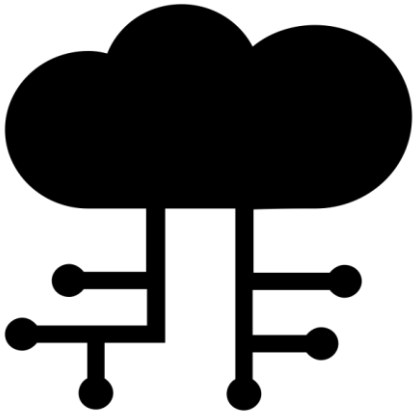
Cloud Computing

Etat des lieux, définitions et concepts



CPE LYON
ÉCOLE SUPÉRIEURE DE CHIMIE PHYSIQUE ÉLECTRONIQUE DE LYON

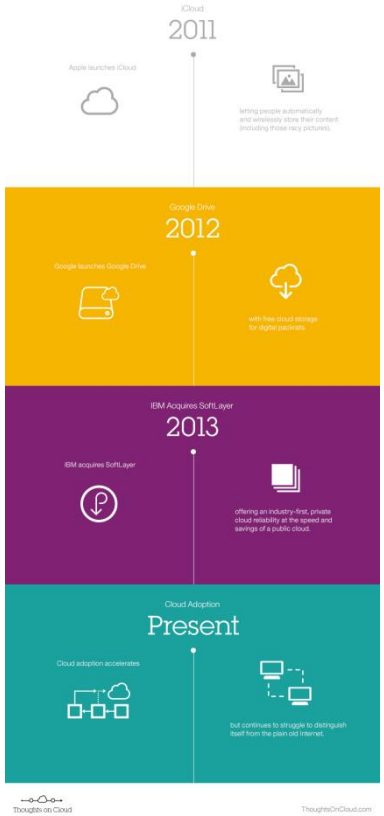
Plan

- Introduction – concepts et définitions
- Architecture type et contraintes
- E.G : Amazone AWS

Qu'est ce que le Cloud Computing ?

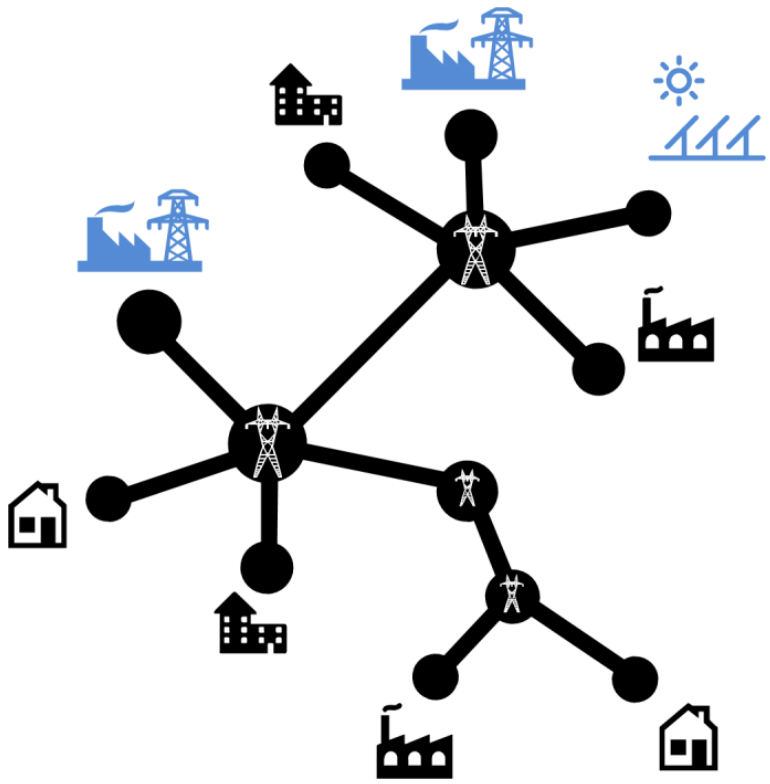
- Un terme Marketing, pour parler d'une évolution plus que d'une révolution

- Principes de bases :
  - On utilise un service
  - L'infrastructure sous jacente est mutualisée
  - On paye en fonction de ce qu'on utilise réellement



- Les équivalents hors informatique
  - L'électricité
  - L'eau

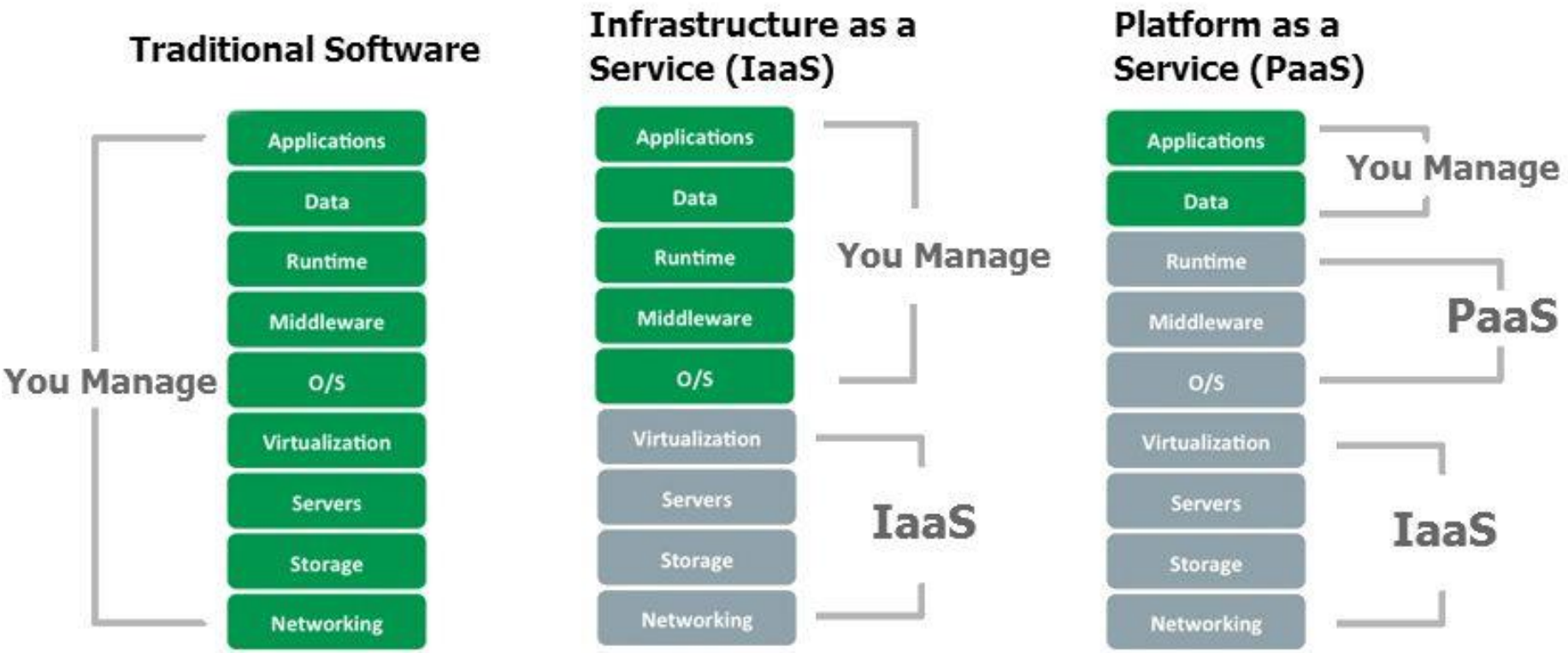
- Je paye ce que j'utilise réellement
- L'infrastructure est mutualisée



- En distingue plusieurs catégories de services

  - **SAAS** – Software As A Service
  - **PAAS** – Plateforme As A Service
  - **IAAS** – Infrastructure As A Service

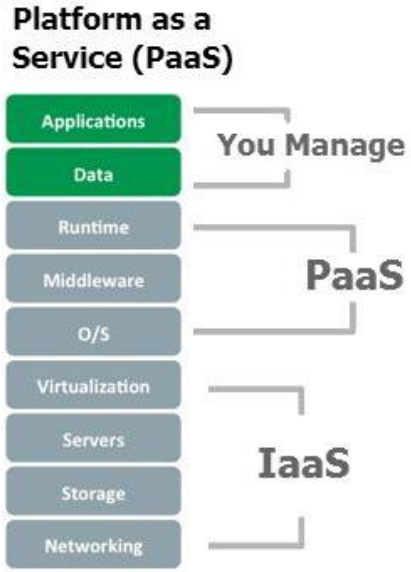
Qu'est ce que le Cloud Computing ?



http://cloud-computing.softwareinsider.com/

SAAS – Software As A Service

- On vend l'utilisation d'un soft en fonction de l'usage que l'on en fait.
  - Payement à l'utilisateur,
  - au nombre de pages vues
  - Bande passante etc ..
- SalesForce.com – Outil de CRM As A Service
  - +150 000 Clients
  - 1.5 Millions d'utilisateurs
  - +1000 Serveurs, 1,5 milliards de requêtes /jour
- Tous les outils dit Web 2.0
  - Gratuit, Payant, ou en Freemium (free + premium)



PAAS – Plateforme As A Service

- Fourniture d'une plateforme applicative
  - PHP, Java, Python etc …
- Payement en fonction du nombre d'utilisateurs
- Evolution des hébergements mutualisés

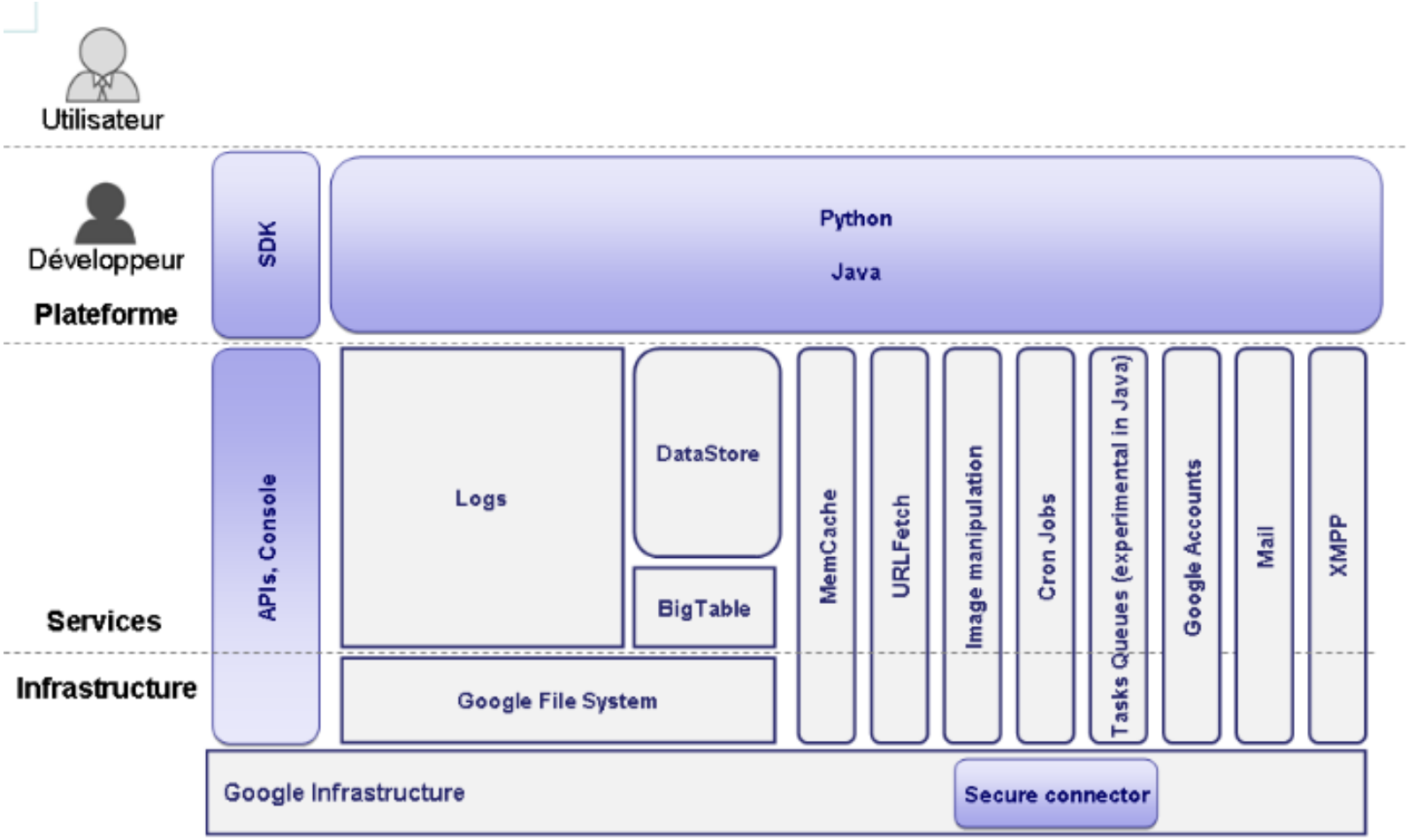
- Google AppEngine
  - Java, Python, Go
- Force.com
  - Apex, Ruby, Java, PHP (Heroku)

