



UNIVERSITÉ
DE LORRAINE



nancy Charlemagne
Département Informatique

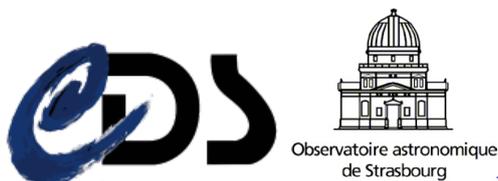
IUT Nancy Charlemagne
Université de Lorraine
2 ter Boulevard Charlemagne
54052 Nancy Cedex
Département Informatique

Étude de l'utilisation du Cloud dans le cadre d'applications et de services astronomiques

Rapport de stage DUT Informatique

Centre de Données astronomiques de Strasbourg

Observatoire Astronomique de Strasbourg



Étudiant : Nicolas VIARD
Année universitaire 2012-2013

IUT Nancy Charlemagne
Université de Lorraine
2 ter Boulevard Charlemagne
54052 Nancy Cedex
Département Informatique

Étude de l'utilisation du Cloud dans le cadre d'applications et de services astronomiques

Rapport de stage DUT Informatique

Centre de Données astronomiques de Strasbourg

Observatoire Astronomique de Strasbourg

11, rue de l'Université

67000 Strasbourg

Étudiant : Nicolas VIARD

Tuteur en entreprise : André SCHAAFF

Parrain de stage : Bernard MANGEOL

Remerciements

Je tiens particulièrement à remercier André SCHAAFF, mon tuteur de stage au sein du CDS, pour avoir proposé ce stage, avoir accepté ma candidature et m'avoir suivi tout au long de ce stage, que ce soit pour définir de nouveaux objectifs ou pour m'aider à réaliser mon rapport de stage et ma soutenance.

Je remercie vivement Bernard MANGEOL, mon tuteur de stage au sein de l'IUT, pour être venu me voir à l'Observatoire et pour avoir répondu à mes questions.

Je remercie chaleureusement Caroline BOT et Cécile LOUP, deux astronomes du CDS dont j'ai partagé le bureau pendant toute la durée de mon stage.

Je tiens également à remercier Thomas BOCH, Gilles LANDAIS, Laurent MICHEL, Anaïs OBERTO et Patricia VANNIER pour avoir présenté les différents services de l'Observatoire à l'arrivée des stagiaires, ce qui m'a permis de bien comprendre le contexte dans lequel mon stage s'inscrivait.

Pour finir, je remercie les autres stagiaires et les membres du personnel de l'Observatoire, qui ont tous contribué à la bonne ambiance dans laquelle j'ai effectué mon stage.

Table des matières

Introduction	4
1 Présentation du stage	6
1.1 L'Observatoire Astronomique de Strasbourg	6
1.1.1 Présentation générale	6
1.1.2 Équipes de recherche	7
1.1.3 Le Centre de Données Astronomiques de Strasbourg	8
1.2 Sujet de stage	11
1.2.1 Étude réalisée deux ans auparavant	11
1.2.2 Énoncé du sujet	11
1.2.3 Organisation du stage	12
2 État de l'art du Cloud	13
2.1 Le Cloud : une offre très variée	13
2.1.1 Différents types de services	13
2.1.2 Différents types de Clouds	15
2.2 Panorama des différents fournisseurs	15
2.2.1 Différents types de fournisseurs	15
2.2.2 Qualité de service	16
2.3 Cas d'utilisation	17
2.3.1 Environnement de test	17
2.3.2 Gestion des pics d'utilisation	18
2.3.3 Astronomie	19
2.4 Problèmes liés au Cloud	19
2.5 Des bibliothèques d'abstraction	20
3 Implémentation de serveurs dans le Cloud	21
3.1 Jclouds	21
3.2 VmWare Hybrid Cloud Evaluation	22
3.3 Dasein Cloud	24
3.4 Performances	25

3.4.1	Serveur de fichiers	26
3.4.2	Serveur de calcul	26
3.4.3	Temps de réponse des serveurs	27
3.5	Évaluation des coûts relatifs au Cloud	27
3.5.1	Serveur de calcul	27
3.5.2	Serveur de données	28
Conclusion		29
Bibliographie		31
Table des figures		31

Introduction

Un stage de 10 semaines s'inscrivait dans le cadre de mon DUT Informatique. Sachant que je souhaite continuer mes études une fois diplômé, je voulais profiter de cette occasion pour découvrir le fonctionnement d'un centre de recherches. Grâce au portail des stages hébergé sur Arche, j'ai pu trouver de nombreuses offres provenant de l'Observatoire Astronomique de Strasbourg.

J'ai été particulièrement intéressé par le stage traitant du Cloud Computing / Storage, un sujet d'actualité dans le domaine de l'informatique. En effet, le Cloud est fréquemment mentionné dans les conférences, les magazines et les sites communautaires destinés aux développeurs.

Afin de présenter le travail que j'ai réalisé, j'organiserai mon propos en trois parties. Premièrement, je présenterai mon stage, en montrant ce qui a amené l'Observatoire Astronomique de Strasbourg à proposer un sujet sur le Cloud Computing / Storage. Cette étude du contexte me permettra d'expliquer comment j'ai organisé mon travail, afin de mener à bien ce projet. Ensuite, je ferai un état de l'art du Cloud, dans le but de présenter les possibilités que ce concept nous apporte actuellement. Enfin, j'expliquerai comment j'ai implémenté des serveurs dans le Cloud, dans le but de tester les éléments qui auront été présentés dans la partie précédente.

Chapitre 1

Présentation du stage

1.1 L'Observatoire Astronomique de Strasbourg

Cette présentation a été co-écrite avec Yann BISCH et Maxime BOURGEOIS, réalisant également leur stage à l'Observatoire Astronomique de Strasbourg.

1.1.1 Présentation générale



FIGURE 1.1 – Grande coupole de l'Observatoire

L'Observatoire Astronomique de Strasbourg est un Observatoire des Sciences de l'Univers (OSU), un UFR de l'Université de Strasbourg et une Unité Mixte de Recherche (UMR 7550) entre l'Université et le CNRS. Il est actuellement dirigé par Hervé WOZNIAK.

L'Observatoire avait pour vocation initiale l'astronomie de position et l'observation de comètes, de météorites et d'étoiles variables. Plus tard, on s'y intéressa à la photométrie et à l'observation d'étoiles doubles. Ce n'est qu'à partir de 1965 que les astronomes comprirent que

l'observation au sol avait atteint ses limites, et proposèrent le concept du Satellite Hipparcos à l'Agence Spatiale Européenne.

Historiquement, il y a eu trois Observatoires Astronomiques. Le premier fut construit en 1673 sur les tours d'enceinte de la ville, avec l'aide de l'astronome Julius REICHEL. Le second a quant à lui été bâti en 1828 sur les bâtiments de l'Académie de Strasbourg. L'Observatoire Astronomique obtint son emplacement actuel en 1881, année de son inauguration. Celui-ci fait partie, avec le jardin botanique situé juste à côté, des éléments qui symbolisent la décision politique de faire de Strasbourg une vitrine, prise par l'empereur Guillaume 1^{er} d'Allemagne à la suite de la guerre de 1870.

Il est composé de trois bâtiments dont le plus célèbre est celui de la Grande Coupole (Figure 1.1), qui contient la 3^e plus grande lunette de France par sa taille. L'Observatoire héberge également le Planétarium de Strasbourg dont il a eu la responsabilité de 1986 à 2008. Celui-ci est désormais intégré au Jardin des Sciences de l'Université de Strasbourg. L'Observatoire dispose également d'un riche patrimoine d'instruments et d'ouvrages anciens.

Il est structuré en trois équipes de recherche et deux Services d'Observation de l'Institut National des Sciences de l'Univers (INSU), le Survey Science Centre d'XMM-Newton (SSC-XMM) et le Centre de Données astronomiques de Strasbourg (CDS).

Son statut d'OSU (voir décret n° 85-657 du 27 juin 1985) place l'Observatoire astronomique au cœur du dispositif national mis en œuvre par l'INSU. Par conséquent, l'Observatoire astronomique de Strasbourg a pour mission de contribuer aux progrès de la connaissance par :

- l'acquisition de données d'observation ;
- le développement et l'exploitation de moyens appropriés ;
- l'élaboration des outils théoriques nécessaires.

De plus, l'Observatoire est également chargé :

- de fournir des services liés à son activité de recherche ;
- d'assurer la formation des étudiants et des personnels de recherche ;
- d'assurer la diffusion des connaissances ;
- de prendre part à des activités de coopération internationale.

1.1.2 Équipes de recherche

Galaxies L'équipe « Galaxies » étudie la formation et l'évolution des galaxies au travers de leurs populations stellaires, ainsi que de la dynamique des étoiles et de la matière noire. Elle est impliquée dans la préparation de la mission satellitaire astrométrique Gaia de l'Agence Spatiale Européenne, dont le lancement est prévu pour octobre prochain, ainsi que dans le grand relevé cinématique RAVE, qui fournira des données d'une précision inédite sur la Voie Lactée.

