que les autres services en font, ou non, autant. Cela permet d'éviter certaines variations aléatoires, horaires (notamment si l'entreprise possède des bureaux dans le monde entier) ou liées à la charge de travail. Mais la taille du pool et la difficulté à déplacer les charges d'un système virtuel à un autre (difficulté exacerbée par le manque d'homogénéité des configurations matérielles) empêchent de profiter pleinement des avantages.

C'est là l'une des raisons pour lesquelles même les centres de données virtualisés connaissent une sous-utilisation. Le modèle d'application ne subit aucun changement, de sorte que la création d'applications n'est pas simplifiée.

- **Les clouds privés** vont au-delà de la virtualisation. Les ressources sont désormais mises en commun au niveau de l'entreprise, et non plus par entité organisationnelle individuelle.¹⁵ Par ailleurs, les charges sont facilement déplacées d'un serveur physique à l'autre, d'où une efficacité et une disponibilité optimales. Cela réduit encore l'impact des variations aléatoires, horaires ou liées à la charge de travail. De plus, les nouveaux modèles d'application optimisés pour le cloud (PaaS comme Azure) rendent le développement d'applications plus efficace et diminuent les coûts d'exploitation courants.
- **Les clouds publics** présentent les mêmes éléments d'architecture que les clouds privés, mais sont mieux à même de prendre en charge les différentes sources de variation. Ils constituent en outre le meilleur moyen de dissiper les variations sectorielles, de tenir compte de l'incidence géographique des variations horaires, ou encore de mettre en œuvre les avantages de la mutualisation.

Les clouds privés répondent à certaines des préoccupations mentionnées plus haut. Dans la mesure où ils possèdent un matériel dédié, ils trouvent plus facilement place à l'intérieur du pare-feu de l'entreprise et réduisent les craintes liées à la **sécurité et à la confidentialité**. En intégrant un cloud privé sur site, il peut être plus facile de remédier aux problèmes de **réglementation, de conformité et de souveraineté** touchant les services qui traversent les frontières juridiques. Lorsque ces préoccupations pèsent lourdement dans la décision d'un directeur informatique, l'investissement dans un cloud privé peut constituer la solution optimale.

Les clouds privés ne se différencient quasiment pas des clouds publics en matière de **maturité et de performance**. Les deux technologies sont développées en parallèle et arriveront donc ensemble à maturité.

		Administration automatisée	Matériel homogène	Nouveau modèle applicatif
Cloud public		✓	✓	✓
Cloud privé		✓	✓	✓
Serveur virtuel		✓	✗	✗
Serveur traditionnel		✗	✗	✗

Opérateur: Département, Informatique centralisée, Fournisseur tiers

Figure 20 : Comparaison entre virtualisation, cloud privé et cloud public.

Source : Microsoft. Les coches en gris indiquent une caractéristique facultative.

¹⁵ L'agrégation des différentes entités organisationnelles devient possible via deux technologies clés. D'une part, la migration dynamique permet de déplacer les systèmes virtuels sans interruption de service, d'où une optimisation plus dynamique. D'autre part, la mise en service et la facturation peuvent se faire en libre-service.

De nombreux niveaux de performance sont proposés par les clouds privés et publics, ce qui exclut tout avantage de l'un par rapport à l'autre.¹⁶

Si les clouds privés permettent de lever certaines inquiétudes, voyons dans le prochain paragraphe s'ils permettent de réaliser les économies décrites au préalable.

3.4 Compromis financier

L'analyse précédente démontre clairement que d'un point de vue conceptuel, le cloud public est mieux à même d'exploiter les avantages de la diversification. Il convient cependant d'examiner dans quelle mesure. La figure 21 montre que le cloud public permet de remédier à toutes les sources de variation, tandis que le cloud privé n'en traite que quelques-unes.

	Sources de fluctuation			
	Aléatoire	Heure du jour	Secteur	Ressources
Cloud privé	✓	✓	✗	✓
Cloud public	✓	✓	✓	✓

Figure 21 : Avantages de la diversification.

Source : Microsoft.

Par exemple, la question des variations sectorielles ne peut pas trouver de réponse dans un cloud privé. De même, les variations prévisionnelles dues à la croissance ne peuvent être résolues que partiellement dans une organisation qui centralise l'ensemble de ses ressources internes dans un cloud privé. Après modélisation de ces facteurs, nous obtenons les résultats présentés à la figure 22.