PAAS – Plateforme As A Service

Comparatif des principaux PaaS du marché				
Service	Editeur	Langages	IDE	Bases de données
Amazon Elastic Beanstalk	Amazon Web Services	Containers Docker, Java (Tomcat), .NET, Node.js, PHP, Python, Ruby (Passenger)	Extensions AWS Toolkit pour Eclipse et Microsoft Visual Studio	Amazon RDS, DynamoDB, SimpleDB, SQL Server, Oracle, IDBM DB2, Informix
Google App Engine	Google	Java, Python, PHP 5.4, Go	Google Plugin for Eclipse pour Java	Cloud SQL
Azure Cloud Services	Microsoft	Java, Node.js, PHP, Python, .NET, Ruby	WebMatrix, Visual Studio + Azure SDK	Data Services Microsoft : Blobs, base SQL, SQL Server, MongoDB
Salesforce1 platform	Salesforce	Langage Apex sur Force.com, Java, Ruby, Node.js, Python sur Heroku1	Outil web pour Force.com, client Heroku toolbelt pour Windows, MacOS X, Linux Debian/Ubuntu	Force.com Database, Heroku Postgres + les les SGBD de l'add-on marketplace Heroku, dont MongoDB, Postgres, MySQLÂÂ
Cloud Foundry	Cloud Foundry Community	Java, Grails, Play, Spring, Node.js, Ruby on Rails, Sinatra, Go, Erlang, etc.	Plugin Eclipse	MySQL, PostgreSQL, MongoDB
Source : JDN				

CPE LYON
ÉCOLE SUPÉRIEURE DE CHIMIE PHYSIQUE ÉLECTRONIQUE DE LYON

PAAS – Google App Engine



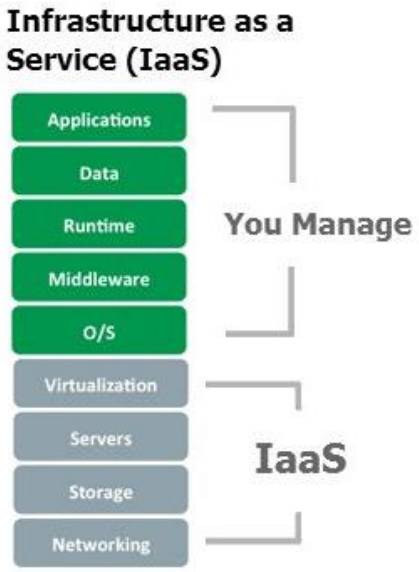
PAAS – Google App Engine

	FREE *LIMIT PER DAY*	PRICE *ABOVE FREE LIMIT*
Instances	28 instance hours	$0.05 / instance / hour
Cloud Datastore *(NoSQL)*	• 50k read/write/small • 1 GB storage	• $0.06 / 100k read or write ops • Small operations free* • $0.18 / GB / month
Network Traffic *(Outgoing)*	1 GB	$0.12 / GB
Network Traffic *(Incoming)*	1 GB	FREE
Cloud Storage	5 GB	$0.026 / GB / month
Memcache	• Free Usage of Shared Pool • No free quota for Dedicated Pool	• Free Usage of Shared Pool • Dedicated Pool: $0.06 / GB / hour
Search	• 1000 basic operations • 0.01 GB indexing documents • 0.25 GB document storage • 100 searches	• 0.50 / 10k searches • 2.00 / GB indexing documents • 0.18 / GB / month Storage
Email API	100 recipients	Contact Sales
Logs API	100 MB	$0.12 per GB
Task Queue	5 GB	$0.026 / GB / month
Logs Storage	1 GB	$0.026 / GB / month
SSL Virtual IPs	-	$39 / virtual IP / month
Bundled Services	Cron, Image Manipulation, Pagespeed, SNI SSL Certificates, Socket API, Task Queue API, URLFetch, Users API	

CPE LYON
ÉCOLE SUPÉRIEURE DE CHIMIE PHYSIQUE ÉLECTRONIQUE DE LYON

IAAS – Infrastructure As A Service

- Fourniture d'un service à la demande
  - Serveurs
  - Espace de stockage
  - Bases de données
  - Etc …
- Payement
  - à l'heure d'utilisation
  - Au Go stocké
  - Au Go transféré



IAAS – Infrastructure As A Service

	Windows Azure	EC2 amazon web services	rackspace	DigitalOcean	linode.com	SOFTLAYER	Google Compute Engine
Low-end	1core/.75gb /20gb	1core/1gb/EBS	1core/1gb/20gb	1core/512mb /20gb	1core/1gb/24gb	1core/1gb/25gb	1 core/.6gb
High-end	16core/112gb/ 382gb	16core/117gb/24x2048	32core/120gb/1.2tb	20core/64gb / 640gb	20core/96gb /1.9TB	custom	16core/104gb
Pricing	$.018 - $4.9/hr	$.013 - $4.6/hr	$.052 - $6.24/hr*	$.007 - $.952/hr	$.015 - $1.44/hr	$.04 - $x.xx/hr	$.012 - $1.184/hr
Bandwidth	5GB incl., then ~$.08/GB	15GB incl.	~$.10/GB	1-9TB Incl.	2-20TB Incl.	5TB incl.	~$.01/GB
Network Out	up to 10GB	**	200MB-10GB	1GB	125MB-10GB	up to 10GB	**
Bare Metal	No	No	Yes	No	No	Yes	No
SLA	99.95%	99.95%	99.9%	99.99%	99.9%	1	99.95%

https://www.scriptrock.com/articles/cloud-service-provider-roundup-the-best-of-the-best

https://en.wikipedia.org/wiki/Cloud_computing_comparison

CPE LYON
ÉCOLE SUPÉRIEURE DE CHIMIE PHYSIQUE ÉLECTRONIQUE DE LYON

+ Avantages

- Diminue le temps de déploiement
- Mutualisation des couts
  - Développement et matériel
- Fiabilité
- Accessibles par tout le monde
- Très forte évolutivité

- Inconvénients

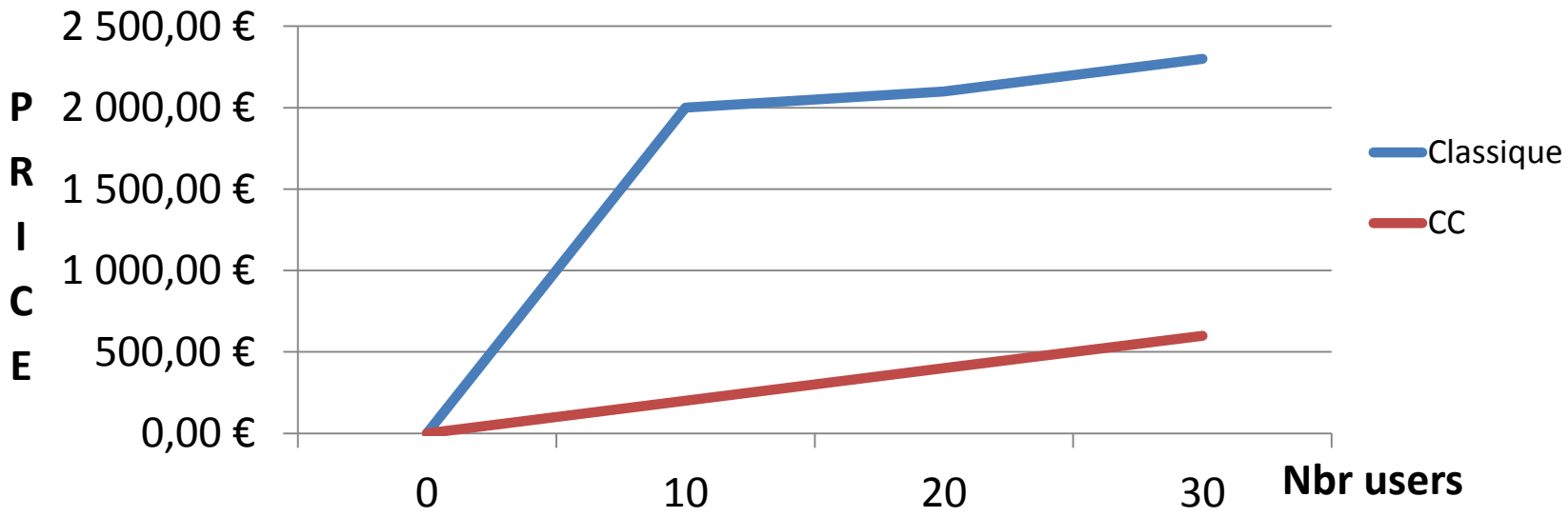
- Sécurité des données
- Rigidité des applications
- Sécurisation de l'accès
- Cycle de vie commun a tous

Classique

- Payement d'une licence
- Achat et de serveurs
- Paramétrage
- Dimensionnement

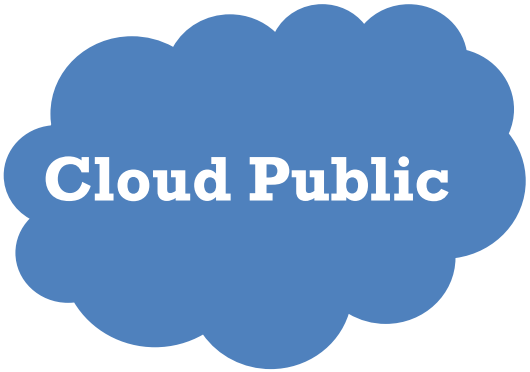
Cloud Computing

- Sécurité des données
- Rigidité des applications
- Sécurisation de l'accès
- Cycle de vie commun a tous