Les recherches sont concentrées sur les caractéristiques et les propriétés physiques de notre galaxie et de ses plus proches voisines. Afin de pouvoir étudier la dynamique gravitationnelle qui régit les mouvements du gaz et des étoiles, on analyse les évolutions de la population stel-

laire de ces galaxies. Les résultats obtenus permettent de reconstituer les événements clés de la vie d'une galaxie.

L'équipe mène également d'autres recherches sur les amas stellaires, qui sont des composants fondamentaux d'une galaxie. Grâce aux connaissances accumulées concernant les mouvements gravitationnels des étoiles, il est possible de simuler numériquement l'évolution d'un amas stellaire, afin de mieux comprendre la formation des galaxies.

Hautes Énergies L'équipe « Hautes Énergies » s'intéresse aux sources galactiques et extragalactiques émettrices en rayons X, objets compacts (étoiles à neutrons, naines blanches, etc.) et noyaux actifs de galaxies. Elle est impliquée dans le SSC-XMM, un consortium international de laboratoires sélectionnés par l'ESA, labellisé par l'INSU comme Service d'Observation, qui est en charge de fournir des catalogues complets de sources X observées par le satellite XMM-Newton à la communauté internationale.

CDS Le Centre de Données astronomiques de Strasbourg (CDS) est à la fois une équipe de recherche et un Service d'Observation. Les services de bases de données (Simbad, VizieR) et de visualisation (Aladin), développés par le CDS, sont utilisés par l'ensemble de la communauté astronomique mondiale. Celui-ci est l'un des acteurs majeurs du développement de l'Observatoire Virtuel International en astronomie. Fin 2008, le CDS a été labellisé « Très Grande Infrastructure de Recherche » (TGIR), ce qui le place au même niveau que des infrastructures internationales comme l'European Southern Observatory ou RENATER à l'échelon national. J'ai effectué mon stage dans ce service, c'est pourquoi il va maintenant vous être présenté plus en détail.

1.1.3 Le Centre de Données Astronomiques de Strasbourg

Créé en 1972 sous le nom de « Centre de Données Stellaires », il devint le Centre de Données astronomiques de Strasbourg en 1983. Il héberge maintenant la base de données de référence mondiale pour l'identification d'objets astronomiques. Par conséquent, ses missions consistent :

- à rassembler des informations utiles concernant les objets astronomiques sous forme informatisée ;
- à mettre à jour ces données en les critiquant et en les comparant ;
- à les distribuer à la communauté internationale ;
- à mener des recherches utilisant ces données.

Le CDS emploie actuellement 8 chercheurs/astronomes-adjoints, 1 chercheur postdoctoral, 11 informaticiens, 10 documentalistes et 3 administratifs, soit 33 personnes au total. Les services principaux du CDS se nomment Simbad, VizieR et Aladin (Figure 1.2).



FIGURE 1.2 – Logos des principaux services du CDS

Simbad Simbad est la base de données de référence pour la nomenclature et la bibliographie des objets astronomiques situés hors du système solaire, qu’il a accumulé depuis plus de 40 ans. Ce service permet aux astronomes de trouver facilement des informations sur plus de 7 millions d’objets, grâce à plus de 280 000 références bibliographiques. De plus, la base de données Simbad contient un résolveur de noms permettant de retrouver l’objet désigné par l’un des 18 millions d’identifiants qu’elle contient. Des liens peuvent bien évidemment être faits avec d’autres services fonctionnant sur Internet. Simbad recevait en 2011 une moyenne de 284 000 requêtes par jour. Pour le mois dernier, il a dû traiter 18 985 164 requêtes soit environ 7 requêtes par secondes.

VizieR VizieR est une base de données qui regroupe à l’heure actuelle plus de 11 000 catalogues d’objets astronomiques. Ces catalogues sont constitués de données relevées durant des missions d’observation de l’espace puis ajoutées à VizieR par les documentalistes du CDS. Ce procédé permet d’homogénéiser les données astronomiques afin de pouvoir les comparer et de les exporter sous différents formats. Les interrogations sur la base de catalogues peuvent porter sur de multiples critères comme la longueur d’onde avec laquelle les objets ont été observés ou encore le nom de la mission correspondante. Entre 2011 et 2012, les requêtes de VizieR ont augmenté d’une moyenne de 370 000 requêtes par jour à une moyenne de 377 000 requêtes par jour. Le service connaît des pics d’utilisation pouvant atteindre 2 millions de requêtes en un jour.

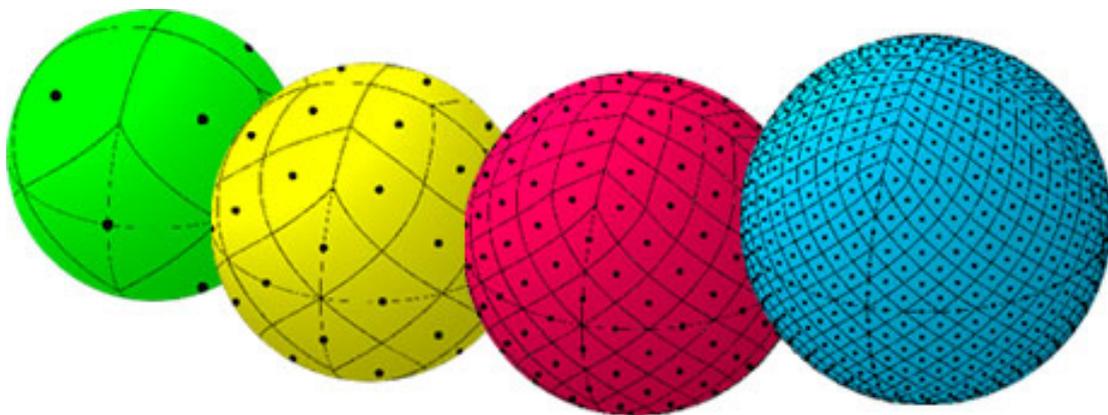


FIGURE 1.3 – Niveaux de précision des boules HEALPix

Aladin Aladin est un atlas interactif du ciel qui permet de visualiser et de comparer des images astronomiques. Il a été entièrement développé en Java pour l'ensemble de la communauté internationale par Pierre FERNIQUE, Thomas BOCH et François BONNAREL. Ces images proviennent d'observatoires sol et spatiaux et peuvent être utilisées pour former des cartes du ciel en trois dimensions, les boules HEALPix. Celles-ci sont des ensembles de losanges de surfaces égales. Ces images disposent de plusieurs niveaux de précision (Figure 1.3), étant chacun constitués de 4 fois plus d'images que le précédent. Elles peuvent être fournies par l'utilisateur, par la base de données d'Aladin ou par d'autres bases de données telles que la NASA Extragalactical Database. Aladin est à la fois une application téléchargeable et une applet Java, toutes deux disponibles sur le site internet du CDS. Aladin recevait en 2012 une moyenne de 18 586 requêtes par jour, ce qui représente une hausse de 11% par rapport à l'année précédente, dont la moyenne était de 16 800 requêtes par jour. Cependant, on peut remarquer que l'usage d'Aladin se resserre sur l'application standalone au détriment de l'applet. Cela témoigne d'une diminution de l'audience « grand public » au profit d'une utilisation scientifique.



FIGURE 1.4 – Logos de l'IVOA et de ses membres

L'International Virtual Observatory Alliance L'International Virtual Observatory Alliance (IVOA) (Figure 1.4) est une organisation internationale regroupant 19 projets d'Observatoire Virtuel (VO) provenant de 17 pays différents : Allemagne, Argentine, Arménie, Australie, Brésil, Canada, Chine, Espagne, États-Unis, France, Hongrie, Inde, Italie, Japon, Royaume-Uni, Russie et Ukraine. Ce projet international a un but similaire au World Wide Web Consortium :

mettre en place des standards et décider quels seront les travaux les plus importants à faire ainsi que les stratégies à adopter. Ceci permet de faciliter les échanges internationaux et la collaboration nécessaire à l'élaboration de standards d'interopérabilité et au déploiement d'outils, de systèmes et de structures organisés nécessaires pour permettre une utilisation internationale des archives astronomiques (images, catalogues, etc.). Telle est la mission confiée à l'IVOA lors de sa création en juin 2002. Le CDS participe activement à ce projet notamment par la réalisation des services cités précédemment.

En tant que Très Grande Infrastructure de Recherche (TGIR), le CDS a des besoins conséquents en terme de puissance de calcul et de stockage de données. Par ailleurs, on peut remarquer que l'utilisation de ses services est en augmentation, ce qui confirme la tendance des années précédentes. Cela permet d'expliquer que des redimensionnements de l'architecture serveur soient envisagés. Dans cette optique, le CDS a donc proposé un stage consistant à évaluer les possibilités offertes par le Cloud Computing / Storage.