La courbe inférieure correspond au coût d'un cloud public (elle est identique à la courbe de la figure 15). La courbe supérieure représente le coût d'un cloud privé. La première est plus basse à tous les niveaux en raison du fort impact de l'agrégation de la demande et de l'effet de la mutualisation. Les clouds publics de portée internationale ont de fortes chances de s'étendre et de compter jusqu'à 100 000 serveurs, tandis que la taille du cloud privé d'une entreprise dépendra de sa demande et de son budget informatique.

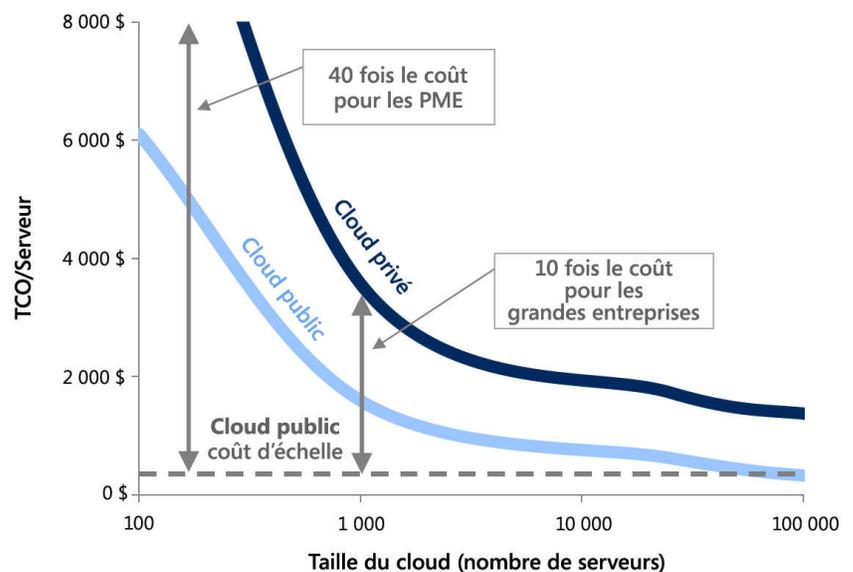


Figure 22 : Avantage financier du cloud public.

Source : Microsoft.

La figure 22 montre également que pour les entreprises possédant une très faible base de serveurs installés (<100), les clouds privés reviennent nettement plus cher qu'un cloud public. Ces entreprises ou services de petite taille n'ont d'autre solution que d'opter pour le cloud public si elles veulent exploiter les avantages du cloud computing. Pour les grandes entreprises dotées d'une base installée d'environ 1 000 serveurs, le cloud privé est envisageable mais son coût est environ dix fois supérieur à celui du cloud public pour une même unité de service. Cela est dû à l'effet conjugué de l'échelle, de la diversification de la demande et de

¹⁶ Les clouds privés permettent néanmoins un degré de personnalisation plus élevé que les clouds publics, d'où des gains de performances lors de certaines opérations de calcul. La personnalisation exige toutefois une forte implication, à la fois humaine et financière, du service Recherche et Développement, ce qui rend difficile toute comparaison prix/performance.

la mutualisation.

Outre une augmentation du coût total de possession, les clouds privés demandent un coût de déploiement initial - investissement qui doit tenir compte des pics de demande. Cela implique une budgétisation et un engagement séparés, et donc une augmentation du risque. Les clouds publics, en revanche, peuvent généralement être mis en service uniquement sur la base d'un paiement à l'utilisation.

3.5 Quel équilibre ? Mise en balance des avantages et des coûts d'un cloud privé

La figure 23 schématise la comparaison entre clouds privés et clouds publics. L'axe vertical mesure l'avantage financier du cloud public. Nous avons vu précédemment que le cloud public présente des avantages économiques inhérents, liés en partie à la taille du client. La position verticale de la bulle dépend par conséquent de la taille de la base de serveurs installés. L'axe horizontal représente la préférence de l'entreprise pour le cloud privé. La taille des cercles reflète la base totale de serveurs installés des entreprises de chaque type. Le quart inférieur droit du schéma représente ainsi les situations où le cloud privé est le plus intéressant (coût relatif faible et préférence élevée).

