

Analyse de données à grande échelle : application au dépôt RIPE Atlas

Mémoire réalisé par Hayat BELLAFKIH
pour l'obtention du diplôme de Master en Sciences Informatiques

Année académique 2017 – 2018

Directeurs : B. Quoitin
M. Goeminne

Service : Département de l'Informatique

Rapporteurs : D. Hauweele

Table des matières

| | | |
|-----------|--|-----------|
| I | RIPE Atlas | 4 |
| I.1 | Introduction | 4 |
| I.2 | A propos RIPE NCC | 4 |
| I.3 | Présentation du projet RIPE Atlas | 4 |
| I.3.1 | Les mesures actives et passives de l'Internet | 5 |
| I.3.2 | Généralités sur les sondes Atlas | 5 |
| I.3.3 | Les générations des sondes Atlas | 6 |
| I.3.4 | Les versions du firmware des sondes Atlas | 7 |
| I.3.5 | La connexion des sondes Atlas à Internet | 7 |
| I.3.6 | Architecture du système RIPE Atlas | 8 |
| I.3.7 | Les sondes Atlas et la vie privée | 11 |
| I.3.8 | La sécurité dans RIPE Atlas | 11 |
| I.3.9 | Les ancres VS sondes Atlas | 12 |
| I.3.10 | Les mesures intégrées : Built-in | 12 |
| I.3.11 | Le système de crédits Atlas | 13 |
| I.3.12 | Les mesures personnalisées : User Defined measurement | 15 |
| I.3.13 | La sélection des sondes Atlas | 15 |
| I.3.14 | Les sources de données Atlas | 15 |
| I.3.15 | Les limitations du RIPE Atlas | 16 |
| I.3.16 | Confiance aux données Atlas | 17 |
| I.4 | Projets existants de mesures d'Internet | 18 |
| I.4.1 | Test Traffic Measurement Service | 18 |
| I.4.2 | ProbeAPI | 18 |
| I.4.3 | Archipelago | 19 |
| I.4.4 | DIMES | 19 |
| I.4.5 | SamKnows | 20 |
| I.5 | Quelques cas d'utilisation des données collectées par les sondes Atlas | 20 |
| I.5.1 | Détection des coupures d'Internet | 20 |
| I.5.2 | Aide à la prise de décision | 20 |
| I.5.3 | Le suivi des censures | 21 |
| I.5.4 | Le suivi des performances d'un réseau | 21 |
| I.5.5 | Le suivi des détours dans un trafic local | 24 |
| I.5.6 | Visualisation : indicateurs et dashboard | 24 |
| I.6 | Conclusion | 24 |
| II | La détection des anomalies dans les délais d'un lien | 26 |
| II.1 | Présentation générale | 26 |
| II.2 | Pourquoi analyser les délais des liens réseaux | 27 |
| II.3 | L'étude des délais des liens | 28 |

| | | |
|------------|---|-----------|
| II.3.1 | Les données utilisées dans l'analyse des délais | 28 |
| II.3.2 | Le principe de la détection des changements des délais | 29 |
| II.4 | L'étude des délais des liens en pratique : l'évolution du RTT différentiel des liens | 30 |
| II.4.1 | Les étapes principales de détection | 30 |
| II.4.2 | Description des paramètres de l'analyse des délais | 31 |
| II.4.3 | L'évolution du RTT différentiel d'un lien et la détection des anomalies . | 32 |
| II.4.4 | Processus de détection des anomalies | 33 |
| II.4.5 | Vue globale des étapes de la détection des anomalies à travers l'évolution des RTTs différentiels | 35 |
| II.5 | Conclusion | 36 |
| III | Introduction au Big Data | 38 |
| III.1 | Introduction | 38 |
| III.2 | Processus d'analyse de données massives | 38 |
| III.3 | Quelques concepts associés au Big Data | 39 |
| III.3.1 | Définition du Big Data : Volume, Vitesse, Variété et Véracité | 39 |
| III.3.2 | L'architecture standard du Big Data | 40 |
| III.3.3 | Les bases de données NoSQL (Not Only SQL) | 42 |
| III.3.4 | Schema on Write VS Schema on Read | 46 |
| III.3.5 | L'informatique distribuée et l'analyse de données massives | 47 |
| III.4 | Parcours de quelques technologies du Big Data | 48 |
| III.4.1 | MongoDB | 48 |
| III.4.2 | Amazon DynamoDB | 48 |
| III.4.3 | Amazon S3, Amazon Glue et Amazon Athena | 49 |
| III.4.4 | Apache Spark | 51 |
| III.5 | Conclusion | 54 |
| IV | Application de quelques technologies Big Data sur l'analyse des traceroutes | 55 |
| IV.1 | Introduction (to adapt) | 55 |
| IV.2 | Critères d'évaluation des technologies du Big Data | 55 |
| IV.3 | MongoDB | 56 |
| IV.4 | Amazon DynamoDB | 56 |
| IV.5 | Amazon S3, Amazon Glue et Amazon Athena | 56 |
| IV.6 | Spark Apache avec Scala | 58 |
| IV.6.1 | Application sur traceroutes | 58 |
| IV.7 | Comparatif des performances | 62 |
| IV.7.1 | Caractéristique de l'environnement de test | 62 |
| IV.7.2 | Spark Apache | 62 |
| IV.8 | Récapitulatif des technologies Big Data | 62 |
| IV.8.1 | Notes sur les données | 64 |
| IV.9 | Conclusion | 64 |
| A | Amazon Athena | 66 |
| A.1 | Création de la table traceroutes | 66 |
| A.2 | Partitionnement de données dans un compartiment AWS S3 | 67 |
| A.2.1 | Présentation du partitionnement | 67 |
| A.2.2 | Application du partitionnement sur les traceroutes Atlas | 67 |
| A.2.3 | L'interrogation de données sur Amazon Athena | 68 |

