



## White Paper

# Une nouvelle génération de solutions scale-out pour la transformation numérique

Commandité par : IBM

Peter Rutten  
September 2016

## L'AVIS D'IDC

---

La transformation numérique n'est pas un simple effet de mode. L'informatique est passée du back office au front office dans presque tous les domaines de l'entreprise, menée par ce qu'IDC appelle la 3e Plateforme, dont les piliers sont le mobile, le social, le cloud, le big data et l'analytique. Dans ce nouvel environnement, les dirigeants d'entreprise doivent améliorer leur compétitivité par la transformation numérique - tirer profit de l'ensemble des technologies numériques grâce à une innovation du modèle organisationnel, opérationnel et commercial afin de mettre en œuvre de nouvelles stratégies de croissance.

Du point de vue informatique, il existe une série plutôt simple d'applications que les entreprises peuvent adopter au fur et à mesure de leur transformation et qui les aidera à savoir quelle direction prendre. IDC donne un aperçu de cette série d'applications afin d'expliquer les différents stades de maturité de ce parcours. Cette série débute dans un territoire familier du Web et du cloud ; puis elle se dirige vers le terrain moins exploré de l'open source et des applications de nouvelle génération ; et enfin, elle entre dans le nouveau monde de l'analytique avancée des données, de l'analytique prédictive et de l'analytique cognitive. Du point de vue de l'informatique, c'est cela que signifie concrètement la transformation numérique.

Mais il est indispensable de garder en tête l'infrastructure nécessaire pour mettre cette série à exécution. Toutes ces applications seront d'autant plus bénéfiques qu'elles reposent sur une infrastructure de type scale-out solide et à haute performance. Pourtant, lorsqu'elles évaluent les offres scale-out sur le marché, la plupart des entreprises comparent uniquement les systèmes qui sont disponibles sur l'architecture processeur dominante aujourd'hui. Elles pensent effectuer une solide comparaison des fournisseurs en considérant les caractéristiques des différents fournisseurs, alors que ces derniers utilisent les mêmes processeurs pour le moteur de leurs produits.

Mais ces entreprises ne prennent pas en considération les avantages que peuvent offrir d'autres types de processeurs plus puissants, plus précisément la gamme IBM Power Systems. Les serveurs OpenPOWER reposant sur la technologie POWER8 sont hautement évolutifs, faciles à mettre en place dans des datacenters, fonctionnent bien avec les applications émergentes et souvent pour des coûts moins élevés. De plus, de nombreuses technologies développées par les membres de la fondation OpenPOWER permettent d'accélérer d'autant plus le débit I/O et la vitesse de traitement.

IDC estime que le fait d'exclure les produits OpenPOWER de l'évaluation comparative des offres scale-out impacte la qualité et l'objectivité de cette évaluation et peut, au bout du compte, empêcher une entreprise de sélectionner l'infrastructure qui serait la plus adaptée et la plus rentable pour entamer son parcours vers la transformation numérique.

## CONTENU DE CE LIVRE BLANC

---

Ce livre blanc aborde le voyage vers la transformation numérique dans lequel se sont embarquées la plupart des entreprises (volontiers ou contre leur gré) aujourd'hui, explique pourquoi une solide infrastructure scale-out est un des éléments essentiels de ce parcours, et donne quelques indications pour la mise en place du scale-out. Ce livre blanc définit une feuille de route concernant l'adoption du scale-out dans le cadre de la transformation numérique comportant trois phases, elles-mêmes constituées de plusieurs étapes, allant de l'exécution de simples applications Web sans-état sur une infrastructure scale-out, puis le cloud, jusqu'à la modélisation prédictive. Il présente ensuite différents paramètres requis pour la mise en place d'une infrastructure scale-out et examine les serveurs de type scale-out du marché actuel. Enfin, ce livre blanc s'intéresse au nouveau genre de serveurs scale-out d'IBM sur POWER8 et à certains des éléments différenciateurs du portefeuille POWER8 scale-out d'IBM. Juste avant la conclusion, le livre aborde les défis et les opportunités qui se présentent à IBM et aux clients.

## SITUATION GENERALE

---

### L'ère de la transformation numérique

La transformation numérique n'est pas qu'un effet de mode, mais désigne l'approche qu'adoptent les entreprises pour mettre en œuvre des changements dans leurs modèles commerciaux et leurs écosystèmes en tirant profit des compétences numériques. IDC identifie cinq stades de maturité en ce qui concerne les progrès que les entreprises ont faits vis-à-vis de la transformation numérique (les pourcentages sont issus de l'étude *Digital Transformation MaturityScape Benchmark Survey* d'IDC, réalisée en Février 2015).

Le groupe des réfractaires (14,2 %), en queue de peloton, offre une expérience clients peu satisfaisante et adopte une attitude défensive en ce qui concerne le numérique. La catégorie suivante est celle des explorateurs numériques (31,8 %) qui proposent des produits, des services et des expériences numérisés, mais intégrés de manière inégale et inconsistante. Le troisième groupe est constitué des joueurs numériques (32,4 %) qui fournissent de manière constante des produits, services et expériences, sans être novateurs. Le quatrième segment est celui des transformateurs numériques (13,6 %) qui sont des leaders sur leurs marchés, offrant des produits, services et expériences novateurs. Enfin, en première ligne se trouvent les perturbateurs du numérique (9 %), qui restaurent les marchés existants et en créent de nouveaux à leur propre avantage.

Aujourd'hui donc, plus de 50 % des entreprises font désormais partie des catégories Joueurs, Transformateurs ou Perturbateurs. L'autre moitié, elle, est de plus en plus menacée de perdre la bataille de la compétitivité.

## *En quoi une infrastructure de type scale-out solide est-elle indispensable pour la transformation numérique ?*

Le scale-out, c'est la création de réseaux de serveurs et de matériel de stockage qui sont massivement parallèles et évolutifs de manière linéaire, présentant une redondance intégrée de sorte que si une unité tombe en panne, une autre prend le relais. Le passage au scale-out à grande échelle a commencé par le calcul de haute performance (CHP), porté notamment par Google et Amazon au début des années 2000, puis il s'est rapidement imposé dans l'ensemble des centres de données du monde entier.

Les volumes de données étaient en augmentation, et les anciens systèmes scale-up sur Unix et ordinateurs centraux étaient considérés comme trop onéreux, trop complexes et incapables de prendre en charge le nombre croissant de logiciels ouverts (ils ont cependant considérablement rattrapé leur retard depuis). L'explosion de la bulle internet, et la récession qui suivit, ont marqué l'entrée dans une ère de budgets informatiques réduits augmentant la demande pour un matériel bon marché exécutant Windows comme Linux. Depuis, les volumes de données ont explosé et le besoin pour les entreprises d'analyser ces grandes quantités de données est devenu critique. Autre évolution majeure : l'infrastructure de type scale-out basée sur Linux et les logiciels open source est devenue le moyen le plus novateur et le plus rentable de se lancer dans la transformation numérique.

Mais alors que les clients commencent à se lancer dans cette transformation, l'incertitude autour de l'infrastructure qu'il leur faut, pas seulement pour les 12 ou 24 prochains mois, mais pour les 3 à 5 ans à venir, demeure. Les entreprises qui font partie des catégories Explorateurs et Joueurs numériques, par exemple, cherchent peut-être à investir dans une infrastructure de type scale-out pour quelques nouvelles applications commerciales qu'elles développent actuellement et qu'elles ne souhaitent pas exécuter sur leurs systèmes scale-up existants. Les services informatiques de ces entreprises doivent garder en mémoire que tous les boîtiers scale-out ne sont pas créés de la même manière et que les applications de transformation numérique évolueront de manière phénoménale, allant bien au-delà des quelques nouvelles applications qu'ils ont l'intention de lancer aujourd'hui.

