

# Introduction au cloud computing

# Qu'est-ce que le cloud computing

# Qu'est-ce que le cloud computing

## D'après vous ?

- Quels services de cloud-computing connaissez-vous ?
- Les utilisez-vous ?

# Qu'est-ce que le cloud computing

D'après vous ?

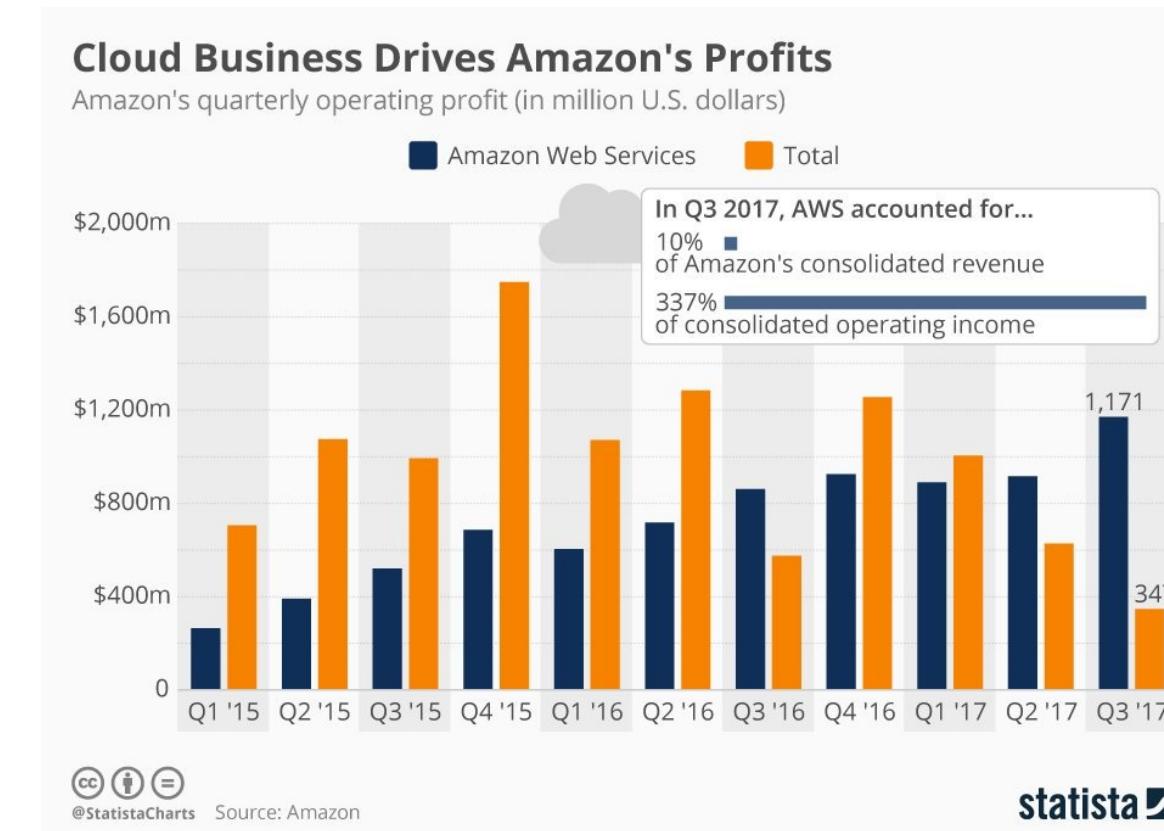
- Quels services de cloud-computing connaissez-vous ?
- Les utilisez-vous ?



# Qu'est-ce que le cloud computing

## D'après Amazon Web Services, un des pionniers du marché

Le cloud computing consiste à fournir des ressources de technologie de l'information à la demande à travers l'internet en ne facturant que les ressources consommées.



# Qu'est-ce que le cloud computing

## D'après le NIST, cinq caractéristiques définissent le cloud:

Le cloud computing est un modèle offrant un accès réseau ubiquitaire, pratique, à la demande, à un ensemble partagé de ressources informatiques configurables (ex: réseaux, serveurs, stockage, applications et services...) qui peut être rapidement mis à disposition puis libéré avec le moins possible d'efforts de gestion ou d'interactions avec le fournisseur de service.



# Qu'est-ce que le cloud computing

D'après le NIST, cinq caractéristiques définissent le cloud:

1) Les ressources sont fournies à la demande en self-service

- Pas d'interaction humaine avec le fournisseur
- Allocation de tout type de ressources:
  - Logiciel (accès à une application, mise en place d'un service ...)
  - Infrastructure virtualisée (machine virtuelle, stockage, réseau, ...)
  - Infrastructure physique (serveur dédié, ...)

packet

The Promise of the Cloud,  
Delivered on Bare Metal

Improve performance and minimize cost with single tenant, dedicated servers by the hour.

# Qu'est-ce que le cloud computing

D'après le NIST, cinq caractéristiques définissent le cloud:

2) Les ressources sont facilement accessible à travers le réseau

- Mise à disposition d'API flexibles pour créer, contrôler et accéder aux ressources
- Accès à distance voir en mobilité
- Utilisation à travers des clients hétérogènes



# Qu'est-ce que le cloud computing

D'après le NIST, cinq caractéristiques définissent le cloud:

3) Les capacités sont mutualisées entre les consommateurs

- Le fournisseur repartit dynamiquement ses capacités (de calcul, stockage ...) en fonction des besoins des consommateurs
- Le consommateur ne se préoccupe pas de savoir comment ces capacités sont allouées
- Il ne connaît pas leur localisation précise
  - Abstraction à gros grain de type région ou datacenter

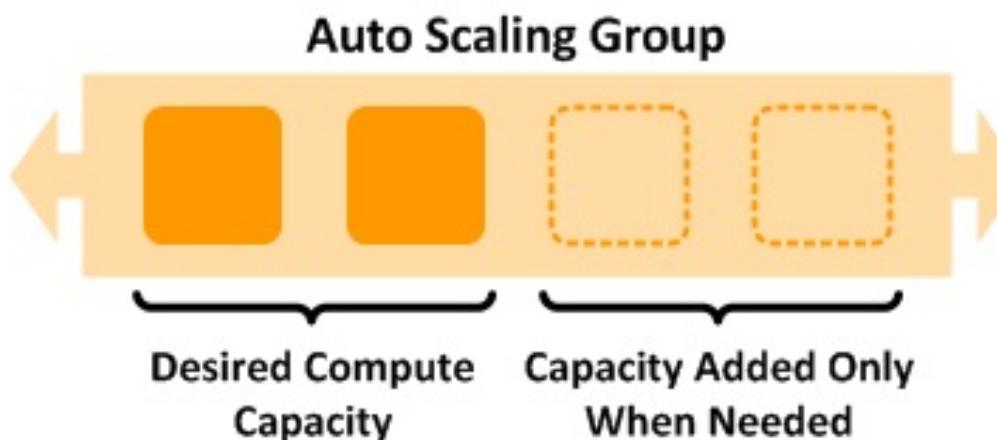


# Qu'est-ce que le cloud computing

D'après le NIST, cinq caractéristiques définissent le cloud:

4) Les ressources fournies sont rapidement élastiques

- Possibilité de faire varier la quantité de ressources allouée:
  - à la hausse ou à la baisse
  - à tout moment et sans délai
  - sans avoir à se préoccuper de la quantité de ressource disponible
  - de façon automatisable en fonction du taux d'utilisation par exemple



# Qu'est-ce que le cloud computing

D'après le NIST, cinq caractéristiques définissent le cloud:

5) La consommation de ressource est mesurée

- Des métriques adaptée à chaque type de ressource permettent d'en quantifier la consommation
  - Quantité de donnée stockée, transférée
  - Nombre de requêtes effectuées
  - Durée d'exécution d'une machine virtuelle
  - Nombre d'utilisateurs actifs
  - ...
- La facturation est transparente et basée sur la consommation réelle
- Le client a les données en main pour optimiser son utilisation

▼ Elastic Compute Cloud	\$526.93
► EU (London) Region	\$0.00
▼ US East (Northern Virginia) Region	\$526.93
Amazon CloudWatch	\$30.30
\$0.00 per alarm-month - first 10 alarms	3.495 Alarms
\$0.00 per metric-month - first 10 metrics	10 Metrics
\$0.00 per request - first 1,000,000 requests	53,309 Requests
\$0.30 per metric-month for the first 10,000 metrics - US East (Northern Virginia)	\$0.00
	101.011 Metrics
	\$30.30

# Une technologie incontournable

# Une technologie incontournable

## Pour les particuliers

- Les données personnelles y sont de plus en plus souvent stockées
  - Accessibles partout
    - À son domicile, en déplacement, au travail
  - Consultables tout support
    - PC fixe ou portable
    - Tablette, téléphone portable ...
  - Facilement partageables
    - Nombreuses API
    - Contrôle d'accès intégré
  - Stockées de façon fiable



# Une technologie incontournable

## Pour les particuliers

- Le navigateur web est devenu un client léger universel
  - Client mail, réseaux sociaux
  - Streaming de vidéos et musique
  - Édition de documents bureautique
  - Visionnage et retouche de photo
  - ...
- Manque de puissance des périphériques mobiles et objets connectés
  - Mise à contribution de serveurs distants

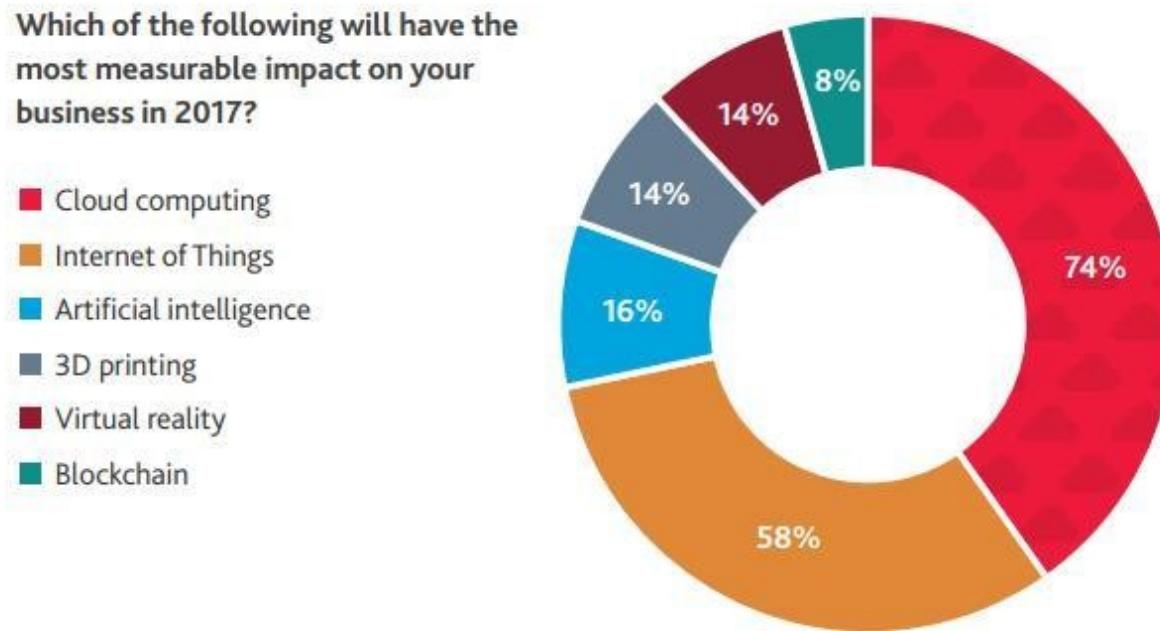


chromebook

# Une technologie incontournable

## Pour les entreprises

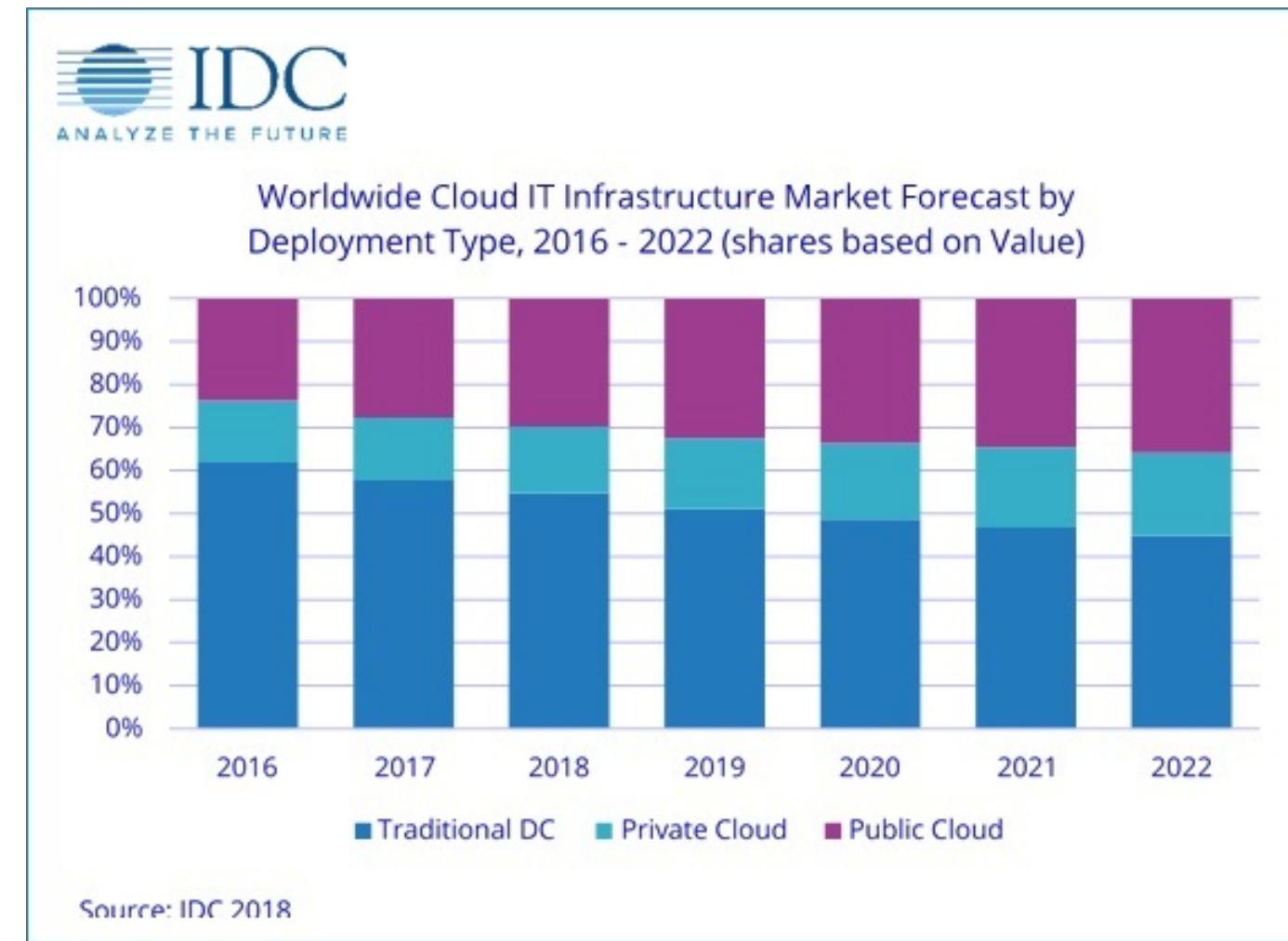
- Le cloud computing suscite beaucoup d'attentes dans les entreprises