Introduction

L'analyse de données, en particulier les données à grande échelle, attire de plus en plus les entreprises à s'y investir. Ces analyses peuvent affecter potentiellement la stratégie de ces entreprises. Les défis de l'analyse de données massives varient suivant le processus suivi. Par exemple, les défis peuvent être liés à la définition des objectifs d'une analyse, le choix de données, la collecte de données, etc.

L'idée de ce travail est d'exploiter l'existence d'un dépôt de données en vue d'évaluer la mise en place ainsi que les performances de quelques technologies conçues pour la manipulation des données massives, appelées aussi Big Data. Les données considérées sont des données collectées par des dispositifs appelés sondes Atlas, et l'accès à ces données est publique. La manipulation de ces dernières nécessite l'utilisation des outils convenables dépassant les capacités des outils traditionnels ; l'exemple des bases de données relationnelles. Ainsi, le choix d'utilisation de telle ou telle technologie Big Data dépend de plusieurs critères.

L'objectif du présent mémoire est de montrer la capacité des nouvelles technologies du Big Data à fournir des solutions efficaces capables d'assurer le stockage des données massives et d'effectuer des tâches de traitement sur de grandes quantités de données. Dans notre cas, ce sont des données collectées par les sondes Atlas. Ces données apportant des informations utiles et pertinentes de l'état des réseaux informatiques.

Dans un premier temps, nous avons étudié le projet RIPE Atlas afin de maîtriser le contexte général de données d'une part, et de bien choisir les données de travail. Le choix a été fait sur les données générées par l'utilitaire Traceroute, en particulier, nous allons réutiliser un outil conçu dans le cadre d'un travail basé aussi sur les données collectées par les sondes Atlas. C'est un outil permettant de détecter les anomalies dans les délais des liens dans les réseaux informatiques. Ensuite, nous avons passé en revue les technologies du Big Data disponibles afin de sélectionner celles que l'on puisse évaluer leurs performances et convenances pour les données choisies. À l'issue de cette étape, nous avons abordé l'évaluation des technologies Big Data sélectionnées pour réutiliser l'outil de la détection des anomalies.

Le présent document est organisé sur quatre chapitres. Le premier chapitre présente le projet RIPE Atlas : la présentation des caractéristiques techniques et fonctionnelles des sondes Atlas ainsi qu'une liste non exhaustive des cas d'usage de ces sondes. Le deuxième chapitre reprend l'algorithme de la détection à évaluer par les technologies Big Data. Le troisième chapitre énumère quelques concepts liés au Big Data ainsi qu'une liste non exhaustive des technologies Big Data. Pour finir, le quatrième chapitre aborde l'application de l'outil de détection en pratique pour les technologies sélectionnées.