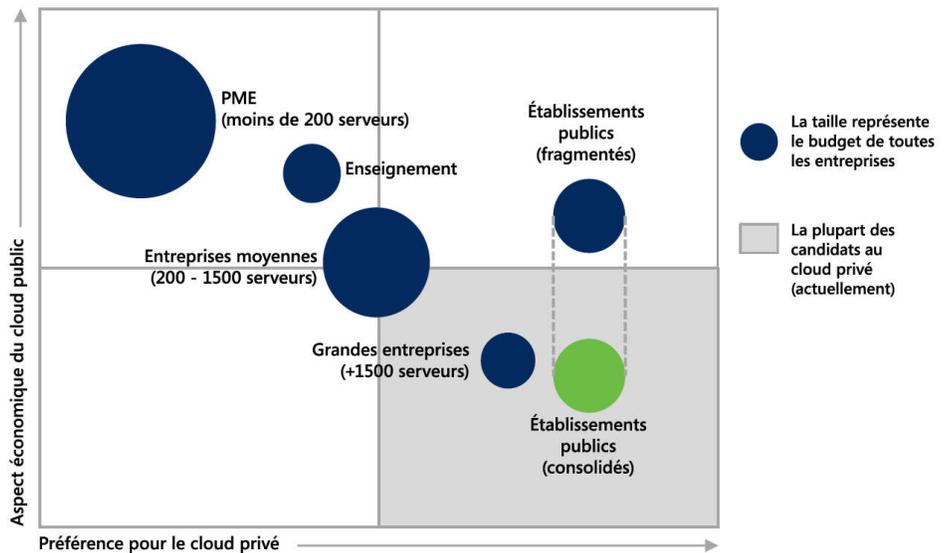


Figure 23 : Coût et avantages des clouds privés.

Source : Microsoft.

Le quart inférieur droit du schéma représente ainsi les situations où le cloud privé est le plus intéressant (coût relatif faible et préférence élevée).

Il va sans dire que la figure 23 est très schématisée. L'informatique n'est jamais monolithique dans ces secteurs d'activité. Les activités informatiques d'une entreprise sont segmentées en fonction du type des charges de travail, telles que messagerie ou progiciel de gestion intégré. Chacune de ces charges présente différents niveaux de sensibilité et d'échelle. Plusieurs études révèlent que l'attrance pour les solutions de cloud public varie grandement selon la charge de travail (figure 24).

Le fait est que de nombreux

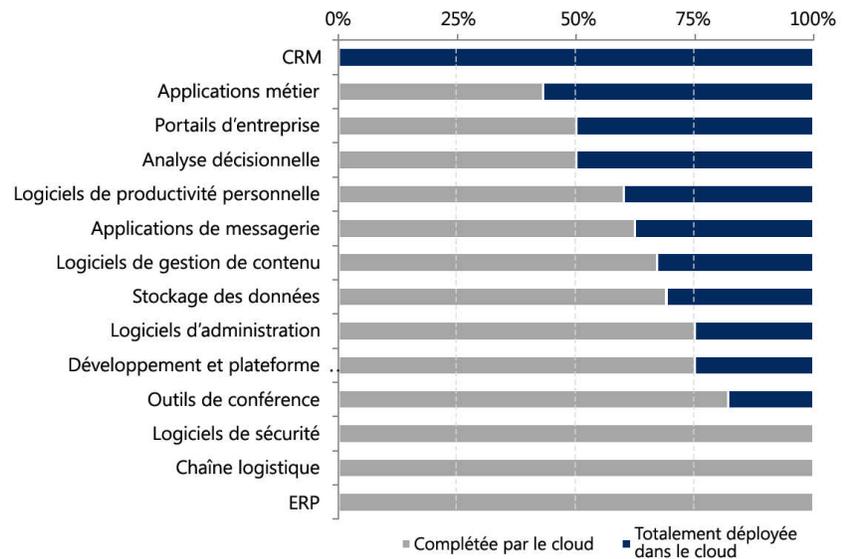


Figure 24 : Charges de travail prêtes pour le cloud (2010).

Source : Enquête Microsoft posant la question suivante : « Dans les 12 à 24 prochains mois, indiquez si une offre dans le cloud compléterait l'offre sur site ou la remplacerait entièrement ».

portefeuilles d'applications ont été développés au cours des 15 à 30 dernières années et sont étroitement imbriqués. Cela est particulièrement vrai pour les progiciels de gestion intégré (ERP) et les applications personnalisées connexes, dans les grandes entreprises disposant d'importants portefeuilles d'applications. Les applications de gestion de la relation client ou de collaboration et les nouvelles applications personnalisées, plus « isolées », sont plus faciles à déployer dans le cloud. Certaines ont parfois besoin d'être réintégréées dans les applications sur sites existantes.