1.2 Sujet de stage

1.2.1 Étude réalisée deux ans auparavant

En mai 2011, André SCHAAFF a réalisé une étude concernant le Cloud. Présentée à l'International Virtual Observatory Alliance, celle-ci avait pour but de s'interroger sur l'intérêt du Cloud pour un Observatoire Virtuel. Les principales conclusions de cette étude ont été les suivantes :

- l'utilisation du Cloud paraît attractive à différents niveaux (réduction des coûts du matériel et de la main-d'œuvre) ;
- les coûts de stockage dégressifs ne compensent ceux occasionnés par l'évolution rapide du volume de données astronomiques ;
- suivre l'évolution de l'offre Cloud peut être intéressant.

1.2.2 Énoncé du sujet

Après deux ans d'évolutions pour le Cloud, le CDS a proposé le sujet de stage suivant :

Le CDS a une forte activité de R&D dans de nombreux domaines (nouvelles interfaces utilisateurs (multitouch, etc.), outils mobiles (tablettes, smartphones), technologies Web (AJAX, JQuery, etc.) / HTML5, virtualisation, haute disponibilité, etc.) Nous avons effectué une petite étude concernant l'utilisation du Cloud pour notre service de catalogues (environ 10 000 actuellement) VizieR il y a 2 ans et nous souhaiterions approfondir cette étude et la généraliser à d'autres cas d'utilisation. Le stagiaire mènera dans un premier temps un travail de réflexion et d'établissement d'un état de l'art concernant le Cloud Computing / Storage. Ce travail sera

ensuite suivi du développement de prototypes dont le contenu sera déterminé par cette étude préalable.

1.2.3 Organisation du stage

Pour répondre à ce sujet, j'ai donc découpé mon travail en deux grandes parties (Figure 1.5). Tout d'abord, j'ai réalisé un État de l'art du Cloud, puis j'ai effectué divers tests afin d'évaluer les coûts et les performances réelles d'un Cloud, ainsi que la facilité d'utilisation de ses services.

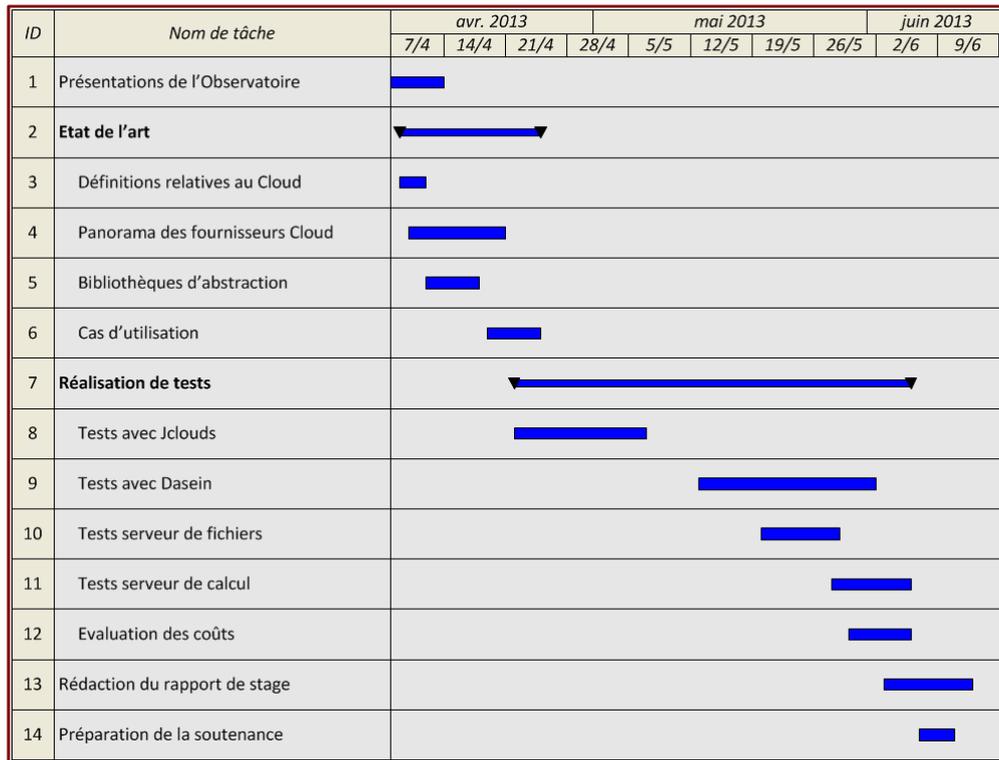


FIGURE 1.5 – Planning du stage

Ce chapitre a permis de présenter le contexte de mon stage de façon générale, afin d'expliquer dans quel but son sujet a été proposé et de pouvoir ensuite présenter le travail réalisé.

Chapitre 2

État de l'art du Cloud

2.1 Le Cloud : une offre très variée

2.1.1 Différents types de services

C'est en 2006, lors du lancement d'Elastic Compute Cloud (EC2), un service proposé par Amazon, que l'on commença à parler d'informatique dans le Cloud. L'effet de mode aidant, ce terme fut ensuite utilisé par une multitude d'offres plus ou moins proches de l'esprit initial. Pour pallier ce problème, le National Institute of Standards and Technology (NIST) définit en 2011 trois niveaux de service (Figure 2.1) dans le Cloud : l'IaaS, le PaaS et le SaaS.

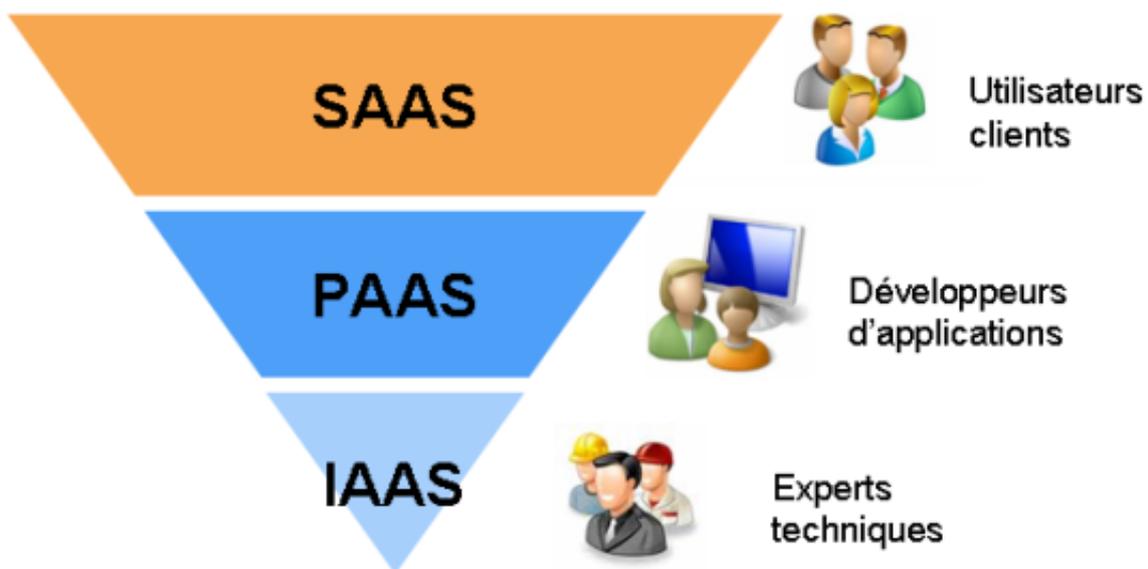


FIGURE 2.1 – Niveaux de service Cloud

Infrastructure as a Service (IaaS) Le modèle de services de Cloud le plus courant est le Cloud d'infrastructure. Ce type de service permet au client de déployer une infrastructure vir-

tualisée et distribuée. L'IaaS peut permettre à une entreprise de mettre en production ses applications, sans investir dans des serveurs physiques locaux. Celle-ci est constituée d'un ensemble de serveurs de calcul ou de données reliés entre eux, qui peuvent être connectés à l'Internet par le biais d'une adresse IP fixe. L'unité de consommation d'un Cloud d'infrastructure est la machine virtuelle.

Platform as a Service (PaaS) Le travail des développeurs peut être simplifié grâce à l'IaaS. Cependant, il faut toujours gérer l'infrastructure mise à leur disposition, ce qui nécessite le travail d'experts techniques. Le modèle PaaS permet de fournir un niveau d'abstraction supplémentaire au Cloud. En effet, il lui ajoute une couche applicative contenant les éléments principaux permettant de déployer des applications : système d'exploitation, système de gestion de bases de données, serveur web... Le PaaS permet donc aux développeurs de mettre en production des applications sans se préoccuper de l'installation des technologies utilisées pour les exécuter.