L'investissement dans un matériel standard d'entrée de gamme aujourd'hui donnera certainement lieu à l'apparition d'obstacles dans la transformation numérique de l'entreprise. Il est recommandé de prendre en considération un investissement dans le scale-out comme une expansion d'un environnement de scale-out existant pour un ensemble spécifique d'applications d'entreprise actuelles et futures nécessaires dans la prise en charge de la transformation numérique. Chacune des applications qui, ensemble contribuent à la transformation numérique, est intensément axée sur les données. Ces applications reposent sur les bases de données ou sur les données de diffusion, et doivent traiter des volumes de données toujours plus importants de manière toujours plus complexe et rapide afin que l'entreprise tire un avantage concurrentiel de ces informations détaillées.

Ce livre blanc présente une feuille de route pour les applications de la transformation numérique en trois étapes qui démontrent qu'un environnement scale-out doit être extrêmement solide et hautement performant tout en étant rentable.

## ***Mise en garde : les dépenses d'exploitation du scale-out***

Le fait que le scale-out ait été un choix d'infrastructure populaire, même avant que la transformation numérique ne devienne une tendance forte, s'explique par de nombreuses raisons. L'une d'entre elles est son extensibilité illimitée. La capacité à ajouter rapidement de nouvelles charges de travail en ajoutant quelques boîtiers en est une autre. La création de clouds privés ou hybrides et la connexion fluide à des clouds publics constituent un aspect important. La flexibilité du scale-out par rapport à la nature plus complexe du scale-up joue un rôle, sans oublier, bien sûr, l'engagement financier moins important en termes de dépenses d'investissement (CAPEX) que représente le scale-out à une époque caractérisée par une évolution technologique rapide.

Mais ce que l'on a ignoré, ce sont les dépenses d'exploitation (OPEX). La plupart des infrastructures de type scale-out installées aujourd'hui ont un très faible taux d'utilisation, même si elles sont virtualisées. La virtualisation commune dans le cadre de laquelle le système d'exploitation (OS) n'est pas partagé a eu un impact sur les OPEX, car chaque système d'exploitation (OS) dans chaque machine virtuelle (MV) doit être fourni, déterminé par un cycle de vie et géré, sauvegardé et protégé par une garantie de reprise après sinistre. IDC estime qu'entre 2010 et 2020, la prolifération des serveurs multipliera par quatre les coûts de gestion et d'administration de ces derniers, et se traduira également par une augmentation des dépenses en matière d'énergie, de refroidissement et d'empreinte environnementale des centres de données. En d'autres termes, les OPEX peuvent atteindre un niveau qui pèse bien plus que prévu sur l'entreprise et sur ses coûts.

Pour parer des OPEX hors de contrôle, les services informatiques doivent envisager des systèmes de scale-out à haute performance ayant forts taux d'utilisation et qui nécessitent un nombre moins important de systèmes physiques dans le centre de données. Ces systèmes fonctionneront en réduisant les coûts liés à la gestion et à l'administration des serveurs, à l'énergie et au refroidissement, et à l'espace du centre de données. Ils offrent également d'autres avantages, comme une baisse des tâches de réplication et une réduction de la complexité pour la reprise après sinistre. Enfin, ils permettent aux employés déchargés des opérations d'entretien de se concentrer sur les tâches qui génèrent des revenus et qui améliorent les performances.

## **Une feuille de route pour l'adoption du scale-out dans le cadre de la transformation numérique**

### ***Les notions élémentaires***

La transformation numérique nécessite qu'une redéfinition de la manière que l'entreprise à de générer des revenus et de monétiser des produits et des services, tout en mettant un fort accent sur une expérience clients contextualisée et personnalisée et sur l'efficacité des opérations commerciales. Elle demande également de centrer l'innovation en matière de produits et de services sur l'expérience des clients et de modifier le fonctionnement de l'entreprise afin qu'elle soit la plus souple et la plus dynamique possible ; d'abord en ayant une meilleure compréhension et en s'adaptant au marché avec des produits novateurs, puis en développant des produits en s'appuyant sur sa capacité à anticiper les tendances du marché, et enfin en créant des produits qui réorganisent complètement le marché, pouvant être un facteur déstabilisant.

En ce qui concerne les applications et l'infrastructure nécessaires à la transformation numérique, la première étape (que la plupart des sociétés ont déjà intégrée) est l'adoption d'applications cloud.

### Le scale-out pour les applications sans état

L'infrastructure de type scale-out est très bien adaptée aux applications sans état, comme le VDI et de nombreuses applications Web. « L'état » est la capacité d'une application à conserver des informations personnalisées, comme l'identité d'un utilisateur ou le bouton sur lequel l'utilisateur vient de cliquer. Les applications *avec état* doivent trouver un moyen de conserver ces informations soit dans un client, dans la couche Web, dans une base de données ou dans une mémoire cache. Les applications *sans état* ne nécessitent pas de telles données. Les applications sans état sont donc adaptées au scale-out, car elles n'ont pas de difficulté à partager l'état de l'utilisateur entre les serveurs. Cela permet de faciliter l'activation d'un plus grand nombre de serveurs, selon les besoins pour ces applications.

### Le scale-out pour les applications Web

La principale-préoccupation concernant les applications Web est généralement la réactivité et l'extensibilité. Les utilisateurs exigent une réactivité immédiate, et pour les applications Web avec un volume de trafic élevé, cela représente un réel défi. L'extensibilité aide à améliorer le temps de réponse en ajoutant plus de serveurs (scale-out) ou en ajoutant plus d'unités centrales (ou des plus puissantes) dans un serveur à nœud unique (scale-up). Les deux approches sont aujourd'hui utilisées avec succès.

Le scale-out peut être avantageux pour les applications Web, car le matériel peut être attribué par petits paliers. Si les applications Web étaient auparavant considérées comme étant moins critiques que les charges de travail de traitement des données de base d'une entreprise, cela tend également à changer, et les applications Web sur un système de scale-out doivent être protégées grâce à des solutions d'équilibrage des charges, de virtualisation et/ou de mise en grappe.

Une autre situation préoccupante est que la manière de travailler des développeurs d'applications les incite à adopter un style de développement plus flexible et souple à l'aide d'une architecture composite. L'utilisation de microservices, à savoir des services qui sont créés de manière indépendante et détaillée et qui peuvent être assemblés en solutions applicatives, constitue une évolution majeure du secteur. L'infrastructure de développement et d'exécution de ces applications doit prendre en charge cette approche modulaire, présenter une structure semblable au cloud et être pleinement activée sur mobile, des exigences qu'une infrastructure de type scale-out disposant de la bonne plateforme de développement et des bons outils peut satisfaire.