Avant de tirer des conclusions définitives, il convient d'éviter le « syndrome de la voiture sans cheval » et de prendre en compte une évolution possible le long des deux axes (aspects économiques et préférence pour le cloud privé).

3.6 Transition vers le cloud sur le long terme

Comme nous en avons fait la remarque dans l'introduction de ce document, il est risqué de prendre des décisions dans les premières phases d'un changement majeur sans avoir une vision claire de la situation finale. Les responsables informatiques doivent concevoir leur architecture en fonction d'une vision à long terme. Il nous faut par conséquent déterminer quel sera l'impact des différentes forces à long terme sur la position des bulles de la figure 23.

Nous prévoyons deux mouvements importants. D'abord, **l'avantage économique du cloud public augmentera au fil du temps**. Plus la quantité de travail effectuée dans le cloud sera importante, plus les économies d'échelle décrites à la deuxième section seront réelles, et plus le surcoût des clouds privés augmentera. Les clients seront plus nombreux à profiter des économies mentionnées plus haut, en termes d'approvisionnement, de demande et de mutualisation. Comme l'illustre la figure 25, cela entraîne un déplacement vers le haut le long de l'axe vertical.

Dans le même temps, nous assistons à la **chute de certains freins à l'adoption du cloud**. Plusieurs études informatiques révèlent qu'au fil du temps, les problèmes de compatibilité, de sécurité, de fiabilité et de confidentialité seront résolus. Ce sera également le cas pour le cloud, ce qui se traduira par un déplacement vers la gauche sur la figure 25. Étudions plus en détail les facteurs à l'origine de ce dernier déplacement.

Évolution de la sécurité du cloud

Nous en sommes encore aux premiers stades de développement des clouds publics. Les domaines stratégiques que sont la sécurité et la fiabilité ne peuvent donc que s'améliorer. Des études suggèrent déjà que la messagerie dans le cloud est plus fiable que la plupart des systèmes sur site. Dans une solution de PaaS, l'application automatique des correctifs et des mises à jour dans les systèmes du cloud renforce notablement la sécurité des données et des applications. En effet, la majorité des vulnérabilités tirent profit des systèmes périmés. De nombreux experts de la sécurité soutiennent qu'il n'y a aucune raison de croire

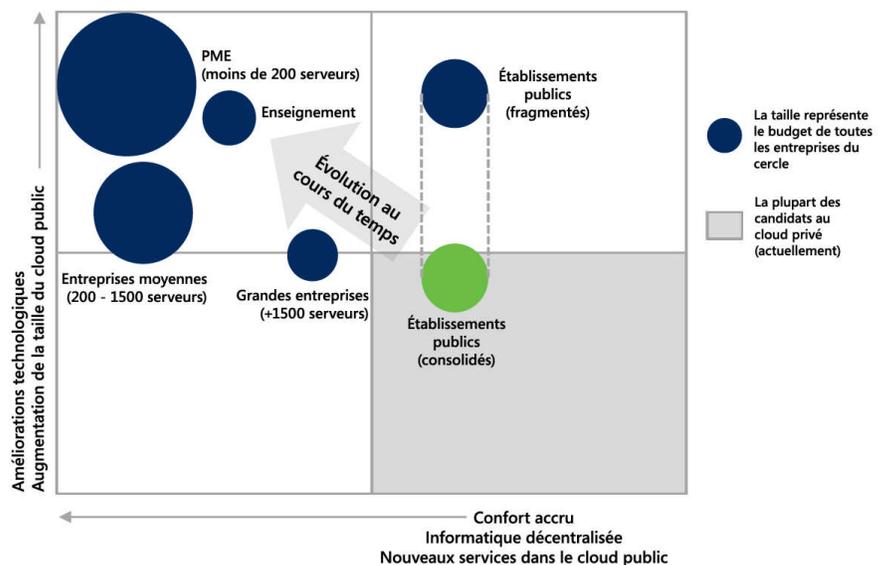


Figure 25 : Évolution prévisionnelle des préférences pour le cloud public et privé.

Source : Microsoft.

que les clouds publics sont moins sécurisés. En réalité, ils ont de fortes chances de devenir plus sûrs que les systèmes sur site en raison de la surveillance particulière des fournisseurs dans ce domaine et du haut niveau d'expertise qu'ils développent.