Chapitre I

RIPE Atlas

I.1 Introduction

Le présent chapitre est une présentation détaillée du projet RIPE Atlas créé par l'organisme RIPE NCC. RIPE Atlas a introduit l'utilisation des dispositifs pour effectuer des mesures des réseaux informatiques dans le monde. Ce chapitre présente dans un premier temps les caractéristiques des sondes Atlas, ensuite, il reprend une liste non exhaustive de quelques outils similaires aux sondes Atlas en matière d'objectifs. Dans la suite, ce chapitre expose quelques limites du système RIPE Atlas. Enfin, ce chapitre liste brièvement quelques travaux basés sur le projet RIPE Atlas illustrant les cas d'usage de ce projet.

I.2 A propos RIPE NCC

Le RIPE NCC est un organisme qui alloue les blocs d'adresses IP et des numéros des Systèmes Autonomes dans l'Europe et une partie de l'Asie, notamment au Moyen-Orient.

Un *Système Autonome*, appelé AS, est un ensemble de réseaux et de routeurs sous la responsabilité d'une même autorité administrative. Chaque Système Autonome est identifié par un code sur 16 bits uniques. Les protocoles qui tournent au sein d'un Système Autonome peuvent être différents.

RIPE NCC assure différents services relatifs à la gestion des réseaux informatiques. Il maintient multiples projets pour un nombre de protocoles comme DNS (DNSMON¹), BGP (Routing Information Service ou RIS²) et d'autres projets et services. En particulier, nous sommes intéressés par projet RIPE Atlas géré aussi par RIPE NCC. L'objectif du projet RIPE Atlas est de déployer des dispositifs dans le monde qui sont capables de collecter des données relatives aux réseaux informatiques.

I.3 Présentation du projet RIPE Atlas

RIPE NCC a créé le projet RIPE Atlas en 2010. Le nombre de sondes déployées est en augmentation constante, sachant qu'elles sont déployées par des volontaires.

1. Source : <https://atlas.ripe.net/dnsmon/>, consultée le 27/12/2018.

2. Source : <https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris>, consultée le 27/12/2018.

I.3.1 Les mesures actives et passives de l'Internet

Il existe plusieurs approches pour analyser l'état d'un réseau informatique. Les deux approches les plus répandues sont : passive et active. L'approche passive fait référence au processus de mesure d'un réseau, sans créer ou modifier le trafic sur ce réseau. L'approche active repose sur l'injection des paquets sur le réseau et surveiller le flux de ce trafic. Cette injection a pour objectif la collecte des données relatives aux performances du réseau en question. Par exemple, la mesure du temps de réponse, le suivi du chemin des paquets, etc.

Les données collectées permettent de surveiller les réseaux pour ensuite proposer des améliorations de l'Internet. Le projet RIPE Atlas est un des outils s'inscrivant dans l'approche active. Ce sont des dispositifs, appelés sondes, hébergés par des volontaires, ils sont distribués et maintenus par RIPE NCC. Les données collectées par ces dispositifs sont disponibles au public sur le dépôt de RIPE Atlas [4].

Actuellement, plus de 10, 000 sondes Atlas sont actives, ces dernières produisent environ 450 millions de mesures par jour, ce qui correspond à 5, 000 résultats par seconde [24].

I.3.2 Généralités sur les sondes Atlas

- Les sondes Atlas mesurent les performances de la couche IP. Une sonde envoie des paquets réels et observe la réponse en temps réel indépendamment des applications en dessus de la couche IP.
- Les sondes Atlas ne sont pas des observatrices des données comme le trafic du routage BGP, ainsi, elles n'observent pas le trafic de leurs hébergeurs.
- Les sondes Atlas se situent dans différents emplacements dans le monde, cette répartition permet de diversifier les mesures (voir les sections des mesures I.3.10 et I.3.12).
- Les sondes Atlas sont déployées volontairement dans une maison, un bureau, un entrepôt de données, etc.
- Les mesures peuvent être lancées à tout moment et pour n'importe quelle période³.
- La participation au projet RIPE Atlas est ouverte à toute personne qui s'y intéresse, cela inclut les résultats de mesures, les outils d'analyse, l'hébergement des sondes elles-mêmes, les travaux, etc.
- RIPE Atlas simule le comportement de la couche IP. Par exemple, avec RIPE Atlas, il est possible de :
 - Suivre l'accessibilité d'une destination⁴ depuis différents emplacements dans le monde et depuis différents réseaux. Car les sondes Atlas sont réparties dans plusieurs pays et déployées dans différents réseaux.
 - Étudier des problèmes du réseau remontés en effectuant des vérifications de connectivité ad-hoc via les mesures effectuées par les sondes Atlas.
 - Tester la connectivité IPv6.
 - Vérifier l'infrastructure DNS.