### Le scale-out pour le cloud

Le cloud permet de réaliser des économies d'échelle. Un cloud privé ou hybride qui a été conçu sur une infrastructure de type scale-out peut être extrêmement avantageux s'il est créé avec de solides outils de gestion du cloud pour la surveillance, l'entretien, la haute disponibilité (HD) et la reprise après sinistre (RS). Sans ces outils, les avantages du scale-out peuvent vite s'évaporer. Un cloud classique se compose d'un grand nombre de serveurs à bas coût identiques (dans les centres d'hyperconvergence, leur densité est souvent optimisée, soit 0,5 unité par socket et aucun matériel redondant) qui sont installés en réseau. En attribuant de manière dynamique les ressources

demandées à un utilisateur final et en instaurant un tarif à l'utilisation, les clouds parviennent à de très hautes performances, plus hautes, par exemple, que les serveurs non virtualisés, les serveurs virtualisés ou les environnements répartis qui ne sont pas gérés comme un cloud. Ce constat vaut pour les clouds privés, hybrides et publics.

Il subsiste toujours un doute sur le marché quant à la définition exacte du cloud et aux différences entre les clouds privés, hybrides et publics. Le service informatique se retrouve souvent confronté à un mandat de « mise dans le cloud », généralement pour des raisons de coûts perçus. Pour faciliter la discussion qui a lieu dans de nombreuses entreprises, les paragraphes suivants expliquent comment IDC définit le cloud (voir *IDC's Worldwide IT Cloud Services Taxonomy, 2015*, IDC n° 258348, septembre 2015). Les lecteurs qui s'estiment être suffisamment informés sur le sujet peuvent passer cette section :

- **Les services de cloud public** sont partagés entre des entreprises et/ou des clients qui n'ont pas de lien, sont ouverts à un univers très peu limité d'utilisateurs potentiels et sont conçus pour un marché, non pour une seule entreprise.
- **Les services de cloud privé** sont partagés au sein d'une même entreprise ou d'une entreprise élargie, avec un accès restreint et un certain niveau de ressources mises à disposition, et sont définis / contrôlés par l'entreprise, loin du contrôle proposé par les offres de cloud public. Dans le monde des services de cloud privé, il existe deux principales options :
  - **Cloud privé d'entreprise.** Dans ce scénario de cloud privé, soit une entreprise acquiert un système de services cloud pré-intégrés, soit elle intègre les éléments des logiciels et du matériel et exploite le service cloud pour sa propre utilisation. L'entreprise fait parfois appel à un tiers pour les services d'intégration et/ou opérationnels. Un cloud privé d'entreprise peut être exécuté dans le propre centre de données de l'entreprise ou peut être co-localisé dans une installation tierce.
  - **Cloud privé hébergé (CPH).** Dans cette situation de cloud privé, les prestataires de services cloud commerciaux tiers proposent aux clients d'accéder à des services de cloud privé que les prestataires ont créés, possèdent et font fonctionner. Dans le monde CPH, IDC identifie deux différents modèles de déploiement CPH :
    - **Le CPH dédié**, dans lequel les prestataires de services s'occupent d'un système de cloud privé qui est entièrement dédié au client pour une période prolongée. Ce modèle est essentiellement une version cloud d'offres d'hébergement géré classiques.
    - **Le CPH sur demande**, dans lequel les prestataires de services fournissent dynamiquement des ressources pour une utilisation dédiée depuis un pool partagé, souvent depuis le même pool que leurs offres de cloud publiques.

- **Les services de cloud hybride** constituent l'intégration et la gestion consolidée des services cloud avec d'autres services cloud et/ou des ressources autres que le cloud (systèmes, applications et bases de données). Les services de cloud hybride englobent les combinaisons « public-public », « public-privé » et « privé-privé », ainsi que les combinaisons « cloud-non cloud ». L'inclusion de ressources externes au cloud nécessite généralement de recourir à des interfaces de services cloud (par ex. API RESTful).
- **Les clouds scale-out** ne sont pas adaptés à tous les types de charges de travail, car toutes les charges ne sont pas adaptées à une exécution sur un scale-out. Le cloud permet aux applications de niveau entreprise d'exécuter et de contrôler depuis un point unique. Il n'est pas approprié aux ensembles de données critiques. En outre, les clouds créés sur une architecture standard à bas coût ne peuvent pas fournir les conditions HD que certaines charges de travail requièrent.

## *La prochaine étape pour les applications open source et de nouvelle génération*

### Le scale-out sur Linux

Le scale-out et Linux sont nés au cours de la même ère technologique, au début des années 2000. Le scale-out sur le matériel x86 standard était considéré comme étant plus abordable en termes de coûts, car il permettait de faire des économies sur les CAPEX, et l'OS Linux open source était gratuit ou bien, en tant que distribution commerciale, bien plus abordable qu'Unix. Depuis, Linux a connu des performances et des gains de part de marché sans précédents. Il s'agit désormais d'un système d'exploitation largement utilisé pour le cloud, les applications open source, les bases de données open source et le CPH.

Tous les clouds de grande envergure exécutent Linux, et Linux ne fonctionne plus uniquement sur x86 ; il est désormais disponible sur les ordinateurs centraux uniquement sur Linux et sur les serveurs scale-up et scale-out IBM POWER8. IDC prévoit que 24 % des logiciels du monde entier s'exécuteront sur Linux en 2020, alors qu'ils représentent 16 % en 2016. Sur le marché des serveurs, la part de Linux a augmenté sur les revenus des fournisseurs de serveurs mondiaux pour atteindre 31,2 % en 2015.

Linux a rencontré tellement de succès que SAP a décidé d'exécuter HANA, la base de données sur laquelle la société avait exclusivement parié, sur cet environnement. Avec SAP HANA, les sociétés peuvent désormais exécuter de puissants systèmes analytiques et systèmes transactionnels en utilisant SAP HANA comme une base de données en mémoire déployée sur les serveurs Linux.

### Le scale-out pour les applications open source

L'infrastructure scale-out et les logiciels open source vont de pair grâce à la combinaison coût-efficacité. Pendant les années au cours desquelles le scale-out est devenu l'infrastructure dominante dans les centres de données du monde entier, les logiciels open source ont totalement bouleversé l'environnement logiciels. D'après Sonatype, qui a récemment publié son rapport sur la chaîne d'approvisionnement en logiciels 2016, l'augmentation du nombre de demandes de téléchargement de composants open source a atteint 31 milliards en 2015, alors qu'il s'élevait à 17 milliards en 2014, soit une augmentation de 82 % d'une année sur l'autre. Sonatype affirme que 10 000 nouvelles versions de composants sont introduites chaque jour dans les écosystèmes de développement. Les entreprises adoptent aujourd'hui le modèle open source dans plusieurs secteurs pour des raisons financières, pour pouvoir impliquer une large communauté de

développeurs qui « parlent » le même langage, pour une innovation rapide et pour rendre possibles les connexions entre les actifs via des API ouverts qui conduisent à de nouvelles perspectives.

## Le scale-out pour les bases de données open source

L'une des ressources particulièrement précieuses pour les entreprises modernes lors du parcours vers la transformation numérique est la disponibilité de nombreuses bases de données open source (effectivement, des centaines), qu'elles soient relationnelles ou non. Avant l'émergence du scale-out comme infrastructure alternative et de l'open source comme modèle de logiciel, l'environnement des bases de données se composait principalement d'Oracle, de Microsoft SQL Server et d'IBM DB2. Ces bases de données relationnelles propriétaires sont devenues onéreuses, facturant d'importants frais de licence par cœur. Les volumes de données augmentant et les données étant de plus en plus complexes et déstructurées, avec notamment des flux de streaming, des tweets, des vidéos, des images et des données issues de capteurs, de plus en plus d'entreprises adoptent actuellement, de manière partielle ou complète, des bases de données open source. Ces dernières sont peut-être un peu moins conviviales, mais elles présentent des fonctionnalités similaires et une disponibilité et une sécurité de plus en plus comparables.