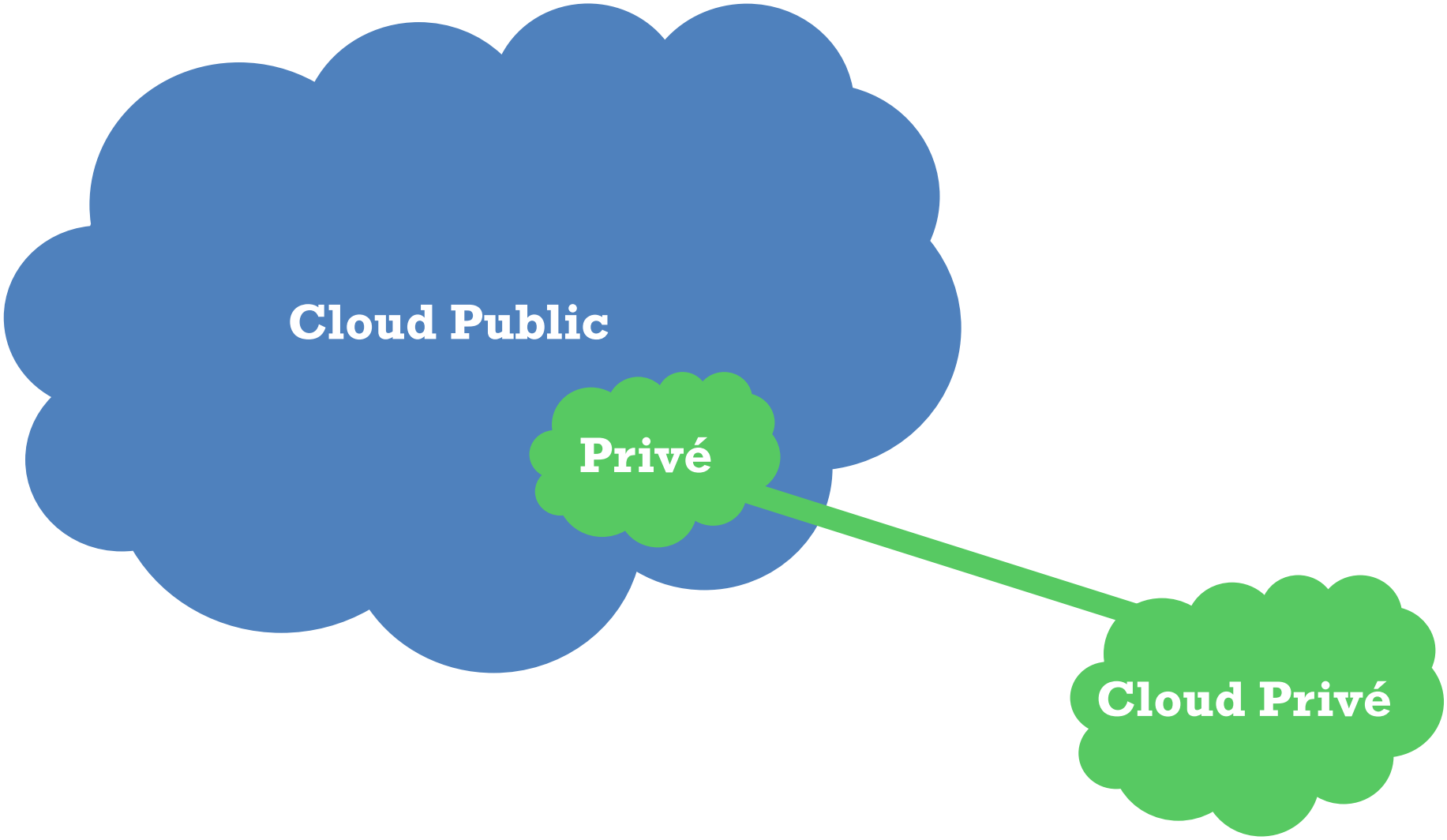
Public

- Exécution chez un prestataire public
- google, amazon, go grid
- Public car accessible en théorie par tout le monde …

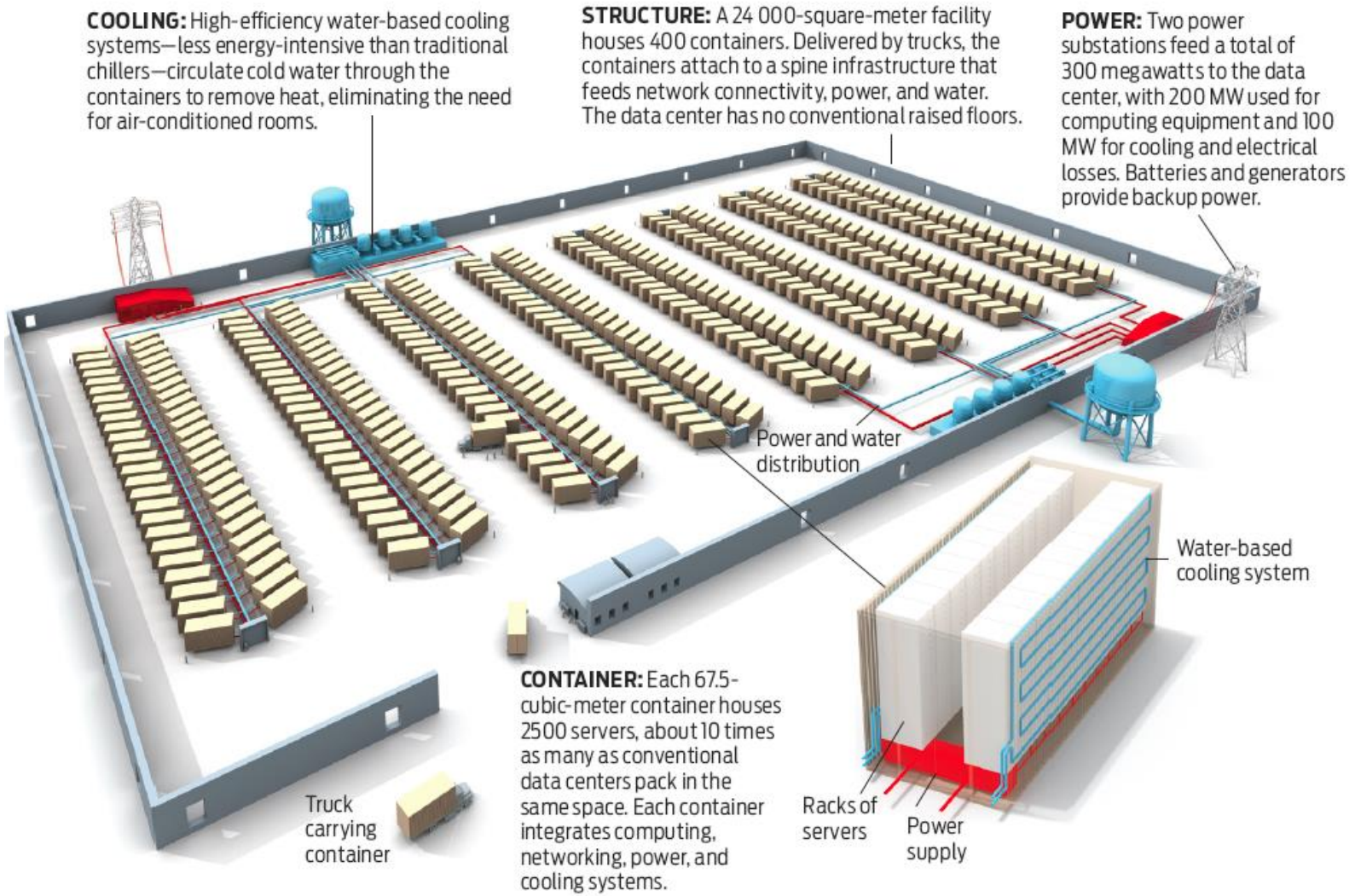


Privé

- Exécution sur les infrastructures d'une entreprise, ou d'un sous groupe de machines
- Associé a plus de sécurité

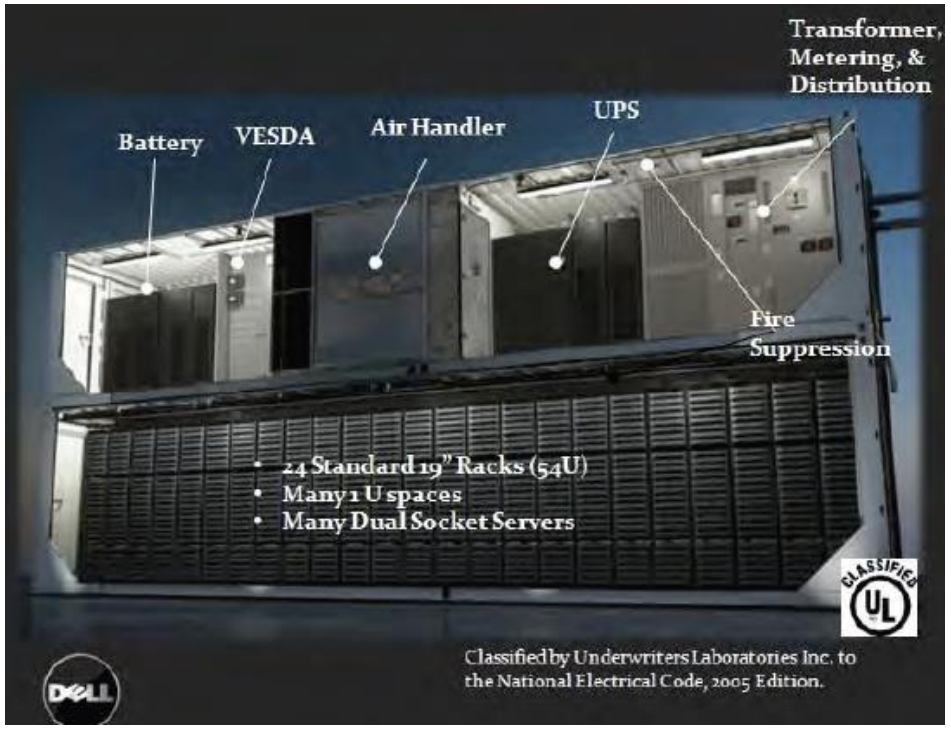


Exemple de datacenter



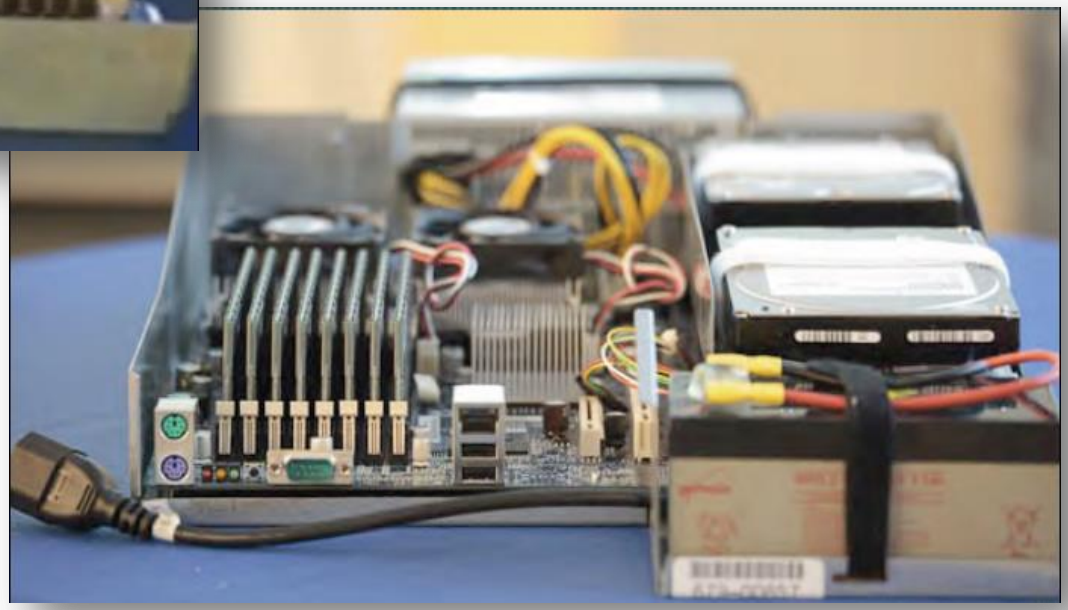
Containerisation du Cloud

Cloud Microsoft Chicago



Louis Naugès, Chief Cloud Evangelist - Revevol International

Home made Server (Google)