Software as a Service (SaaS) Actuellement, le modèle de Cloud SaaS est celui qui génère le plus de revenus. Celui-ci permet à l'entreprise cliente d'accéder à une application hébergée dans le Cloud sans avoir à se préoccuper de l'infrastructure, de la plate-forme ou de l'application en elle-même. Cela permet de ne pas avoir à mettre à jour ou à installer des correctifs, car ceux-ci relèvent du fournisseur SaaS.

Autres types de services Bien qu'ils ne fassent pas partie des standards mis en place par le NIST, certains fournisseurs proposent également les services suivants :

- Data as a Service (DaaS) : mise à la disposition via le Cloud de données récoltées par le fournisseur ;
- Desktop as a Service (DaaS) : externalisation d'une Virtual Desktop Infrastructure (VDI) auprès d'un fournisseur de services ;
- Storage as a Service (STaaS) : stockage de fichiers chez un fournisseur à des fins de sauvegarde ou de partage de fichiers ;

Points communs Le Cloud permet donc d'offrir des services très différents. Cependant, ceux-ci ont quelques points communs essentiels qui les différencient réellement de l'offre existante. Tout d'abord, le Cloud permet de ne payer que les ressources qui sont réellement consommées, et permet ainsi d'éviter de sur-dimensionner son architecture. Ensuite, le Cloud permet d'augmenter et de diminuer les ressources utilisées de façon extrêmement rapide, et offre ainsi des capacités théoriquement infinies.

2.1.2 Différents types de Clouds

L'organisation des serveurs utilisés dans le Cloud sont souvent séparés en quatre catégories : le Cloud public, le Cloud privé, le Cloud communautaire et le Cloud hybride.

Cloud public Ce terme permet de désigner des serveurs dans le Cloud qui sont partagés avec les autres utilisateurs. Cela permet bien évidemment de réduire les coûts relatifs à leur utilisation, mais peut compromettre la sécurité de ses serveurs et diminuer leurs performances.

Cloud privé Le Cloud privé peut désigner deux architectures différentes. En effet, on peut désigner ainsi des serveurs dans le Cloud qui seraient physiquement isolés de ceux utilisés par d'autres clients de son fournisseur, mais également une architecture virtualisée hébergée dans les locaux de l'entreprise.

Cloud communautaire Le Cloud communautaire est un Cloud public un peu particulier. En effet, quelques fournisseurs permettent de limiter le partage des machines utilisées à une communauté, par exemple un groupement d'entreprises. Cela permet de mieux protéger les données présentes sur son serveur.

Cloud hybride Le Cloud hybride est un terme décrivant un ensemble de plusieurs types de Clouds interagissant ensemble. Cette association offre plusieurs possibilités, dont la plus utile est le « Cloud Bursting ». Ce concept est caractérisé par le déploiement de serveurs identiques à l'intérieur et à l'extérieur de l'entreprise. Une bonne répartition de charge permet ensuite de n'utiliser le Cloud que pour gérer les pics d'utilisation, et de réaliser ainsi d'importantes économies.

2.2 Panorama des différents fournisseurs

2.2.1 Différents types de fournisseurs

Actuellement, le nombre d'entreprises fournissant des services de Cloud est impressionnant (Figure 2.2). Celles-ci peuvent être réparties en plusieurs catégories, car leur mode de fonctionnement est bien différent.

Fournisseur multi-services Certains fournisseurs proposent à la fois des services d'IaaS, de PaaS et de SaaS. Les plus connus sont bien évidemment Amazon Web Services, Google Cloud et Windows Azure. Leur présence dans tous les secteurs du Cloud leur permet d'être considérés comme de véritables piliers du Cloud.

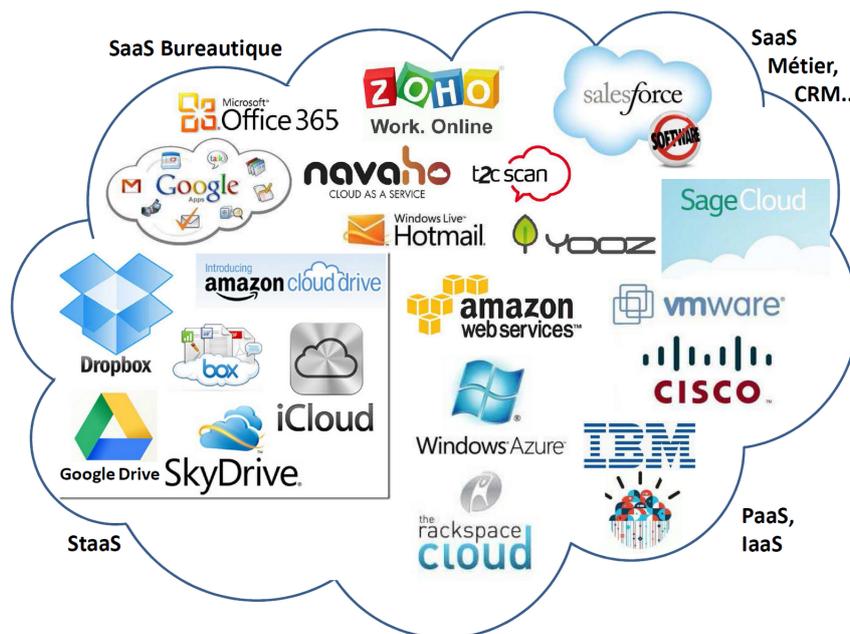


FIGURE 2.2 – Le Cloud : une grande variété de fournisseurs

Fournisseurs intermédiaires Le modèle IaaS / PaaS / SaaS peut permettre à une entreprise de déployer une plate-forme PaaS, sur une infrastructure IaaS fournie par une entreprise tierce. De même, des services SaaS peuvent être mis à la disposition des utilisateurs finaux via une plate-forme PaaS. De très nombreuses offres de SaaS et de PaaS sont ainsi construites sur d'autres services de Cloud. On peut par exemple citer des PaaS tels que CloudBees, permettant de développer des applications Java et de profiter de logiciels d'intégration continue, qui est hébergé sur les serveurs d'Amazon Web Services et de HP Cloud, ou encore PHP Fog, qui utilise l'infrastructure de Rackspace et de Windows Azure.

Fournisseurs de modèles Cloud Certaines sociétés proposent des logiciels permettant de virtualiser des serveurs sur son infrastructure locale, afin d'offrir des services de Cloud. Cela peut avoir deux objectifs : devenir un fournisseur et proposer ainsi simplement l'utilisation de ses serveurs à ses futurs clients, mais aussi construire en interne une architecture semblable à celle de divers Clouds publics afin d'utiliser le « Cloud Bursting » à son avantage. Des solutions de ce type sont proposées par deux leaders de la virtualisation : Citrix et VmWare, qui éditent également des applications permettant de bien gérer un Cloud hybride. Par ailleurs, le projet libre OpenStack, mené par Rackspace, fournisseur de services de Cloud, permet également de déployer simplement son propre Cloud.

2.2.2 Qualité de service

La qualité de service des différents fournisseurs peut être inégale. Cependant, ceux-ci la définissent Service Level Agreement (SLA). Un fournisseur assure en général une disponibilité de

son service de 99,95%, ce qui correspond à environ 4 heures d'indisponibilité par an. Ce chiffre peut atteindre 99,99%. Le SLA permet également de donner des indications sur la pérennité des données, en précisant par exemple si elles seront redondantes ou non. Enfin, le SLA permet de définir la manière dont sera dédommagé le client si ces conditions ne sont pas respectées. Ce dédommagement n'est en général qu'un avoir sur les prochaines consommations du client. Par ailleurs, CloudSleuth (Figure 2.3), un service d'expertise[3] proposé par Compuware, permet d'évaluer gratuitement la disponibilité et le temps de réponse de bon nombre de fournisseurs Cloud, et ce depuis différentes grandes villes du monde. Les résultats peuvent être présentés sous forme de carte, de graphique ou de tableau.



FIGURE 2.3 – Outil de diagnostic de différents fournisseurs

2.3 Cas d'utilisation

Afin d'économiser des ressources, une stratégie Cloud peut être mise en place suivant divers cas d'utilisation. Certains concernent d'ailleurs déjà les données astronomiques. Afin d'en présenter certains, je me suis intéressé aux client d'Amazon Web Services[2].

2.3.1 Environnement de test

Certaines entreprises se servent du Cloud comme d'un environnement de test et de recette. En effet, le Cloud permet de démarrer une phase de test avant d'acheter les serveurs qui serviront ensuite à déployer un projet. De plus, si un projet en test est finalement abandonné, l'entreprise qui utilise le Cloud pour ses tests n'aura pas investi inutilement dans des serveurs. Certains fournisseurs, comme Cloudbees (PaaS Java) par exemple, fournissent d'ailleurs des serveurs de

test à prix réduits, qui, limités en nombre de connexions simultanées, ne permettraient pas une mise en production.