Vers une plus grande conformité des clouds

Les exigences de conformité peuvent émaner d'une entreprise, d'un secteur d'activité ou d'un gouvernement (directive européenne sur la protection des données, par exemple). Elles sont parfois difficiles à atteindre dans le cloud en l'absence de plateforme de développement robuste adaptée aux besoins de l'entreprise. À mesure que les technologies du cloud évolueront et que les exigences de conformité s'adapteront à ses architectures, le cloud deviendra de plus en plus conforme à ces exigences. Il deviendra ainsi accessible à un plus grand nombre d'entreprises aux charges de travail diverses. Prenons le cas des signatures électroniques. Aux débuts d'Internet, elles étaient rarement acceptées pour les contrats et les documents. Mais, alors que les technologies d'authentification et de chiffrement se sont améliorées et que les exigences de conformité ont évolué, elles sont tout naturellement devenues acceptables. Aujourd'hui, la plupart des contrats (y compris les contrats d'ouverture de compte bancaire ou d'emprunt) acceptent les signatures électroniques.

Le vaste groupe d'utilisateurs qui montrent une confiance croissante dans les clouds publics (PME et consommateurs de solutions de SaaS) constituera un levier formidable dans ce domaine. Cette force croissante continuera à exiger des gouvernements qu'ils s'adaptent au passage vers le cloud en modernisant la législation. Une telle révolution de la réglementation ferait du cloud public une solution plus viable pour les grandes entreprises. Cela entraînerait un déplacement des segments d'activité sur l'axe horizontal, vers une préférence pour les clouds publics.

L'informatique décentralisée restera majoritaire

De nombreuses transitions technologiques étaient auparavant menées non pas par les directeurs informatiques, mais par les divisions, les décideurs de l'entreprise, les développeurs et les utilisateurs. Qui plus est, elles allaient souvent à l'encontre des recommandations des directeurs informatiques. Par exemple, les utilisateurs et les administrations ont été les premiers à adopter les PC et les serveurs, avant que les stratégies informatiques des grandes entreprises ne les incluent officiellement. Plus récemment, le même phénomène s'est produit avec les téléphones portables. C'est l'adoption du public qui a poussé les entreprises informatiques à prendre en charge ces équipements. Nous observons un cas de figure semblable avec le cloud. Les développeurs et les différents services ont commencé à recourir aux services du cloud, souvent sans que le service informatique soit au courant (d'où le nom de « clouds voyous »). De nombreux utilisateurs professionnels ne veulent pas attendre que le service informatique leur fournisse un cloud privé. Pour eux, efficacité et commodité priment sur les politiques.

L'impatience n'est pas le seul facteur à l'origine des « clouds voyous ». Des contraintes budgétaires sans cesse croissantes conduisent également les utilisateurs et les services à adopter les solutions du cloud public moins onéreuses, qu'ils ne pourraient se permettre via les canaux traditionnels. Par exemple, lorsque Derek Gottfrid a voulu traiter les 4 To d'archives du New York Times pour les héberger en ligne, il a fait appel au cloud sans demander l'avis du service informatique du Times.¹⁷ De la même manière, du fait de l'incroyable transparence tarifaire qu'offre le cloud, les directeurs généraux et financiers feront pression sur les directeurs informatiques pour les inciter à passer dans le cloud public.

Les directeurs informatiques doivent admettre que ces comportements sont monnaie courante au début d'un changement majeur. Soit ils développent et implémentent rapidement un cloud privé offrant les mêmes fonctionnalités, soit ils mettent en place des stratégies qui incorporent ces comportements dans les habitudes informatiques

¹⁷ <http://open.blogs.nytimes.com/211/07/01/self-service-prorated-super-computing-fun/>

Évolution rapide des perceptions

L'adoption massive du modèle SaaS par les grandes entreprises témoigne des changements de perception (figure 26). Elle indique que même les entreprises exigeantes de grande taille évoluent. Il y a à peine quelques années, très peu de grandes entreprises étaient prêtes à faire passer leur messagerie dans le cloud, avec toutes les données confidentielles qu'elle contenait. C'est pourtant ce qui se passe aujourd'hui.

Plus les expériences positives renforceront l'intérêt dans le cloud, plus ce cycle vertueux s'intensifiera et favorisera une acceptation plus généralisée du concept de cloud computing.