# Une technologie incontournable

## Pour les entreprises

- Forte croissance des investissements dans les infrastructures cloud



# Une technologie incontournable

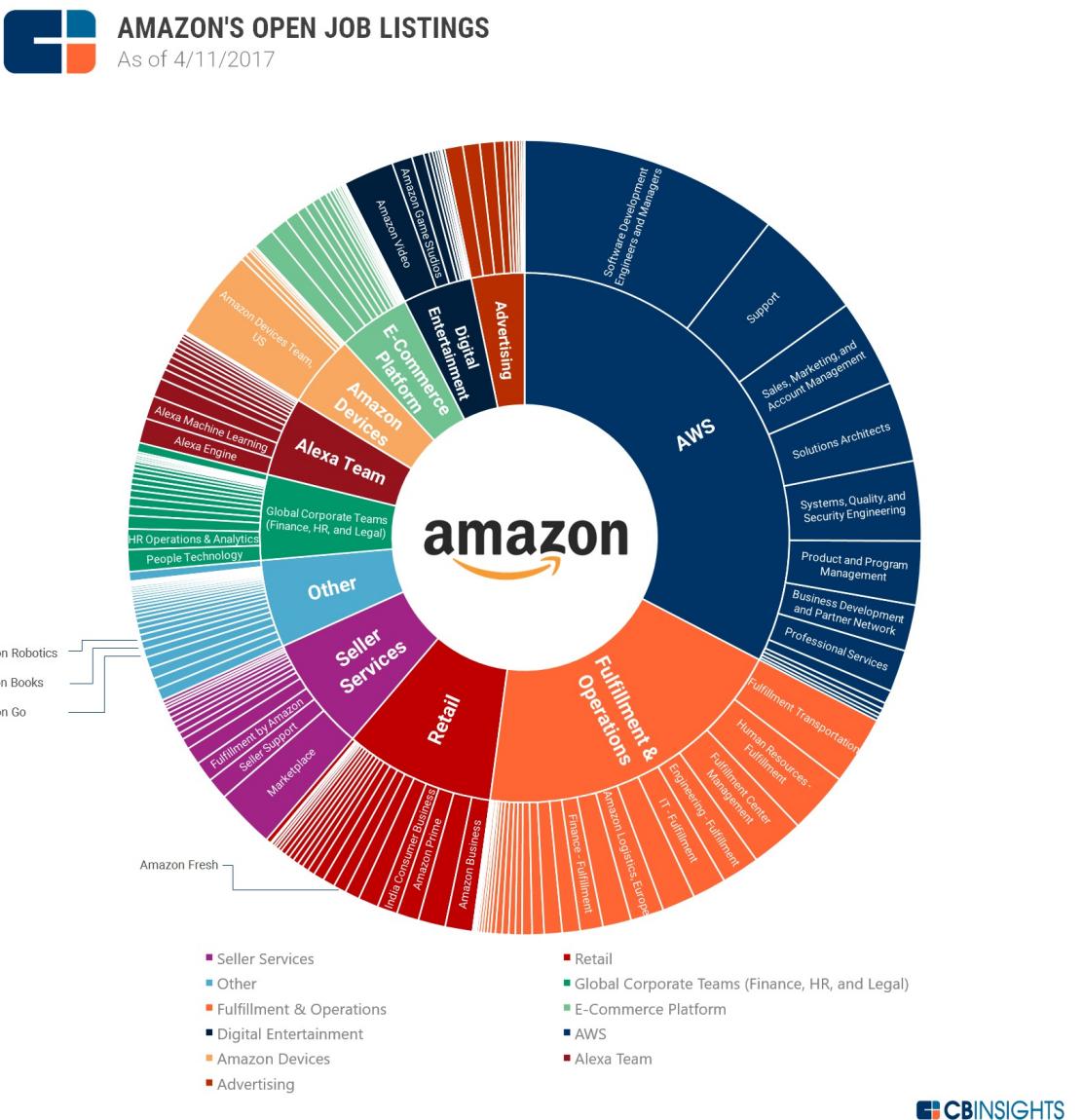
## Pour les entreprises

- Les logiciels d'entreprise sont de plus en plus fournis dans le cloud
  - D'après l'IDC 25% du marché en 2016, 40% en 2021



# Une technologie incontournable

Pour un futur ingénieur



# Historique

# Historique

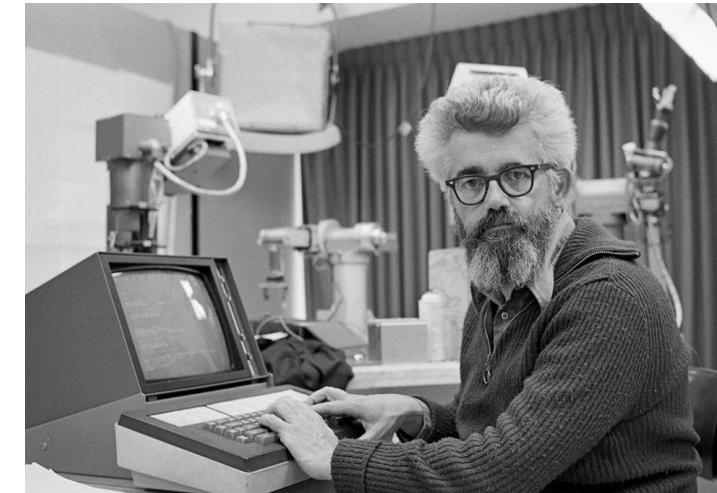
## Des concepts qui ne datent pas d'hier

“L'informatique sera peut-être utilisée un jour comme un service comme celui proposé par les opérateurs téléphoniques. Chaque abonné ne paie que pour les capacités qu'il utilise réellement mais a accès à toutes les possibilités de programmation d'un système de grande envergure”

Prof. John McCarthy, MIT, 1961

“Ce qui est intéressant avec le cloud computing c'est qu'on a redéfini ce terme pour inclure tout ce qu'on faisait déjà. L'industrie informatique est encore plus sensible aux effets de mode que l'industrie des vêtements féminin.”

Larry Ellison, co-fondateur et CTO de Oracle, 2008



# Historique

## Aux origines de l'informatique

- Les mainframes ('60s/'70s/'80s)
  - Des ordinateurs centralisés et puissants
  - Utilisés à distance via des terminaux légers
  - Partagés entre de multiples clients et utilisateurs
  - Compartimentés dans des machines virtuelles
    - Une technologie présente dans les tous premiers systèmes d'exploitations !



# Historique

## La micro-informatique ('80s/'90s)

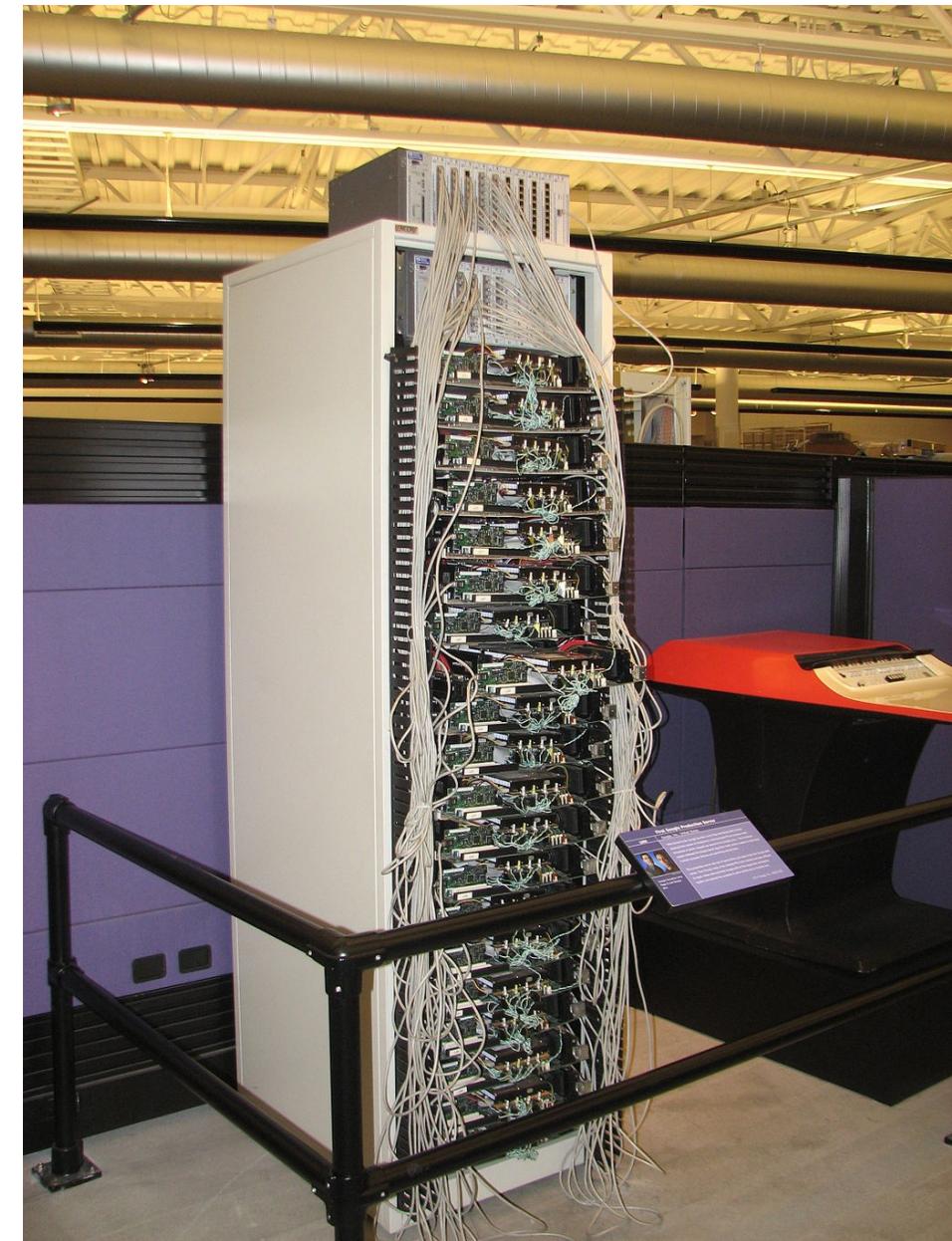
- Une abondance d'ordinateurs peu coûteux
  - Les ordinateurs peuvent être déployés partout sans contraintes (alimentation électrique, refroidissement...)
  - Les applications sont exécutées directement sur le poste de l'utilisateur



# Historique

## La micro-informatique ('80s/'90s)

- Les centres de données
  - Internet révolutionne les usages
  - Besoin de serveurs fiables et connectés pour chaque entreprise
  - Les premiers datacenters apparaissent
  - Multiplication rapide avec la bulle des dotcom



# Historique

## La consolidation de serveurs ('00s)

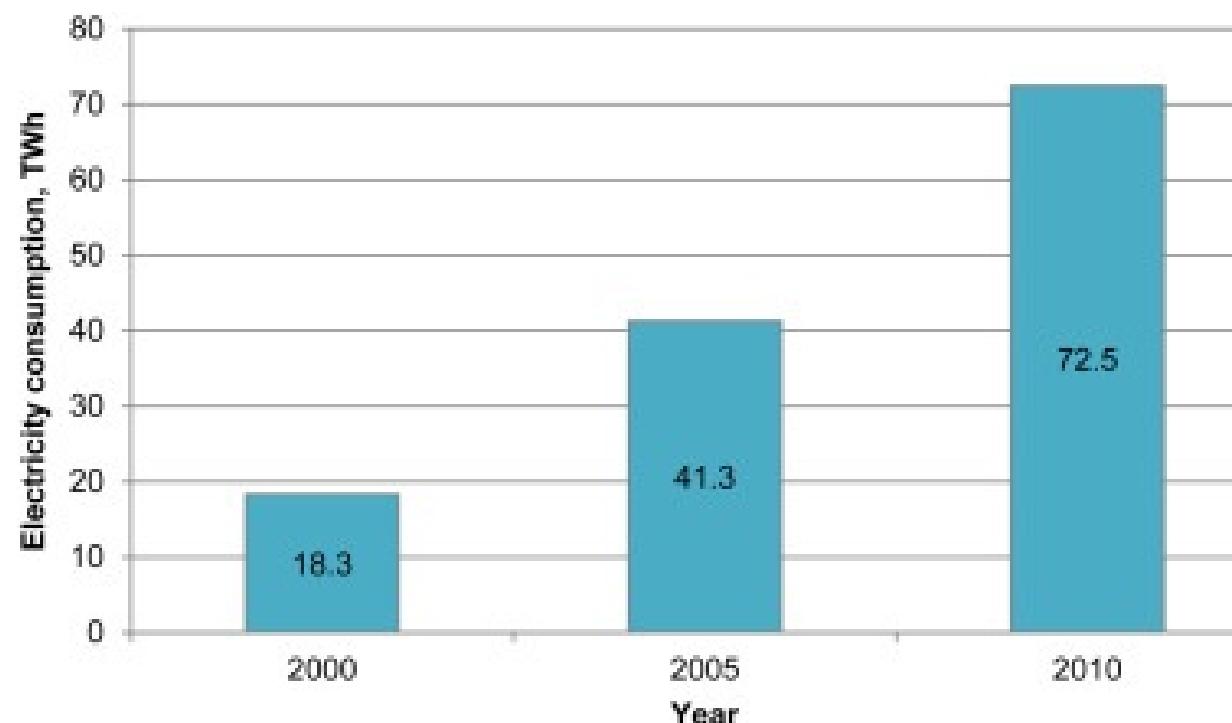
- L'essor des technologies Web et client/serveur
  - L'industrie du Web repart sur des bases plus saines
  - De nouvelles technologies apparaissent (Ajax, HTML5...)
  - Les possibilités du Web sont décuplées (Web 2.0)
    - Le navigateur Web devient un client léger efficace et universel
  - Le développement des réseaux ouvre de nouvelles possibilités
    - (Très) Haut débit
    - Internet sur périphériques mobiles
  - L'ensemble des industries se convertit au numérique
- Le besoin en serveurs continue de croître



# Historique

## La consolidation de serveurs ('00s)

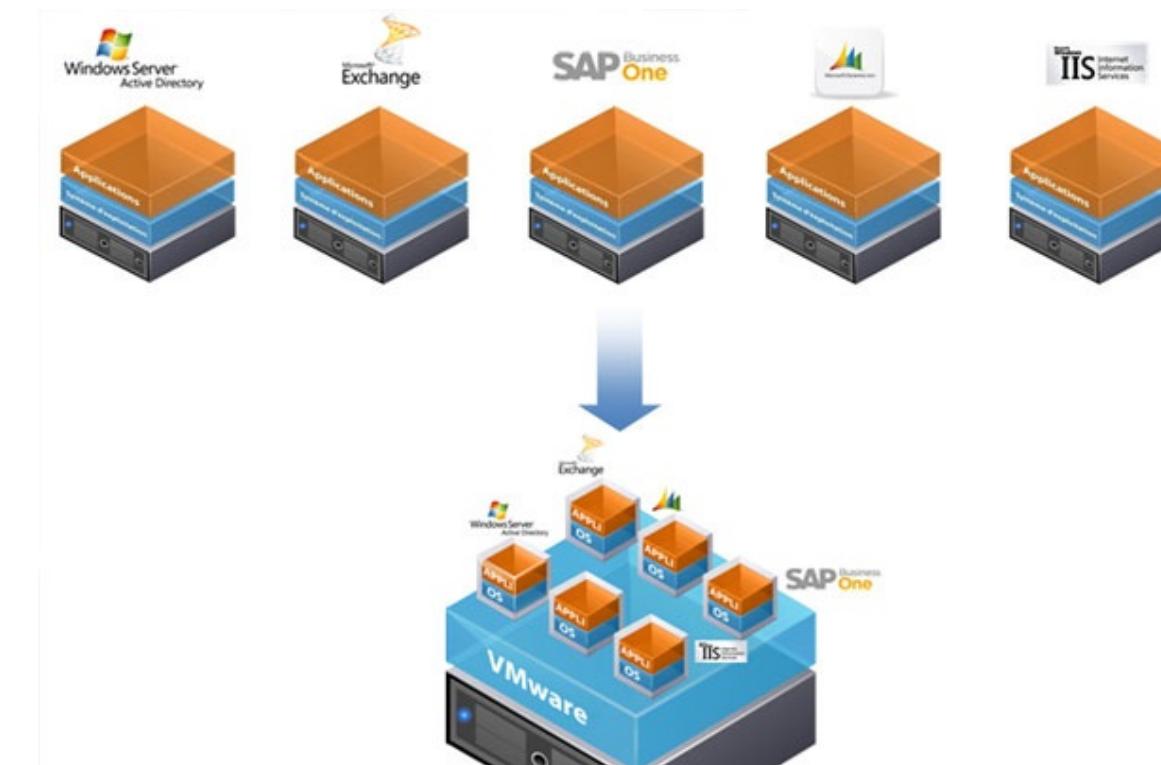
- Une nécessité pour rationaliser les coûts
  - Des parcs informatiques de plus en plus complexes et pléthoriques
  - Un consommation énergétique toujours plus importante
  - Des serveurs de plus en plus puissants souvent sous-exploités
    - Chaque serveur est dédié à une organisation ...
    - ...et souvent à une unique application
  - D'après McKinsey le taux d'utilisation moyen d'un serveur est de 15%
  - D'après une étude de Stanford 30% des serveurs ne sont jamais utilisés