Régulièrement améliorées par une vaste communauté de développeurs, les bases de données open source comme MySQL et Postgres ont atteint une qualité de niveau professionnel et sont idéales pour le scale-out, car l'ajout d'autres cœurs dans le but de les développer n'explosera pas le budget informatique d'une entreprise. De plus, les bases de données non relationnelles ou NoSQL, comme MongoDB et Cassandra, sont de plus en plus populaires en raison de leur capacité à stocker tout type de données déstructurées, et sont conçues pour déployer et développer massivement de nouvelles applications. DB-Engines arrivait deuxième d'un classement par ordre de priorité des bases de données répertoriées MySQL mis à jour tous les mois, MongoDB en quatrième place après Microsoft SQL Server, et Cassandra en septième position après DB2 en août 2016.

## Le scale-out pour les applications de nouvelle génération

IDC prédit qu'en 2018, les entreprises adoptant des stratégies de transformation numérique vont plus que doubler leurs performances en matière de développement de logiciels et que deux tiers de leurs codeurs se concentreront sur les applications et services liés à la transformation numérique stratégiques. Les applications de nouvelle génération, un élément clé de l'effort de transformation numérique, sont nettement différentes des applications classiques. Elles utilisent des langages de programmation différents et sont conçues autrement. Les développeurs d'applications de nouvelle génération ont besoin d'une capacité de traitement, d'une extensibilité, d'une redondance d'infrastructure, des performances, d'une capacité de stockage, d'une bande passante et d'un temps de disponibilité extrêmement flexibles, ainsi que d'un coût bas par unité de traitement, stockage et bande passante. Voici certaines caractéristiques spécifiques des applications de nouvelle génération :

- Elles sont conçues grâce à ce que nous appelons des « microservices », ce qui signifie qu'elles ne constituent pas une application monolithique mais un ensemble de petits services qui sont chacun responsables d'un processus distinct, elles peuvent être déployées de manière indépendante et elles peuvent se connecter les unes aux autres via des API. Un des avantages majeurs concernant l'augmentation des performances d'une application créée avec des microservices est qu'un développeur peut adapter uniquement les microservices qui ont besoin de ressources supplémentaires, tandis que les autres peuvent rester inchangés. A la différence des applications monolithiques qui doivent être intégralement adaptées.

- Elles sont souvent sans état et partent du postulat que la durée de vie de l'infrastructure sur laquelle elles sont exécutées n'est pas garantie. Puisqu'elles sont sans état, elles peuvent simplement se déplacer vers un autre nœud dans un environnement de scale-out en cas de panne, sans perte d'état. Si l'application elle-même dysfonctionne, elle est simplement remplacée, et non réparée.
- Elles s'adaptent en fonction de la demande. Lorsque la demande augmente, d'autres instances s'activent automatiquement. De plus, elles sont réactualisées bien plus souvent que les applications classiques, parfois plusieurs fois par jour.

## **L'étape ultime vers les systèmes cognitifs**

### **Le scale-out pour les applications cognitives (apprentissage machine)**

IDC définit les systèmes cognitifs comme une technologie qui utilise un traitement du langage naturel et un apprentissage machine approfondis, permettant de comprendre et de répondre aux questions, ainsi que de fournir des conseils. Le système émet des hypothèses et formule d'éventuelles réponses selon les preuves disponibles. Il peut être formé par l'intermédiaire de grandes quantités de contenu, s'adapte automatiquement et apprend de ses erreurs et de ses échecs (voir *Worldwide Cognitive Software Platforms Forecast, 2015-2019: The Emergence of a New Market*, IDC n° 258781, septembre 2015).

IDC prédit qu'en 2018, plus de 50 % des équipes de développeurs intégreront des services cognitifs dans leurs applications (contre environ 1 % aujourd'hui). Les applications cognitives reposent essentiellement sur les données et ont des caractéristiques qui sont comparables à des applications de traitement haute performance (HPC). IDC remarque que l'infrastructure semblable au HPC, qui repose sur les grappes de scale-out, intègre de plus en plus les clouds et les centres de données pour les applications cognitives.

### **Le scale-out pour l'analytique des données haute performance**

L'analyse des données haute performance (HPDA) fait référence à une technologie de traitement haute performance pour résoudre des problèmes liés au big data très complexes ou critiques dans le temps, notamment des tâches que les serveurs d'entreprise ne peuvent gérer seuls (voir *IDC FutureScape: Worldwide High-Performance Data Analysis 2016 Predictions – Science and Industry Turn to HPC for Mission-Critical Big Data Challenges*, IDC n° 259921, novembre 2016). L'HPDA est, tout comme les systèmes cognitifs, issue du HPC où les grappes de serveurs scale-out sont la norme.

L'utilisation de grandes quantités de données trouve son origine dans la communauté HPC où elles étaient utilisées pour la modélisation et la simulation par le secteur public, le monde universitaire et de très grosses entreprises, comme des constructeurs d'automobiles, des producteurs de biens de consommation et des distributeurs d'envergure nationale. L'émergence du big data est née de l'Internet et des comportements de consommation mobile et sociale. Aujourd'hui, la simulation et l'analytique ont commencé à converger vers le monde des entreprises car celles-ci commencent à utiliser le HPC pour résoudre des problématiques big data, une démarche considérée comme critique lorsqu'il s'agit de garder leur niveau de compétitivité.

## Le scale-out pour l'analytique avancée et prédictive

L'analytique avancée et prédictive (APA) est la capacité à extraire des données structurées / déstructurées dans le but de prendre des décisions d'entreprise à l'aide d'un logiciel d'analyse statistique qui repose sur l'analyse de grands volumes de données, et est donc souvent utilisée conjointement avec un entrepôt de données dans un environnement scale-out. Il est intéressant de noter que la modélisation prédictive, comme SPSS, peut également s'appliquer aux transactions sur un environnement scale-up (par ex. détection de fraude par carte de crédit dans une application bancaire qui s'exécute sur un ordinateur central).

L'APA utilise une série de techniques permettant de créer, tester et exécuter des modèles statistiques. Certaines techniques utilisées sont l'apprentissage machine, la décrémentation, les réseaux neuronaux, l'induction de règles et la mise en grappe. L'analytique avancée et prédictive sert à découvrir des relations entre les données et à faire des prédictions qui sont masquées, non apparentes ou trop complexes pour être extraites à l'aide d'un logiciel de requête, de transmission de données et d'analyse multidimensionnelle (voir *Advanced and Predictive Analytics Software: Market Segments and User Types*, IDC n° APA51X, juin 2015).

Alors que l'APA nécessitait auparavant les compétences de spécialistes, il est aujourd'hui de plus en plus utilisé par les entreprises qui ne sont pas forcément des expertes, utilisant les résultats APA pour prendre des décisions éclairées en s'appuyant sur les données. L'APA se compose de plusieurs types de logiciels : des langages de développement pour les programmes d'écriture et les routines de programmation ; un logiciel d'amélioration de la productivité qui aide les scientifiques grâce à des algorithmes modulaires, des opérations de préparation des données et des solutions d'apprentissage machine ; des outils permettant de faciliter le travail de l'analyste métier en masquant les modèles analytiques ; des applications APA pour les directions métiers et opérationnelles, comme la maintenance préventive et la détection de fraude bancaire ; et l'APA intégrée dans les applications (par ex. intégrée dans un logiciel de centre d'appels pour prédire le risque de perte de clientèle).