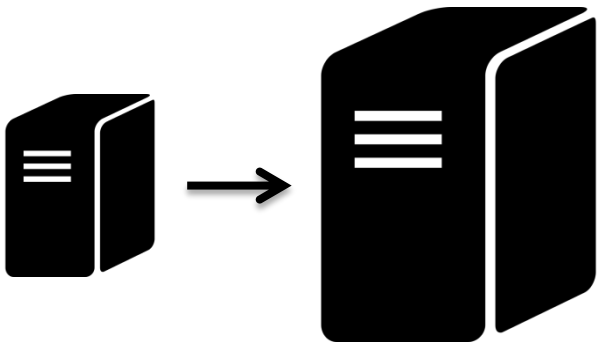


Louis Naugès, Chief Cloud Evangelist - Revevol International

World Data Centers Distribution Google

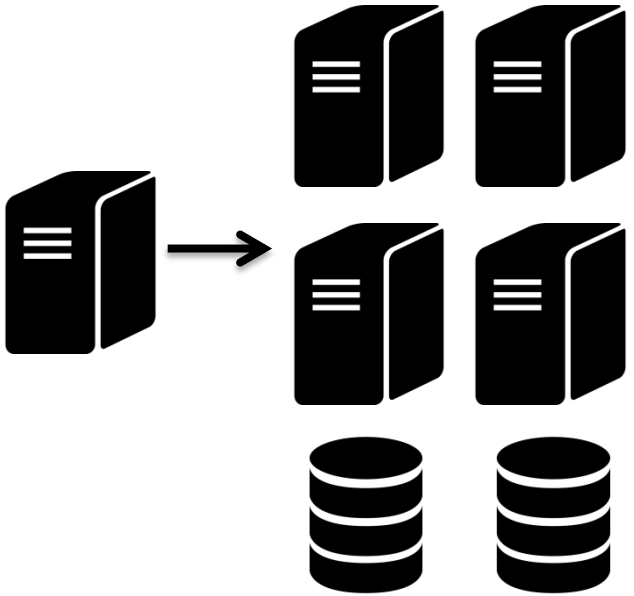
Vertical

- Augmentation des ressources du système
  - Rapide mais limité

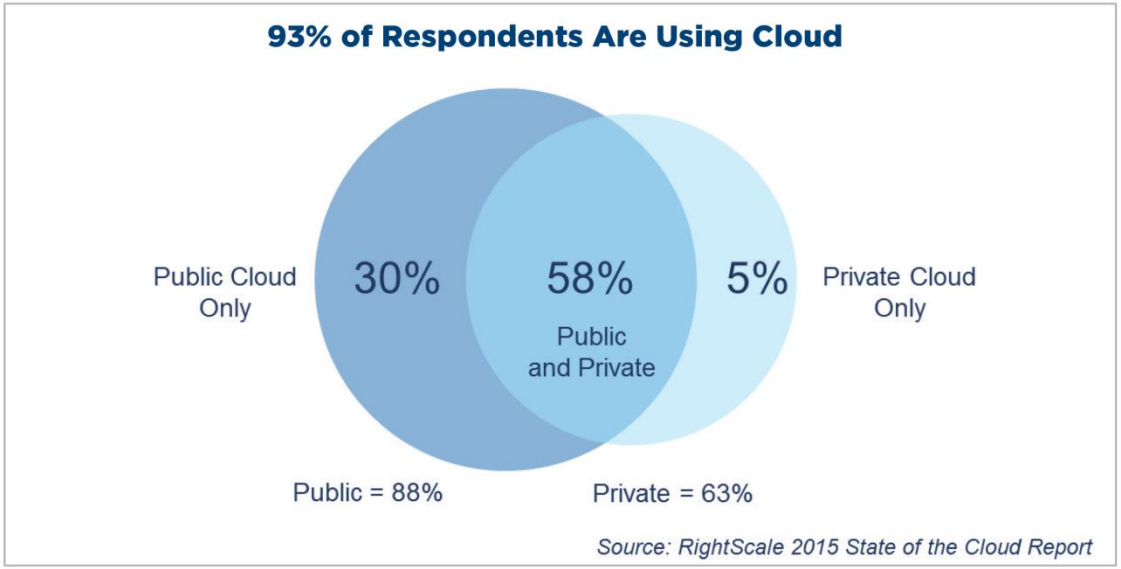


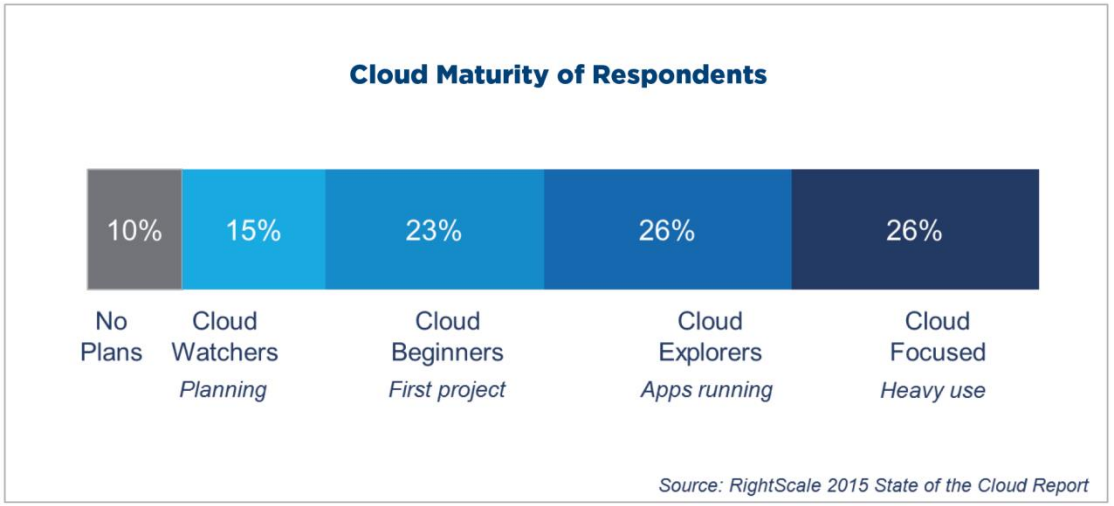
Horizontale

- Augmentation du nombre de machines
  - Nécessite une compatibilité logiciel,
  - potentiellement illimité