# Historique

## La consolidation de serveurs ('00s)

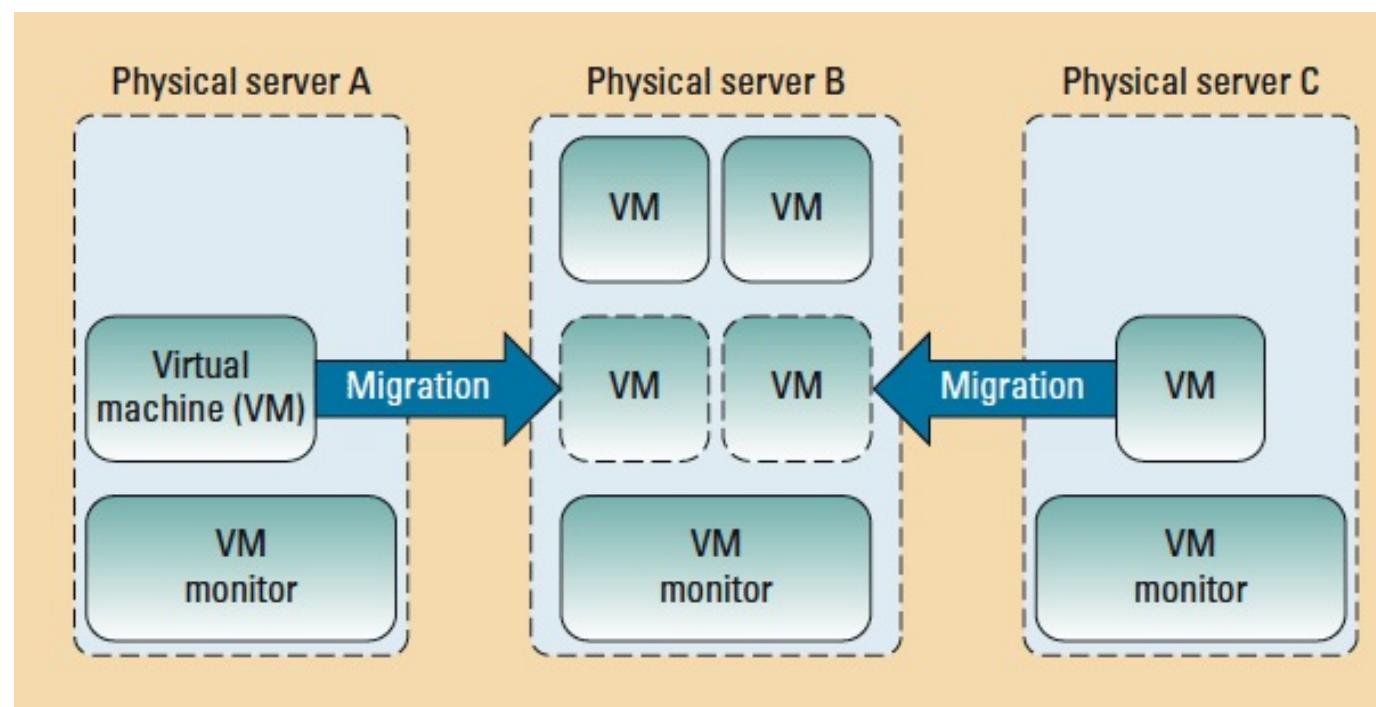
- Le retour de la virtualisation du matériel
  - Permet de partager les serveurs en plusieurs machines virtuelles
    - Découpler le logiciel du matériel
    - Indépendance des piles logicielles dans chaque machine virtuelle
    - Meilleure isolation entre les utilisateurs
  - Techniques de virtualisation pour les processeurs x86
    - Emulation/para-virtualisation efficace
    - Jeux d'instruction dédiés



# Historique

## La consolidation de serveurs ('00s)

- Le retour de la virtualisation du matériel
  - Permet d'augmenter le taux d'utilisation des serveurs
  - Facilite la gestion du parc de serveurs
    - Plus de réactivité pour disposer d'une nouvelle machine
    - Facilité pour réassigner un serveur à de nouvelles tâche
    - Même à chaud via une migration de machine virtuelle



# Historique

## L'arrivée du cloud computing



- Amazon Web Services (2006)
  - Besoin interne de d'API permettant de déployer efficacement des infrastructures

“The executive team expected a project to take three months, but it was taking three months just to build the database, compute or storage component.”

- Savoir faire acquis pour le site e-commerce
- Besoin commun à toute l'industrie
  - Difficulté de déployer une application à l'échelle du web
- Lancement d'un service commercial d'infrastructure à la demande
  - Services faisant d'internet un OS pour application web

“Any organization or company or any developer [can] run their technology applications on top of our technology infrastructure platform.”, Andy Jassy AWS CEO

# Historique

## L'arrivée du cloud computing

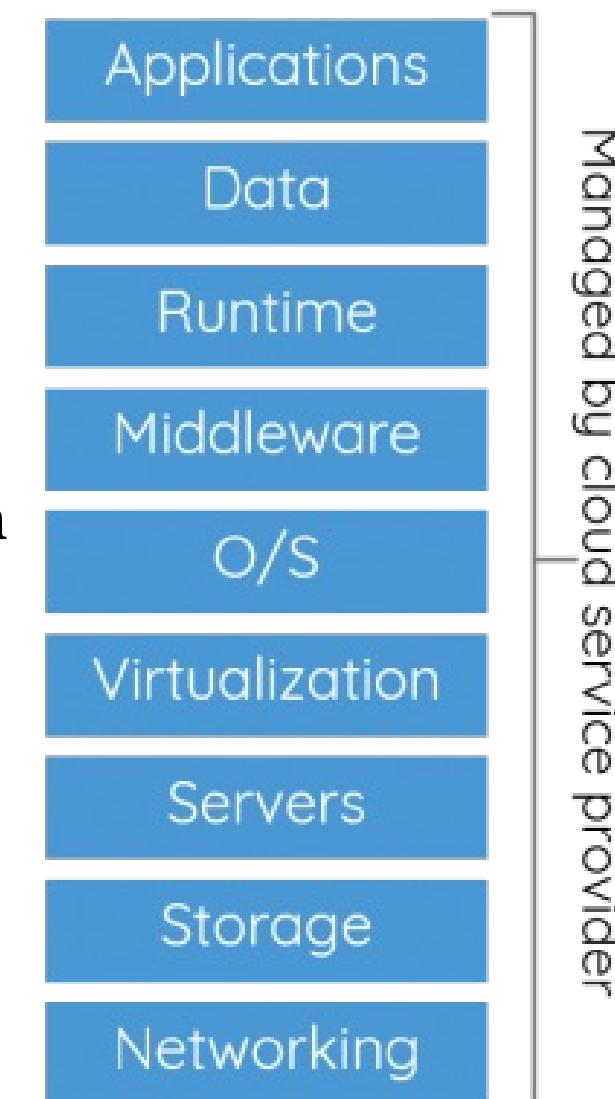
- Les services d'AWS connaissent un fort succès
  - Ajout constant de nouvelles fonctionnalités
- De nombreux concurrents se lancent progressivement sur le marché
  - 2008: Google lance son App Engine
  - 2008: Dell tente de mettre un copyright sur le terme cloud computing
  - 2010: Microsoft lance le service Azure

# Les différentes formes de cloud

# \*aaS: Everything-as-a-Service

## Le SaaS: Software as a Service

- Logiciel à la demande
- Accès à distance
  - Typiquement via un navigateur web
- Paiement par abonnement flexible
- Hébergement centralisé et mutualisé
- Adaptation de l'application à ses besoins par les paramétrages prévus
  - Favorise la configuration à la personnalisation
  - Console d'administration conviviale
- Une seule base de code pour tous les clients
  - Intégration de personnalisation spécifiques à un client si besoin
  - Prise en compte de l'ensemble pour toute mise à jour
- APIs pour intégration dans un système complet



# \*aaS: Everything-as-a-Service

## Le SaaS: Software as a Service

- Exemples de services logiciels proposés
  - Bureautique, messagerie, calendrier
  - Logiciels d'entreprise (RH, CRM, ERP, ...)
  - Outils de création (développement logiciel, retouche photo, ...)
  - Outils collaboratifs (visioconférence, chat ...)
  - Composants logiciels (synthèse vocale, encodage, analyse de données ...)

# \*aaS: Everything-as-a-Service

## Le SaaS: Software as a Service

- Remplace progressivement les logiciels à licences
  - Modèle par défaut pour les nouvelles entreprises vendant du logiciel
  - Transition progressive des gros acteurs existants
    - Portage des applications phares en SaaS
    - Achat de concurrents nés dans le cloud
- Exemples:
  - Oracle HR
    - Les logiciels à installer sur site sont toujours maintenus
    - Les nouvelles fonctionnalités sont en priorité intégrées à la version cloud
  - “On-premise HR users risk being left behind”
  - Adobe
    - Conversion son produit phare en mode SaaS (Creative Suite)
    - Licence à 1800\$ -> Abonnement à 50\$/mois
    - Forte croissance des revenus et de la capitalisation de l'entreprise

# \*aaS: Everything-as-a-Service

## Le SaaS: Software as a Service



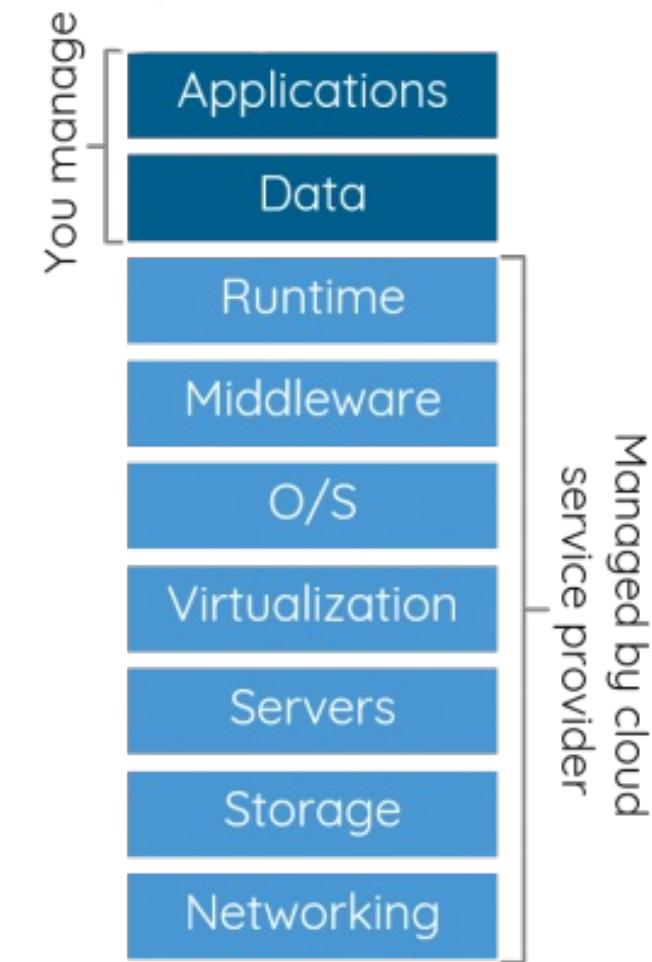
- Un modèle commercial pour le développement open source
  - Hébergement d'un logiciel open source en mode SaaS
  - Contributions open source pour répondre aux besoin des clients
  - Le client n'est pas prisonnier du modèle SaaS
    - Possibilité de déployer le logiciel open source autrement
- Exemple: Automattic / WordPress
  - Société fondée par des développeurs du projet open source WordPress
  - Fourni un service d'hébergement web basé sur WordPress
  - Contribue au développement du projet open source
  - Emploie aujourd'hui 800 personnes
  - 30 % des sites Internet sont estimés fonctionner avec WordPress
    - Wordpress.com: 147M de visiteurs uniques US par mois  
(Facebook 210M)

# \*aaS: Everything-as-a-Service

## Le PaaS: Platform as a Service

Service de plateforme d'exécution d'applications

- Le client ne gère que le code de son application
- Le fournisseur:
  - Met en place l'environnement complet d'exécution
    - OS, configuration, réseau, support exécutifs ...
    - Mises à jour, sécurité
  - Héberge et exécute le code de l'application
  - Adapte les ressources matérielles/virtuelles à la charge
  - Propose un workflow et des services facilitant le développement et le déploiement



# \*aaS: Everything-as-a-Service

## Le PaaS: Platform as a Service

- L'exemple d'Heroku

- Une des premières offres de PaaS à succès (fondé en 2007)
- Applications à petite/moyenne échelle
- Support de nombreux langages (Ruby, Python, Java, Go, ...)
- Déploiement à partir d'un "git push" d'un dépôt contenant
  - Code source
  - Liste de dépendances
  - Recette de compilation
- Compilation automatique au sein de la plateforme
- Instanciation d'un ou plusieurs dynos exécutant l'application
  - Conteneurs assurant l'isolation de l'application
- Applications de type service web
  - Définition d'une adresse DNS pour l'application
  - Répartition des requêtes HTTP reçues entre les différents dynos

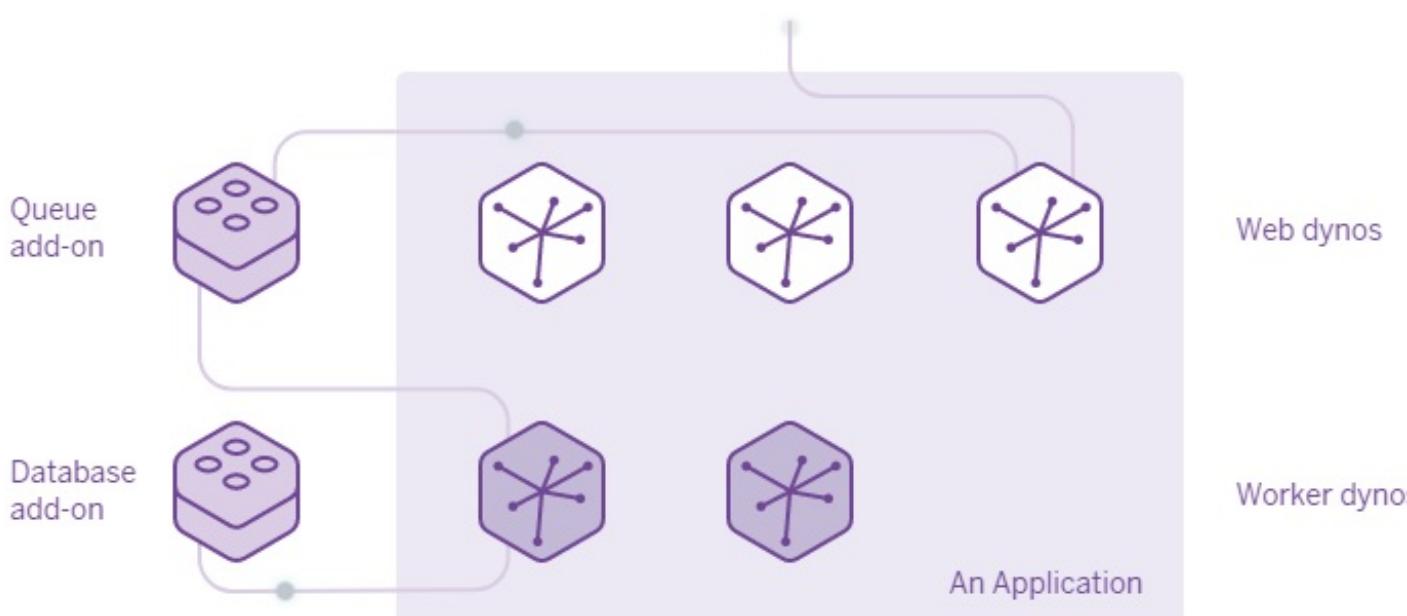


# \*aaS: Everything-as-a-Service

## Le PaaS: Platform as a Service



- Fonctionnement des dynos
  - Entièrement déployés par la plateforme
    - En fonction des dépendances détectées
  - Possibilité de lancer des processus de support (workers)
    - Communication à travers des files
  - Accès à un petit système de fichiers éphémère dans le dynos
    - Le dyno peut être relancé à tout instant
  - Service de base de données hébergée pour le stockage persistant
    - PostgreSQL, Memcached, Redis, ...