## Paramètres requis pour l'infrastructure scale-out

Pour que le département IT déploie une feuille de route de manière efficace, toutes les étapes allant des applications Web sans état aux charges de travail cognitives et prédictives, en passant par le matériel serveur avec lequel l'infrastructure est conçue, doivent respecter un certain nombre de conditions. La feuille de route du scale-out montre qu'en évoluant, ces charges de travail ne peuvent être compatibles avec un matériel standard d'entrée de gamme. Les demandes sur l'infrastructure seraient trop importantes.

Les serveurs doivent être faciles à configurer et doivent se déployer rapidement. Les serveurs doivent également prendre en charge les applications de nouvelle génération qui seront les moteurs de la transformation numérique de l'entreprise. Les serveurs doivent être faciles à gérer et présenter de bonnes performances avec des charges de travail riches en données. Ils doivent être sécurisés. Ils doivent également offrir une haute disponibilité (HD) car il ne peut y avoir de temps d'arrêt.

L'infrastructure du serveur scale-out pour ces charges de travail doit également pouvoir être gérée sans qu'il n'y ait besoin d'un ensemble de compétences inhabituelles. Il est important qu'ils prennent en charge un vaste écosystème de serveurs et qu'ils soient intégralement prêts pour le cloud. Les

serveurs doivent fonctionner sur un OS largement adopté pour la transformation numérique et doivent pouvoir exécuter tout l'éventail des solutions logicielles open source. Enfin, ils doivent être disponibles à de faibles coûts CAPEX et OPEX et doivent pouvoir être facilement remplacés lorsque leurs successeurs de nouvelle génération arriveront.

## Le marché actuel pour les serveurs scale-out sur Linux

Le marché actuel pour l'infrastructure de type scale-out sur Linux, à savoir des serveurs à 1 et 2 sockets exécutant Linux, se développe rapidement. IDC prévoit qu'entre 2016 et 2020, le nombre de livraisons de ce type de systèmes augmentera avec un TCAM (taux de croissance annuel moyen) de +7,5 %. En 2020, la part de Linux sur le marché des serveurs à 1 et 2 sockets atteindra 42,6 % du marché total en termes de revenus des fournisseurs.

Le choix du processeur est crucial dans le processus de décision d'une architecture scale-out des dirigeants d'entreprise, responsables techniques (CTO) et directeurs marketing (CMO). Il est donc important que les clients n'évaluent pas seulement les systèmes x86, mais prennent également en compte l'alternative que représente le processeur OpenPOWER dans leur étude comparative. Étant donné que, avec OpenPOWER, il existe à présent des alternatives scale-out à faibles coûts CAPEX et OPEX sous Linux - Linux semble être un OS très bien adapté aux applications riches en données nécessaires à la prise en charge de la transformation numérique - il serait bénéfique pour les acheteurs de comparer ces solutions aux produits x86 qu'ils envisagent d'utiliser.

## IBM POWER : UN NOUVEAU MODELE DE SOLUTIONS SCALE-OUT

---

De nombreuses entreprises désireuses de mettre en œuvre des applications de transformation numérique et qui ne souhaitent pas les exécuter sur leur environnement scale-up étudient aujourd'hui l'infrastructure scale-out. Mais un grand nombre d'entre elles ne comparent pas les offres scale-out basée sur x86 avec une solution scale-out IBM POWER, celle-ci étant la seule alternative à x86 en termes de performance et de prix.

Du point de vue des performances, les solutions scale-out IBM POWER reposent sur la même conception de processeur IBM POWER8 que les produits scale-up, qui sont de loin des leaders sur le marché des processeurs à plus de 8 sockets. IBM POWER8 présente des performances par cœur exceptionnelles, grâce à son multithreading simultané (SMT) à 8 thread (par rapport à l'hyperthreading à 2 thread du x86), à son sous-système E/S à large bande passante et à sa bande passante mémoire supérieure. Les mêmes fonctionnalités que celles qui permettent aux systèmes IBM POWER scale-up d'être leaders sur le marché ont été intégrées dans les offres IBM POWER scale-out.

Les serveurs IBM POWER ont été créés autour du concept selon lequel les entreprises exécutent avant tout (et de plus en plus) des applications critiques pour exploiter des données. Parmi elles, se trouvent les applications de nouvelle génération. Une application mobile classique, par exemple, est très axée sur les données, même s'il s'agit d'un type de données différent que des lignes et des colonnes. L'architecture IBM POWER est conçue pour s'adapter à de nouvelles formes de données, qu'il s'agisse de données structurées, mesurées en pétaoctets, de masses de données déstructurées ou de flux. Le système présente des taux d'ingestion très élevés pour les données, et ses mémoires

caches ont été conçues pour maintenir le processeur occupé. En revanche, les processeurs x86 sont destinés à des usages très variables, des jeux vidéo au traitement de feuilles de calcul, en passant par l'exécution d'une application. En mettant l'accent sur les applications orientées entreprise, IBM a la capacité d'optimiser IBM POWER pour ces charges de travail.

Lorsque nous nous plaçons du point de vue d'un développeur d'applications, ces caractéristiques lui permettent de développer des applications différenciées qui sont avantageuses en matière de performances d'IBM POWER. Par exemple, lorsqu'un processeur contient de nombreux thread, cela signifie qu'un développeur d'applications n'encode pas simplement les processeurs parallèles (comme sur x86), mais huit threads parallèles dans les processeurs parallèles, ce qui donne lieu à un nombre bien plus élevé de processus simultanés pour optimiser l'application. Les applications existantes qui fonctionnaient sur x86 peuvent, d'après IBM, être simplement déplacées dans 95 % des cas et seront prêtes à l'emploi. Les mémoires caches et le threading, en étant adaptés et exploités, commenceront à réaliser de meilleures performances.

Mais les développeurs peuvent parvenir à une plus grande différenciation grâce à l'accélération sur IBM POWER. Le portefeuille LC (812LC et 822LC) peut notamment être commandé avec CAPI, à savoir Coherent Accelerator Processor Interface (Interface de processeur à accélérateur cohérent), une interface entre le processeur et l'E/S qui réalise de bien meilleures performances qu'un PCIe Interconnect, ainsi que des GPU. La gamme LC et certains produits de la gamme L possèdent un solide portefeuille d'accélérateurs que les développeurs peuvent exploiter pour les applications novatrices.

## Un portefeuille allant du IaaS au scale-out pour l'entreprise

Le portefeuille d'IBM POWER8 s'étend des offres d'infrastructure IaaS à une gamme de produits de scale-out (systèmes à 1 ou 2 sockets) allant du milieu de gamme (4 sockets) au niveau entreprise (8 sockets). IBM POWER8 est disponible comme IaaS dans SoftLayer, ce qui lui permet de bénéficier des performances du cloud. PurePower Systems est une infrastructure convergente d'IBM pour le cloud. De plus, IBM propose des implémentations de clouds à ses clients qui comportent du matériel, des logiciels et des services.