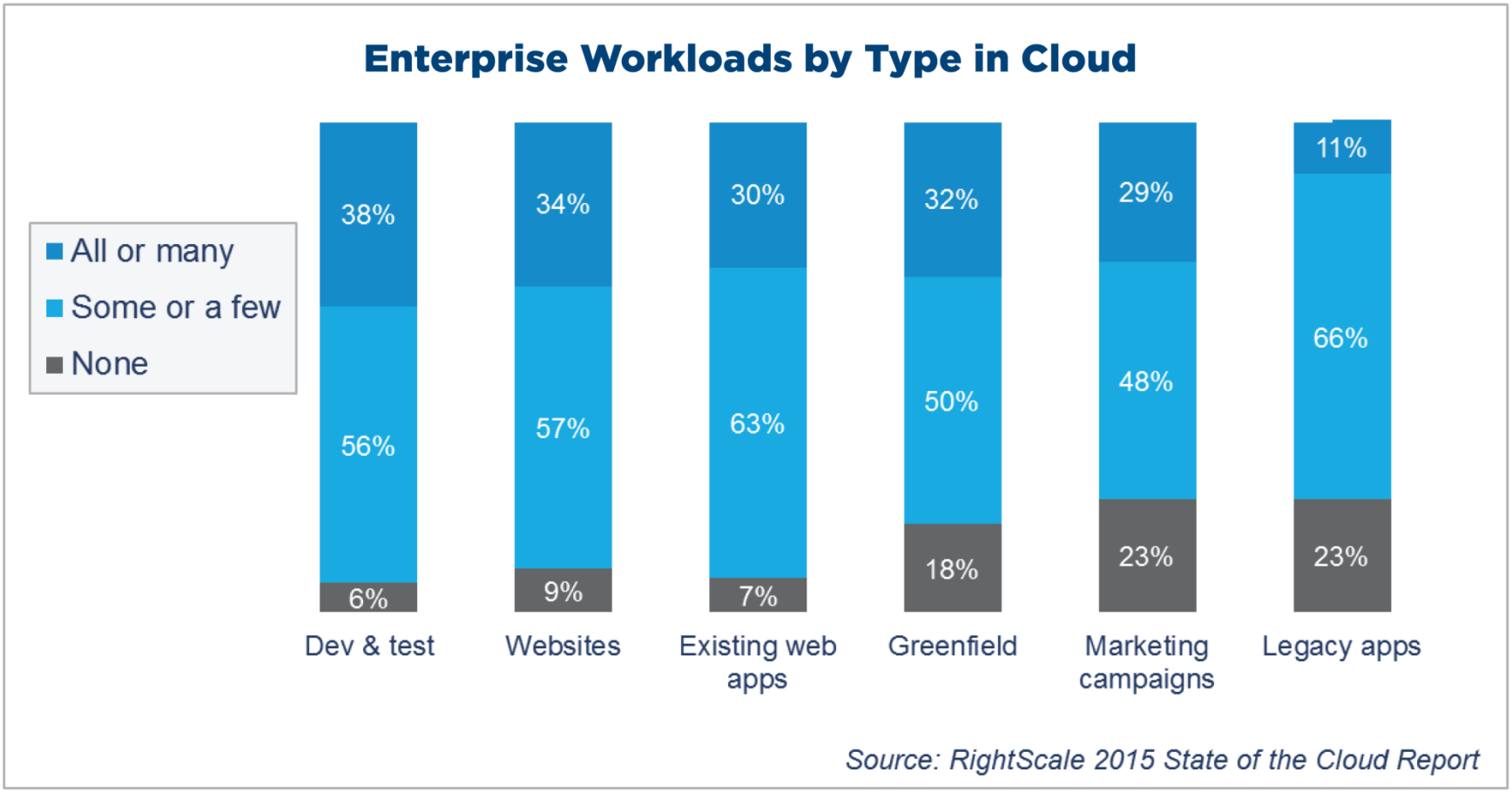


Etat du marché – usage du cloud

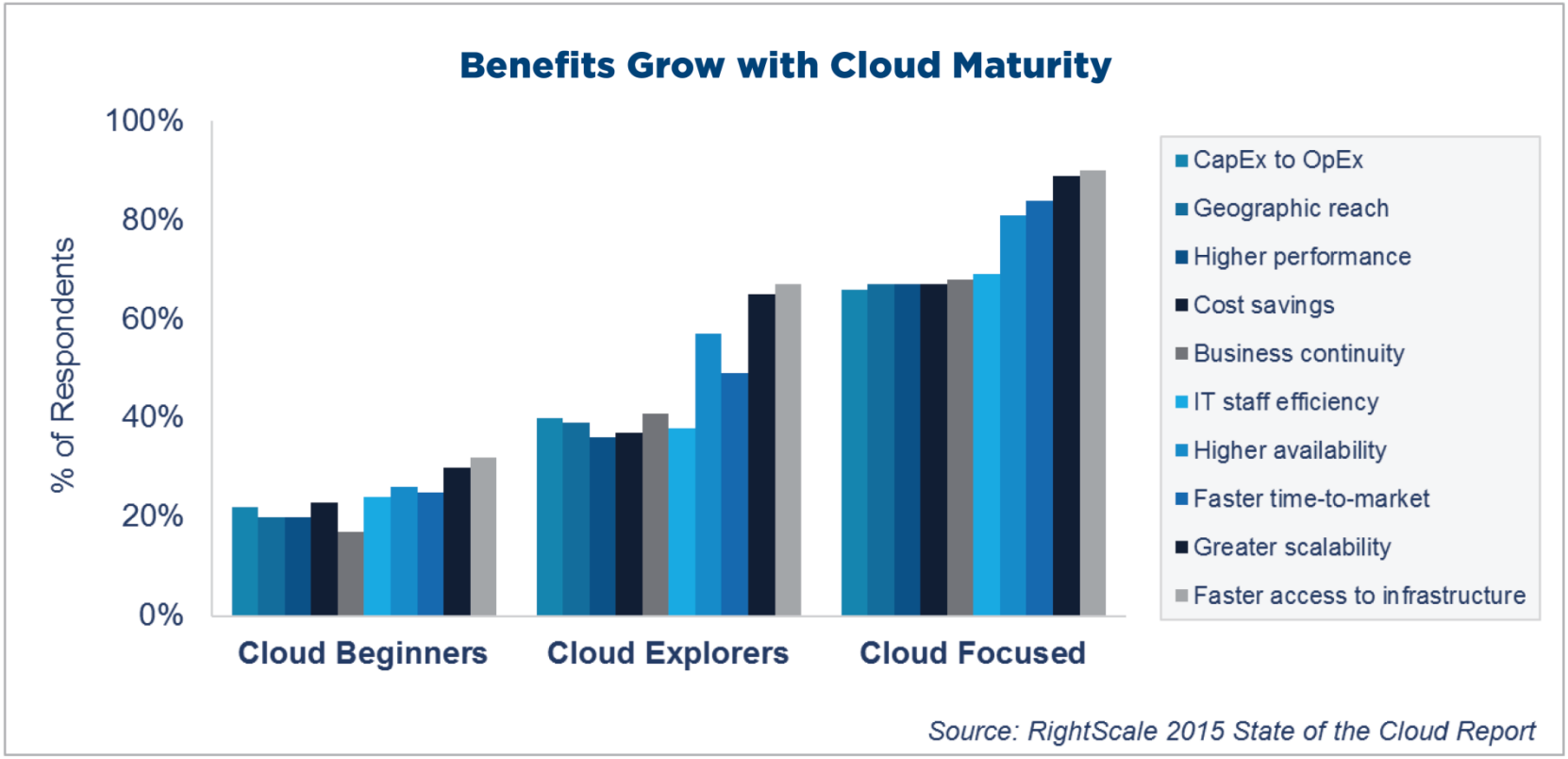




Etat du marché – usage du cloud 2



Etat du marché – avantages liés au cloud



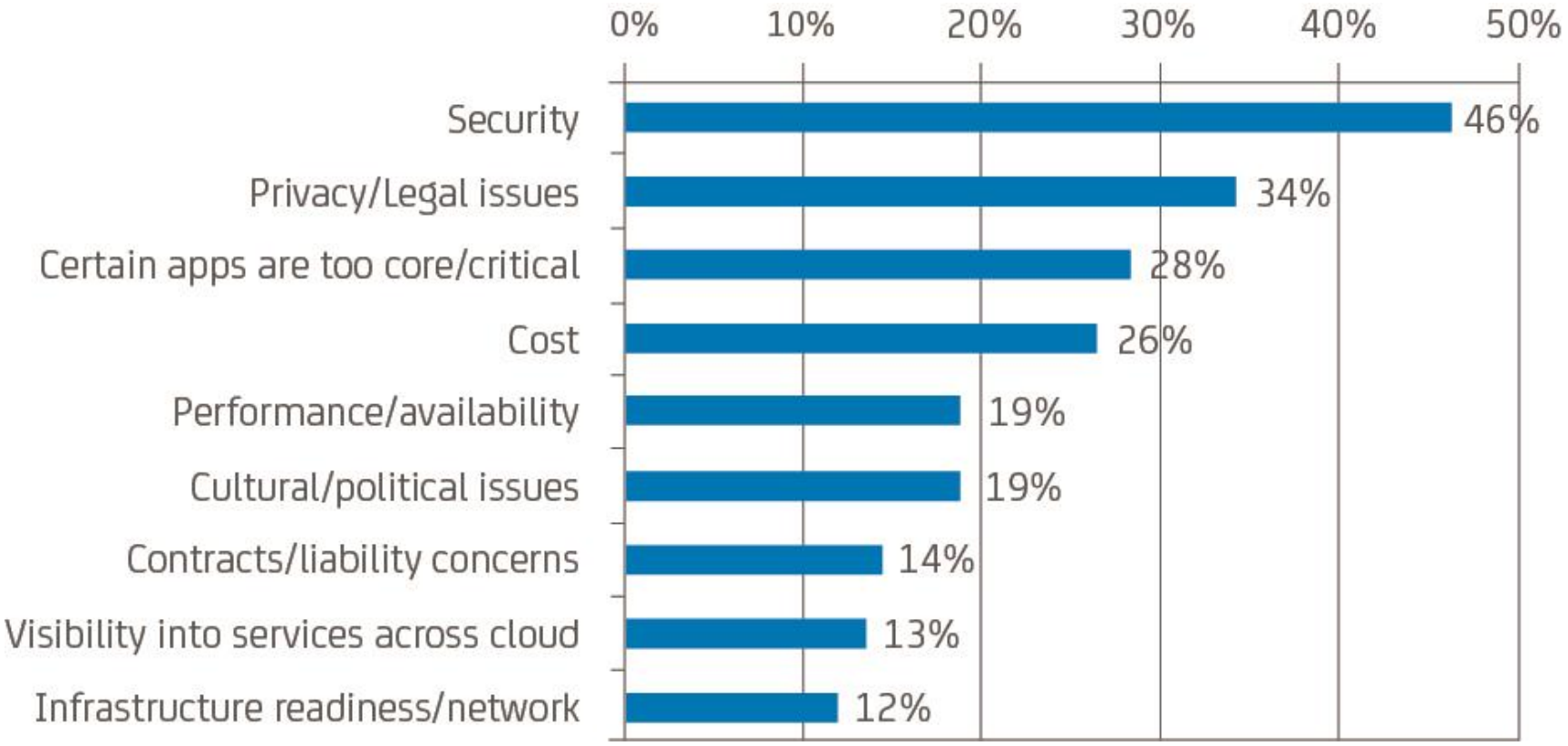
Etat du marché – inquiétudes liées au cloud

Top 5 Challenges Change with Cloud Maturity

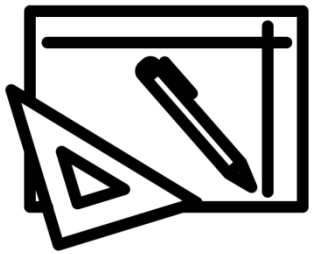
Place	Cloud Beginners	Cloud Explorers	Cloud Focused
#1	Lack of resources/expertise (35%)	**Complexity of building a private cloud (32%)**	Security (19%)
#2	Security (32%)	Security (30%)	Compliance (18%)
#3	Compliance (28%)	**Managing multiple cloud services (30%)**	Managing costs (18%)
#4	Governance/control (28%)	Lack of resources/expertise (26%)	Managing multiple cloud services (17%)
#5	Managing costs (27%)	Compliance/governance/control (24%)	Governance/control (17%)

Source: RightScale 2015 State of the Cloud Report

Etat du marché – inquiétudes liées au cloud



http://cloud-computing.tmcnet.com/features/articles/351722-cloud-adoption-exceeding-expectations-new-global-study-finds.htm

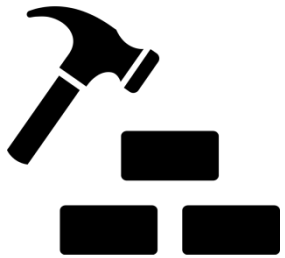


Architecture type et contraintes

Des architectures différentes

- Pour profiter de cette écosystème une application doit avoir été conçue pour ca.

- Toutes les applications ne sont pas capables de passer de 10 à 100 000 utilisateurs sans modifications.

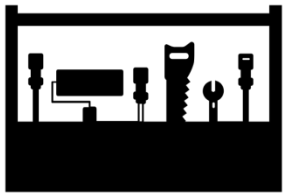


Contraintes à respecter pour un PAAS

- Interdiction d'écrire directement dans un fichier
- Impossible de créer des threads
  - Généralement utilisés pour répartir la charge
- Temps d'exécution limité (e.g 30s)
- Base de données relationnelles à héviter
- Communications réseaux limité

Techniques utilisées

- Evolutivité Horizontale (+ de machine)
- Opérations désynchronisées
- Base de données non-relationnel + MapReduce
- Utilisation d'API
- Bonnes pratiques d'architectures

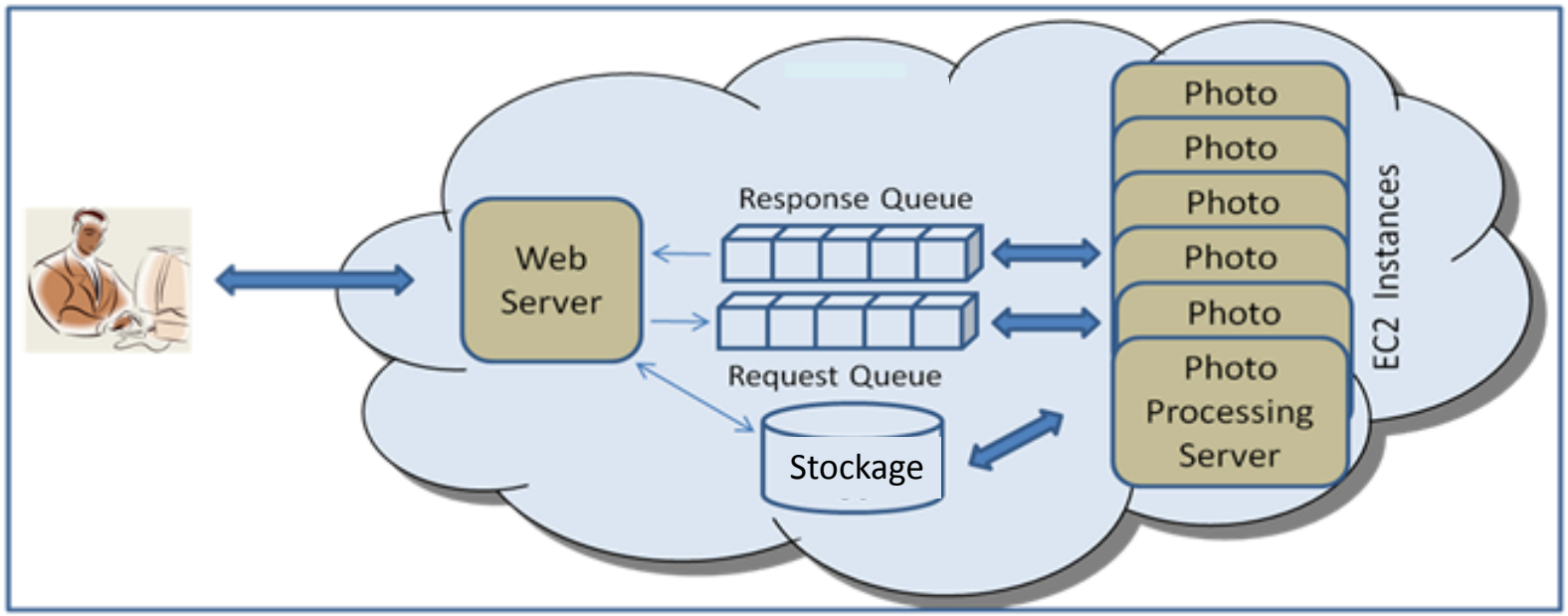


Opérations désynchronisées

- Une chaine de liaison principale qui traite et affiche les pages.
- Des serveurs dédiés pour les taches :
  - D'analyse, de traitement lourd
  - De compilations de données
  - De maintenance
  - De nettoyage des résultats
- Un système de communication entre les serveurs pour synchroniser l'ensemble