# \*aaS: Everything-as-a-Service

## Le PaaS: Platform as a Service



- Google App Engine
  - Lancé en 2008
  - Gros client connu: Snapchat
  - Exécution des applications dans un bac à sable restreint
    - Accès plus restreint à l'OS
    - Lancement des applications à travers un framework (ex: python WSGI)
  - Depuis 2016: *flexible environment*
    - Relâchement des contraintes via exécution dans des VMs dédiées
    - Déploiement et passage à l'échelle moins rapide

# \*aaS: Everything-as-a-Service



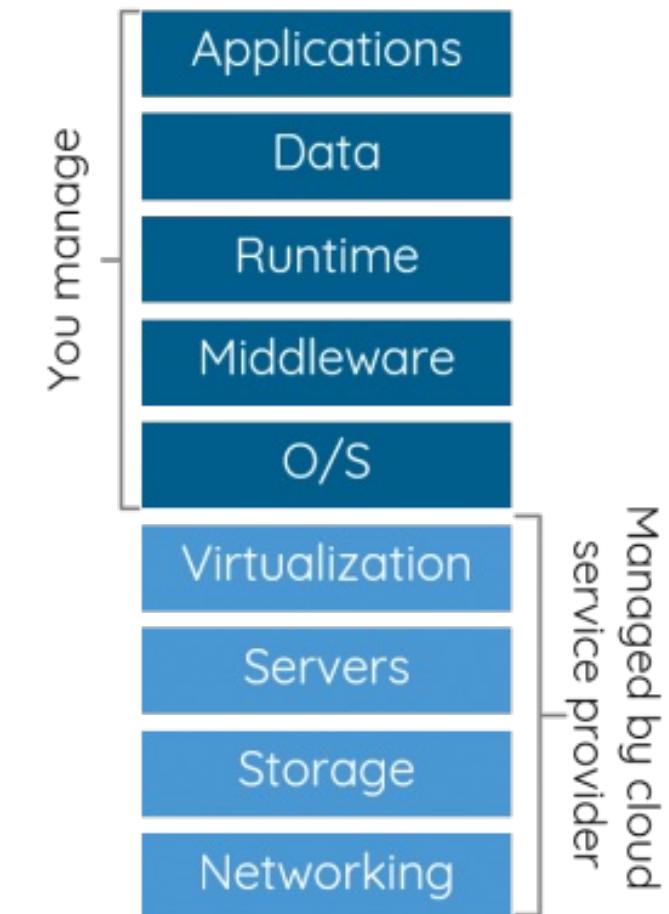
## Le PaaS: Platform as a Service

- RedHat OpenShift
  - Plateforme hébergée par RedHat
    - Basée sur un produit open source: OpenShift Origin
    - Lui-même basé sur l'orchestrateur de conteneurs Kubernetes
  - Met à profit l'écosystème des conteneurs popularisé par Docker
    - Recettes de construction de conteneurs
    - Formats standardisés et dépôts d'images conteneur
  - Implémente les services attendus d'une plateforme de PaaS
    - Construction, mise à jour, déploiement de conteneurs
    - Ordonnancement, routage, équilibrage de charge

# \*aaS: Everything-as-a-Service

## L'IaaS: Infrastructure as a Service

- Composants d'infrastructure à la demande
  - Serveurs physiques ou virtuels
  - Espaces de stockage
  - Interconnexion réseau
- Généralement associé à des services logiciels
  - Base de données
  - Équilibrage de charge
  - DNS
  - Monitoring ...

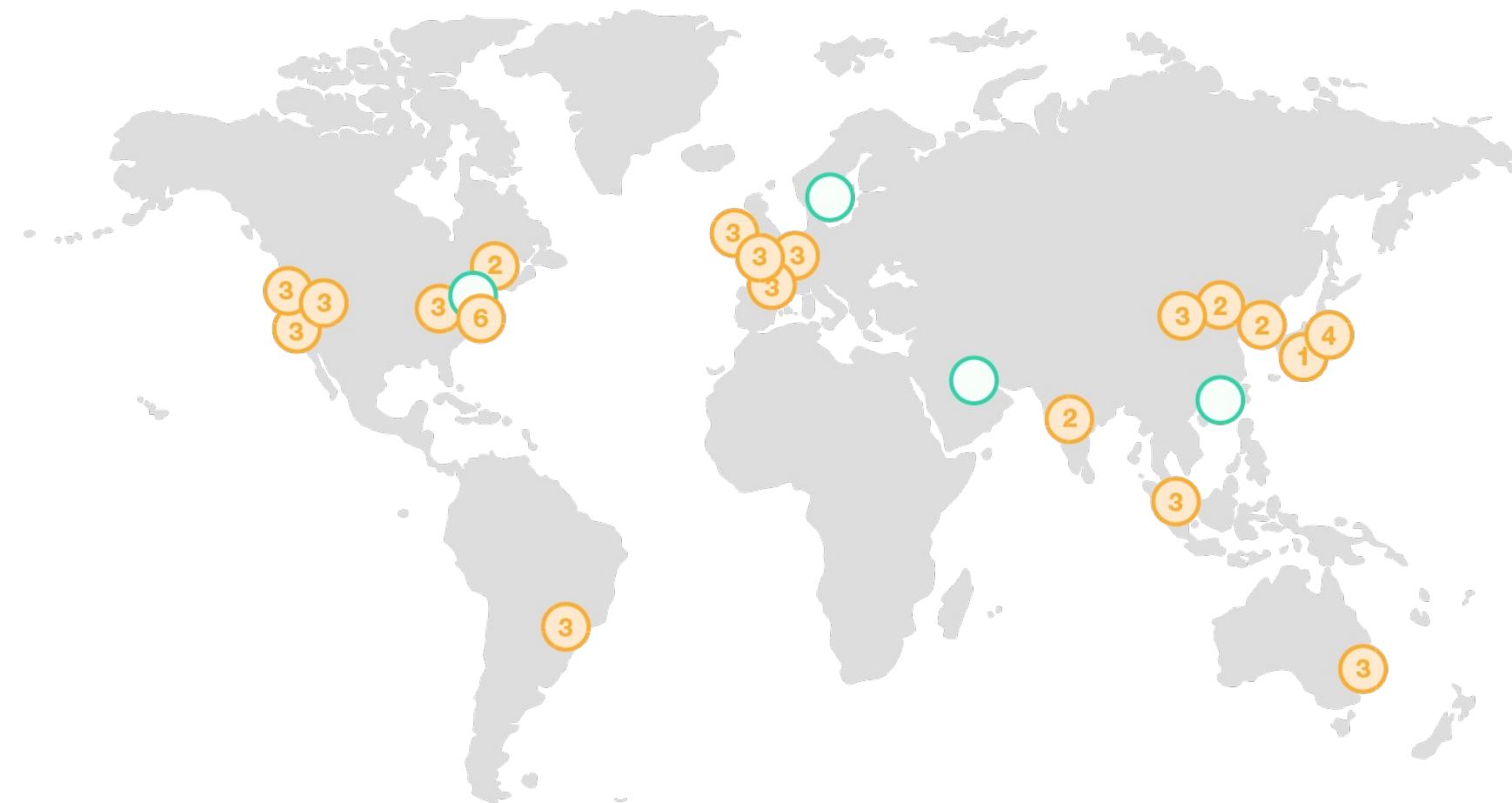


# \*aaS: Everything-as-a-Service

## L'IaaS: Infrastructure as a Service



- Le leader AWS
  - 55 centres de données dans 18 régions du monde
    - Ordre de grandeur: 50000 serveurs / centre de données



# \*aaS: Everything-as-a-Service

## L'IaaS: Infrastructure as a Service



- Le leader AWS
  - Une offre pléthorique répondant à de nombreux besoins

<b>Compute</b> EC2 Lightsail ↗ Elastic Container Service EKS Lambda Batch Elastic Beanstalk	<b>Developer Tools</b> CodeStar CodeCommit CodeBuild CodeDeploy CodePipeline Cloud9 X-Ray	<b>Analytics</b> Athena EMR CloudSearch Elasticsearch Service Kinesis QuickSight ↗ Data Pipeline AWS Glue	<b>Customer Engagement</b> Amazon Connect Pinpoint Simple Email Service
<b>Storage</b> S3 EFS Glacier Storage Gateway	<b>Management Tools</b> CloudWatch AWS Auto Scaling CloudFormation CloudTrail Config OpsWorks Service Catalog Systems Manager Trusted Advisor Managed Services	<b>Security, Identity &amp; Compliance</b> IAM Cognito Secrets Manager GuardDuty Inspector Amazon Macie ↗ AWS Single Sign-On Certificate Manager CloudHSM Directory Service WAF & Shield Artifact	<b>Business Productivity</b> Alexa for Business Amazon Chime ↗ WorkDocs WorkMail
<b>Database</b> RDS DynamoDB ElastiCache Neptune Amazon Redshift	<b>Media Services</b> Elastic Transcoder Kinesis Video Streams MediaConvert MediaLive MediaPackage MediaStore MediaTailor	<b>Desktop &amp; App Streaming</b> WorkSpaces AppStream 2.0	<b>Internet Of Things</b> IoT Core IoT 1-Click IoT Device Management IoT Analytics Greengrass Amazon FreeRTOS IoT Device Defender
<b>Migration</b> AWS Migration Hub Application Discovery Service Database Migration Service Server Migration Service Snowball	<b>Mobile Services</b> Mobile Hub AWS AppSync Device Farm Mobile Analytics	<b>Game Development</b> Amazon GameLift	

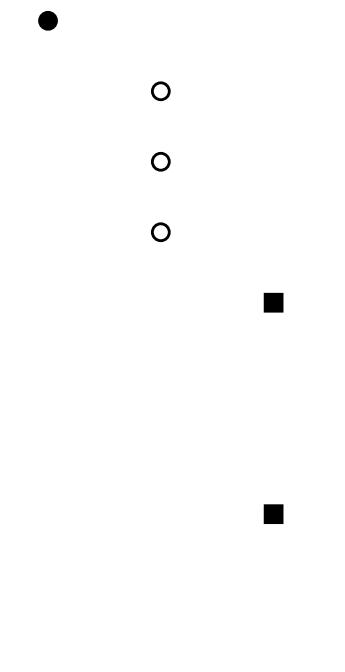
# \*aaS: Everything-as-a-Service

## L'IaaS: Infrastructure as a Service



- Le leader AWS
  - Exemples de services proposés

EC2	Machines virtuelles
VPC	Réseau virtuels
S3	Stockage objet
RDS,DynamoDB	Base de données
Route53	DNS
ELB	Répartiteur de charge
SQS	File de messages
CloudWatch	Monitoring
WAF	Firewall





# \*aaS: Everything-as-a-Service

- D'autres terminologies pour des services qui rentrent plus ou moins dans les principales catégories
  - MBaaS: Mobile Backend (~SaaS)
    - Briques de base pour réaliser un backend d'application mobiles
    - Notifications push, intégration aux réseaux sociaux, stockage de données, ...
  - DBaaS: Base de données (~SaaS)
  - CaaS: Conteneurs (~PaaS)
  - FaaS: Fonctions (~PaaS)
    - Permet de ne fournir que le code d'une fonction
    - Exécution en fonction de différents évènements
    - Aucune gestion de ressource, concurrence infinie
  - WPaaS: Workplace (~SaaS)
    - Ensemble de logiciels de bureau et de communication intégrés
  - DaaS: Desktop (~SaaS)
    - Ordinateur de bureau virtuel (VDI)
  - BPaaS: Business Process (~SaaS)
    - Tout type de tâche récurrente pour une entreprise
  - ...

# Modèles déploiement

## Cloud public

- Infrastructure de cloud ouverte à un large public
- Gérée par une entreprise, une institution académique ou gouvernementale
- Propose un catalogue de services en ligne avec tous les produits, configurations, options et prix
  - Répond aux besoins du plus grand nombre
  - Peu de possibilités de personnaliser les services proposés
- Peut cibler un public particulier
  - AWS GovCloud: adapté aux normes de sécurité du gouvernement américain

# Modèles déploiement

## Cloud privé

- Infrastructure dédiée à une unique organisation
  - Partagée entre unités commerciales, départements etc.
- Hébergée sur site ou loué dans un datacenter
- Services similaires à ceux proposés dans cloud public
- Respecte les modalités du cloud
  - Approvisionnement self-service, élasticité, automatisation, modèles de facturation ...
- Adapté à de grosses organisations ayant déjà une infrastructure IT
  - Ticket d'entrée élevé
  - Réutiliser compétences et infrastructures existantes
  - Contraintes non satisfaites par les offres de cloud publiques
    - Régulations sur les appels d'offre, la sécurité, contraintes opérationnelles etc.

# Modèles déploiement

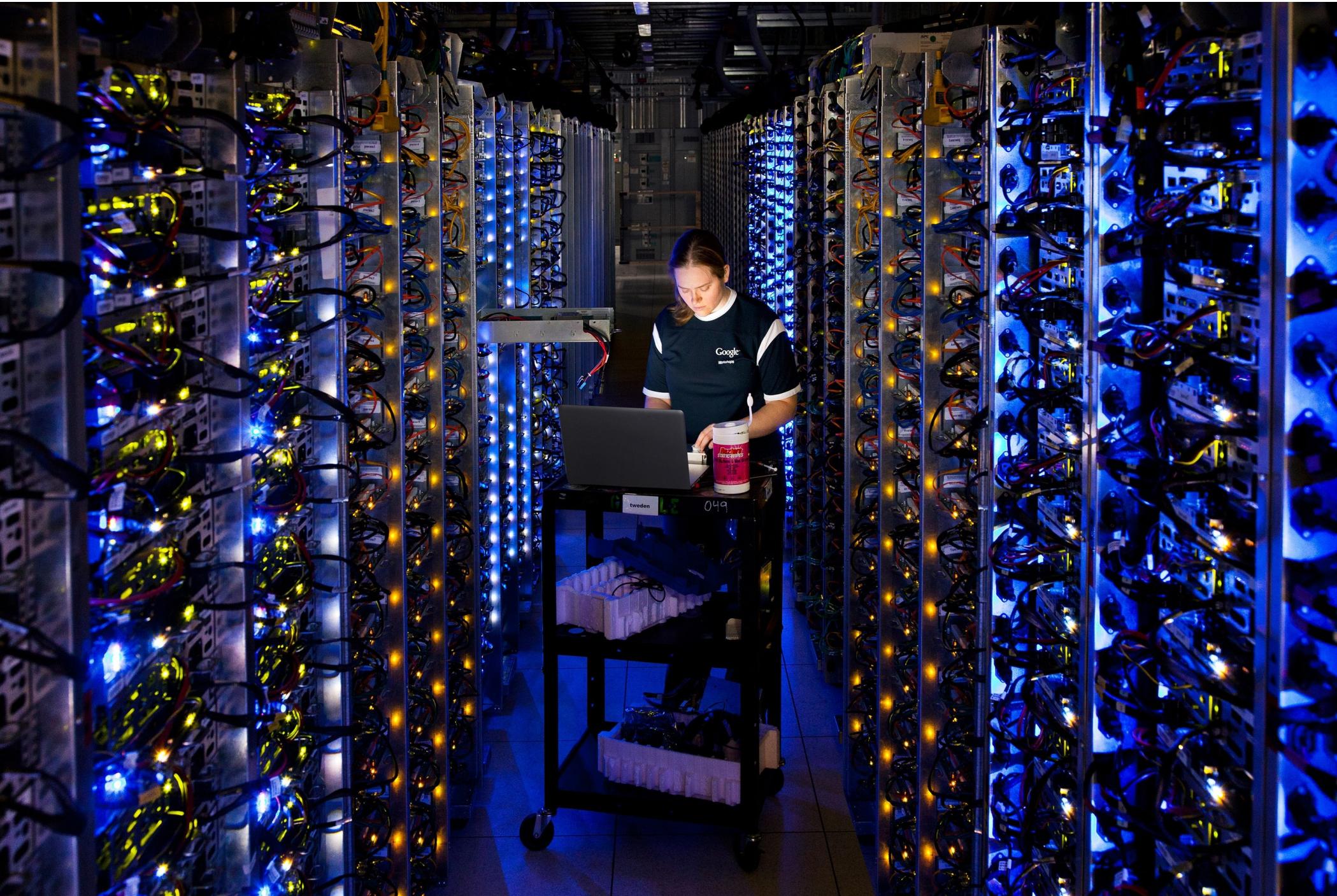
## Cloud hybride

- Agrège de plusieurs infrastructures de cloud
  - Cloud privé et cloud public
- Une interface commune permet de centraliser et uniformiser:
  - l'accès au services cloud et leur exploitation
  - le reporting et la facturation
- Utilisations
  - Cloud bursting: utilisation du cloud public quand le privé est saturé
  - Première étape vers la migration à un cloud public
  - Permet de garder sur site des composants critiques ou trop spécifiques
- Services des opérateurs de cloud public visant ce marché
  - Fédération des systèmes d'authentification
  - Synchronisation des données
  - Interconnexion réseau privée