VSC Technologies Cette filiale de la SNCF, gérant un des plus importants sites commerciaux européens : voyages-sncf.com, a choisi d'utiliser le Cloud pour tester et mettre en recette ses applications. Le Cloud n'est pas utilisé pour mettre en production une application car il est considéré comme trop peu sécurisé par rapport aux données personnelles des clients du site. Cependant, le Cloud leur a par exemple permis d'économiser un mois de tests sur un chantier de 6 mois concernant un projet de calendrier des prix.

Informa Informa est une multinationale basée en Suisse qui fournit compétences et savoir-faire inter-entreprises dans les secteurs de l'édition et de la formation. Le Cloud lui a permis de réduire le temps de déploiement de ses environnements de développement et de test. En effet, cette opération a duré 3 jours au lieu des deux mois habituels lorsqu'ils ont lancé une application de comptabilité permettant de générer et d'envoyer des rapports financiers aux unités commerciales d'Informa.

2.3.2 Gestion des pics d'utilisation

Certains sites internet ont parfois des pics d'utilisation très importants par rapport à leur moyenne de requêtes par jour. Par conséquent, un paiement à la demande peut permettre de réaliser d'importantes économies.

Orange Digital Orange Digital est la société gérant les services numériques de l'opérateur Everything Everywhere. Grâce à une étude interne sur le besoin de redimensionner son infrastructure, cette société a constaté qu'une mise à l'échelle conduirait à des serveurs qui seraient sur-dimensionnés 90% du temps. L'utilisation du Cloud pour répondre aux pics de trafic a permis de réduire les coûts d'infrastructure de 2 millions de livres sterling sur une période de trois ans.

Naughty Dog Célèbre société de développement de jeux vidéos, Naughty Dog a choisi le Cloud pour héberger les parties multijoueurs des jeux Uncharted 2 et Uncharted 3. Ce choix leur offre la possibilité de démarrer et de stopper les serveurs de jeu au fur et à mesure des connexions des joueurs. Cela permet à la fois d'éviter d'avoir des serveurs inutilisés, mais aussi de pouvoir démarrer un nombre théoriquement infini de parties simultanées. En raison du fort succès de ses jeux, le gain réalisé a été estimé à 90 % du coût qu'auraient représenté des serveurs équivalents sur site.

2.3.3 Astronomie

Le Cloud a déjà été utilisé pour des applications relevant du domaine de l'astronomie. En effet, celui-ci peut-être très utile pour traiter de très grande quantités de données.

The Server Labs The Server Labs est une société de services européenne dont l'une des missions actuelles est d'aider l'European Space Agency à construire l'Astrometric Global Iterative Solution (AGIS) afin de traiter les observations produites par le satellite Gaia. The Server Labs a estimé que traiter la série de données entière d'un milliard d'étoiles avec 6 ans de données coûterait 463 000 dollars grâce au Cloud, tandis que l'achat seul des serveurs permettant de réaliser la même tâche coûterait 972 000 dollars.

NASA La NASA a choisi d'héberger dans le Cloud les données envoyées par l'astromobile Curiosity, en mission sur Mars. Cela permet de mettre simplement à la disposition de la communauté scientifique mondiale les images provenant de cet appareil, ainsi que d'automatiser leur traitement, sans avoir à se préoccuper de l'infrastructure mise en œuvre. Cette mission doit permettre aux chercheurs d'identifier des dangers potentiels ou des zones ayant un intérêt scientifique particulier.

2.4 Problèmes liés au Cloud

L'utilisation de serveurs présents dans le Cloud peut cependant être la source de certains problèmes, qu'il faut prendre en compte avant d'envisager de recourir à ce genre de services.

Sécurité des données Lorsque que l'on déploie une infrastructure ou une application dans le Cloud, on peut s'interroger sur la sécurité des données hébergées. En effet, il est difficile d'évaluer la protection qui a pu être mise en place par les différents fournisseurs. Cependant, la plupart des fournisseurs ont obtenu une certification ISO 27000 pour leur Information Security Management System (ISMS). Celles-ci sont des normes de sécurité basées sur l'évaluation de risques périodiques. Elle permet de protéger non seulement la confidentialité, mais aussi l'intégrité et la disponibilité des données appartenant à une entreprise ou à l'un de ses clients.

Interruption de services Ces dernières années, certains fournisseurs Cloud ont brusquement arrêté de fournir leurs services. Ce fut par exemple le cas de GoDaddy, SliceHost et Zumo-Drive. Ces fournisseurs ont à ce moment permis à leurs clients de faire migrer gratuitement leurs services vers l'infrastructure de leurs concurrents. Cependant, cette migration a entraîné une modification des tarifs des ressources consommées par leurs anciens clients, ce qui a pu conduire certaines entreprises à étudier à nouveau l'ensemble des offres proposées.

Des interfaces de programmation différentes La transition d'un fournisseur de Cloud à un autre peut se révéler très difficile. En effet, la plupart des fournisseurs disposent de leur propre interface de programmation permettant d'automatiser la gestion de ses ressources dans le Cloud. Par conséquent, changer de fournisseur peut conduire à devoir développer à nouveau tous les scripts et applications qu'une entreprise aurait pu mettre en place dans le but de contrôler l'utilisation du Cloud.

2.5 Des bibliothèques d'abstraction

Pour pallier ce dernier problème, différentes bibliothèques d'abstraction[1] ont été développées. En effet, celles-ci permettent d'uniformiser les interactions avec le Cloud afin de n'avoir que quelques paramètres à changer si un changement de fournisseur est envisagé. Ces projets sont en général Open Source et concernent de nombreux langages différents.



FIGURE 2.4 – Logos des principales bibliothèques d'abstraction

Les plus connues (Figure 2.4) sont :

- Dasein Cloud (Java) ;
- Jclouds (Java et Clojure) ;
- Deltacloud (Ruby et C++) ;
- The Simple Cloud API (PHP) ;
- Apache Libcloud (Python).

Cette étude m'a permis d'évaluer les possibilités que pouvait offrir le Cloud. Par la suite, j'ai pu implémenter des serveurs dans le Cloud et les commander grâce à des bibliothèques d'abstraction.

Chapitre 3

Implémentation de serveurs dans le Cloud

Afin de tester le déploiement des serveurs dans le Cloud, j'ai choisi d'utiliser une bibliothèque d'abstraction. Mon choix s'est porté sur la bibliothèque Jclouds, car elle semblait être compatible avec de nombreux fournisseurs et était développée en Java, langage à la base de la formation dispensée à l'IUT Nancy-Charlemagne. J'ai donc cherché des fournisseurs compatibles avec Jclouds offrant à leurs clients potentiels une version d'essai qui ne nécessite pas d'utiliser un numéro de carte bancaire. Je me suis donc servi de deux fournisseurs de Cloud : Cloud Sigma et Go2Cloud.

3.1 Jclouds

Jclouds[7] est un projet Open Source fondé par Adrian COLE. Sa dernière mise à jour date d'avril 2013. Cette bibliothèque est divisée en deux parties principales : Blobstore et Compute-Service. Le premier permet de gérer les serveurs de données tandis que le deuxième permet de gérer les serveurs de calcul.

1^{er} problème Ce projet souffre d'un problème de documentation. En effet, bien que des guides d'utilisation soient disponibles, ils utilisent fréquemment des classes devenues obsolètes (deprecated), ou qui ont même été supprimées au fil des versions. De plus, la documentation (javadoc) disponible est très incomplète, ce qui rend très difficile la compréhension du rôle des différentes classes et méthodes. Cependant, le projet est suivi par une communauté communiquant via un Google Group, ce qui permet d'obtenir aisément de l'aide et des informations. De plus, de nombreux exemples de code utilisant la bibliothèques sont disponibles en téléchargement sur le dépôt Github du projet.

2^e problème L'Open Source a eu un effet négatif sur ce projet. En effet, de nombreuses entreprises ont participé à son développement, afin de s'assurer que toutes les fonctionnalités de leur offre de Cloud puissent être gérées par la bibliothèque. Cela entraîne donc une architecture

des classes incohérente et des disparités entre les possibilités de contrôle de chaque fournisseur de Cloud, ce qui est bien évidemment l'inverse des besoins auxquels doit répondre une bibliothèque d'abstraction.

Apports de l'utilisation de Jclouds D'une part, me servir de Jclouds m'a permis d'avoir un élément de comparaison lorsque j'ai testé d'autres outils par la suite. D'autre part, l'utilisation de ce projet m'a permis de me documenter sur trois éléments :

- Maven : un outil comparable au système Make d'Unix, dont le but est de produire un logiciel Java à partir de ses sources, en optimisant les tâches réalisées à cette fin et en garantissant le bon ordre de fabrication ;
- Ant : un outil similaire à Maven, mais plus procédural. Il faut lui indiquer ce qu'il doit faire et quand il doit le faire, contrairement à Maven qui exécute des actions par défaut ;
- Log4J : une API configurable permettant de journaliser des informations.