Dans la catégorie du scale-out, IBM propose un portefeuille de systèmes à un ou deux sockets très performants dont le prix d'achat commence juste en dessous de 5 000 \$. La gamme scale-out pour les charges de travail critiques se compose de Power S822, Power S814 et Power S824 (le deuxième chiffre indique le nombre de sockets). Le scale-out sur Linux se compose de six systèmes (L désigne Linux, C désigne Cluster [grappe]) : Power S812LC, Power S822LC pour l'informatique de gestion, Power S822LC pour l'informatique haute performance, Power S812L, Power S822L et Power S824L (avec la technologie NVIDIA). Avec cette gamme, IBM a insufflé un nouveau souffle dans le segment des serveurs de volume et offre une alternative distincte aux serveurs de volume sur x86, et même une série d'avantages du point de vue de la valeur. Ses trois serveurs d'entreprise sont Power E850, Power E870, Power E870C, Power E880, et Power E880C.

Le portefeuille s'élargit encore davantage lorsque l'on recherche des systèmes IBM POWER créés par d'autres fabricants de serveurs qui font partie du Consortium OpenPOWER. Le Consortium cède la technologie IBM POWER sous licence à des tiers et présente d'autres fournisseurs de technologie qui offrent des innovations pour l'accélération IBM POWER sous une seule désignation. Les serveurs

POWER sont actuellement commercialisés ou en cours de création par Tyan, Inspur, Supermicro, Inventec, Wistron, Cirrascale, ChaungHe et Zoom Netcom. Rackspace devrait lancer Barreleye, son serveur cloud IBM POWER8 bare-metal, d'un jour à l'autre, et Rackspace et Google fonctionnent tous les deux sur Zaius, qui est sur POWER9.

## La différenciation du scale-out POWER

Comme mentionné précédemment, le scale-out dans le cadre de la transformation numérique nécessite certains paramètres d'infrastructure. Cette section détermine les performances du portefeuille scale-out d'IBM POWER sur ces paramètres, à l'exception des performances du système qui ont été étudiées précédemment.

### *Vitesse de déploiement*

La vitesse de déploiement se mesure à l'aide de deux éléments : le déploiement du matériel et de l'environnement d'exploitation et le déploiement d'applications. Pour le déploiement du matériel, les systèmes scale-up et scale-out IBM POWER prennent en charge les mêmes types d'outils pour le déploiement x86 et à des vitesses comparables. IBM fournit POWER VC, qui signifie Virtualization Center (Centre de virtualisation). Il s'agit d'un outil créé sur OpenStack pour la gestion de la virtualisation et les déploiements d'environnements clouds. Il sert à automatiser et à simplifier la gestion des VM sur les serveurs scale-out POWER avec une interface utilisateur "facile à naviguer" et permettant d'optimiser l'affectation des ressources. Les serveurs scale-out POWER peuvent également tirer profit de ces outils open source, comme le JuJu, pour un déploiement rapide. En outre, IBM possède un programme appelé Rapid Build Solution pour accélérer le déploiement en proposant des systèmes préconfigurés et préchargés qui sont prêts à être branchés et à être utilisés par les clients. En ce qui concerne la vitesse de déploiement des applications, lorsque la configuration d'une application nécessite un important volume de traitement, les avantages architecturaux de POWER se révèlent intéressants.

### *Prise en charge des applications fondamentales*

En ce qui concerne la prise en charge des applications cœurs de métier d'une entreprise, comme la gestion des données, tous les systèmes IBM POWER, du plus petit au plus important, présentent le même environnement d'exploitation intégré et les mêmes performances intégrées. Un client peut exécuter la même application tout en bas et tout en haut de l'échelle, quelle que soit l'application. Par exemple, si par le passé, un client a exécuté sa base de données sur une très grande machine POWER5 pour des raisons de puissance, aujourd'hui, cette même base de données peut fonctionner sur un serveur POWER8 scale-out avec, peut-être 4 sockets au lieu de 16, grâce aux mesures des performances de POWER8. En bref, POWER prend en charge ces applications sur l'ensemble de la gamme.

C'est également vrai pour les systèmes x86 plus importants. Une application riche en données, par exemple, qui fonctionne sur un système x86 à 4 sockets migrera facilement vers un système POWER à 1 ou 2 sockets, grâce à la densité de charge de travail du système POWER. SAP HANA, par exemple, qui fonctionne sur x86 depuis plusieurs années, est disponible sur POWER depuis 2015. HANA sur POWER a vu son taux d'adoption dans les centres de données augmenter de manière significative, comme l'a indiqué IBM lors de sa récente publication trimestrielle. IDC s'est entretenu avec des clients HANA qui

sont passés de x86 à POWER, car ils étaient confrontés à certaines limites avec leurs appareils x86 HANA. Ils ont consolidé HANA et plusieurs applications SAP sur un seul système POWER scale-out.

### *Prise en charge des applications de nouvelle génération*

Les développeurs d'applications de nouvelle génération bénéficient des performances techniques d'un système POWER8 scale-out. Les fournisseurs de logiciels et les développeurs d'applications ont indiqué qu'un des avantages du codage sur POWER8 est qu'il ne nécessite rien de nouveau. Ils peuvent choisir le Linux qu'ils souhaitent, travailler avec les outils qu'ils préfèrent et utiliser les techniques d'encodage qu'ils ont toujours utilisées. En d'autres termes, un développeur peut simplement commencer à développer pour POWER sans avoir à acquérir de nouvelles compétences.

Au-delà des similitudes entre le codage sur POWER8 et les autres systèmes, POWER8 offre des avantages certains, permettant ainsi aux développeurs d'applications une optimisation directement sur l'architecture et d'offrir de meilleures performances à leurs clients à l'aide du même code de base. Par exemple, le développement d'applications sur Linux et POWER scale-out offre les avantages combinant performance, extensibilité flexible, résilience et fiabilité des environnements critiques. Parmi ces avantages, se trouvent également les 8 thread précédemment mentionnés. Il est communément reconnu que l'étape la plus décisive pour un développeur d'applications est de passer d'un codage simple thread à un codage multithread. Passer de deux à, par exemple, huit thread est considéré comme une tâche supplémentaire dans la mise en parallèle des processus. De plus, les développeurs d'applications qui aimaient écrire pour des processus uniques sur x86 peuvent toujours écrire des processus destinés à un thread unique sur POWER8, ils fonctionneront plus rapidement.

### *Haute disponibilité*

De nombreuses personnes se demandent si IBM proposera réellement des fonctions haute disponibilité comparables sur sa gamme scale-out que sur les systèmes scale-up de classe entreprise avec lesquels il a bâti sa réputation en termes de fiabilité. Ce qui différencie POWER8 en matière de haute disponibilité réside dans son architecture et a été transféré de ses systèmes de classe entreprise à sa gamme scale-out.

Les mémoires tampons qui sont intégrées au système aident, par exemple, à éliminer les erreurs. Les erreurs peuvent se produire au niveau d'une puce ou d'un système, elles sont causées par des particules ou du bruit et modifieront une instruction dans un programme ou dans une valeur de donnée. Une telle erreur pourra généralement mettre à mal une architecture moins robuste, ce qui nécessitera un redémarrage. POWER8 possède également des processus de reprise automatique conçus pour permettre la reprise après des défaillances détectées en interne. En d'autres termes, le système ne repose pas sur un processus consistant à détecter la défaillance et à laisser ensuite un logiciel tiers la prendre en charge, le système POWER8 prend en charge la défaillance et la corrige. POWER8 comporte des contrôleurs de mémoire intelligents avec des mémoires de reprise et une détection des erreurs qui lui permettent de savoir quand il y a un problème. Il peut ensuite corriger le problème sur le bus ou entre le contrôleur et le module DIMM. Il possède également des modules DRAM de secours.