3. Si le nombre de crédits (voir la section des crédits I.3.11) disponibles le permet et qu'il n'y a pas de dépassement du nombre de mesures autorisé.

4. Une destination représente une adresse IP.

La section I.5 reprend quelques cas d'utilisation et les sujets qu'on peut étudier en se basant sur le projet RIPE Atlas.

I.3.3 Les générations des sondes Atlas

Depuis leur création en 2010, les sondes Atlas ont connu trois générations du matériel : v1, v2 et v3. Le tableau I.1 reprend quelques caractérisations de ces trois générations et la figure I.4 montre le matériel utilisé dans chaque génération.

| | v1 | v2 | v3 |
|------------------------------|-------------------------|-------------------------|-------------------------|
| Matériel informatique | Lantronix XPort Pro [8] | Lantronix XPort Pro [8] | tp-link tl-mr3020 |
| Début d'utilisation | 2010 | 2011 | 2013 |
| Mémoire RAM | 8 Mo | 16 Mo | 32 Mo |
| Mémoire Flash | 16 Mo | 16 Mo | 4 Mo |
| CPU | 32-bit | 32-bit | 32-bit |
| Support du Wi-Fi | Non | Non | oui |
| Support du NAT | oui | oui | oui |
| Vitesses supportées | 10 Mbit/s et 100 Mbit/s | 10 Mbit/s et 100 Mbit/s | 10 Mbit/s et 100 Mbit/s |

TABLE I.1 – Les caractéristiques du matériel des trois générations des sondes Atlas



FIGURE I.1 – Génération 1

FIGURE I.2 – Génération 2

FIGURE I.3 – Génération 3

FIGURE I.4 – Les trois générations des sondes Atlas

Source : <https://atlas.ripe.net/docs/>, consultée le 05/08/2018.

Pour précision, les générations 1 et 2 présentent une très faible consommation d'énergie, cependant, elles ont un temps de redémarrage et coûts de production élevés.

En 2015, plusieurs utilisateurs des sondes Atlas ont montré un intérêt aux sondes virtuelles. Les sondes virtuelles présentent des avantages et aussi des inconvénients. Parmi les avantages, la conception des sondes virtuelles permet d'explorer des emplacements qui sont difficilement accessibles. En effet, cela permet d'étendre le réseau des sondes Atlas. D'autre part, les sondes virtuelles peuvent être installées sans contraintes physiques ou organisationnelles. Parmi les inconvénients, une complexité sera ajoutée au système RIPE Atlas, plus de ressources seront demandées. De plus, il peut y avoir le problème de la qualité de données ; le manque de données peut faire référence à une perte de paquets ou bien la machine qui héberge la sonde n'est plus disponible pour continuer les mesures.

I.3.4 Les versions du firmware des sondes Atlas

En principe, toutes les sondes Atlas collectent la même information, indépendamment de leur version du firmware. On trouve les mêmes attributs⁵ dans toutes les versions sauf de légers changements : ajout d'un ou de plusieurs attributs, la modification des noms des attributs, etc. Pour la simplification, nous donnons un identifiant entier pour chaque version, entre les parenthèses. Cet identifiant sera utilisé dans la suite de ce document.

Il existe plusieurs versions du firmware :

- version 1 est identifié par 1 (1);
- version 4400 est identifiée par une valeur entre 4400 et 4459 (2);
- version 4460 est identifiée par une valeur entre 4460 et 4539 (3);
- version 4540 est identifiée par une valeur entre 4540 et 4569 (4);
- version 4570 est identifiée par une valeur entre 4570 et 4609 (5);
- dernière version du firmware⁶ est 4610 (6).

I.3.5 La connexion des sondes Atlas à Internet

Les générations 1 et 2 des sondes Atlas ont une interface Ethernet (RJ-45). La génération 3 dispose techniquement des capacités Wi-Fi. Cependant, ces sondes ne sont pas suffisamment prêtes au niveau logiciel pour supporter le Wi-Fi.

Une fois la sonde se connecte au port d'Ethernet, elle acquiert une adresse IPv4, un résolveur DNS en utilisant DHCP et la configuration IPv6 via *Router Advertisement*. Ensuite, elle essaie de rejoindre l'infrastructure du RIPE Atlas. Pour ce faire, elle utilise le résolveur DNS et se connecte à l'infrastructure à travers SSH sur le port TCP de sortie 443 comme il est illustré dans la figure I.5. L'architecture du système RIPE Atlas est détaillée dans la section I.3.6.

