

VILLES INTELLIGENTES : DÉFIS TECHNOLOGIQUES ET SOCIÉTAUX

1. Introduction
2. Infrastructures réseaux des systèmes urbains
3. Infrastructures logicielles des systèmes urbains
- 4. Gestion des données urbaines dans les nuages informatiques**
5. Gestion des données et vie privée du citoyen urbain

Christine Morin



Les données sont au cœur de la réalisation des « villes intelligentes ». Quand on parle de données, dans ce contexte, il peut s'agir aussi bien des données ouvertes mises à disposition par les collectivités que des données générées par les citoyens ou collectées depuis différents capteurs fixes ou mobiles dans l'espace urbain.

De multiples applications traitent ces données pour générer de l'information utile aux citoyens, par exemple pour leur faciliter la vie quotidienne, mais aussi utile aux collectivités qui sont en charge de la gestion des services de la ville et en charge d'imaginer les infrastructures et services du futur.

Cette semaine, je vais vous présenter les technologies sous-jacentes à la gestion des données urbaines dans les nuages informatiques ou clouds.

Même s'ils sont invisibles pour les citoyens de la ville intelligente, les clouds sont omniprésents pour le stockage, le partage de données urbaines comme pour le support pour l'exécution des applications de traitement des données.

4. Gestion des données urbaines dans les nuages informatiques

- Brève histoire des nuages informatiques
- Modèles de service et de déploiement
- Technologie clé : la virtualisation
- IaaS : les points de vue utilisateur et fournisseur
- PaaS : programmation et déploiement des applications
- Stockage de données
- Traitement de données
- Traitement de flux de données

Christine Morin VILLES INTELLIGENTES : DÉFIS TECHNOLOGIQUES ET SOCIÉTAUX

Cette semaine, le cours est structuré en 8 séquences et dresse un **panorama des technologies sous-jacentes à la gestion et au traitement des données urbaines dans les nuages informatiques** :

- Point de vue historique sur la genèse des nuages informatiques,
- Les modèles de services et de déploiements des applications dans le nuage,
- La virtualisation, une technologie qui est essentielle pour les nuages informatiques,
- Description de deux des modèles de services :
 - le modèle d'infrastructures (**IaaS**) en tant que service, tant du point de vue du fournisseur que des utilisateurs,
 - le modèle de plates-formes (**PaaS**) en tant que service qui permet aux programmeurs de programmer et de déployer simplement les applications dans les nuages informatiques,
- Les 3 dernières séquences traiteront respectivement du **stockage**, du **traitement des données** et des **traitements de flux de données** dans les clouds.

4. Gestion des données urbaines dans les nuages informatiques

- Brève histoire des nuages informatiques
- Modèles de service et de déploiement
- Technologie clé : la virtualisation
- IaaS : les points de vue utilisateur et fournisseur
- PaaS : programmation et déploiement des applications
- Stockage de données
- Traitement de données
- Traitement de flux de données

Christine Morin VILLES INTELLIGENTES : DÉFIS TECHNOLOGIQUES ET SOCIÉTAUX

Donc nous allons commencer par une brève histoire des nuages informatiques.

Un vieux rêve : le calcul comme un service public

“If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry. ”

John McCarthy, intervention au MIT en 1961



4

La vision du calcul offert en tant que service public n'est pas nouvelle. Elle a été exprimée par John McCarthy, lauréat du Prix Turing en 1971, lors de son intervention à la célébration du centenaire du MIT en 1961 : *“If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry.”*

Il imagine déjà offrir un service public de clouds, analogue au service de téléphonie. Donc un service dont il prophétise qu'il serait source d'une nouvelle et importante industrie.

L'histoire va lui donner raison une quarantaine d'années plus tard. Mais avant d'en arriver au cloud, il y a eu des précurseurs.

Le mainframe : un précurseur (dès la fin des années 50)



5

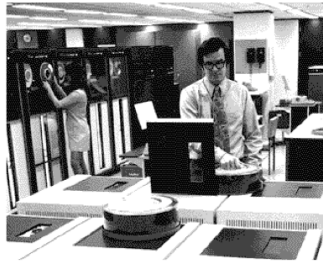
À partir de la fin des années 50, les moyens de calcul sont concentrés dans ce qu'on appelle des mainframes, comme l'IBM 704 auxquels les utilisateurs accèdent à partir de terminaux.

Ces ordinateurs coûteux sont utilisés dans des secteurs ayant besoin de puissance de calcul comme les domaines militaires et scientifiques, par exemple.