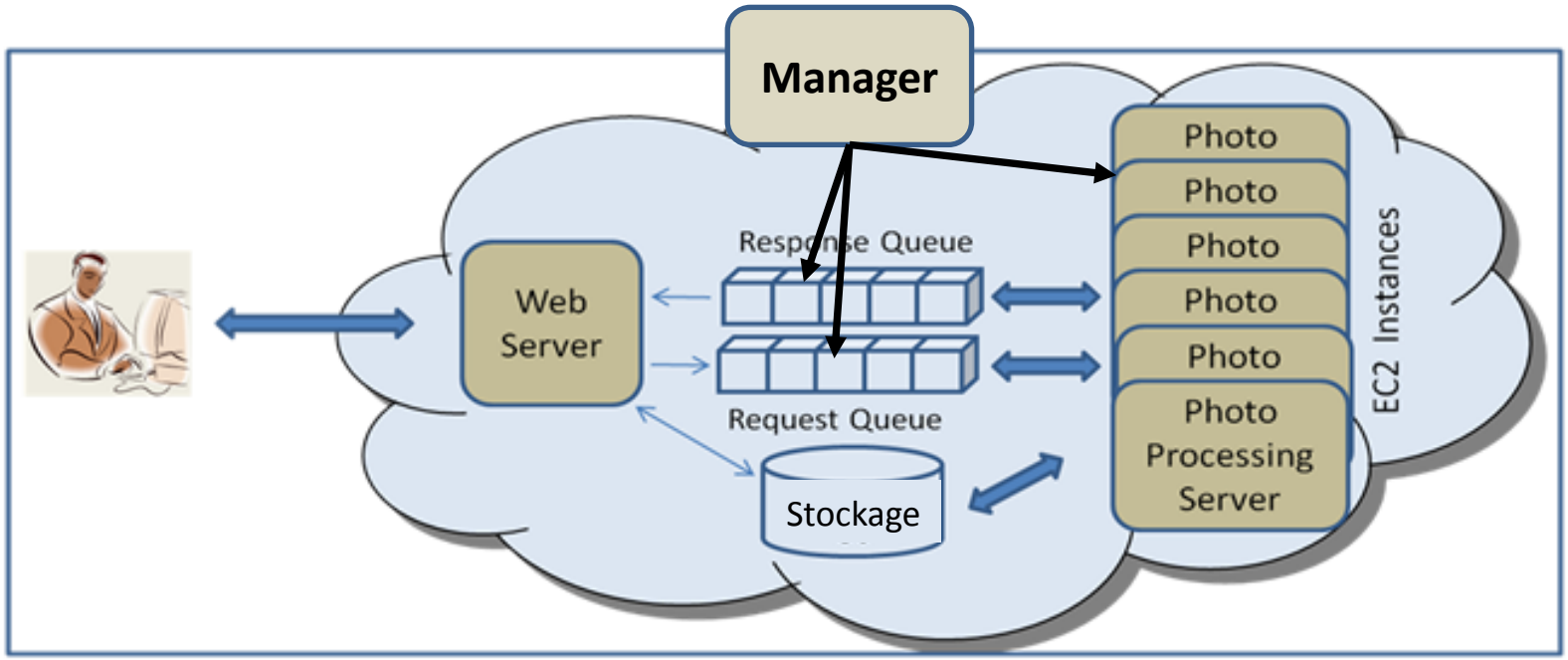
Opérations désynchronisées

- Utilisation de files de messages pour communiquer entre les serveurs
- Le nombre de « Photo processing server » varie en fonction de la taille de la file



Opérations désynchronisées

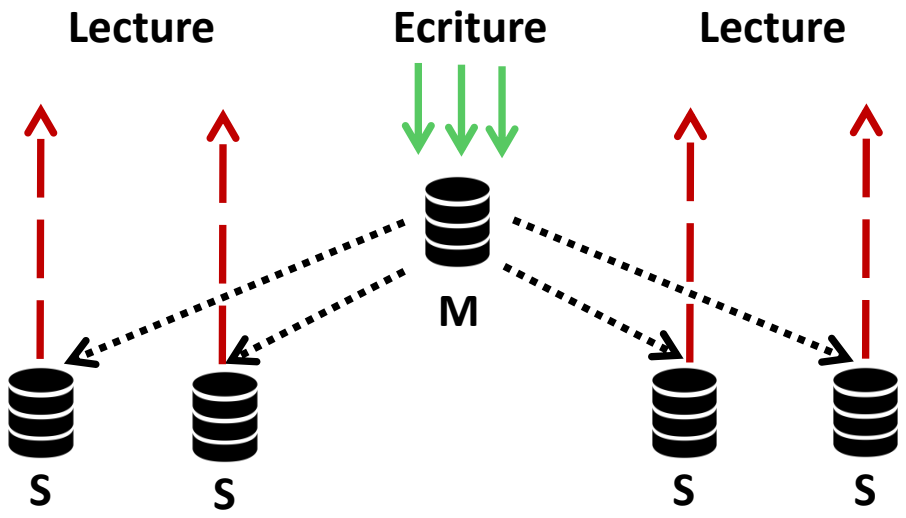
- Un manager surveille les opérations, il démarre ou arrête des « photo processing server » en fonction de la taille de la file
- Il surveille également qu'il n'y a pas de blocage sur un traitement



Problèmes liés aux BD relationnelles

- Pour les grosses applications, la base de données est souvent le point qui limite le plus l'évolutivité de l'application

  - Il ne peut y avoir qu'un **seul point d'écriture** (Master)
  - Il peut y **avoir plusieurs points de lecture** (Slave)



Problèmes liés aux BD relationnelles

- Cette difficulté d'évolution vient de la propriété ACID que l'on souhaite appliquer aux données.

  - **A**tomicité : une transaction doit soit être complètement validée ou complètement annulée
  - **C**onsistance : une transaction ne peut pas laisser la base de données dans un état incohérent
  - **I**solation : une transaction ne peut voir aucune autre transaction en cours d'exécution
  - **D**urabilité : après que le client a été informé du succès de la transaction, les résultats de celle-ci ne disparaîtront pas.

Différentes approches BDD

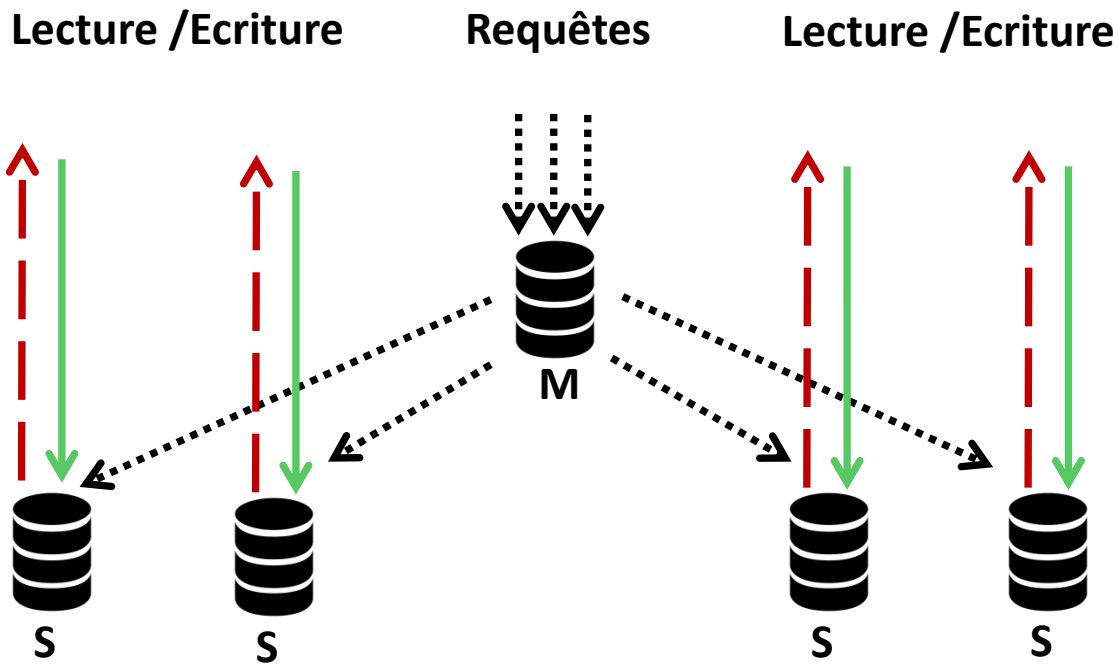
Epaisse	Relationnelle	Non relationnelle
▪ Stocke les données ▪ Valide l'intégrité des données ▪ Traite les données	▪ Stocke les données ▪ Valide l'intégrité des données	▪ Stocke les données

❑ L'application prend en charge les autres taches

Base de données non-relationnelle

- Pas de méthode ACID
- Pas de schéma
- Pas de langage de requêtage évolué comme SQL
- Lecture et écriture sur tous les nœuds
- Un master pour répartir les requêtes

Base de données non-relationnelle



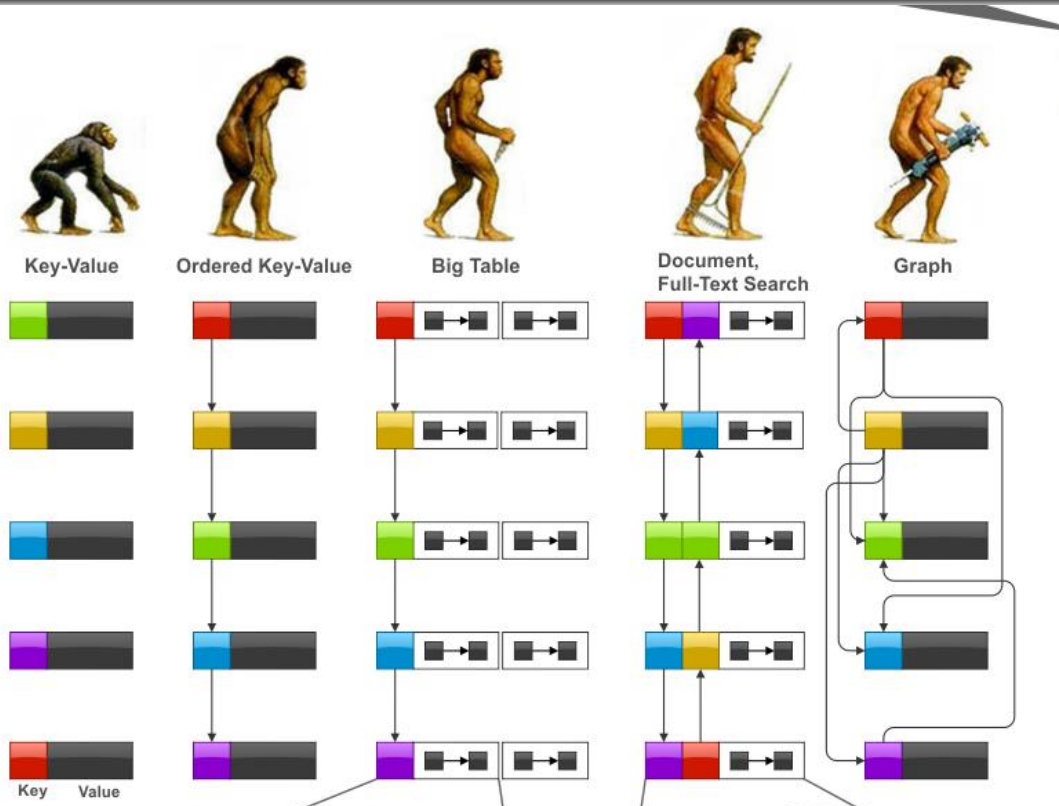
NoSql Base

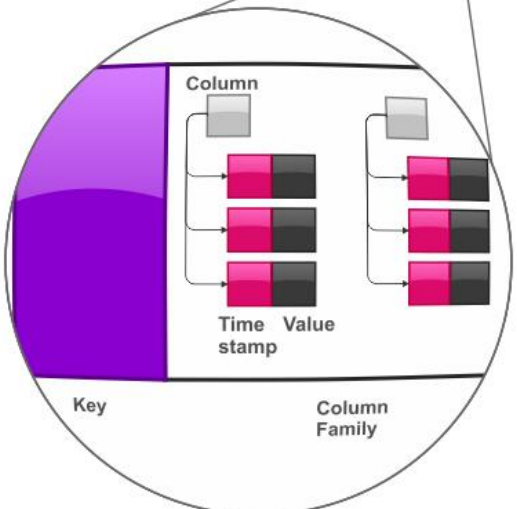
- **N**ot **O**nly **SQL**
  - Non relational
  - Distributed
  - Schema-less
  - Horizontally scalable