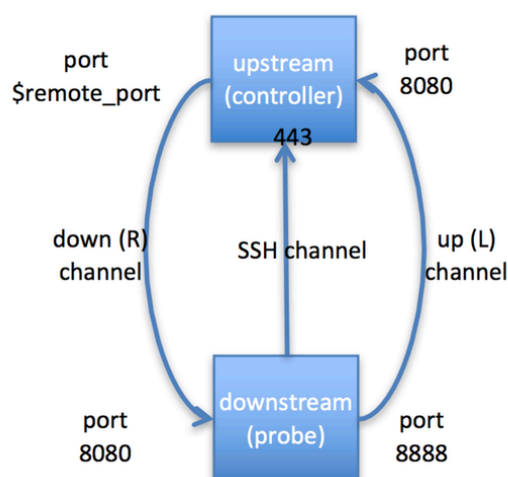


FIGURE I.5 – La connexion des sondes Atlas à l'infrastructure RIPE Atlas [23]

5. Attribut dans le sens du JSON : chaque résultat de mesure est enregistré comme étant un objet JSON.

6. A la date de consultation 25/01/2018.

I.3.6 Architecture du système RIPE Atlas

Il existe deux catégories d'outils de surveillance du réseau : des outils matériels et d'autres logiciels. Les sondes Atlas sont parmi les outils matériels. Le choix d'utilisation d'un outil matériel au lieu d'un outil logiciel dépend de plusieurs facteurs, par exemple l'indépendance du système d'exploitation, la facilité de déploiement, la disponibilité des sondes tout le temps (au lieu d'être dépendante de la machine qui l'héberge) et d'autres facteurs liés à la sécurité.

Le système RIPE Atlas est conçu pour qu'il soit opérationnel de façon distribuée. La plupart des composantes ont assez de connaissances pour remplir leurs rôles, sans nécessairement avoir besoin de connaître les états des autres composantes du système. Cela assure que le système soit capable d'assurer la plupart des fonctionnalités en cas d'un problème temporaire. Par exemple, si une sonde est déconnectée de l'infrastructure, elle continue les mesures planifiées et les données sont renvoyées dès sa reconnexion au système.

La figure I.6 montre une vue d'ensemble de l'architecture du RIPE Atlas.