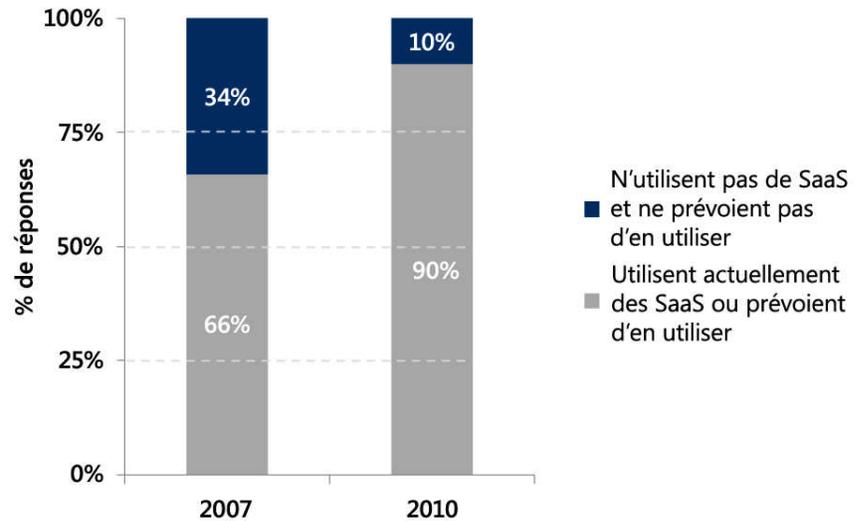


Figure 26 : Adoption croissante des solutions de SaaS.
Source : Gartner.

En résumé, s'il existe indéniablement des obstacles à l'adoption du cloud aujourd'hui, ils sont amenés à tomber progressivement. De nouveaux freins à cette adoption sont susceptibles de surgir, mais l'avantage économique du cloud public progressera à mesure que les fournisseurs de cloud computing dévoileront les aspects économiques avantageux analysés à la section précédente. Le penchant pour les clouds privés est motivé principalement par les questions de sécurité et de conformité concernant les charges de travail actuelles. Or, la rentabilité et l'agilité du cloud public donnera justement lieu à de nouvelles charges.

Reprenons l'analogie de la « voiture sans cheval ». Si les automobiles ont connu un tel succès, ce n'est pas seulement parce qu'elles étaient plus rapides et plus pratiques (et éventuellement plus abordables) que les voitures à cheval. Le système de transport entier devait changer. Pour permettre cette transition, il a fallu développer un réseau routier, des écoles de conduite, des cartes et une signalisation précises, des règles de sécurité et un réseau international de stations d'essence. Chaque avancée successive révélait les points forts de la voiture. Au final, l'automobile a changé jusqu'aux habitudes de vie des populations. C'est ainsi que vers le milieu du XX^e siècle, les banlieues ont fait leur apparition. Elles ont donné naissance à une nouvelle classe de travailleurs, qui effectuent des déplacements journaliers, et avec elle, à une nouvelle demande de voitures. Ce changement de comportement a engendré un tel enchaînement de réactions positives que l'automobile est inexorablement devenue une composante essentielle et irremplaçable de la vie moderne.

Nous pensons que le cloud connaîtra un essor similaire, et pas seulement en raison des aspects économiques, des développements technologiques qualitatifs et de l'évolution des perceptions. Il sera porté par une série de réorientations de la part des informaticiens, organismes de régulation, opérateurs téléphoniques, éditeurs de logiciels, intégrateurs de systèmes et fournisseurs de plateformes cloud. La valeur du cloud augmentera parallèlement à son adoption.

4. LA TRANSITION VERS LE CLOUD

Nous ne sommes qu'au début de la transition vers le cloud, d'où une certaine confusion quant à l'orientation de cette transformation progressive. Dans ce document, nous ne nous cantonnons pas aux technologies actuelles. Nous examinons plus particulièrement les facteurs économiques sous-jacents du cloud pour définir une destination : ce vers quoi ce bouleversement et cette innovation conduiront notre industrie. Notre analyse fait ressortir une adoption du cloud à long terme favorisée par trois importantes économies d'échelle : (1) les centres de données de plus grande envergure peuvent déployer des ressources de calcul pour un coût plus faible que les plus petits centres de données ; (2) l'agrégation de la demande optimise l'utilisation de ces ressources, notamment dans les clouds publics ; et (3) la mutualisation réduit les coûts de personnel liés à la maintenance des applications pour les grands clouds publics. Enfin, le cloud fournit un niveau sans précédent d'élasticité et d'agilité qui feront naître de nouvelles solutions et applications.