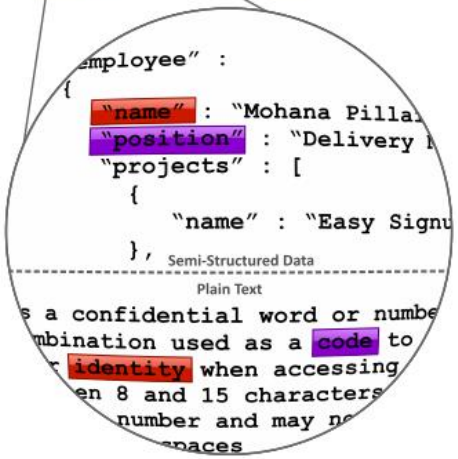
- NoSql databases families (http://www.nosql-database.org/)
  - **Key-value Storage**: Membase, Redis, Riak
  - **Wide-column Storage**: Cassandra, Hadoop
  - **Document Storage**: MongoDB, CouchDB
  - **Graph Storage**: Neo4j, InfoGrid, Bigdata
  - **Eventually Consistent Key-Value Storage**: Amazon Dynamo, Voldemort

https://www.credera.com/blog/technology-insights/java/a-managers-guide-to-nosql/

NoSql







https://highlyscalable.wordpress.com/2012/03/01/nosql-data-modeling-techniques/

Google BigTable

- L'une des première base de données non-relationnelle
  - Utilisé par google pour la majorité de ses services
  - A inspiré de nombreux autres projets :
    - Cassandra (facebook),
    - Dynamo (Amazon)
- Utilise un système de fichier distribué (Google File System)
- Stocke les données en colonnes plutôt qu'en lignes
- Le traitement complexe des données est assuré par MapReduce

Map Reduce

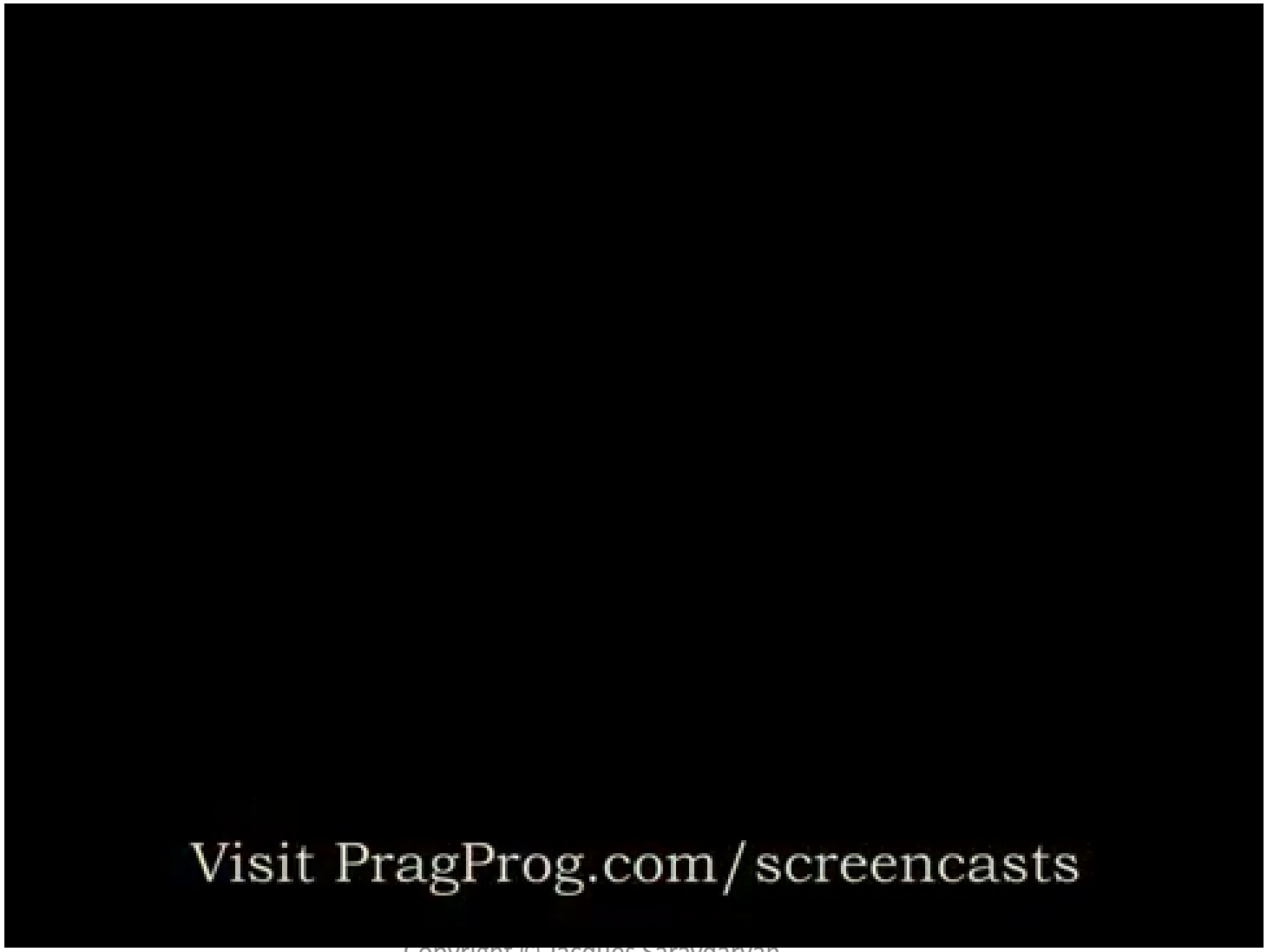
- Traitement distribué de gros volumes de données par un cluster de machine. Exemple :
  - Compter le nom d'occurrence d'un mot dans plusieurs documents.
  - Analyse de logs
  - Etc …

- **Initialement développé par Google, ré-implementé dans la communauté Open Source (hadoop Map Reduce)**

Map Reduce

- MapReduce gère nativement le traitement en parallèle sur un groupe de machine.
- L'utilisateur peut se concentrer sur le traitement des données
- Le traitement s'effectue en deux temps
  - **Map** : traite les données.
    → Exemple : compte le nombre de mot par document
  - **Reduce** : regroupe les données traitées
    → Exemple : effectue la somme pour tous les documents
- Il faut parfois plusieurs passage pour traiter complètement les données

Map Reduce

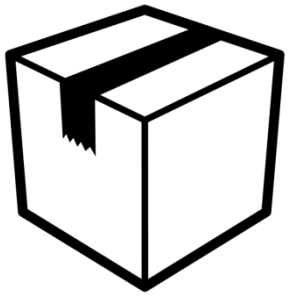


Map Reduce

Find Max Temperature of each city

Best practice

- Optimisation du code
- Mécanisme d'optimisation des requêtes
  - **Compression**
  - **Cache**
- Externalisation des taches sans valeurs ajoutées (SSL)
- Mise en place d'un Content Delivery Network (CDN)



– Exemple – Amazon Web Service

Les Services

- Calcul
  - **AMAZON EC2**
  - AWS Elastic Beanstalk
- Stockage et diffusion de contenu
  - **Amazon S3**
  - **Amazon CloudFront**
- Base de données
  - **Amazon RDS**
  - Amazon DynamoDB
- Analyse
  - **Amazon Elastic Map Reduce**
  - Amazon Machine Learning
- Autres:
  - Amazon Workspace
  - AWS IoT

CPE LYON
ÉCOLE SUPÉRIEURE DE CHIMIE PHYSIQUE ÉLECTRONIQUE DE LYON

Les Services

- Tous les services AWS sont accessibles au travers d'une API permettant d'effectuer l'ensemble des opérations
- Tous les services sont facturés en fonction de l'utilisation réelle qu'on en fait.
- AWS proposent ses services dans différentes régions (US et EU), la localisation est importante lorsqu'on souhaite diminuer le temps de réponse d'une application.

AWS S3

- Service de stockage en ligne, de type NAS (fichier)
- Sans limite de taille (1Mo -> Plusieurs Peta ..)
- Les données sont automatiquement répliquées
- Les fichiers sont accédés en HTTP (GET, DELETE, PUT etc …)
- Actuellement, il y a plus de 64 milliards de fichiers sur S3

- Le payement se calcul en fonction
  - du nombre de Go stocké
  - du nombre de requête
  - du volume de données transférées (IN et OUT)

AWS S3

- Service de stockage en ligne, de type NAS (fichier)
- Sans limite de taille (1Mo -> Plusieurs Peta ..)
- Les données sont automatiquement répliquées
- Les fichiers sont accédés en HTTP (GET, DELETE, PUT etc …)
- Actuellement, il y a plus de 64 milliards de fichiers sur S3

- Le payement se calcul en fonction
  - du nombre de Go stocké
  - du nombre de requête
  - du volume de données transférées (IN et OUT)

AWS S3

- Les fichiers sont organisés dans des dossiers appelés : BUCKET
  - Un bucket est unique sur l'ensemble de la plateforme

- Chaque fichier a une adresse HTTP distincte permettant d'y accéder facilement.
  - http://s3.amazonaws.com/bucket/file

- L'accès à chaque fichier peut être contrôlé par un système d'autorisation en rajoutant une clef dans l'URL

AMAZON EC2

- Terminologies :
  - Une machine virtuelle est appelée une **INSTANCE**
  - Un image d'une machine virtuelle est appelée une **AMI**
  - Un disque de stockage en mode « block » est appelé **EBS**

- Une **INSTANCE** se lance à partir d'une **AMI** existante
- On peut associer un **EBS** avec une **INSTANCE**
- Une **INSTANCE** peut être sauvegardée en **AMI**