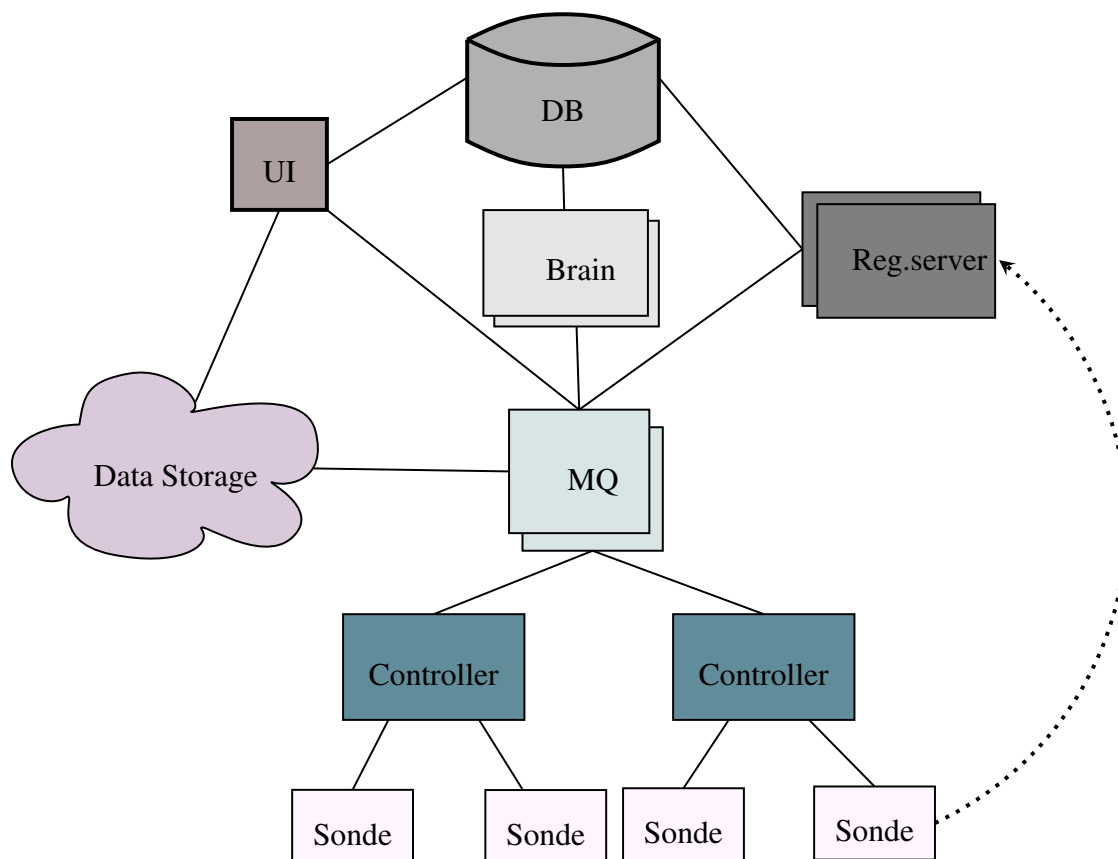


FIGURE I.6 – L'architecture du système RIPE Atlas

Source : Schéma repris du travail de Kisteleki [24]

L'architecture du système RIPE Atlas est constituée par les composantes suivantes :

Registration server (Reg.server) : c'est le seul point d'entrée de confiance pour les sondes Atlas. Son rôle est de recevoir toutes les sondes désirant se connecter au système RIPE Atlas. Ensuite, il redirige chaque sonde vers le contrôleur adéquat, celui le plus proche de la sonde et que ce contrôleur soit suffisamment non occupé. Le serveur d'enregistrement a un aperçu de haut niveau du système.

Controller : un contrôleur accepte d'établir une connexion avec une sonde parmi celles dont il a reçu leurs clés du serveur d'enregistrement (Reg.server). Une fois la connexion est établie entre une sonde et un contrôleur, ce dernier garde cette connexion active pour recevoir les résultats et prévenir la sonde des mesures à effectuer. Le rôle du contrôleur est de communiquer avec les sondes, associer les mesures aux sondes en se basant sur la disponibilité de la sonde et sur autres critères, enfin, collecter les résultats intermédiaires des mesures.

Message Queue (MQ) : Tout d'abord définissons MQ :

« Message Queue ou file d'attente de message : est une technique de programmation utilisée pour la communication interprocessus ou la communication de serveur-à-serveur. Les files d'attente de message permettent le fonctionnement des liaisons asynchrones normalisées entre deux serveurs, c'est-à-dire de canaux de communications tels que l'expéditeur et le récepteur du message ne sont pas contraints de s'attendre l'un l'autre, mais poursuivent chacun l'exécution de leurs tâches^a. »

^a. Source : https://fr.wikipedia.org/wiki/File_d'attente_de_message, consultée le 05/08/2018.

Un cluster de serveurs MQ agit comme un système nerveux central au sein de l'architecture du RIPE Atlas. Il gère la connectivité entre les composantes de l'infrastructure et assure l'échange de messages avec un délai minimal. C'est cette composante qui élimine le besoin que les autres composantes de l'infrastructure soient au courant des états des autres composantes de l'infrastructure. En plus, chaque composante peut être ajoutée ou retirée sans devoir synchroniser cette information à l'infrastructure entière. Si c'est le cas d'une déconnexion d'une composante, les messages seront sauvegardés sur différents niveaux jusqu'au moment de la reconnexion.

UI (User Interface) : elle s'occupe des interactions de l'utilisateur. Elle sert les pages pour l'interface graphique de mesures [3]. Elle traite les appels en provenance de l'API⁷ et sert les demandes de téléchargement en provenance de l'API.

Brain : il effectue des tâches de haut niveau dans le système, notamment la planification des mesures. Cette planification est basée sur les demandes reçues via l'interface graphique web de mesures (UI) ou bien via l'API. La planification passe par la présélection des sondes Atlas et la négociation avec les contrôleurs pour voir la disponibilité des sondes Atlas.

DB : c'est une base de données SQL contenant toutes les informations du système RIPE Atlas : les informations sur les sondes et leurs propriétés, les meta-data des mesures, les utilisateurs, les crédits, etc.