Pour les entreprises de toutes tailles, le cloud représente une opportunité inouïe. Il offre la possibilité de sortir de la longue tradition des professionnels de l'informatique qui consacrent 80 % de leur temps et de leur budget à maintenir les systèmes en état de marche, laissant peu de place à l'innovation. **Les services de cloud computing permettront aux divisions informatiques de privilégier l'innovation et de laisser les activités annexes aux mains de fournisseurs fiables et rentables.** Grâce à ce nouveau modèle, les responsables informatiques seront à même de proposer de nouvelles solutions considérées auparavant comme trop coûteuses ou trop complexes à mettre en œuvre. Cela est d'autant plus vrai pour les plateformes de cloud (PaaS), qui accélèrent et simplifient considérablement la création de nouvelles applications tirant pleinement parti des avantages du cloud.

Cela ne se fera pas du jour au lendemain. Les responsables informatiques doivent développer une nouvelle vision de l'avenir sur 5 à 10 ans et admettre qu'avec leur entreprise, ils jouent un rôle majeur. Il leur faut tracer une voie entre la situation actuelle et l'avenir. La première étape est essentielle. Elle consiste à segmenter le portefeuille d'applications existantes (figure 27). Les avantages liés aux facteurs économiques et à l'agilité sont plus importants pour certaines applications. Leur migration doit donc avoir lieu en premier. Il existe cependant bel et bien des obstacles à l'heure actuelle. S'il apparaît évident que nombre d'entre eux tomberont au fil du temps, comme nous l'avons souligné dans la section précédente, le cloud ne convient pas encore à certaines applications. La migration d'applications étroitement intégrées présentant un schéma d'utilisation relativement stable n'est sans doute pas justifiée. De la même manière que certaines applications pour mainframes ne sont jamais passées à l'architecture client/serveur. Si les nouvelles applications personnalisées ne posent pas le problème de l'ancienneté, leur conception selon un modèle robuste et capable de monter en charge n'est pas chose facile. Les plateformes optimisées pour le cloud (PaaS) simplifient considérablement cette tâche.

Cette transition est délicate. Si le service informatique agit trop vite dans des domaines pour lesquels le cloud n'est pas prêt, il prend le risque de compromettre la continuité métier, la sécurité et la conformité. S'il

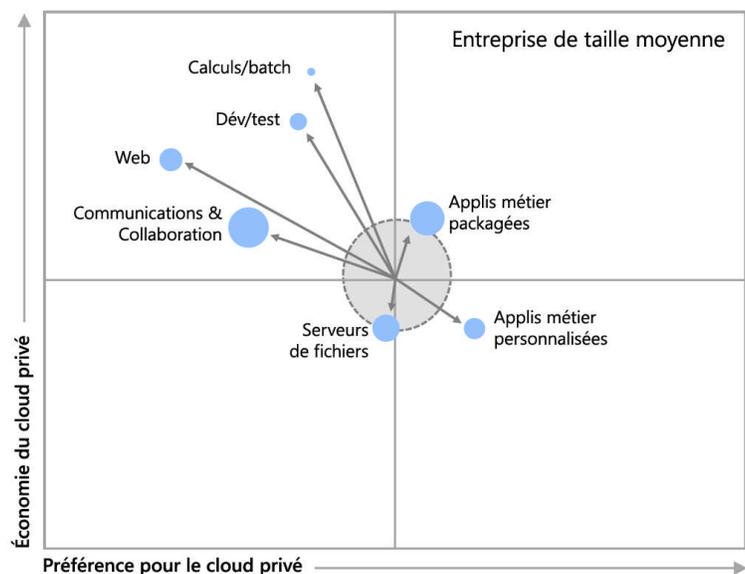


Figure 27 : Segmentation du portefeuille informatique.

Source : Microsoft.

progresses trop lentement, en revanche, l'entreprise sera désavantagée en termes de coût, d'agilité ou de valeur face à des concurrents qui profitent pleinement des fonctionnalités du cloud. Une transition trop lente fait également courir le risque à l'entreprise de voir se développer plusieurs solutions de cloud de manière désordonnée (clouds voyous), auquel cas le directeur informatique perdrait le contrôle de l'infrastructure informatique. En somme, les responsables informatiques qui mèneront le train du cloud seront en mesure de contrôler et de façonner cette transition ; ceux qui prendront le train en marche perdront progressivement tout contrôle.