Ce type d'ordinateurs, qui possède un processeur, puissant est utilisé en temps partagé par une communauté d'utilisateurs au sein de l'organisation propriétaire.

Les techniques de virtualisation voient le jour à cette époque pour permettre le partage de l'unique processeur de la machine entre différents utilisateurs. Plus tard le système Multics permettra aussi la gestion des mainframes en temps partagé.

Exemple : le centre de calcul au service de la communauté scientifique



6

Dans les années 70, avec le développement du réseau Arpanet, l'ancêtre du réseau Internet, les scientifiques, dans les universités américaines, vont pouvoir accéder à distance à des supercalculateurs. Il s'agit en quelque sorte d'une mise en œuvre pionnière de l'idée de la puissance de calcul fournie comme un service.

Les systèmes en mode batch permettent d'ordonnancer l'exécution des applications soumises par les utilisateurs sur ces machines en fonction de priorités. Donc les utilisateurs voient leurs applications mises dans des files d'attente qui peuvent être très longues.

Des mainframes aux grappes de calculateurs



Dans les années 80, les mainframes monoprocesseurs vont être remplacés par des calculateurs parallèles comportant plusieurs processeurs au sein d'une même machine.

Ces mêmes multiprocesseurs vont être supplantés, à partir du milieu des années 90, par les grappes de calculateurs qui vont progressivement les remplacer pour fournir la puissance de calcul.

Donc d'une architecture centralisée, on va ainsi passer à une architecture distribuée, constituée de serveurs standards interconnectés par des réseaux haut débit faible latence, également standard.

L'avantage de ce nouveau type d'architecture matériel est son extensibilité, la possibilité de rajouter des machines, et sa haute disponibilité puisqu'il y a beaucoup de redondances. Et elles sont très faciles à faire évoluer. Comparativement aux mainframes, elles sont aussi très avantageuses sur le plan financier.

La grille de calcul : précurseur dans la communauté scientifique (à partir de la fin des années 90)



8

Avec les grilles informatiques qui se développent à partir de la fin des années 90, surtout dans la communauté scientifique, on n'est plus très loin du service public de calcul.

Le terme grille provient de l'analogie avec le réseau électrique qu'on appelle grid aux États-Unis.

Donc lorsque vous branchez votre cafetière électrique sur le secteur? Vous obtenez la puissance électrique nécessaire à son fonctionnement sans savoir où et par quel type de centrale électrique l'électricité a été produite. Et vous ne connaissez rien des caractéristiques du réseau de transport.

La grille de calcul : précurseur dans la communauté scientifique (à partir de la fin des années 90)



9

L'idée est analogue avec les grilles informatiques. La puissance de calcul et de stockage est fournie par des centres de calcul et de données qui sont de différents types et situés dans différents endroits. Et l'utilisateur accède à ces ressources à travers le réseau Internet.

Dans cette vision, les terminaux d'accès ont été remplacés par les ordinateurs personnels et les ressources informatiques, auxquels un utilisateur peut accéder, peuvent être situées dans un ou plusieurs centres de calcul qui ont conclu des accords de mutualisation des ressources.

Les utilisateurs et les ressources sont donc gérés au sein d'organisations virtuelles qui définissent les conditions d'utilisation des ressources au sein des différents projets.

Même si de nombreux travaux de recherche se sont penchés sur les modèles économiques pour les grilles, les grilles n'ont pas du tout percé sur le marché et sont restées l'apanage des scientifiques. Elles souffrent encore des inconvénients des machines gérées dans les centres de calcul, à savoir les files d'attente pour l'accès aux ressources et puis les applications doivent s'adapter à l'environnement matériel et logiciel disponible, ce qui pose bien sûr des problèmes de portabilité des applications.

Emergence des nuages informatiques - *cloud* (2000)

- Google pionnier avec le service *Gmail* pour le *cloud* grand public (2004)
- Amazon pionnier avec le service *EC2* pour le *cloud* d'entreprise (2006)



10

C'est finalement dans le monde commercial que sont nés les nuages informatiques. Le terme nuage informatique ou cloud est apparu au milieu des années 2000, très probablement en 2006 dans la bouche d'Éric Schmidt lors d'une conférence.

Ce terme a immédiatement été repris par Amazon dans le nom de son service EC2, pour Elastic Compute Cloud.