# Fonctionnement d'un opérateur de cloud



**Un des plus grands centre de données de Google en Europe, à Mons**

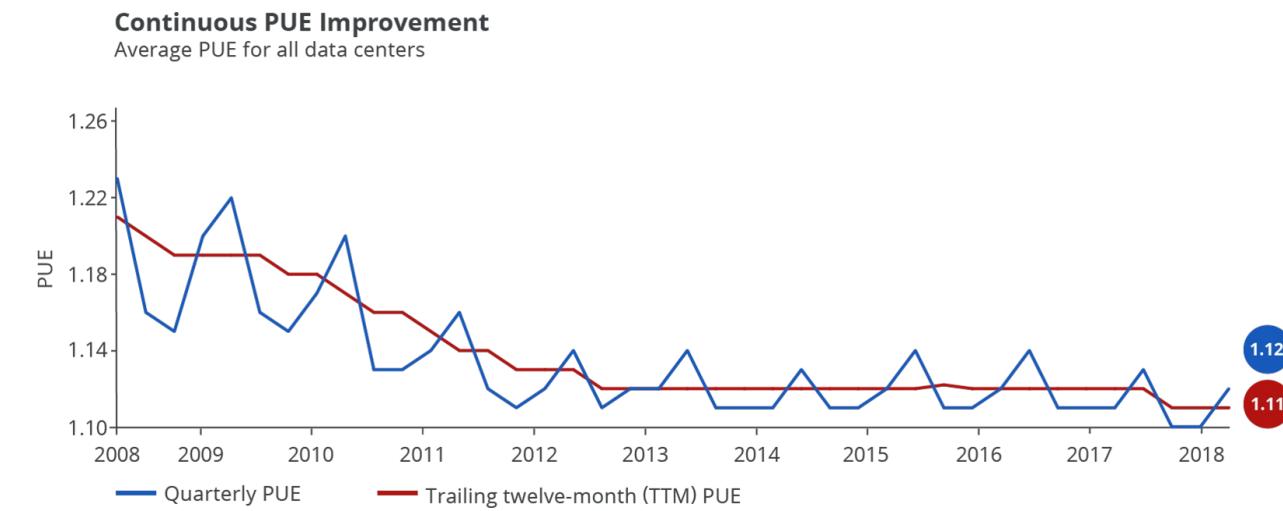
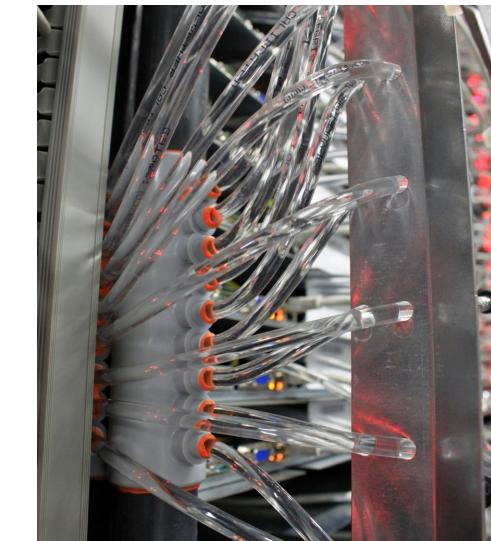


A l'intérieur des salles machines

# Infrastructures à grande échelle

## Choix du lieu d'implantation

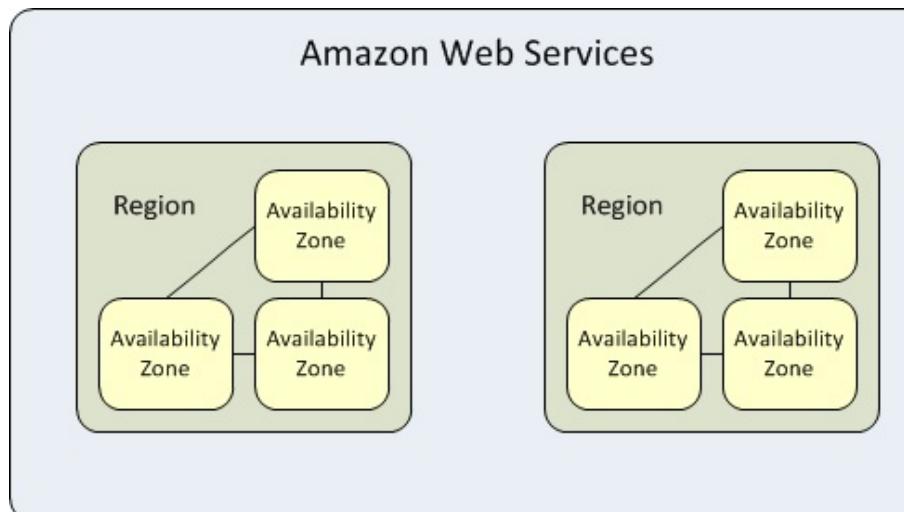
- Alimentation électrique (Amazon: ~30MW par datacenter)
  - Sources d'énergie bon marché (énergie hydraulique)
  - Coût équivalent à celui du matériel sur sa durée de vie
- Optimisation du refroidissement
  - Minimisation du PUE
    - $(\text{Energie totale consommée}) / (\text{Energie consommée par l'informatique})$
  - Climat
  - Proximité de sources d'eau
  - Refroidissement à eau à travers échangeur de chaleur
- Connectivité réseau
  - Peering, Transit
  - Distance aux clients
- Compétences



# Infrastructures à grande échelle

## Régions et zones de disponibilité

- Choix de localisation des ressources offert aux utilisateurs de la plateforme
  - Disposer de ressources partout dans le monde
  - Construire des applications résilientes
- Région: zone géographique large (ex: côte Est, côte Ouest)
  - Divisée en plusieurs zones de disponibilités
  - Au moins un datacenter par zone de disponibilité (AZ)
- Deux AZ au sein d'une région sont:
  - suffisamment éloignées pour minimiser les risques de pannes simultanées
    - Coupure électrique, alea naturel ...
  - suffisamment proches pour une interconnexion réseau faible latence
    - RéPLICATION de données, applications distribuées



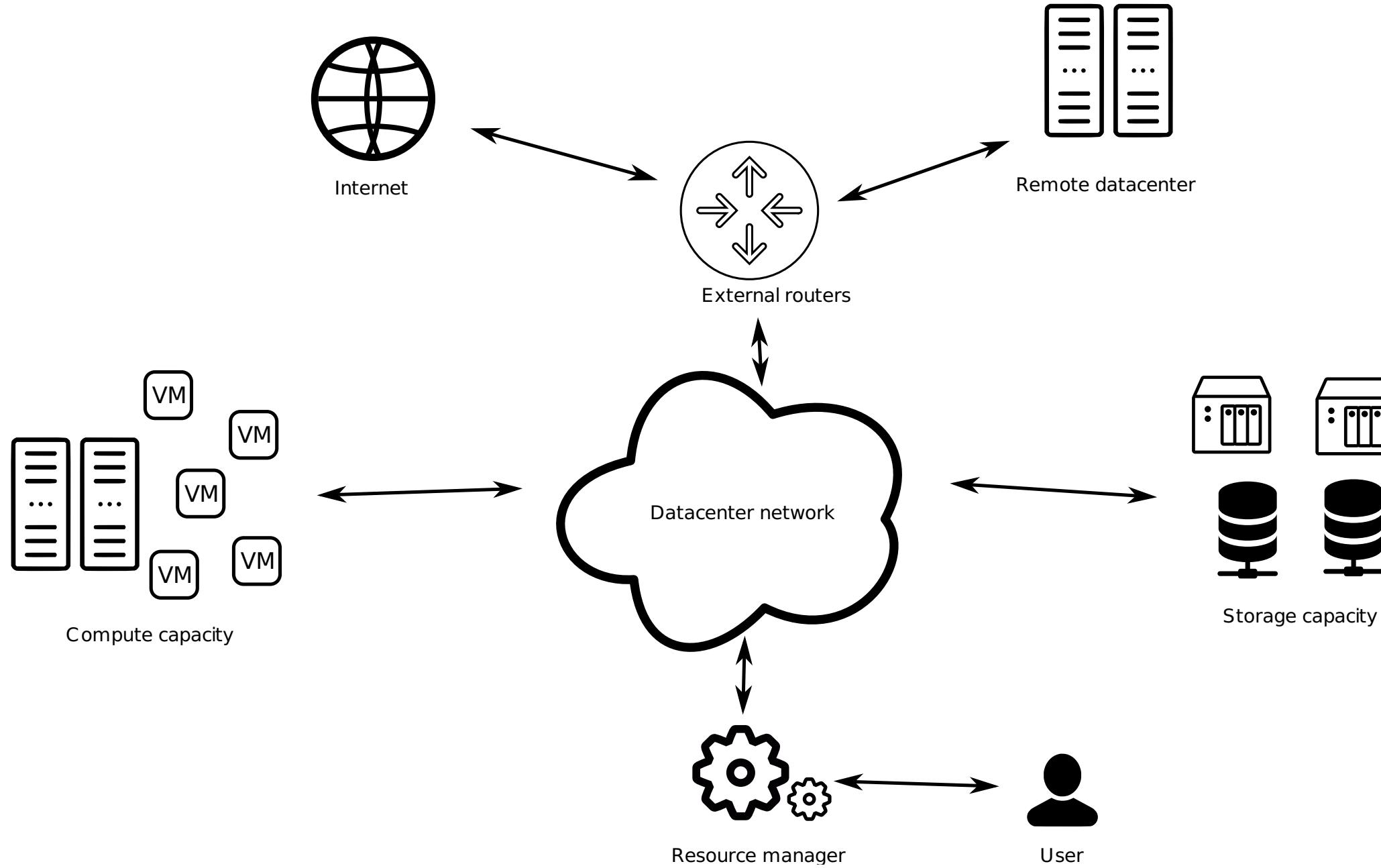
# Infrastructures à grande échelle

## Avantages compétitifs

- Expérience accumulée incomparable

“Every day, Amazon adds enough new server capacity to support all of Amazon’s global infrastructure when it was a \$7 billion annual revenue enterprise,”

- Grosse capacité à investir dans la R&D pour améliorer sa plateforme
  - Réduction des coûts
  - Nouveaux produits répondant aux besoins des clients
  - Fiabilisation de la plateforme
    - Redondance, élimination des SPOF (Single Points Of Failure)
    - Automatisation
    - Surveillance
  - Sécurisation
- Relations privilégiées avec les fournisseurs
  - Influence sur le développement des produits
  - Informations exclusives
    - Roadmaps
    - Problèmes de sécurité



Architecture d'un centre de données

# Capacité de calcul

- Serveurs
  - Majoritairement à base de composants pris sur l'étagère (COTS)
  - Serveurs optimisés pour l'usage spécifique auquel ils sont destinés
    - Suppression des composants non-nécessaires
    - Interfaces uniformisées de gestion à distance
    - Procédures de maintenance standardisées
  - Ordre de grandeur: 50-100k serveurs par DC (Amazon)



# Capacité de calcul

## Machines virtuelles

- Vision d'un matériel virtuel dédié
  - CPU, mémoire
  - Périphériques (réseau, stockage...)
- Un système d'exploitation par VM
  - Indépendant du système de l'hyperviseur

# Capacité de calcul

## Machines virtuelles

- Isolation complète des utilisateurs
  - Sécurité
- Système d'exploitation dédié
  - Indépendance complète de la pile logicielle
- Brique de base des systèmes IaaS

# Gestion des ressources

- Deux objectifs à priori incompatible
  - Éviter les serveurs inutilisés
  - Disposer de la capacité à tout moment pour satisfaire les clients
    - Caractéristique principale du cloud: élasticité
- Tirer profit du nombre élevé de clients
  - Diversité dans les schémas d'utilisation de la plateforme
  - Pics d'activité différents
    - Heures ouvrées ou non
    - Saisonnalité
  - Traitements par lot (batch)
    - Peut être exécuté au moment le plus opportun
    - Incitation financière à exploiter les creux de demande

# Gestion des ressources

- Offre de différents niveaux de qualité de service
  - Ressources pré-réservees
    - Le client planifie ses besoins à l'avance
      - Ex: 10 VMs pendant 1 an
    - Plus la période est longue, plus le coût horaire est bas

# Gestion des ressources

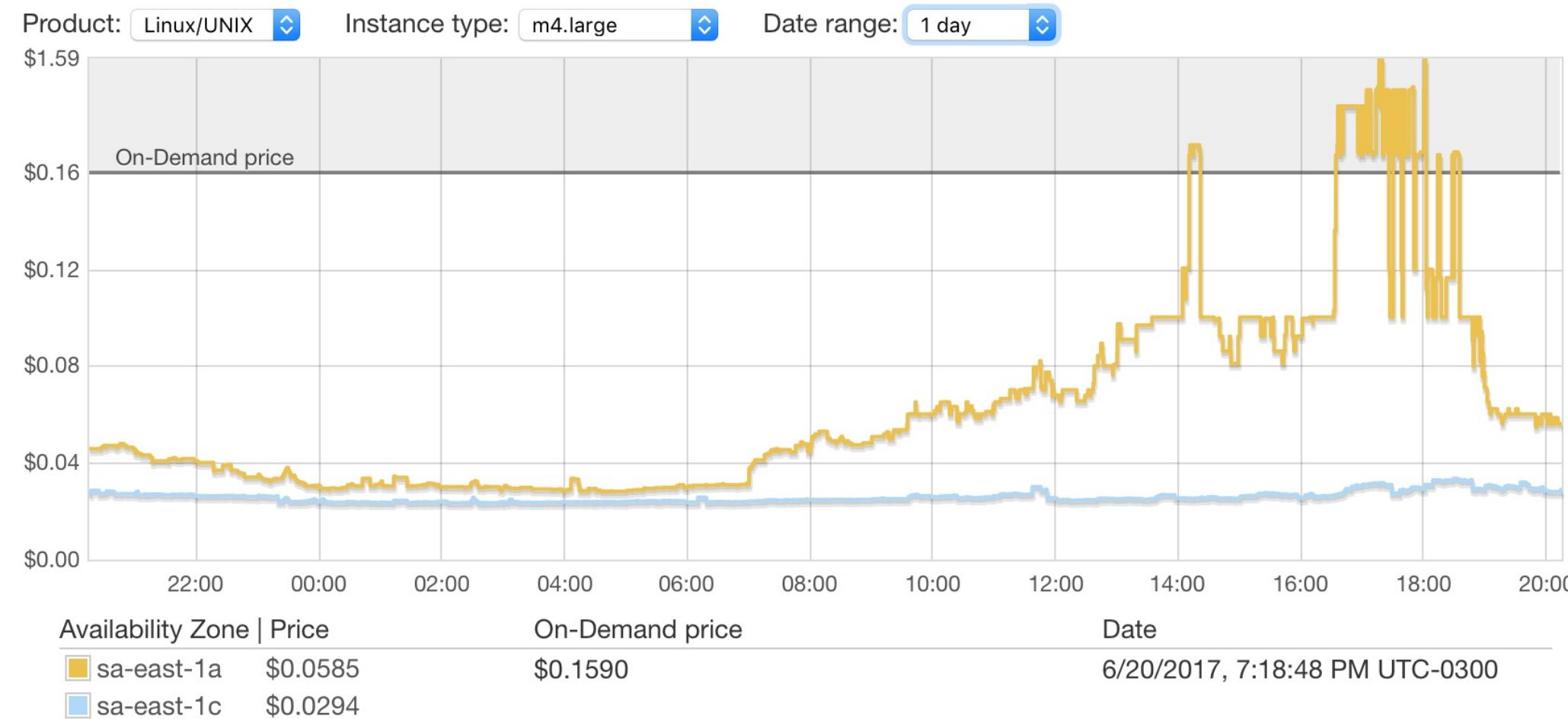
- Offre de différents niveaux de qualité de service
  - Ressources pre-réservées
    - Le client planifie ses besoins à l'avance
      - Ex: 10 VMs pendant 1 an
    - Plus la période est longue, plus le coût horaire est bas
  - Ressources à la demande
    - Des ressources sont allouées au client instantanément
    - Coût le plus élevé

# Gestion des ressources

- Offre de différents niveaux de qualité de service
  - Ressources pre-réservees
    - Le client planifie ses besoins à l'avance
      - Ex: 10 VMs pendant 1 an
    - Plus la période est longue, plus le coût horaire est bas
  - Ressources à la demande
    - Des ressources sont allouées au client instantanément
    - Coût le plus élevé
  - Instances spot
    - Vente aux enchères de la capacité restante
    - Le client place une enchère pour une quantité de ressource
    - Elles sont accordées cours actuel est inférieur au prix de l'enchère
    - Les ressources sont retirées dès que le cours dépasse l'enchère
      - Destruction des VM avec délai de grâce de 2 minutes

# Gestion des ressources

## Spot Instance Pricing History



Exemple de variation des cours spot

# Stockage

- Très gros volumes de données
  - Facebook: 1000 peta-octets par DC
- Entièrement accessible par le réseau
  - Serveurs internes au centre de données
  - Serveurs d'autres centres de données du fournisseur
  - Internet
- Durable et hautement disponible
- Capacité extensible de façon transparente
  - Par simple ajout de matériel identique
  - Croissante du volume stocké et de la bande passante
  - Sans limite d'échelle ou de goulot d'étranglement

# Stockage défini par le logiciel

- Architecture typique:
  - Ensemble de serveurs classiques
    - Relié à un grand nombre de disques
    - Baies internes ou externes
  - Disques rotatif
    - Peu performant (stagnation à 7.2k RPM)
    - Toujours utilisé pour le stockage générique
    - Reste moins coûteux que les SSDs
    - Le coût au GB continue de baisser
    - Bientôt le principal marché pour ce type de disque
  - SSD
    - Indispensable dès qu'il y a un besoin de performance
    - 10k-100k IOPS (~100 IOPS pour un disque rotatif)
    - ~500MB/s (~100MB/s pour un disque rotatif)
    - Performances encore plus élevées en NVMe