Pour diriger cette transition, les responsables informatiques doivent adopter une vision à long terme de leur architecture informatique. Certains y voient l'émergence d'un nouveau rôle. Celui d'un architecte des services du cloud, chargé d'identifier les applications et services à déplacer vers le cloud, ainsi que le moment, en s'appuyant sur un script commercial et une connaissance approfondie des possibilités offertes par le cloud. Il convient avant tout de faire l'inventaire des ressources et des stratégies de l'entreprise. Cela implique une classification des applications et des données, visant à définir les exigences à leur appliquer en termes de stratégie et de performance (par exemple, stratégie de conservation des données confidentielles ou top secret). En fonction de cette classification, les responsables informatiques jugeront certaines activités informatiques adaptées à un cloud public et considéreront que d'autres méritent un investissement dans un cloud privé. Cette approche permet de tirer pleinement parti des opportunités du cloud, tout en cherchant un équilibre entre les critères économiques d'une part, et la sécurité, la performance et le risque d'autre part.

Pour y parvenir, les responsables informatiques ont besoin d'un partenaire qui prenne position pour la vision à long terme du cloud et de ses opportunités, un partenaire qui accepte de s'éloigner des architectures informatiques actuelles. Le partenaire idéal est en même temps tenu de connaître en profondeur les réalités de l'informatique d'aujourd'hui, ses enjeux et le meilleur moyen d'assurer le voyage vers le cloud. Il ne provoquera pas un changement trop rapide ni ne favorisera le statu quo. Les clients ont besoin d'un partenaire ayant lui-même étudié le meilleur moyen de marier les structures informatiques actuelles et le cloud plutôt que de laisser cette tâche au client en ignorant les complexités de cette transformation.

Chez Microsoft, nous sommes de fervents supporters du cloud. Nous fournissons à la fois des logiciels sous la forme de services commerciaux (Office 365) et une plateforme cloud computing (Windows Azure). Office 365 comprend des applications auxquelles les utilisateurs sont habitués, telles que la messagerie Exchange et le logiciel de collaboration SharePoint, fournies via le cloud Microsoft. Windows Azure, notre plateforme de cloud computing, permet aux utilisateurs de créer leurs propres applications et activités informatiques dans le cloud, dans un environnement sécurisé et capable de monter en charge. L'écriture d'applications pour le cloud qui soient robustes et aptes à monter en charge n'est pas chose aisée. C'est pourquoi nous avons créé Windows Azure. Cette plateforme nous a permis d'exploiter l'expertise Microsoft afin de créer des applications optimisées pour le cloud, telles que Office 365, Bing et Windows Live Hotmail. Plutôt que de déplacer simplement des systèmes virtuels vers le cloud, nous avons construit une solution PaaS afin de simplifier la tâche des développeurs et des administrateurs système.

Microsoft amène également dans le cloud l'une des plus riches communautés de partenaires au monde. En effet, nous possédons plus de 600 000 partenaires dans plus de 200 pays, à l'écoute de plusieurs millions d'entreprises. Nous collaborons déjà avec des milliers de partenaires sur la transition vers le cloud. Ensemble, nous aspirons à créer l'un des clouds les plus sécurisés, les plus fiables et les plus disponibles du monde.

Depuis une trentaine d'années, Microsoft tisse d'étroites relations avec les entreprises informatiques, leurs partenaires et leurs conseillers. Nous sommes donc en mesure de comprendre les défis auxquels les services informatiques sont aujourd'hui confrontés. Microsoft adhère à la vision du cloud et possède en outre l'expérience nécessaire pour aider les responsables informatiques à franchir le pas.

Depuis longtemps, Microsoft a su donner vie à sa conception de l'avenir. Bill Gates a fondé Microsoft dans le dessein de mettre un PC dans chaque foyer et sur chaque bureau, à une époque où seules les très grandes entreprises avaient les moyens de posséder des ordinateurs. Aujourd'hui, grâce entre autres à Microsoft et à ses partenaires, plus d'un milliard de foyers et de bureaux possèdent un ordinateur. Plusieurs millions de

développeurs et d'entreprises vivent de leurs ordinateurs et nous sommes fiers d'avoir participé à cette réussite.

Nous avons aujourd'hui le projet de faire en sorte que chaque foyer, bureau et appareil mobile bénéficie du cloud computing. Les puissants aspects économiques du cloud nous entraînent tous dans cette voie. Rejoignez Microsoft et ses partenaires dans leur voyage et faites de cette vision une réalité.