Les deux premiers m'ont servi à comprendre comment étaient construits les exemples et à récupérer les différents composants de Jclouds. Log4J m'a quant à lui montré que les fonctions que j'utilisais ne fonctionnaient pas parce que je n'étais pas connecté au Cloud, alors qu'aucun autre message ne me le signalait.

En réalité, les versions d'essai que j'avais utilisées ne permettaient de contrôler ses serveurs dans le Cloud qu'avec l'aide de leur interface web. Il était donc impossible de contrôler ses serveurs en s'affranchissant des outils liés au fournisseur. Par ailleurs, elles duraient toutes deux moins d'une semaine, ce qui m'a poussé à utiliser d'autres outils pour interagir avec le Cloud.

3.2 VmWare Hybrid Cloud Evaluation

Comme dit précédemment, VmWare produit un logiciel de virtualisation permettant de fournir des services de Cloud. L'un des avantages de cet outil est qu'il n'est contrôlé que par une unique interface de programmation, et qu'un vaste écosystème de fournisseurs est disponible. Bon nombre d'entre eux proposent une version d'essai, mais celles-ci requièrent fréquemment un numéro de carte bancaire. Afin d'encourager plus de clients à utiliser les services de ses partenaires, VmWare a donc créé L'Hybrid Cloud Evaluation[8].

Possibilités Cette version permet de tester les possibilités offertes par le Cloud pendant 90 jours, ce qui est largement supérieur à ce que tous les autres fournisseurs de services proposent. De plus, il est possible de lancer deux serveurs simultanément. Ces deux raisons m'ont poussé à utiliser cette plate-forme pendant le reste de mon stage.

vApp et Template Les serveurs déployés par l'intermédiaire d'un partenaire de VmWare ne sont pas organisés de manière classique. En effet, les composantes principales de son Cloud

ne sont plus des machines virtuelles, mais des vApp, qui regroupent plusieurs machines virtuelles interagissant entre elles. Ces vApp peuvent être créées à partir d'un modèle (Template), comme par exemple celui d'un serveur LAMP (Linux Apache Mysql Php). L'utilisateur peut bien évidemment créer ses propres vApps, afin de pouvoir par la suite déployer très facilement un serveur identique à un autre créé auparavant. Il ne faut pas confondre vApp et PaaS : bien que les vApps puissent contenir une plate-forme de développement, les logiciels qui y sont pré-installés ne seront pas mis à jour par le fournisseur.

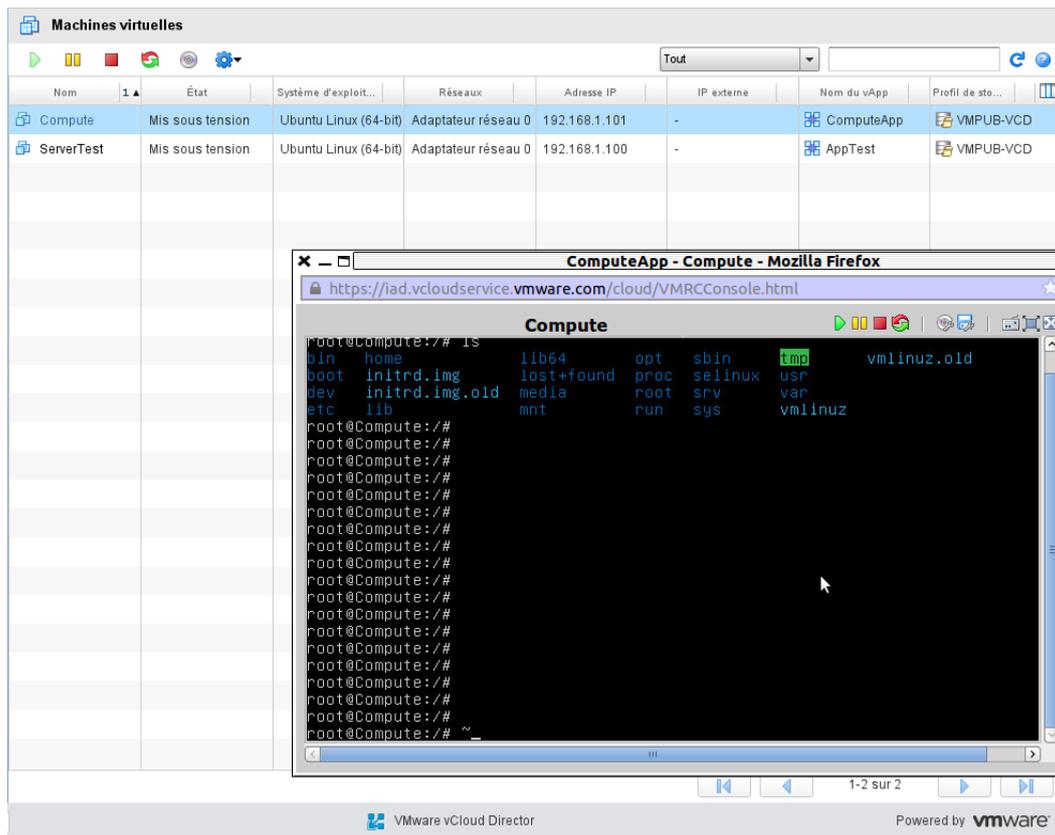


FIGURE 3.1 – Interface de vCloud Director

vCloud Director vCloud Director (Figure 3.1) est l'interface web payante permettant de gérer ses serveurs dans le Cloud. Elle est cependant fournie avec l'Hybrid Cloud Evaluation. Celle-ci permet de lancer et supprimer des serveurs, mais surtout de les contrôler grâce au plugin Virtual Remote Machine Control (VRMC). En effet, il permet d'accéder au terminal du serveur, et donc d'effectuer n'importe quelle tâche sur la machine. vCloud Director permet également de gérer le pare-feu protégeant les serveurs du Cloud, afin d'autoriser des connexions entrantes et/ou sortantes.

Interruptions de service Cette version d'évaluation connaît cependant quelques problèmes. En effet, il m'est arrivé plusieurs fois de ne pas pouvoir me connecter au service sans raison particulière (Figure 3.2). Cependant, ces coupures sont probablement dues au fait qu'il ne s'agisse

Login to your Cloud

We were unable to log you in: You need an active account to login

***If you need assistance with logging in please contact our Support department**

Username & Organization: (Ex: admin@1234)

admin@5411

Enter your Password:

[Forgot your username or password?](#)

LOGIN NOW

Keep me logged into the Simple Console

 Don't have an account? [Sign-up now!](#)

FIGURE 3.2 – Erreur de connexion à l'évaluation

que d'une version d'évaluation.

3.3 Dasein Cloud

Après avoir utilisé Jclouds, j'ai cherché une autre bibliothèque d'abstraction performante. Je me suis tout d'abord orienté vers deltacloud[5], car celle-ci était disponible pour le langage C++, pour lequel j'ai de l'expérience grâce au projet tutoré que j'ai réalisé pendant mon DUT. Cependant, ce projet ne me permettait pas d'utiliser la version d'évaluation de vmWare, dont la durée de la période d'essai était un avantage non négligeable.

Par ailleurs, le site officiel de Dasein Cloud[4] n'étant plus mis à jour depuis longtemps, j'ai fini par découvrir dans son forum qu'il n'était plus utilisé et que tout était maintenant géré depuis le dépôt Github[6] du projet. Dasein Cloud est également un projet Open Source, fondé par George REESE, co-fondateur de la société Enstratus, appartenant désormais à DELL. Cette société produit un logiciel permettant de gérer facilement des serveurs virtuels appartenant à l'infrastructure de plusieurs fournisseurs de Cloud. La dernière version de ce projet date d'avril 2013 et la suivante est prévue pour septembre prochain.

Fournisseurs compatibles Le nombre de fournisseurs pouvant être compatibles avec Dasein Cloud est nettement inférieur à celui de Jclouds. Cependant, les dernières mises à jour ont principalement concerné l'ajout du support de nouvelles interfaces de programmation. Contrai-

rement à Jclouds, Dasein Cloud offre une vraie normalisation de l'utilisation du Cloud. En effet, les fournisseurs ne sont ajoutés aux versions finales du projet qu'une fois que toutes les fonctions attendues sont ajoutées au projet. Par conséquent, il est beaucoup plus simple d'utiliser le même code pour différents fournisseurs. On notera également la possibilité d'utiliser un fournisseur fictif (mock) dans le but de réaliser des tests. Celui-ci se comporte comme si les actions effectuées étaient réellement mises en œuvre dans une infrastructure de Cloud.