Data Storage : c'est un cluster Hadoop/HBase pour le stockage à long terme de tous les résultats. Cette technologie permet aussi d'effectuer des calculs d'agrégation périodiques et d'autres tâches.

7. Source : <https://atlas.ripe.net/docs/api/v2/manual/>, consultée le 05/08/2018.

Hadoop MapReduce est un modèle de programmation qui permet de traiter les données massives suivant une architecture distribuée dans un cluster.

HBase est une base de données non relationnelle et distribuée. Elle est adaptée au stockage de données massives.

La figure I.7 illustre les étapes d'établissement de la connexion entre une sonde Atlas et l'infrastructure RIPE Atlas.

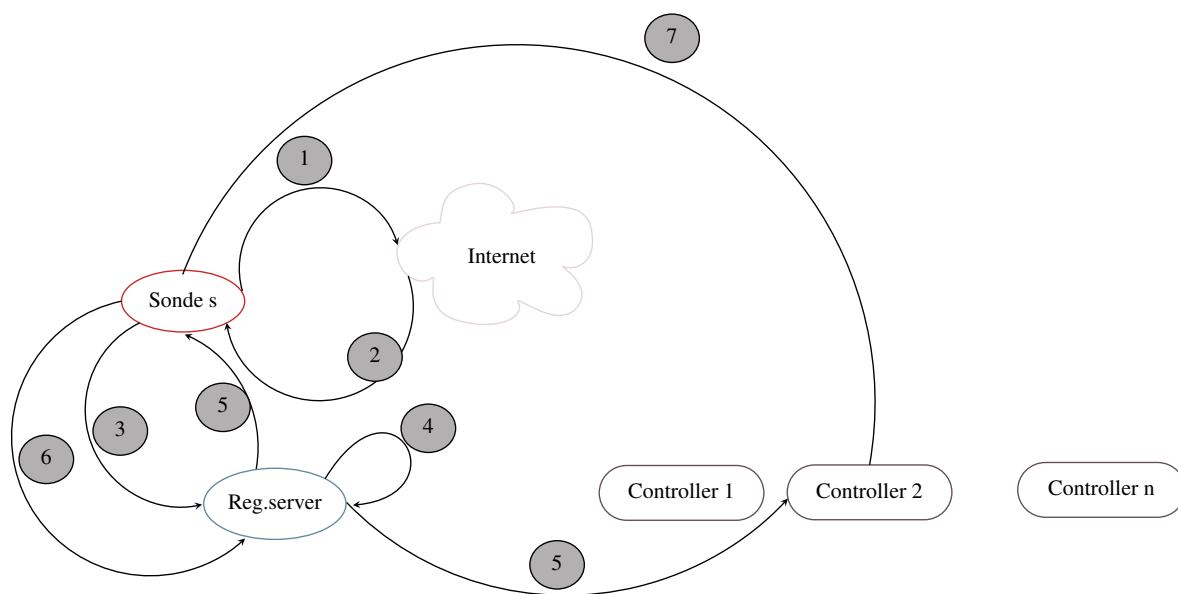


FIGURE I.7 – Les étapes d'établissement d'une connexion entre la sonde Atlas et l'architecture RIPE Atlas

Les étapes suivantes illustrent le déroulement de la connexion d'une sonde Atlas *s* à l'infrastructure RIPE Atlas.

- La sonde Atlas se connecte à Internet via le câble Ethernet RJ45 (1).
- La sonde Atlas acquiert différentes informations : une adresse IPv4, une adresse IPv6 via Router Advertisement et les informations du résolveur DNS via DHCP (2).
- Les informations précédemment acquises permettent à la sonde Atlas de se connecter au serveur d'enregistrement (Reg.server). C'est la première entrée vers l'infrastructure (3).
- En se basant sur la géolocalisation de la sonde Atlas, la charge des différents contrôleurs et d'autres options, le serveur d'enregistrement décide le contrôleur qui va être associé à la sonde Atlas (4).
- Suite à la décision du serveur d'enregistrement, le contrôleur reçoit l'identifiant de la sonde Atlas à gérer et la sonde Atlas reçoit l'identifiant du contrôleur avec à qui elle sera associée (5).
- Une fois l'association entre la sonde Atlas et le contrôleur est faite, la sonde Atlas se déconnecte du serveur d'enregistrement (6).