On peut dire, en quelque sorte, que le nuage informatique a démocratisé la puissance de calcul et le stockage en permettant à tout un chacun d'utiliser les ressources situées dans les grands centres de données accessibles à travers Internet, avec un modèle de tarification, selon la consommation.

Avec les nuages informatiques, vous ne payez que pour ce que vous utilisez.

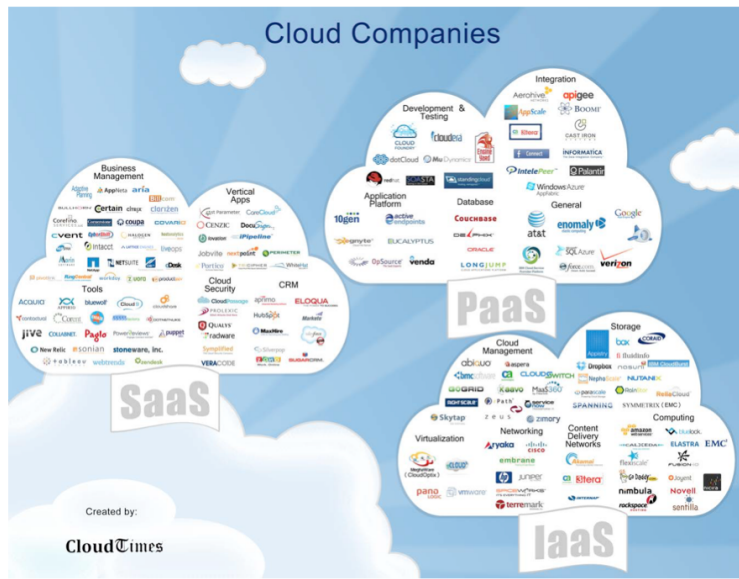
Donc l'avènement du cloud computing a été rendu possible grâce à un regain d'intérêt pour les techniques de virtualisation qui datent de l'époque des mainframes, au moment où VMware a développé la virtualisation sur les processeurs X86 utilisés dans les ordinateurs personnels.

La virtualisation a donc permis de s'affranchir des contraintes logiciel et architecturale, que l'on rencontrait précédemment notamment sur les grilles, en permettant d'exécuter simultanément plusieurs systèmes d'exploitations sur un même serveur et donc d'adapter l'environnement logiciel aux besoins de chaque utilisateur.

Alors, comme nous le verrons dans la séquence suivante, il y a plusieurs modèles de services dans le cloud :

- Google a été pionnier dans le domaine du cloud grand public avec la fourniture dès 2004 d'un service de courrier électronique gratuit utilisable par tous.
- L'entreprise Amazon, quant à elle, qui est connue donc pour son site de commerce électronique, a démocratisé le cloud d'entreprise avec un service pionnier de calcul à la demande, le service EC2 créé en 2006. Pour l'anecdote, il s'agissait pour Amazon, qui dispose de grands centres de données afin de pouvoir traiter les pics de commandes par exemple à Noël, de réduire ses coûts en louant les ressources inutilisées en dehors des périodes de pointe. En effet, les serveurs sont très gourmands en énergie lorsqu'ils sont allumés même s'ils ne font rien. Donc, c'était intéressant pour Amazon de les louer.

Marché foisonnant des nuages informatiques



11

Depuis lors, le marché du cloud computing est foisonnant avec de très nombreux acteurs, petits et grands, qui proposent différents types de services à une clientèle très variée, qui va du commun des mortels jusqu'aux entreprises et autres institutions en passant par la communauté scientifique.

Donc les différents acteurs des villes intelligentes sont des utilisateurs actifs du cloud pour stocker et traiter des volumes croissants de données urbaines mises à disposition par les collectivités ou générées continuellement par les citoyens et les capteurs présents dans les espaces urbains.

Illustrations & photos : crédits

p. 4 : "John McCarthy Stanford" by "null0". CC BY-SA 2.0 via Wikimedia Commons

p. 5 : Image courtesy of Computer History Museum,
<http://www.computerhistory.org/collections/catalog/102641437>

p. 6 : droits réservés, <http://www.multicians.org/6180pix.html>

p. 7 : © IRISA, http://www.irisa.fr/myriads/software/grid5000/index_html

p. 8-9 : droits réservés, http://www-igm.univ-mlv.fr/~dr/XPOSE2006/Jolly_Laskri/historique.html

p. 10 : © Google, © Amazon Web Services

p. 11 : droits réservés, Cloud Times,
<http://cloudtimes.org/wp-content/uploads/2011/11/Clouds.cloudtimes.png>