Organisation du projet La manière dont sont organisés les divers éléments du projet est beaucoup plus simple. En effet, les binaires à utiliser sont regroupés en archives correspondant à chaque fournisseur, plus une autre contenant le cœur du projet. Les diverses fonctions sont regroupées en différents groupes de services, comme par exemple les « ComputeServices » ou les « DataCenterServices ». Pour chacun des fournisseurs, un simple test permet de vérifier qu'un groupe est supporté. Pour choisir quel Cloud utiliser, il suffit de le mentionner lors de la connexion au Cloud. Par la suite, les opérations sont effectuées très simplement grâce à l'utilisation du polymorphisme et de la liaison dynamique. Le développeur n'a alors plus à se préoccuper du fournisseur utilisé.

Problèmes rencontrés La normalisation pose cependant un problème : certaines fonctionnalités spécifiques à un fournisseur ne peuvent pas être utilisées. Dans le cas de vmWare, les vApps sont tout simplement considérées comme une seule machine virtuelle. Cela empêche donc de reconfigurer son fonctionnement interne, en lui ajoutant par exemple une nouvelle instance de calcul.

D'autre part, j'ai pu constater une fois qu'une fonction qui aurait du être supportée ne faisait rien en réalité. Ce problème était mineur car il s'agissait de la fonction permettant de redémarrer un serveur. Une alternative possible était donc bien évidemment de l'éteindre pour ensuite l'allumer à nouveau. Cependant, ce genre de problèmes peut être présent pour d'autres fonctions d'autres fournisseurs. Il convient donc d'être vigilant à ce sujet.

Enfin, il m'a été impossible de lancer un serveur avec Dasein. En effet, la bibliothèque ne reconnaît comme types de serveurs pouvant être lancés que les vApps créées par l'utilisateur. Or, la version d'essai vmWare ne permettait que d'utiliser des modèles pré-configurés. J'ai donc du créer les serveurs depuis l'interface web fournie avec l'évaluation. Ce problème ne devrait cependant plus en être un pour un véritable client d'un fournisseur de Cloud appartenant à l'écosystème mis en place par vmWare.

3.4 Performances

Ayant déployé deux serveurs dans le Cloud, j'ai voulu tester si leurs performances correspondaient bien à celles annoncées. Dans ce but, j'ai réalisé de simples tests permettant de tester la vitesse de téléchargement, la rapidité de calcul et le temps de réponse à une requête Ping.

3.4.1 Serveur de fichiers

Sur l'un de mes serveurs, j'ai stocké les 7 premiers niveaux d'une carte du ciel HEALPix. Ces images représentent environ 23 Go. De ce fait, ma capacité de stockage limitée à 30 Go ne m'aurait pas permis de stocker également le niveau suivant. Le dépôt de ces images sur le serveur a pu bénéficier d'un débit variant entre 250 et 300 Ko par seconde, et a ainsi duré un peu moins d'une journée. Les données ont été structurées de manière à pouvoir y accéder avec l'application Aladin. La connexion au serveur n'a posé aucun problème, et la vitesse de téléchargement des images a pu atteindre environ 1 Mo par seconde, ce qui est tout à fait satisfaisant étant donné que les serveurs sont basés à Ashburn en Virginie.

3.4.2 Serveur de calcul

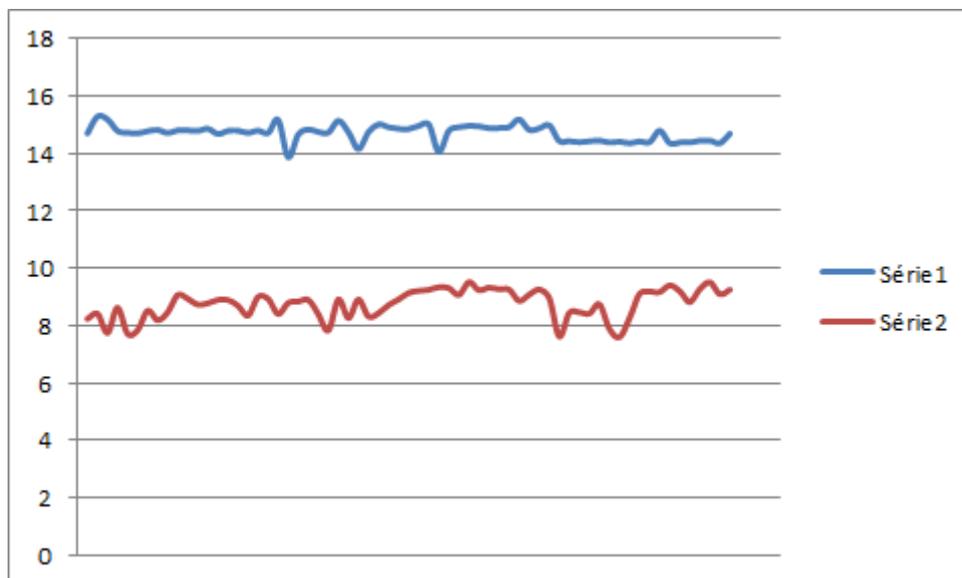


FIGURE 3.3 – Temps d'exécution d'une application Java nécessitant beaucoup de calculs (min)

Afin de tester le serveur de calcul, j'ai utilisé une application Java que j'ai développée il y a quelques temps qui permet d'effectuer des tests sur une grande quantité de données inscrites dans un fichier. En effectuant une grande quantité de calculs, je souhaitais ainsi vérifier que les performances du serveur virtuel étaient conformes à ses caractéristiques, et qu'elles ne variaient pas trop à cause de l'influence de paramètres sur lesquels je n'avais pas de contrôle, comme par exemple l'utilisation que pourraient faire les autres clients de la machine physique sur laquelle était mon serveur de test. Celle-ci prend plus de 8 minutes à s'exécuter sur la machine que j'utilisais à l'Observatoire et plus de 14 minutes dans le Cloud. La différence entre ces performances est cohérente avec la différence existant entre les deux processeurs utilisés. En effet, l'application, qui n'utilise qu'un seul processus à la fois, a fonctionné sur un processeur double cœur cadencé à 2,2 Ghz et sur un processeur virtuel simple cœur théoriquement cadencé à 1,2 Ghz. Lancer cette application toutes les heures pendant près de trois jours m'a permis d'observer que

le temps d'exécution reste stable au fil du temps (Figure 3.3).

3.4.3 Temps de réponse des serveurs

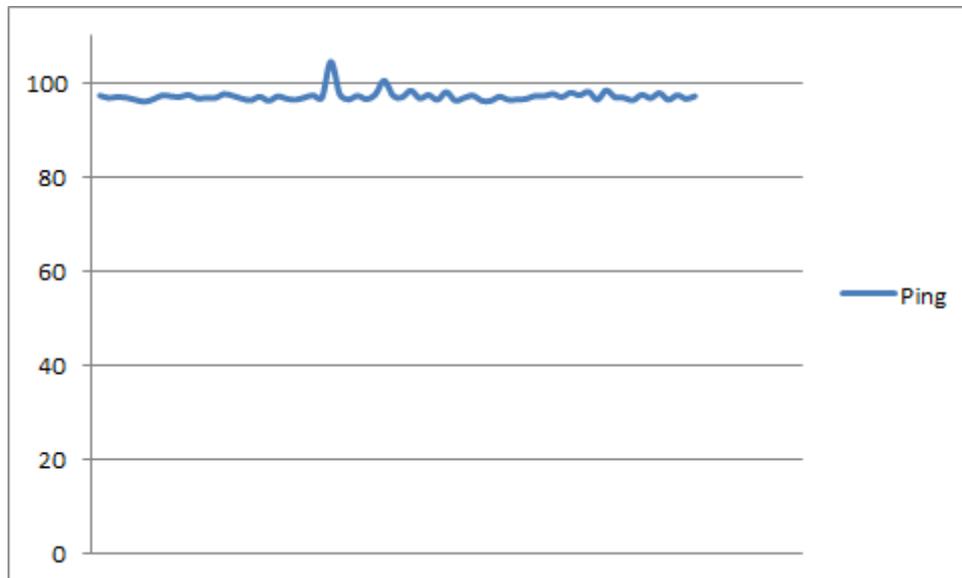


FIGURE 3.4 – Temps de réponse du serveur (ms)

Dans le même ordre d'idée, j'ai voulu tester si les serveurs étaient toujours aussi réactifs. Pour cela, j'ai lancé une requête Ping toutes les heures pendant près de deux jours, ce qui m'a permis d'assurer la stabilité du temps de réponse en dessous de 100 ms (Figure 3.4).