# Stockage défini par le logiciel

## Ressources de stockage créées à la demande

- Caractéristiques définies par l'utilisateur
  - Volume
  - Performance (IOPS, bande passante, latence)
  - Niveau de fiabilité
  - Type d'interface: bloc, objet, fichier
- Allocation des ressources nécessaires
  - RéPLICATION/distribution des données
    - Plusieurs disques
    - Plusieurs serveurs
    - Plusieurs centres de données
- Accès transparent via l'interface choisie

# Interfaces de stockage

## Stockage bloc

- Utilisation: disque virtuel
- Interface: iSCSI, ...
- Granularité: petits blocs de taille fixe
- Métagdonnées: aucune
- Mise à jour: partielle, en place

## Système de fichiers

- Utilisation: application distribuée
- Granularité: octet
- Interface: Posix, ...
- Métagdonnées: fixées
- Mise à jour: partielle, en place

## Stockage objet

- Granularité: objet
- Interface: HTTP



### BLOCK STORAGE

Data stored and managed as blocks within sectors and tracks



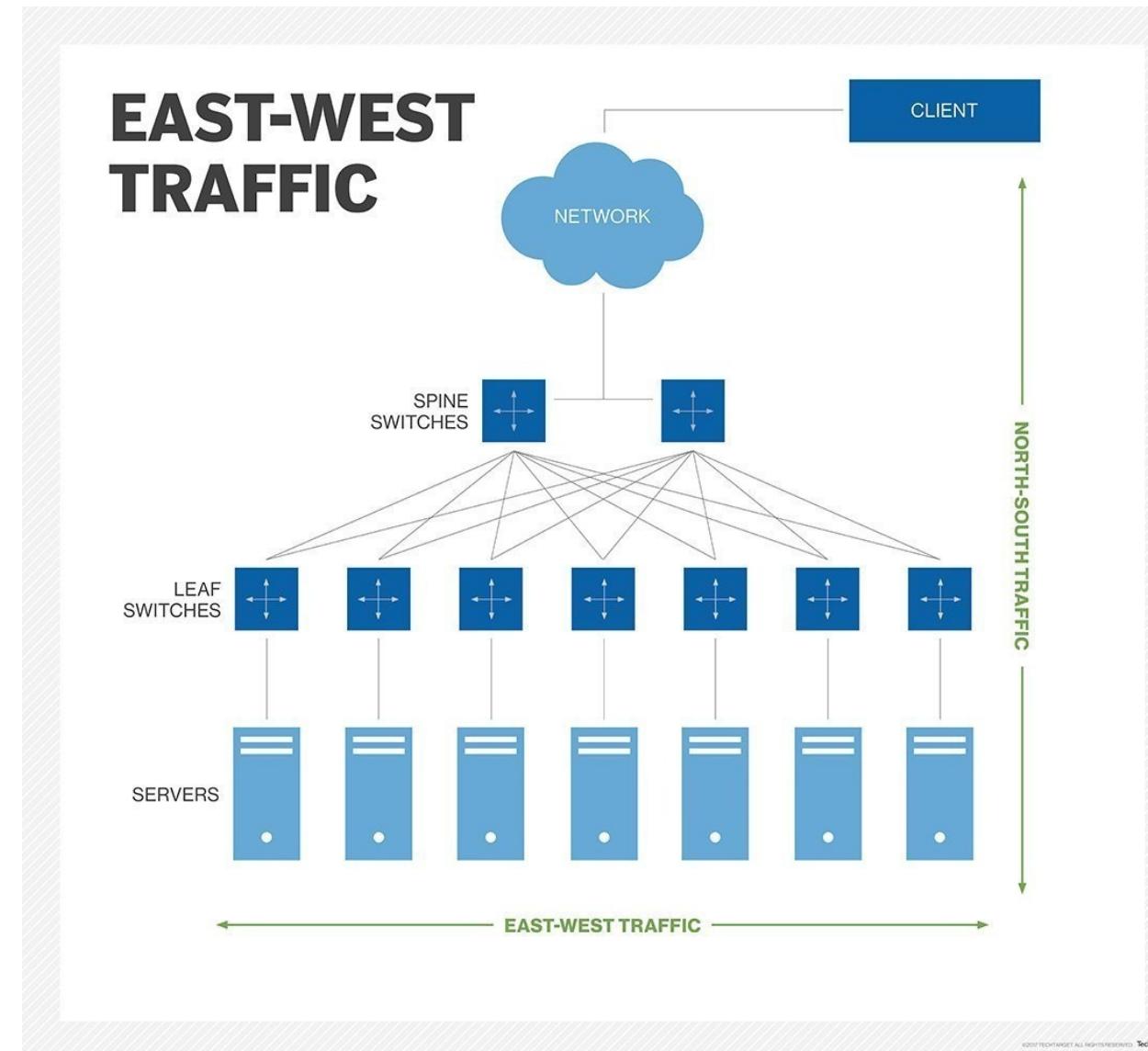
### FILE STORAGE

Data stored as files  
manages files organized into hierarchical file systems

OBJET  
Data  
mas...  
65 with a  
ins

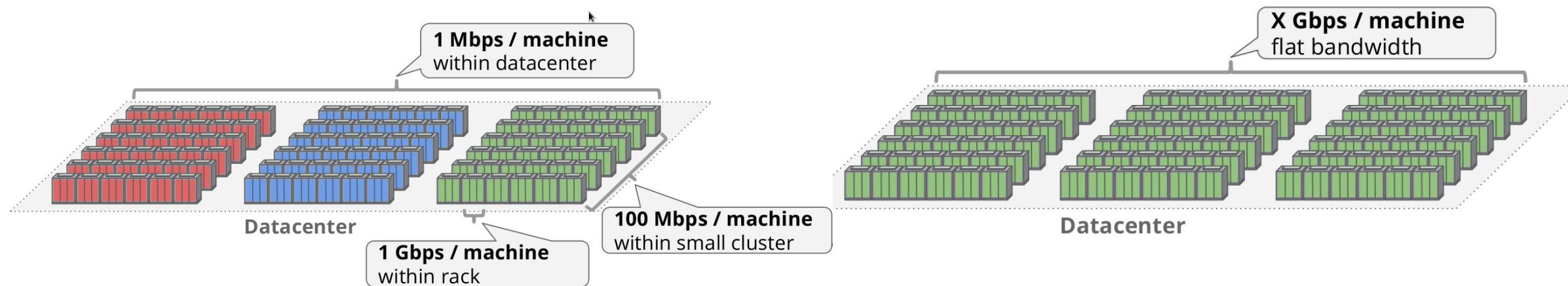
# Les réseaux de centre de données

- Nouveaux défis liés au cloud
  - Charge de travail imprévisible
    - Impossible d'optimiser le réseau pour un type de traffic pré-défini
  - Traffic pas seulement nord-sud mais aussi est-west
    - Applications de plus en plus distribuées (microservices)



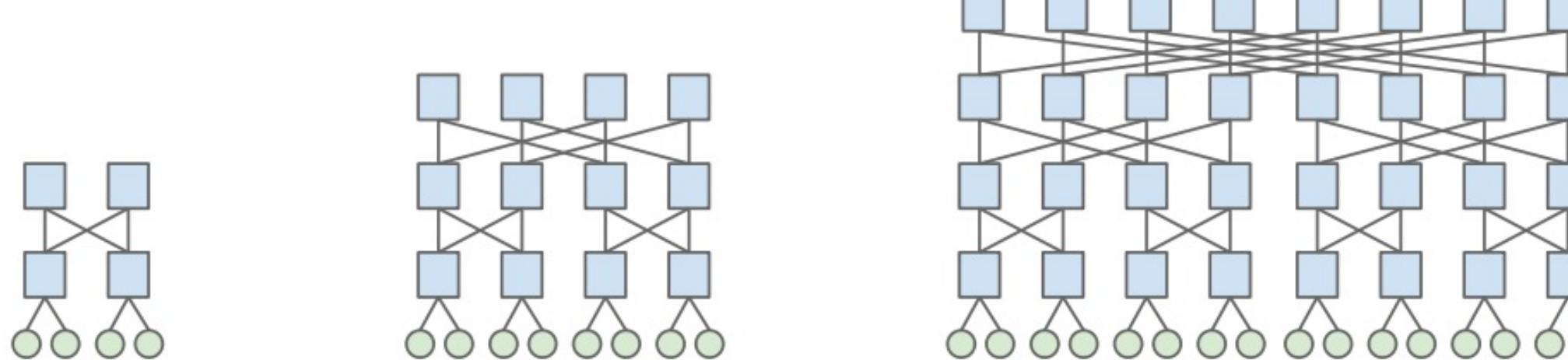
# Les réseaux de centre de données

- Très grande échelle
  - Réseau unique pour l'ensemble du centre données
  - Partitionner empêcherait de bien utiliser les ressources
    - Ex: une sous partie n'ayant plus que des capacités de stockage libre, l'autre n'ayant que du calcul
  - Equilibrer le centre données
    - Aucune ressource ne doit être plus rare qu'une autre



# Les réseaux de centres de données

- Haute performance
  - Loi d'Amdahl moins connue: 1Mb/s d'I/O nécessaires par Mhz
    - Interfaces 100Gb/s pour les serveurs les plus performants
    - Le 10Gb/s se démocratise
  - Pleine bande passante de bisection
    - $100k \text{ serveurs} * 10\text{Gb/s} = 1\text{Pb/s}$
- Comment construire un switch virtuel de 1Pb/s ?
  - Aucun chassis de switch ne permet de gérer autant de serveurs
  - Utilisation de topologies de Clos
    - Grand nombre de switchs génériques bon marché



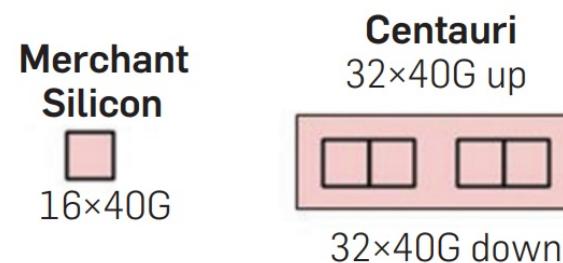
# Google: réseau Jupiter

- Brique de base:

- Circuit intégré pilotant un switch 16 ports
- Chaque port est configurable 1x40Gb/s ou 4x10Gb/s
- Centauri: chassis regroupant 4 switchs indépendants

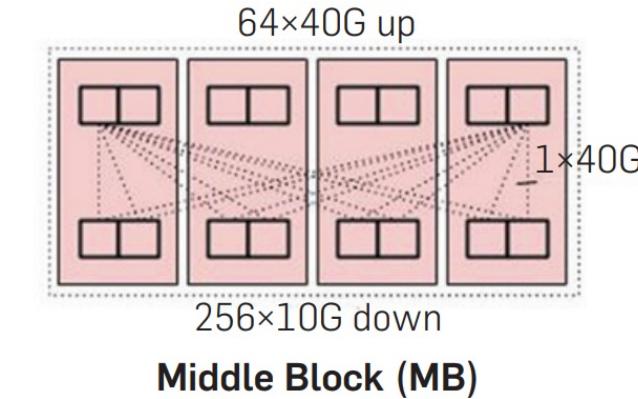
- Top of Rack (TOR):

- Premier niveau de switchs
- Connecte les serveurs
- Modulable suivant la performance souhaitée
  - 10Gb/s bloquant: 48 ports descendants / 16 ports montants
  - 10Gb/s non bloquant: 32 ports descendants / 32 ports montants
  - 40Gb/s non bloquant: 8 ports descendants / 8 ports montants

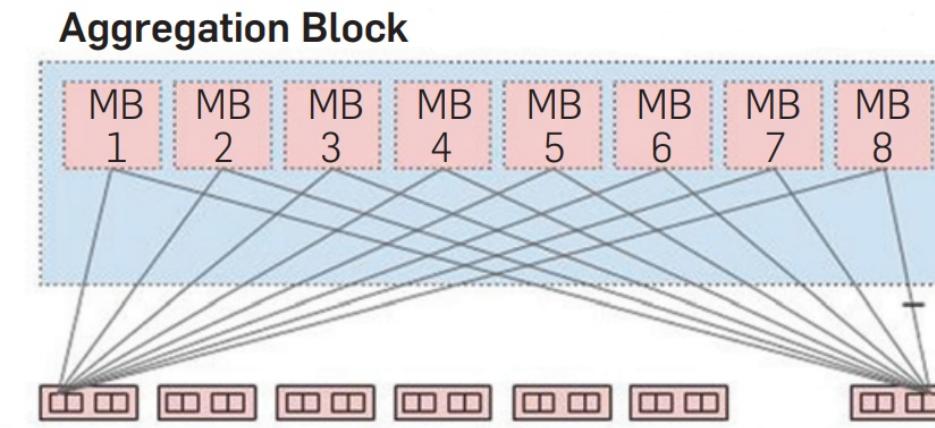


# Google: réseau Jupiter

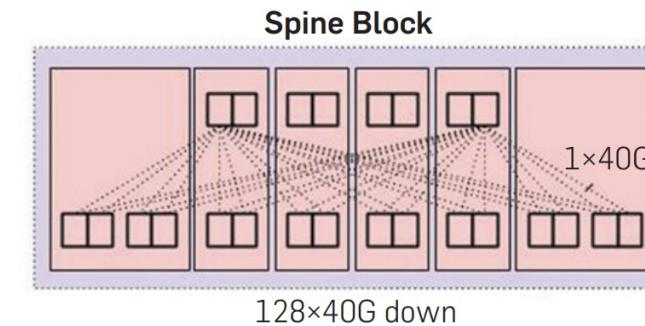
- Middle block (MB)
  - Switch virtuel
  - 256 ports descendants à 10Gb/s
  - 64 ports montants à 40 Gb/s
  - Réalisé à partir de 4 Centauris



- Aggregation Block
  - Relie 8 MB à des switchs TOR



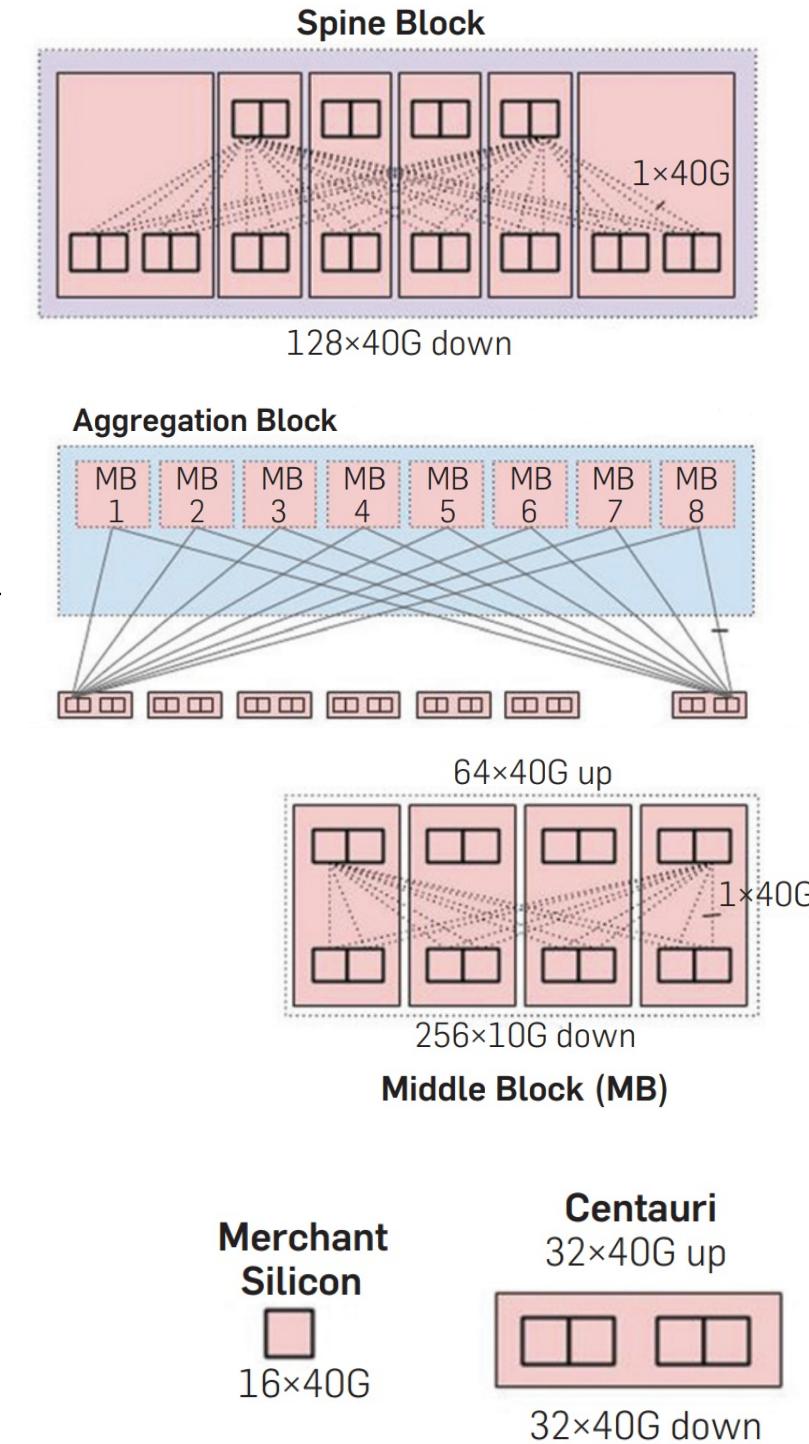
- Spine block
  - Switch virtuel
  - 128 ports descendants à 40Gb/s
  - Réalisé à partir de 6 centauris
  - Interconnecte les aggregation block