IBM affirme qu'un grand nombre de ses clients achètent POWER en raison de la mémoire non volatile (NVM) et de la capacité à corriger les défaillances de mémoire avant qu'elles n'aient lieu. Les clients

qui exécutaient HANA sur x86 et qui ont adopté HANA sur POWER en particulier, ont procédé à ce changement car leurs anciens systèmes étaient tombés en panne suite à des défaillances de mémoire.

La principale mesure d'IBM a été d'intégrer un continuum Haute Disponibilité à sa gamme scale-out. La gamme de produits LC d'entrée de gamme comporte un nombre moins important de ces fonctions Haute Disponibilité intégrées, car elle est conçue et ses coûts sont optimisés pour des clusters déployés dans le cloud. Dans un cluster, si un nœud tombe en panne, un autre nœud prend le relais et le traitement continue comme avant. Les autres produits de la gamme scale-out, comme le S824 ou le S822, possèdent un plus grand nombre de ces fonctions RAS pour les charges de travail critiques, car ils sont souvent déployés en tant que système unique. Les produits haut de gamme possèdent des composants redondants, mais sur la gamme LC, l'accent est mis sur la conception afin de rester compétitif en termes de prix : un meilleur composant, un meilleur fournisseur et un meilleur processus de fabrication.

## *L'écosystème logiciel*

L'écosystème logiciel pour la gamme IBM POWER8 scale-out est aussi vaste et varié que l'écosystème logiciel pour tous les environnements Linux. L'affirmation « S'il fonctionne sur Linux sur x86, il fonctionne sur Linux sur POWER » est en grande partie vraie, avec toutefois une mise en garde : si une application tire profit d'une extension propriétaire, quel que soit le type de matériel, un client ou un fournisseur indépendant de logiciel devra déterminer comment gérer cet élément pour POWER8. Cette situation n'est pas propre à POWER ; la même situation se produira sur x86 en cas de transfert d'un logiciel qui utilise une extension propriétaire d'un fabricant de matériel à un autre. Outre cela, IBM est convaincu que le code peut être simplement transféré vers POWER8. S'il s'agit d'un code compilé, il doit simplement être recompilé ; s'il est scripté, il fonctionne normalement.

D'après IBM, lorsque Canonical a transféré Ubuntu sur POWER, l'ensemble du code scripté a été transposé et exécuté. Et sur l'ensemble du code compilé de la société, 95 % a été exécuté. Ubuntu a simplement recompilé le code, et il s'est exécuté sans erreurs. Les 5 % qui ne fonctionnaient pas s'expliquaient par l'utilisation d'extensions propriétaires et par quelques manœuvres d'encodage qui ont dû être retravaillées pour que les codes fonctionnent sur POWER8.

## *Prêt pour le cloud*

La gamme POWER8 scale-out est conçue et ses coûts optimisés pour un déploiement en environnement cloud. Les performances par cœur et la densité de VM avec POWER8 constituent un avantage significatif pour un environnement cloud car, ensemble, elles permettent de surmonter un obstacle important auquel les centres de données sont confrontés, à savoir le coût de la surface au sol. Les performances par cœur et la densité de VM facilitent la consolidation du matériel. IBM affirme qu'en raison de la densité de VM de POWER8, les centres de données peuvent exécuter beaucoup plus de VM dans le même espace. Il est possible d'obtenir d'autres avantages liés à la densité grâce à la conteneurisation, par exemple sur Docker. Pour les MSP, la densité de VM de POWER8 se traduit, assez simplement, par une augmentation des revenus au mètre carré ; pour les centres de données, elle donne lieu à une baisse des OPEX.

Lorsque des clients déploient des applications dans le cloud, ils ont besoin qu'elles soient disponibles. Les fonctions HD POWER8 mentionnées précédemment, comme la mémoire non volatile et la

protection de la mémoire, permettent de s'assurer qu'un cloud sur POWER8 scale-out est intrinsèquement fiable. En ce qui concerne les compétences, il n'y a pas de différence : un administrateur qui peut exécuter un cloud sur x86 peut également l'exécuter sur POWER8, à l'aide des mêmes outils.

Enfin, l'un des principaux avantages de l'utilisation de POWER dans le cloud est que l'architecture est totalement ouverte et peut faire l'objet d'une licence, du matériel au code, essentiellement comme le modèle ARM. Les fournisseurs de services cloud de grande envergure, comme Google et Rackspace, ont déclaré qu'ils développaient et fabriquaient des éléments de l'architecture POWER pour leurs centres de données, pour lesquels ils avaient obtenu une licence d'IBM via le Consortium OpenPOWER. IBM est convaincu que le Consortium OpenPOWER, qui permet aux partenaires d'accorder une licence pour tous les composants de la technologie POWER, est indispensable à la réussite du cloud sur le long terme.

### ***Le coût des environnements POWER***

L'opinion générale est que POWER est toujours plus onéreux ; néanmoins, aujourd'hui, tout le monde peut aller sur le site d'IBM et acheter un POWER S812LC pour 4 800 \$. IDC a discuté avec plusieurs entreprises et directions informatiques des prix de l'ensemble de la gamme scale-out : leur perception change. IBM a récemment annoncé de nouveaux environnements au portefeuille LC qui amélioreront encore plus les niveaux de prix. Si IDC ne compare pas les systèmes selon les prix ou les rapports qualité-prix, les mesures réalisées pour POWER8 laissent sous-entendre qu'il serait dommage pour les acheteurs de produits informatiques, et pour leur entreprise, de ne pas réaliser ces comparaisons qualité-prix dans le cadre de leurs évaluations des infrastructures de type scale-out.

### **Intégrer POWER scale-up classe entreprise à un environnement scale-out**

Enfin, les clients qui envisagent POWER8 pour le scale-out devraient également jeter un œil à la gamme POWER scale-up. Utiliser les deux leur offrira des avantages architecturaux, et leur permettra de profiter des nombreuses années de développement technologique contenues dans les portefeuilles scale-out et scale-up. Les systèmes POWER sont pleinement capables d'exécuter des charges de travail open source, avec Linux disponible dans le portefeuille des systèmes scale-out et scale-up. Les processeurs Integrated Facility for Linux (IFL) de POWER permettent aux clients d'exécuter même les charges de travail les plus petites, par exemple une application de nouvelle génération, de manière très économique.

Les clients peuvent lancer des applications de nouvelle génération sur des systèmes scale-out plus petits, puis les faire migrer vers le système POWER de classe entreprise lorsqu'ils commencent à augmenter leur puissance de traitement et nécessitent les mêmes caractéristiques d'entreprise que les systèmes d'enregistrement classiques. Par exemple, les clients POWER exécutent de plus en plus de bases de données open source sur Linux, comme MongoDB, MariaDB et Postgres, en plus, voire à la place, de leurs systèmes d'enregistrement classiques.

IBM POWER propose trois systèmes scale-up appropriés à la réalisation du parcours vers la transformation numérique : l'E850, l'E870 et l'E880. Ils sont fabriqués autour du nouveau processeur très rapide IBM POWER8 et sont dotés d'une mémoire allant jusqu'à 16 To pour l'E870 et 32 To pour l'E880. Ces systèmes possèdent des cœurs renforcés et de plus grandes mémoires caches que ceux

des générations POWER précédentes, ainsi que deux contrôleurs de mémoire intégrés dotés d'une latence améliorée et d'une plus grande bande passante. Ces trois systèmes possèdent un sous-système E/S intégré et des fentes E/S PCIe Gen3 à branchement chaud. Plusieurs améliorations ont permis d'augmenter la fiabilité, la facilité de maintenance et la disponibilité déjà reconnues du système, notamment des blocs d'alimentation redondants, une protection avancée de la mémoire et des processeurs redondants.