3.5 Évaluation des coûts relatifs au Cloud

Afin d'évaluer les coûts d'une infrastructure dans le Cloud, j'ai relevé les coûts de différents fournisseurs pour des configurations semblables, que ce soit pour un serveur de calcul (Table 3.1) ou de données (Table 3.2). Cela peut constituer une première indication qui déterminera le choix d'un fournisseur de services de Cloud. Les fournisseurs de Cloud sélectionnés pour ces comparatifs ont été choisis parmi ceux compatibles avec la bibliothèque Dasein Cloud, à l'exception de LunaCloud. Ce dernier est présenté à titre de comparaison, car il offre la meilleure disponibilité et le meilleur temps de réponse pour la ville de Paris, selon l'outil CloudSleuth[3]. Par ailleurs, des serveurs européens ont été choisis dès que cela était possible, bien que ceux-ci soient de manière générale plus chers que s'ils étaient hébergés aux États-Unis.

3.5.1 Serveur de calcul

L'évaluation du coût d'un serveur de calcul a été compliquée car tous les fournisseurs de Cloud ne proposent pas la même variété de configurations. Cependant, j'ai choisi d'essayer de m'approcher au plus d'une machine ayant les caractéristiques suivantes :

Fournisseur	Type	vCPU	RAM	Stockage	Coût mensuel
The Joyent Cloud	Public	4	15 Go	1467 Go	267,24 €
Microsoft Azure	Public	8	14 Go	?	275,40 €
Amazon Web Services	Public	8	15 Go	1690 Go	289,51 €
LunaCloud	Public	8	16 Go	1020 Go	293,76 €
The HP Cloud	Public	4	16 Go	480 Go	316,12 €
SoftLayer	Public	8	16 Go	100 Go	385,86€
GoGrid	Privé	16	16 Go	800 Go	448,50 €
CloudSigma	Privé	8	16 Go	1024 Go	506,82 €
Rackspace	Privé	6	15 Go	620 Go	508,04€
OpSource Cloud	Privé	8	16 Go	1024 Go	614,28 €
AT&T Synaptic Storage	Privé	8	16 Go	1024 Go	630,99 €
Tata InstaCompute	Privé	8	16 Go	1024 Go	672,16€
TerreMark	Privé	8	16 Go	1024 Go	716,86 €

TABLE 3.1 – Coût d'un serveur de calcul

- 8 processeurs virtuels (vCPU) ;
- 16 Go de mémoire vive (RAM) ;
- 1024 Go de stockage.

Précisions Les tarifs d'un serveur de calcul ayant cette configuration varient donc entre 260 et 390 euros par mois pour un Cloud public, et de 440 à 720 euros pour un Cloud privé. Bien que le Cloud soit caractérisé par un paiement à l'utilisation, on peut remarquer qu'un serveur de calcul qui ne fonctionne pas, mais qui est encore configuré est en général facturé comme s'il était en fonctionnement. De plus, le choix d'une adresse IP fixe pour pouvoir accéder à son serveur et les requêtes que celui-ci reçoit sont en général facturés en supplément. Cependant, ces coûts additionnels sont négligeables par comparaison avec le coût du serveur lui-même.

3.5.2 Serveur de données

Pour évaluer le coût d'un serveur de données, j'ai choisi de baser mes recherches sur un serveur de 25 To de données, dont environ 1 Go par jour sont téléchargés, ce qui correspond au serveur utilisé pour la base de données images d'Aladin.

Précisions Le coût de dépôt des données est nul sur l'ensemble de ces plate-formes. Le coût du téléchargement des données est payant mais ne représente que 3 euros maximum dans les cas présentés dans cette étude. On peut remarquer que le coût d'un serveur de données peut varier entre 1 000 et 4 000 euros environ. Cette grande différence s'explique par la stratégie de sauvegarde des données mise en place par le fournisseur de Cloud. En effet, les fournisseurs les plus coûteux permettent en général de dupliquer le contenu des serveurs de leurs clients dans différentes régions du Monde. Par ailleurs, le coût du Go de stockage est en général dégressif

Fournisseur	Coût mensuel
Google Storage (Reduced Availability)	1 078,89 €
Microsoft Azure (localement redondant)	1 245,39 €
Amazon Web Services (redondance réduite)	1 279,13 €
Google Storage	1 514,40 €
Microsoft Azure (géographiquement redondant)	1 537,23 €
Amazon Web Services	1 598,24 €
The HP Cloud Object Storage	1 784,32 €
LunaCloud	1 793,60 €
GoGrid	2 187,23 €
AT&T Synaptic Storage	2 971,70 €
AT&T Synaptic Storage (redondant)	3 961,49 €

TABLE 3.2 – Coût d'un serveur de données

en fonction de la quantité totale de volume de stockage nécessaire.

Ainsi, j'ai pu constater que les performances effectives des serveurs présents dans le Cloud étaient bien conformes à celles annoncées. Bien évidemment, des tests de ce genre devraient être effectués à nouveau en cas de changement de fournisseur. De plus, j'ai pu remarquer que l'utilisation du Cloud pouvait être grandement simplifiée grâce aux bibliothèques d'abstraction. Parmi les deux que j'ai testées, je recommande bien évidemment Dasein Cloud, dont la meilleure organisation simplifie la compréhension de son fonctionnement, et des possibilités que peut offrir cet outil. Les coûts d'un serveur dans le Cloud sont relativement élevés par rapport à un serveur physique, et il peut être plus difficile de trouver un financement mensuel qu'un financement ponctuel. Cependant, ces serveurs ne nécessitent aucun frais d'entretien et très peu de coûts de main-d'œuvre, ce qui peut leur permettre d'être tout de même rentables.

Conclusion

Pendant ces 10 semaines, j'ai pu réaliser un état de l'art du Cloud Computing / Storage et tester les possibilités que peut offrir ce type d'infrastructure, ainsi que les moyens permettant de la contrôler. Ce rapport devrait permettre d'apporter à celui qui envisagerait d'utiliser des services de Cloud des informations sur les différents services existants et les moyens de les mettre en œuvre simplement.

Ce stage m'a apporté énormément de connaissances dans le domaine de l'informatique. En effet, afin de bien évaluer les possibilités offertes par les différents fournisseurs de Cloud, j'ai fréquemment fait des recherches sur l'intérêt des outils proposés par les différents fournisseurs pendant que je réalisais l'état de l'art du Cloud. De plus, j'ai eu l'occasion d'assister à deux présentations organisées par les membres de la liste de diffusion « Infusion », regroupant les informaticiens de l'Observatoire. Celles-ci étaient basées sur l'utilisation des bases de données NoSQL et les arbres d'indexations R(*)-Tree et M-Tree, adaptés à des requêtes en base de données concernant des éléments positionnés dans le plan ou dans l'espace. Celles-ci étaient orientées dans le but de présenter leurs applications dans le domaine astronomique, mais les informations présentées peuvent être utiles pour bon nombre de projets.

Par ailleurs, ce stage m'a permis de découvrir le milieu de la recherche comme je le souhaitais. En effet, j'ai eu la chance d'assister à quelques présentations de différents thèmes de recherches en astronomie, ainsi qu'à la soutenance HDR (Habilitation à Diriger des Recherches) de Benoît FAMAËY. Pour finir, je suis également très heureux de revenir travailler ici cet été, ce qui constituera mon premier emploi dans le secteur de l'informatique, comme me l'a proposé mon tuteur André SCHAAFF.

Bibliographie

- [1] 7 ways to beat the glut of cloud apis. <http://blog.programmableweb.com/2010/03/18/7-ways-to-beat-the-glut-of-cloud-apis/>, Avril 2013.
- [2] Case studies - amazon web services. <http://aws.amazon.com/fr/solutions/case-studies/>, Avril 2013.
- [3] Cloudsleuth - global provider view. <https://cloudsleuth.net/global-provider-view>, Juin 2013.
- [4] The dasein cloud api. <http://dasein-cloud.sourceforge.net/>, Mai 2013.
- [5] Deltacloud api. <http://deltacloud.apache.org/>, Mai 2013.
- [6] Greese dasein cloud - github. <https://github.com/greese/dasein-cloud/>, Mai 2013.
- [7] Jclouds. <http://jclouds.incubator.apache.org/>, Avril 2013.
- [8] Vmware hybrid cloud evaluation. <https://vcloudservice.vmware.com/starteval/info>, Mai 2013.

Table des figures

1.1	Grande coupole de l'Observatoire	6
1.2	Logos des principaux services du CDS	9
1.3	Niveaux de précision des boules HEALPix	9
1.4	Logos de l'IVOA et de ses membres	10
1.5	Planning du stage	12
2.1	Niveaux de service Cloud	13
2.2	Le Cloud : une grande variété de fournisseurs	16
2.3	Outil de diagnostic de différents fournisseurs	17
2.4	Logos des principales bibliothèques d'abstraction	20
3.1	Interface de vCloud Director	23
3.2	Erreur de connexion à l'évaluation	24
3.3	Temps d'exécution d'une application Java nécessitant beaucoup de calculs (min)	26
3.4	Temps de réponse du serveur (ms)	27