# Google: réseau Jupiter

## Exercice:

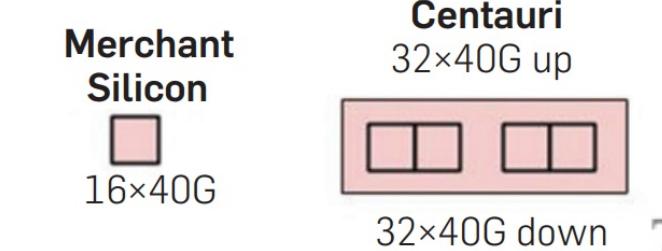
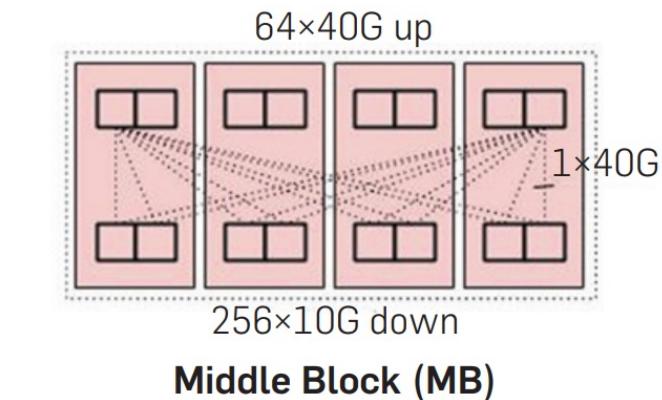
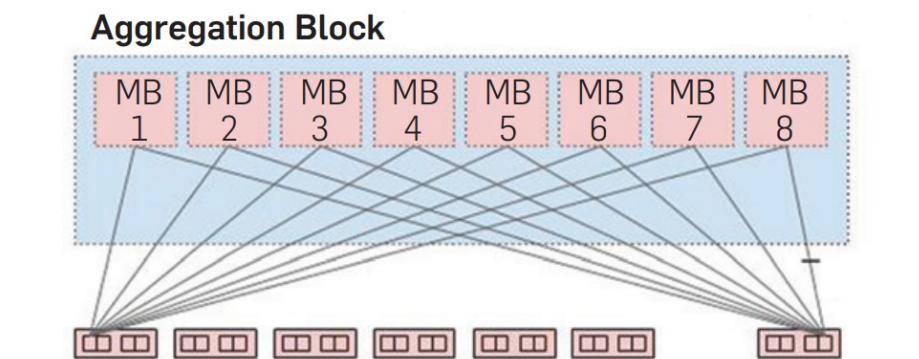
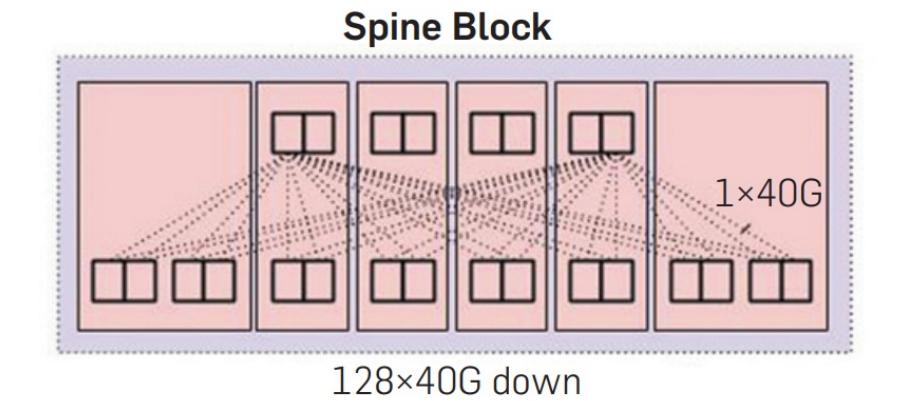
- Calculer le nombre maximal de serveurs supportés par cette topologie
- 10Gb/s, pleine bande passante de bisection
- Rappel:
  - Chaque switch doit être connecté une fois à chaque switch (virtuel) du niveau supérieur
  - Pour améliorer la fiabilité, agrégation de deux liens pour ces connexions



# Google: réseau Jupiter

## Réponse:

- 32 ports up par switch TOR, 8 middle blocks
  - => 4 liens L1 vers chaque middle block
- 256 ports down par middle block, 4 ports par switch L1
  - => 64 switches L1 par aggregation block
- 512 ports up par aggregation block
  - => 256 spine blocks maximum (agrégation de deux liens)
- 128 ports down par spine block, 2 ports par aggregation block
  - => 64 aggregation blocks maximum
- 64 L2 x 64 L1 x 32 noeuds
  - => 131072 serveurs



# Google: réseau Jupiter

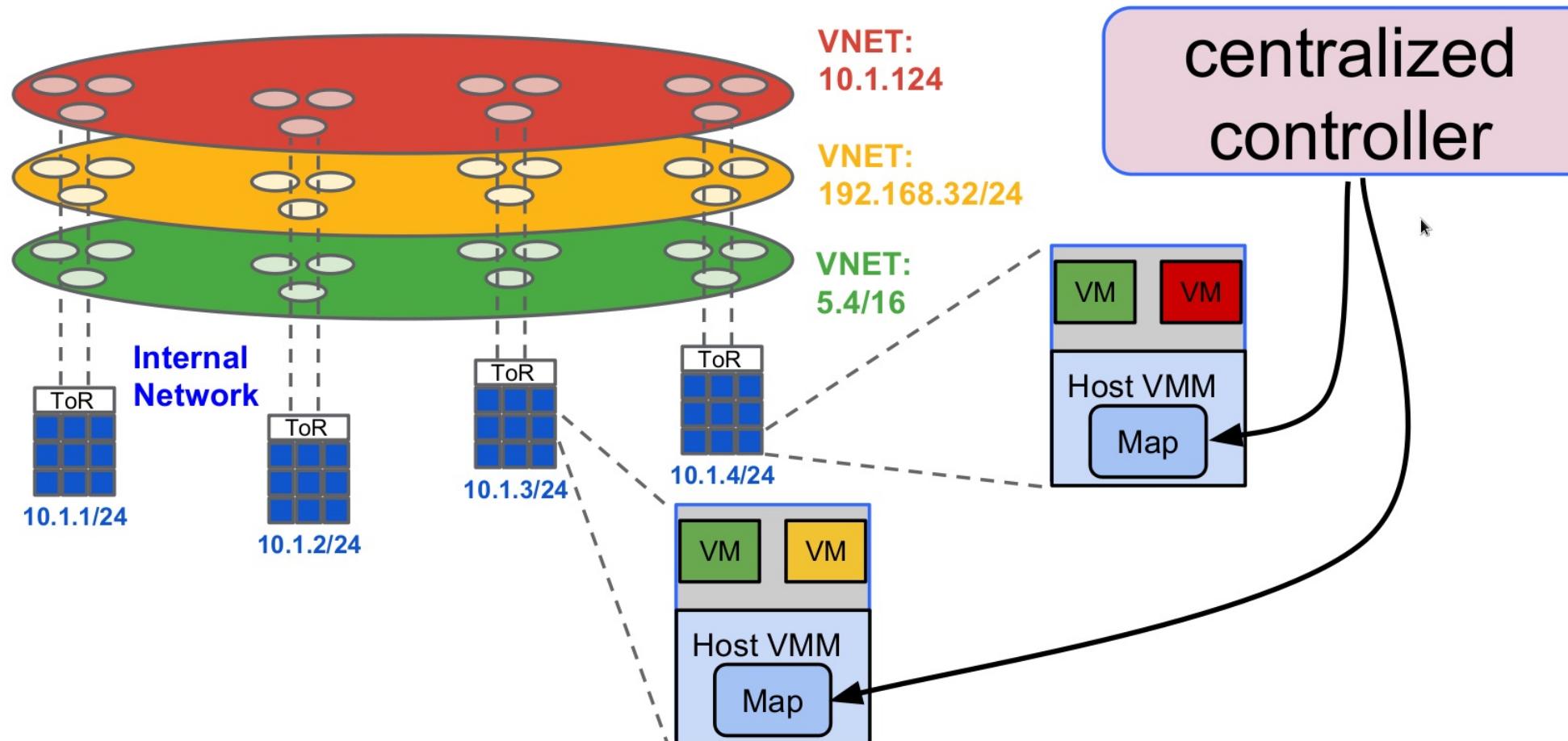
- Routage (Firepath)
  - Domaine de broadcast layer 2 (Ethernet) limité au ToR
    - Spanning Tree inadapté à grande échelle
    - Incapacité d'exploiter les multiples chemins sur le réseau
  - Communication purement layer 3 (IP) en dehors
  - Routage multi-path (ECMP)
  - Recalcul automatique des routes en cas de panne
    - Chaque switch connaît la topologie idéale
    - Un contrôleur (élu) collecte les pannes survenant sur le réseau
    - Il informe les autres switchs des pannes qui mettent à jour leurs routes

# Virtualisation des réseaux

- Infrastructure partagée entre l'ensemble des utilisateurs
- Fournir à chaque utilisateur des réseaux virtuels isolés et dédiés
  - Applications distribuées sur plusieurs VMs qui doivent communiquer
  - Donner l'illusion qu'elles sont sur le même réseau
    - Indépendement du serveur qui les héberge
  - S'assurer qu'elles sont isolées des VMs des autres utilisateurs
    - Sécurité
    - Performance
- Permettre à l'utilisateur d'agrémenter ses réseaux avec des fonctions usuelles
  - Routeur
  - Firewall
  - Répartiteur de charge

# Virtualisation des réseaux

- Virtualisation des adresses dans des espaces de noms distincts
  - Les adresses assignées à un utilisateur n'ont de sens qu'au sein de son réseau
  - Une même adresse peut être assignée à plusieurs utilisateurs
- Implémentation grâce au Software Defined Networking (SDN)



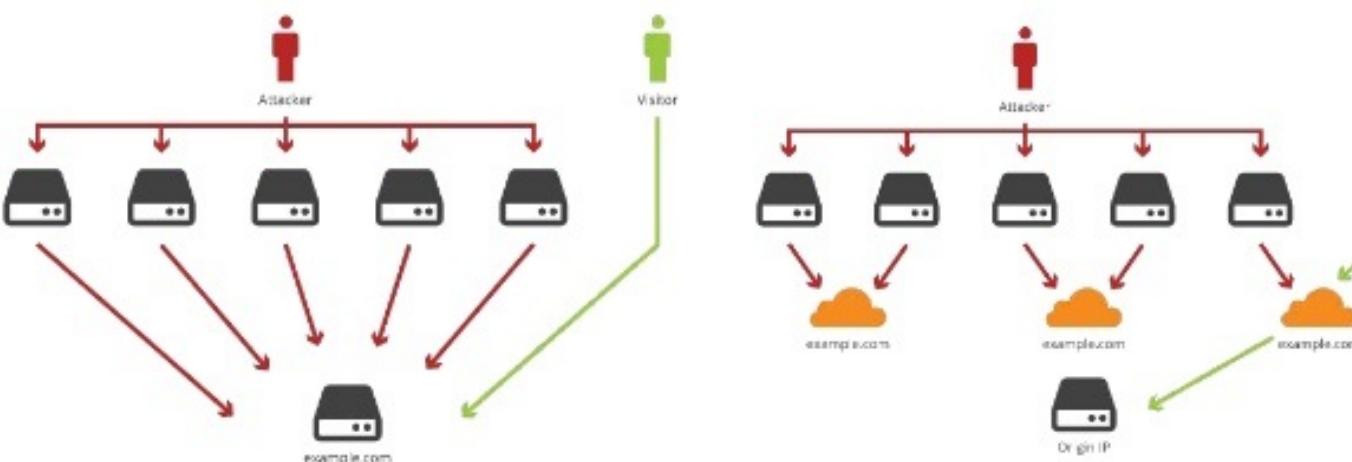
# Software defined networking

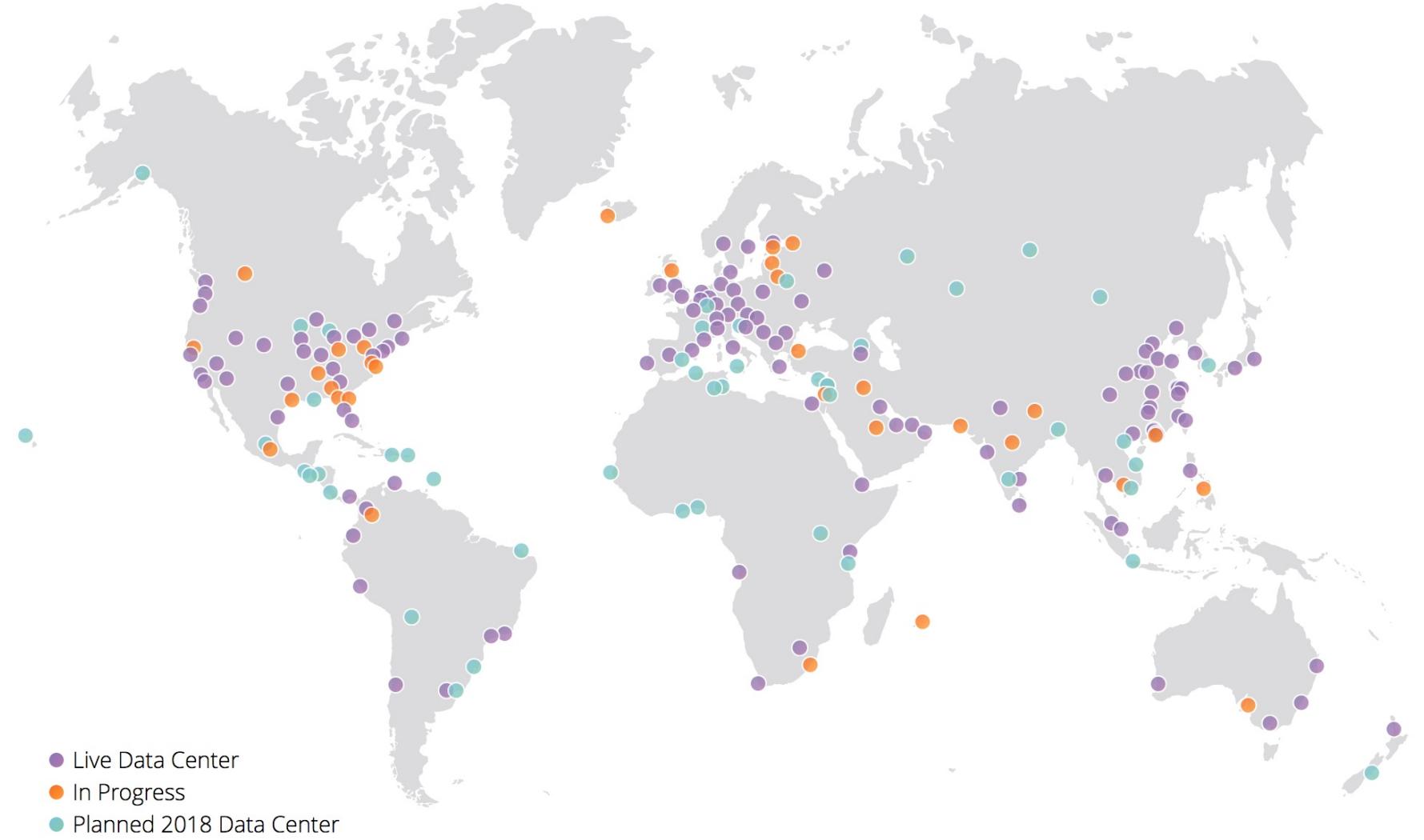
- Séparation entre plan de contrôle et plan de données
  - **Plan de données:** ensemble de règles appliquées par le matériel ou logiciel à chaque paquet
  - **Plan de contrôle:** logiciel qui configure les règles appliquées par le plan de données
- Déport du plan de contrôle vers des serveurs dédiés
  - Standardisation des protocoles de communication entre contrôle et données
    - Notamment le protocole OpenFlow standardisé en 2011
  - Centralisation logique du contrôle l'ensemble du réseau

# Les Content Delivery Networks (CDN)

- Les services hébergés dans le cloud ont souvent un public mondial
- Nécessité de transférer des données efficacement partout dans le monde
  - Faible latence
    - Très important pour les applications Web
    - Impact direct sur la fréquentation d'un site
  - Bande passante élevée
    - Données volumineuses (media)
    - Nombre d'utilisateurs
    - Capacité de résister à des attaques de déni de service (DDoS)
- Solution
  - Disposer de points de présence (PoP) partout dans le monde
  - Répliquer les données localement via des mécanismes de cache

## Products - DDoS Protection





## Les points de présence de CloudFlare dans le monde

- Acteur spécialisé dans le CDN
  - 200 points de présence dans le monde fin 2018
  - Gros investissement dans la protection contre les DDoS
- Autres exemples de CDN:
  - Akamai, AWS, Google ...