AMAZON EC2

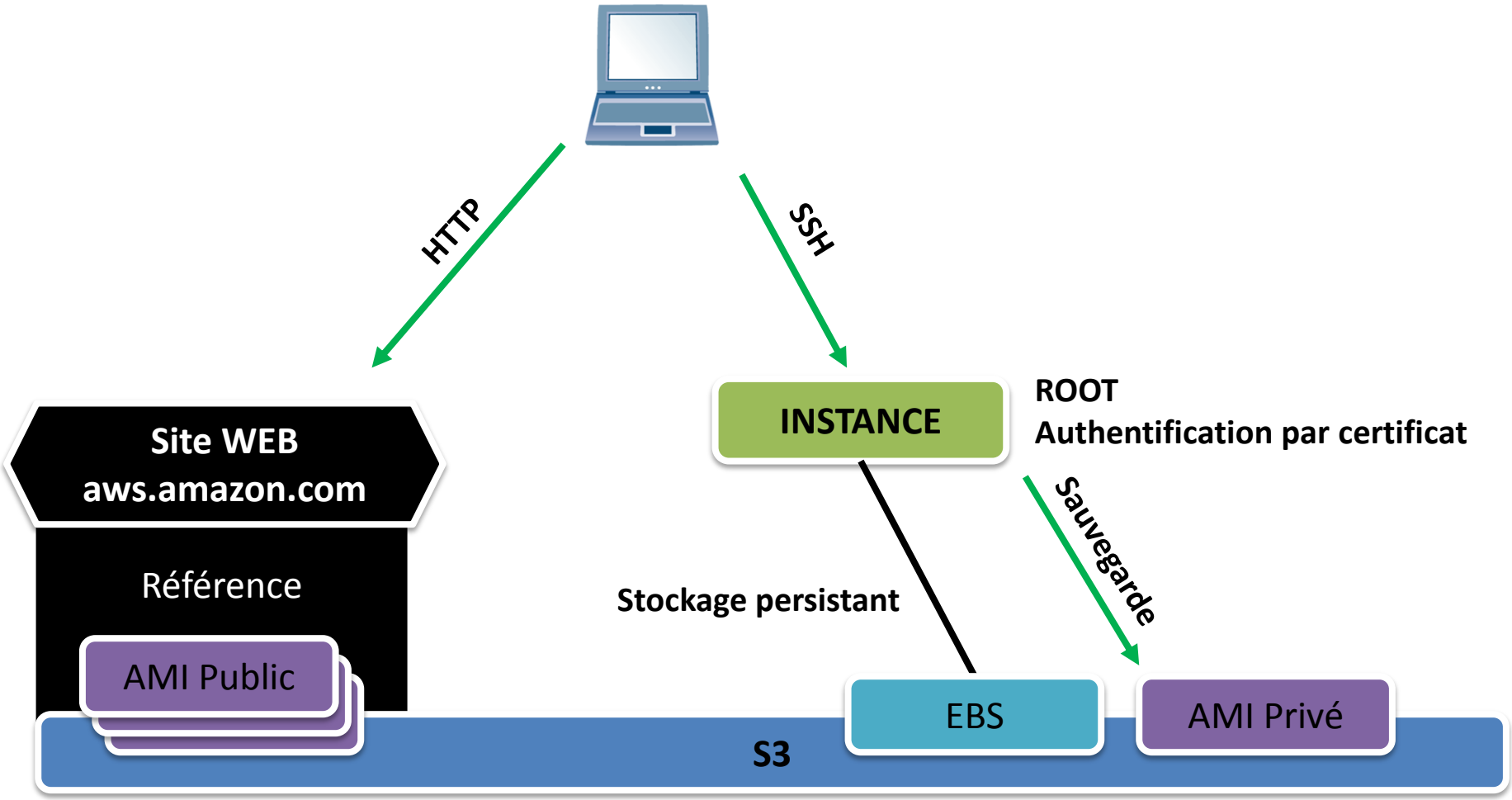
TYPE	EC2 CPU	RAM	Disque	Prix / heure	Prix / mois
Small	1	1.7 G	160 G	$ 0.095	$ 68.4
Medium	5	1.7 G	350 G	$ 0.19	$ 136.8
Large	4	7.5 G	850 G	$ 0.38	$ 273.6
Extra large	8	15 G	1690 G	$ 0.76	$ 547.2
Double Extra Large	13	34.2 G	850 G	$ 1.34	$ 964.8
Quadruple Extra Large	26	68.4 G	1690 G	$2.68	$ 1929.6

AMAZON EC2

- Pour utiliser EC2, il faut
  - Créer un compte + numéro CB
  - Générer un couple de clefs de sécurité (SECRET_KEY et PUBLIC_KEY)
  - Générer un couple de certificats x509

- Les clefs de sécurité sont utilisées dans les API pour authentifier l'utilisateur qui fait les requêtes
- Le certificat est utilisé pour se connecter sur les INSTANCES

AMAZON EC2



AMAZON EMR

- Cluster Map Reduce prêt à l'utilisation
- Utilise des AMI préconfigurées pour cette tache.

- Pour l'utiliser il faut :
  - Développer les fonctions Map et Reduce
  - Rendre disponibles les données INPUT sur S3
  - Choisir le nombre de machines que l'on souhaite avoir dans le cluster
  - Les données traitées sont stockées sur S3

- Comme pour les autres, accessible par une API

AMAZON EMR

https://aws.amazon.com/fr/elasticmapreduce/

AWS: Cloud Front

- Service de Content Delivery Network basé sur S3
- Permet de décharger les serveurs
  - Fichiers statiques
  - Vidéos
  - Binaires
- Payement aux nombres de téléchargement par fichier

AWS: Cloud Front

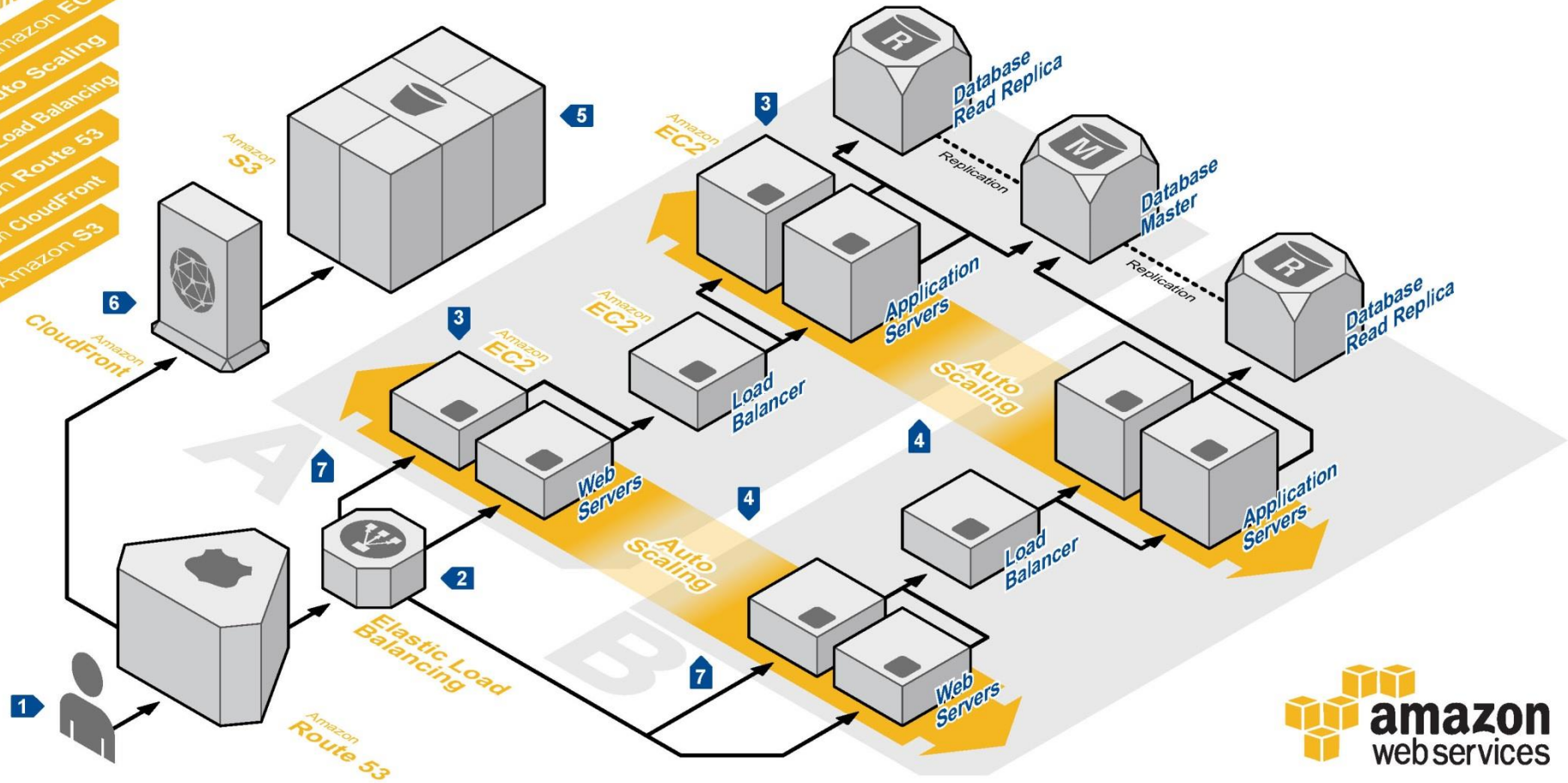
- United States
  - Ashburn, VA
  - Dallas/Fort Worth, TX
  - Los Angeles, CA
  - Miami, FL
  - Newark, NJ
  - Palo Alto, CA
  - Seattle, WA
  - St. Louis, MO
- Europe
  - Amsterdam
  - Dublin
  - Frankfurt
  - London
- Asia
  - Hong Kong
  - Tokyo

AWS Reference Architectures

- Amazon EC2
- Auto Scaling
- Elastic Load Balancing
- Amazon Route 53
- Amazon CloudFront
- Amazon S3

WEB APPLICATION HOSTING

Highly available and scalable web hosting can be complex and expensive. Dense peak periods and wild swings in traffic patterns result in low utilization rates of expensive hardware. Amazon Web Services provides the reliable, scalable, secure, and high-performance infrastructure required for web applications while enabling an elastic, scale out and scale down infrastructure to match IT costs in real time as customer traffic fluctuates.



System Overview

1 The user's DNS requests are served by **Amazon Route 53**, a highly available Domain Name System (DNS) service. Network traffic is routed to infrastructure running in Amazon Web Services.

2 HTTP requests are first handled by Elastic Load Balancing, which automatically distributes incoming application traffic across multiple **Amazon Elastic Compute Cloud (EC2)** instances across Availability Zones (AZs). It enables even greater fault tolerance in your applications, seamlessly providing the amount of load balancing capacity needed in response to incoming application traffic.

3 Web servers and application servers are deployed on **Amazon EC2** instances. Most organizations will select an **Amazon Machine Image (AMI)** and then customize it to their needs. This custom AMI will then be used as the starting point for future web development.

4 Web servers and application servers are deployed in an **Auto Scaling** group. Auto Scaling automatically adjusts your capacity up or down according to conditions you define. With Auto Scaling, you can ensure that the number of **Amazon EC2** instances you're using increases seamlessly during demand spikes to maintain performance and decreases automatically during demand lulls to minimize costs.

5 Resources and static content used by the web application are stored on **Amazon Simple Storage Service (S3)**, a highly durable storage infrastructure designed for mission-critical and primary data storage.

6 Static and streaming content is delivered by **Amazon CloudFront**, a global network of edge locations. Requests are automatically routed to the nearest edge location, so content is delivered with the best possible performance.

7 **Availability zones (AZs)** are distinct geographic locations that are engineered to insulate against failures in other AZs. Multiple AZs are combined into a region. Here, the entire web application is deployed in two different AZs for high availability.

AMAZON Machine Learning

https://aws.amazon.com/fr/machine-learning/

References

References

- Haute disponibilité et datacentre, *Blaise DAVID, Consultant Virtualisation, Quadix Technologies*

- https://aws.amazon.com/fr/
- https://cloud.google.com/appengine/docs
- https://cloud.google.com/solutions/
- RigthScale 2015, State of the Cloud Report

CPE LYON
ÉCOLE SUPÉRIEURE
DE CHIMIE PHYSIQUE ÉLECTRONIQUE
DE LYON

Jacques Saraydaryan

Jacques.saraydaryan@cpe.fr