Pour en savoir plus sur le scale-up dans le cadre de la transformation numérique, voir *Digital Transformation Decisions: Bringing Your Enterprise-Class Scale-Up Solutions into the Next-Gen World* (livre blanc d'IDC).

## DEFIS ET OPPORTUNITES

---

### Pour les entreprises

#### *Défi*

Pour les entreprises, le principal défi est de trouver la bonne stratégie de scale-out qui rendra possible leur voyage vers la transformation numérique. Il y a plusieurs variables à prendre en compte : Quels types d'applications exécutons-nous aujourd'hui ? Quels types d'applications exécuterons-nous dans 12 ou 24 mois ? Quelles sont nos exigences en matière de haute disponibilité ? Souhaitons-nous réaliser un déploiement dans un cloud ? Privé, hybride ou public ? Quelle est la bonne infrastructure ?

#### *Opportunité*

La principale opportunité pour ces entreprises est d'obtenir un avantage concurrentiel important en choisissant la bonne infrastructure. Pour cela, il convient de ne pas simplement comparer différents fournisseurs de l'infrastructure x86, mais également d'inclure l'infrastructure sur OpenPOWER dans l'évaluation. Les produits scale-out actuels qui fonctionnent sur OpenPOWER sont hautement concurrentiels, voire avantageux, sur le plan du rapport qualité-prix.

### Pour IBM

#### *Défi*

IBM POWER est reconnu pour son portefeuille d'entreprise composé de serveurs scale-up dotés de caractéristiques haute performance. Ces systèmes fonctionnaient traditionnellement sur Unix, et sont désormais disponibles sur Linux, ce que le marché commence à reconnaître. Cependant, ce que le marché semble ignorer, c'est qu'IBM propose un portefeuille solide et varié de serveurs scale-out pour le cloud et pour les applications de transformation numérique. Le premier prix pour ces systèmes à un et deux sockets est 4 800 \$. Pourtant, ces systèmes possèdent la même puce POWER8 que celle qui sert de moteur aux systèmes destinés aux grandes entreprises. IBM doit promouvoir ces systèmes, prouver au marché qu'ils ont de meilleures performances, sont plus rentables et sont tout aussi faciles à utiliser que le scale-out sur x86. Un autre enjeu : inciter le marché à inclure les offres POWER dans les évaluations comparatives de scale-out de toutes les entreprises.

## Opportunité

Pour IBM, l'opportunité est importante. Aujourd'hui, le marché des serveurs à un et deux sockets est dans une situation monopolistique du point de vue de l'architecture des processeurs. S'il existe une grande concurrence entre les fournisseurs qui commercialisent des systèmes dotés de la même architecture de processeur, au bout du compte, les utilisateurs ne bénéficient pas d'un réel choix. Si IBM peut montrer au marché que ses offres scale-out alternatives sur POWER sont toutes aussi compétitives, voire avantageuses au niveau du rapport qualité-prix, il pourra acquérir une part de marché significative.

## CONCLUSION

---

La transformation numérique signifie que le modèle d'entreprise et l'innovation de produit tirent profit des nouvelles capacités numériques et des capacités mobiles croissantes afin d'offrir aux clients des expériences qui les raviront et qui satisferont leurs attentes en constante évolution. Les « capacités numériques » décrites précédemment sont, bien entendu, une condition préalable pour réaliser ce voyage. Elles se composent de nombreux éléments, allant des développeurs d'applications qui peuvent coder à l'aide de microservices aux applications qui peuvent prédire le comportement des clients, en passant par les processeurs qui peuvent mettre en parallèle les ensembles d'instructions de process.

IDC estime que l'évolution des capacités numériques d'une entreprise se traduit par l'adoption des bonnes applications fonctionnant sur la bonne infrastructure. Nous pensons également que pour y parvenir, il convient d'adopter une approche par étapes, qui simplifiera et rendra plus efficace l'aboutissement du voyage, à savoir l'adoption d'une analytique cognitive, d'une analytique de données haute performance et d'une modélisation avancée et prédictive, permettant d'obtenir un avantage concurrentiel certain. Ce livre blanc décrit ces étapes en détail.

Nous décrivons également les conditions que l'infrastructure, en tant que fondation à la transformation numérique, doit remplir. Il serait certainement plus bénéfique que cette infrastructure soit de type scale-out sur Linux, même si un grand nombre des étapes indiquées précédemment peuvent également être réalisées sur une infrastructure de type scale-up sur Linux, celles-ci offrant le bon niveau d'usage, de virtualisation et d'extensibilité. Néanmoins, le scale-out présente des avantages certains, comme nous l'avons expliqué précédemment.

Les responsables informatiques doivent cependant faire preuve d'une grande prudence en ce qui concerne l'infrastructure de type scale-out sur laquelle ils choisissent de s'appuyer pour leur transformation. La plupart des serveurs à 1 et 2 sockets avec lesquels les environnements scale-out Linux sont architecturés sont fournis avec un seul type de processeur, le x86, même si de nombreux fournisseurs de différentes marques les proposent. D'après IDC, les entreprises qui évaluent uniquement les produits de serveurs scale-out qui exécutent Linux sur x86 passent à côté de l'opportunité de comparer les mesures de ces derniers avec les serveurs scale-out OpenPOWER, en particulier ceux d'IBM. Si IDC ne réalise pas de comparaisons directes entre les processeurs, nous sommes persuadés que les clients devraient le faire avec minutie avant d'investir dans leur infrastructure informatique pour les cinq prochaines années critiques dans cette ère de changements significatifs.

## À propos d'IDC

International Data Corporation (IDC) est le premier fournisseur mondial en veille stratégique, services de conseil et événements pour les marchés des technologies de l'information et de la communication et de l'électronique grand public. IDC aide les professionnels en informatique, les dirigeants d'entreprise et les investisseurs à prendre des décisions factuelles sur leurs achats de technologies et sur leur stratégie d'entreprise. Plus de 1 100 analystes IDC offrent une expertise mondiale, régionale et locale dans le domaine des technologies et sur les opportunités et tendances de l'industrie dans plus de 110 pays dans le monde. Depuis 50 ans, IDC prodigue des conseils stratégiques pour aider ses clients à atteindre leurs objectifs commerciaux clés. IDC est une filiale d'IDG, leader mondial dans la recherche, les médias et l'événementiel.

## Siège social mondial

5 Speen Street  
Framingham, MA 01701  
États-Unis  
508.872.8200  
Twitter : @IDC  
idc-community.com  
www.idc.com

---

### Copyright

Publication externe des informations et données d'IDC : l'utilisation des informations d'IDC dans le cadre de publicités, de communiqués de presse ou de contenus promotionnels doit être soumise à l'accord écrit préalable du vice-président ou directeur local d'IDC concerné. Une version du document proposé doit accompagner la demande. IDC se réserve le droit de refuser l'utilisation externe de ces informations pour quelque raison que ce soit.

Copyright 2016 IDC. Toute reproduction sans autorisation écrite est interdite.