# Répartition en fonction de l'origine géographique

- Méthodes basées sur le DNS (www.site.com -> 1.2.3.4)
  - GeoDNS:
    - Lors d'une requête DNS, analyse de l'IP source de la requête
    - Réponse en fonction du pays d'origine de l'IP
- Avantage:
  - Simple à implémenter (centralisation au niveau du serveur DNS)
- Inconvénient:
  - La requête DNS est rarement directement faite par le client final
  - L'IP du serveur faisant la requête n'est pas suffisante pour identifier le meilleur PoP

# Répartition en fonction de l'origine géographique

- Méthodes basées sur le routage IP
  - Anycast
    - Tous les serveurs du CDN ont la même adresse IP externe
    - Via BGP différentes routes sont distribuées sur les routeurs de chaque région du monde
- Avantage:
  - Décision basée sur le routage permettant un meilleur choix du serveur de destination
- Inconvénient:
  - Implémentation distribuée, collaboration avec divers acteurs
  - En cas de problème le trafic peut basculer de façon inopinée d'un serveur à l'autre

# Les impacts du cloud sur les entreprises

# Considérations économiques

## Le prix

- Intrinsèquement un service cloud n'est pas forcément bon marché
  - Un fournisseur de cloud bénéficie d'économies d'échelles
    - Mais il facture sa R&D et ses marges
    - La marge opérationnelle d'AWS est de 25%
  - Répliquer son infrastructure dans le cloud peut être plus coûteux
- Difficile de comparer dans l'absolu le *total cost of ownership* (TCO)
  - Structure de coûts variant d'une entreprise à l'autre
- D'après HPE (vendeur de matériel)
  - Une infra. sur site revient 1.7x à 3.4x moins cher que sur AWS
- D'après AWS (vendeur de cloud)
  - Une infra. migrée dans le cloud revient 30% moins cher que sur site

# Considérations économiques

## CapEx vs OpEx

- **CapEx (Capital Expenses)**: investissement en capital visant à acquérir de nouvelles capacités apportant une valeur positive sur le long terme
- **OpEx (Operating Expenses)**: dépenses courantes nécessaires au fonctionnement de l'entreprise et de ses installations
- Historiquement les capacités informatiques étaient nécessairement acquises par CapEx
  - Achat de matériel informatique et d'infrastructures
  - Achat de logiciels à licences perpetuelles
- Le cloud computing offre une alternative basée sur les OpEx
  - Caractéristiques fondamentales du cloud computing
    - Capacité de calcul à la demande
    - Paiement uniquement de ce qui est consommé

# Considérations économiques

## CapEx vs OpEx

- Utiliser des CapEx a quelques avantages pour une entreprise
  - Capacité définitivement acquise
  - Actif amortissable sur plusieurs années
  - Dépenses prévisibles et maîtrisées
  - Indépendance vis à vis du fournisseur
- Cependant cela a aussi plusieurs inconvénients
  - Manque de réactivité
    - Nécessite de pouvoir prévoir ses besoins à long terme
    - Fait intervenir des budgets sujet à plus de validation hiérarchique
  - Génère des coûts supplémentaires d'infrastructures
  - Doit être opéré et maintenu par l'acheteur
  - Problèmes de vieillissement
    - Architectures plus vraiment adaptées mais pas encore rentabilisées

# Considérations économiques

## CapEx vs OpEx

- Les évolutions de l'informatique renforcent le poids de ces inconvénients
  - Évolution de plus en plus rapide des logiciels et des besoins associés
- Transition numérique des entreprises
  - Satisfaire rapidement tout besoin métier lié au numérique
    - Avantage concurrentiel stratégique
    - Prise de décision au plus près de chaque équipe
- Crise financière de 2008-2013
  - Assèchement des financements
  - Difficultés à investir sur le long terme
- Le cloud computing arrive au bon moment pour solutionner ces problèmes

# Considérations économiques

## Impact de ce nouveau mode d'acquisition

- L'avantage économique des grosses entreprises est réduit
  - Un compétiteur peu émerger sans avoir à lever énormément de fonds
  - Mise au point très rapides de produits d'envergure mondiale
- Evolution touchant tous les secteurs économiques
  - Y compris domaines à priori non technologiques
  - Utilisations innovantes du numérique bouleversant des entreprises bien établies
- Exemples grand public:
  - Uber
  - AirBnB
  - Netflix

# Considérations économiques

## L'exemple de Zynga

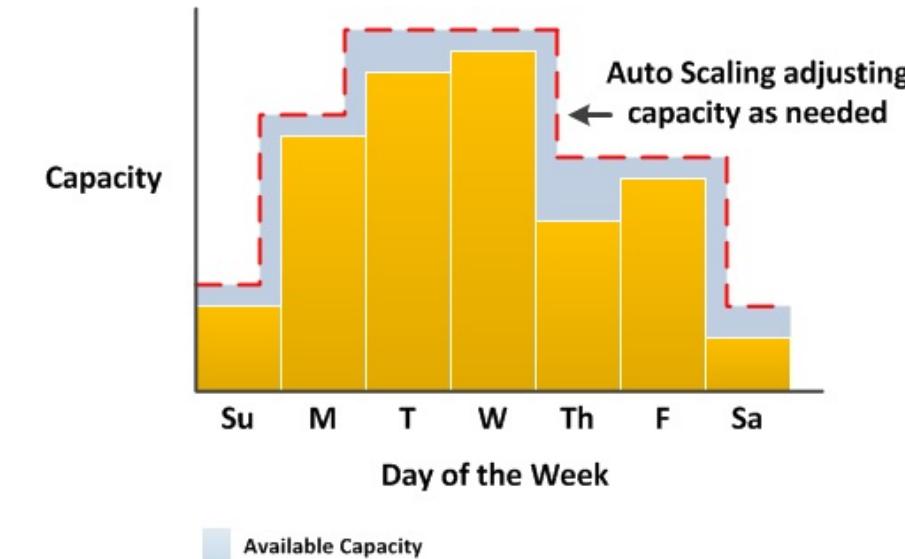
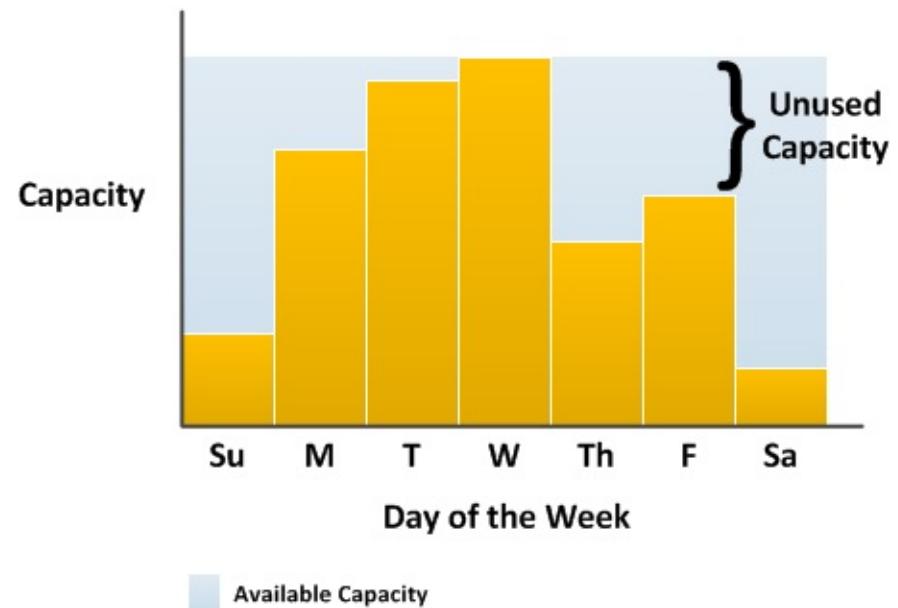


- Succès fulgurant de FarmVille
  - De 0 à 10 millions d'utilisateurs journaliers en quelques semaines
    - 10 000 instances de VMs sur AWS
  - Succès au delà des espérances
    - Impossible d'acheter aussi rapidement autant de serveurs
- Décide d'acquérir ses propres serveurs pour poursuivre son développement
- Rapidement le modèle économique de Zynga s'écroule
  - Dépendance à Facebook qui freine le développement de ces jeux
  - Bascule vers une utilisation mobile
- Infrastructure inadaptée, sur-dimensionnée et en voie d'obsolescence
  - Retour sur AWS

# Considérations économiques

## Adaptation des dépenses aux besoins

- Le besoin en ressources est souvent irregulier
  - Heures ouvrées non ouvrées
  - Saisonnalité
- Nécessité de dimensionner les serveurs selon le pic de demande



# Impacts organisationnels

## Réponse aux nouvelles exigences envers les services informatiques

- Les applications internes doivent:
  - Etre accessible sur tous types de terminaux
    - Mobile
    - BYOD (Bring Your Own Device)
  - Etre utilisables en télétravail
  - Fonctionner 24/7 de façon fiable et sécurisée
  - Disposer de contrôle d'accès étendus
    - Ex: Permettre un partage de données limité avec des prestataires
  - Présenter une interface agréable et intuitive
    - Agacement face à des applications moins modernes que les applications grand public
- Évolution souvent trop lente des intranets
- Le modèle SaaS répond particulièrement bien à ces besoins

# Impacts organisationnels

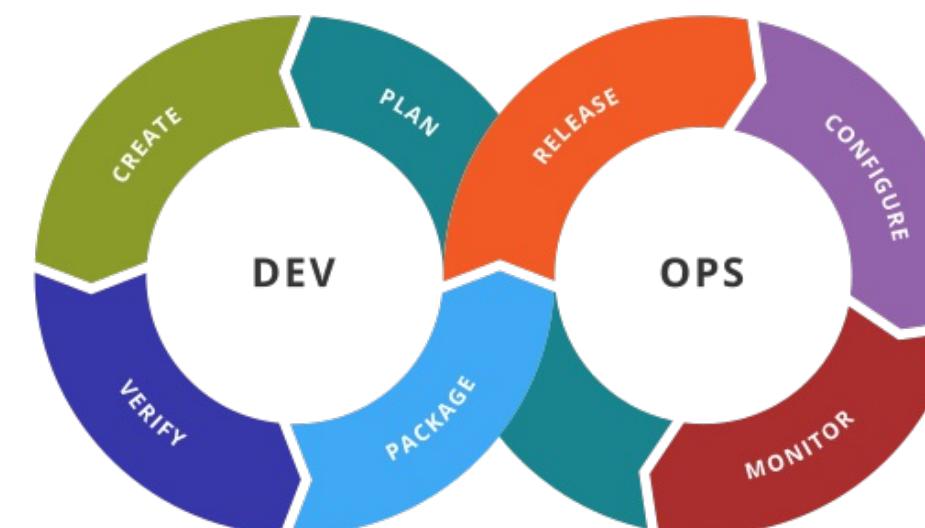
## Rédefinition du rôle des départements informatiques

- Moins de ressources accaparées des tâches génériques
  - Évite les tâches communes à la gestion toute infrastructure informatique
  - Bénéficie des optimisations des opérateurs de cloud
- Formation des équipes en place aux technologies du cloud
- Focalisation sur les besoins métiers
  - Plus de réactivité vis à vis des besoins métiers
    - Différence avec la sous-traitance
  - Proposer des solutions innovantes favorisant la création de valeur
    - Ne plus être seulement un centre de coûts
    - Expérimentation plus facile dans le cloud

# Impacts organisationnels

## Le devops

- Equipes de développement (**dev**)
  - Chargé du développement logiciel
  - Jugé sur sa capacité à réaliser les développements dans les temps
  - Risque sacrifier la fiabilité pour atteindre ses objectifs
- Equipes d'administration (**ops**)
  - Chargé de la gestion des infrastructures, du déploiement des logiciels et de la surveillance de leur fonctionnement
  - Jugé sur sa capacité à assurer la fiabilité de systèmes
  - Risque de limiter l'innovation pour éviter l'introduction de problèmes
- Considérer ces deux catégories de tâches comme un ensemble imbriqué devant être réalisé de concert



# Impacts organisationnels

## Le devops

- Bénéfices:
  - Aligne les responsabilités des équipes sur des objectifs communs
    - Prise en charge de l'intégralité de la chaîne
    - Du développement au déploiement
  - Meilleure prise en compte des contraintes de production dans les développements
    - Uniformisation des environnements de développement et de production
    - Intégration continue
    - Automatisation des déploiements
  - Permet l'application de principes agiles
    - Réduction des cycles de développement
    - Déploiement en production très fréquents

# Impacts organisationnels

## Le devops

- Le cloud en facilite grandement la mise en oeuvre
  - Rapprochement des compétences nécessaires
    - Programmabilité des infrastructures
    - Nécessite moins de compétences systèmes spécifiques
  - Facilité de mise en place d'environnements de test
  - Culture de l'automatisation

# Les problématiques liées au cloud

# Les problématiques liées au cloud

## Sécurité des données et régulation

- Enjeux importants autour des données numériques
  - Données personnelles, industrielles, commerciales ...
  - Protection et/ou valorisation des données
- Confiance dans le prestataire hébergeant les données
  - Sécurité informatique
    - Capacité de sécurisation de la plateforme
    - Possibilité d'être une victime collatérale d'un piratage
  - Respect des réglementations
    - RGPD
    - Hébergement dans un pays tiers
  - Surveillance étatique
    - CLOUD Act
- Responsabilité vis-à-vis de l'utilisateur final



# Les problématiques liées au cloud

## Centralisation d'Internet

- Risques de pannes systémique
  - AWS (Février 2017)
    - Origine: mauvaise commande lors du redémarrage d'un service
    - Panne du service de stockage objet S3 dans une region
    - 54 des 100 premiers sites e-commerce affectés
    - Nombreux objets connectés en panne
  - Dyn (Octobre: 2016)
    - Origine: cyberattaque
    - Service DNS inaccessible en Amerique du Nord
    - Grand nombre de sites inaccessibles aux Etats-Unis
    - Airbnb, Github, Netflix, Reddit, Spotify, Twitter ...

# Les problématiques liées au cloud

## Enfermement propriétaire (lock-in)

- Difficulté de changer de fournisseur
  - API propriétaires
  - Services uniques à un fournisseur
  - Touche particulièrement le PaaS
- Développement de stratégies multi-cloud
  - Eviter de s'enfermer chez un fournisseur
  - Outils facilitant l'exploitation plusieurs plateformes

“Nous ne pouvons pas facilement basculer nos opérations à un autre fournisseur de cloud, et toute disruption de, ou interférence avec, notre utilisation de Google Cloud Platform aurait une incidence défavorable importante sur notre business, nos résultats d'exploitation et notre santé financière”

Spotify

# Les problématiques liées au cloud

## Enfermement propriétaire (lock-in)

- Prix transparents mais difficiles à prévoir
  - Nombreux coûts variables s'additionnant au coût horaire des ressources
    - Quantité de données transférées
    - Nombre de requêtes BDD
    - Appels à une API...
  - Variabilité d'un mois à l'autre
    - Variation du nombre d'utilisateurs
    - DDoS
    - Bug
  - Diagnostiquer l'origine d'une sur-consommation peut être difficile
    - Impossibilité d'instrumenter à sa guise le service
- Risque de s'enfermer sur une plateforme coûteuse à terme

“Toute transition des services fournis actuellement par Google cloud à un autre fournisseur de cloud serait difficile à réaliser et nous causerait des dépenses et une perte de temps significative. Nous nous sommes engagés à dépenser 2 milliards de dollars chez Google cloud sur les cinq prochaines années”

# Les problématiques liées au cloud

## Enfermement propriétaire (lock-in)

- Dépendance au fournisseur
  - Tout système est sujet à des pannes, dysfonctionnements ou ralentissements
  - En cas de problème, manque d'information temps réel sur le diagnostic
    - Problème pour informer ses propres clients sur les services impactés
  - Limites imposées aux services proposés ou fonctionnalités manquantes
    - Peu / pas de personnalisation possible
- Transfert de données
  - Difficulté de transférer de grosses quantités de données par internet

# Les problématiques liées au cloud

## Enfermement propriétaire (lock-in)



Merci de votre attention !

Des questions ?