- La connexion entre la sonde Atlas et le contrôleur est maintenue le plus longtemps possible. Les contrôleurs gardent le contact avec les autres composantes via Message Queue. Dans le cas où une des composantes se déconnecte de l'architecture, les événements sont conservés jusqu'au moment où la connexion est restaurée (7).

La connexion précédemment établie permet aux sondes Atlas d'envoyer leurs rapports de mesures aux serveurs de stockage. C'est la même connexion qui permet de passer les commandes aux sondes pour qu'elles puissent effectuer les mesures et les mises à jour de leur firmware.

I.3.7 Les sondes Atlas et la vie privée

La sonde Atlas n'a pas l'accès au trafic de son hébergeur. Elle maintient sa connexion avec l'infrastructure centrale et elle exécute les mesures planifiées vers les destinations publiques sur Internet.

Les sondes Atlas peuvent révéler l'adresse IP de leur hébergeur. Bien que, les informations personnelles telles que les adresses MAC et les adresses e-mail ne seront jamais affichées. Cependant, l'adresse IPv6 peut exposer l'adresse MAC.

I.3.8 La sécurité dans RIPE Atlas

La connexion entre les composantes de l'infrastructure RIPE Atlas est maintenue le plus longtemps possible comme c'est décrit dans la section I.3.6. De ce fait, la sécurité des différentes connexions est primordiale. Afin de réduire la surface d'attaque contre ces sondes, les précautions suivantes sont prises :

- Les hébergeurs des sondes Atlas ne disposent d'aucun service qui leur permet de se connecter aux sondes (dans le sens de TCP/IP).
- Les sondes Atlas n'échangent aucune clé d'authentification entre elles. En effet, chaque sonde dispose de sa clé qu'elle l'utilise pour se connecter à l'infrastructure.
- Comme les sondes Atlas sont chez les hébergeurs, il est impossible qu'elles soient résilientes au démontage. Cependant, si c'était le cas, cela ne devrait pas affecter les autres sondes Atlas.
- Toutes les communications au sein de l'infrastructure RIPE Atlas se font d'une manière sécurisée. Les connexions entre les composantes sont maintenues grâce aux *secure channels* avec *mutual authentication*.
- Le logiciel qui tourne dans les sondes Atlas peut être facilement mis à niveau ; la sonde Atlas est capable de vérifier l'authenticité d'une nouvelle version du firmware et cela via les signatures cryptographiques.

Le système RIPE Atlas est un système comme les autres, il n'est pas résilient à 100 % aux attaques. Cependant, l'équipe RIPE Atlas propose régulièrement des améliorations et des fixations de bugs surmontées par la communauté RIPE Atlas.

I.3.9 Les ancrs VS sondes Atlas

Les ancrs Atlas sont des dispositifs agissant comme cibles aux différentes mesures lancées par les sondes Atlas. Il est possible de planifier des mesures entre les ancrs RIPE Atlas, ces mesures permettent de vérifier l'état des réseaux qui hébergent ces ancrs. Les ancrs Atlas peuvent être considérées comme cibles aux mesures suivantes :

- Ping.
- Traceroute.
- DNS : les ancrs ont été configurées avec BIND pour qu'elles agissent en tant que serveur DNS faisant autorité.
- HTTP et HTTPS : l'ancre fait tourner un serveur Web, ce dernier utilise un gestionnaire de réponses personnalisé aux requêtes HTTP(S) ayant comme seule option la taille du payload. Cette taille peut prendre une valeur maximale de 4096 et la réponse est fournie sous format JSON. L'exemple d'une requête HTTP avec une taille de 536 depuis une sonde Atlas vers une ancre Atlas est :

```
http://nl-ams-as3333.anchors.atlas.ripe.net/536
```

Les ancrs sont configurées avec un certificat SSL auto-signé en utilisant une clé de 2048 bit et un temps d'expiration de 100 ans. Le tableau I.2 reprend une comparaison de certaines caractéristiques communes entre les sondes et les ancrs Atlas.

| | Sonde Atlas | Ancre Atlas |
|--------------------------|----------------|---------------------------------|
| Mesures originaires de | oui | oui |
| Mesures à destination de | — ⁸ | ping, traceroute, DNS, HTTP(S). |
| Nomination | — | structurée ⁹ |
| Crédit gagnés | N | $10 * N$ |
| Besoin en bande passante | léger | important |
| Coût : gratuite | oui | non ¹⁰ |

TABLE I.2 – Comparaison entre sondes et ancrs RIPE Atlas

I.3.10 Les mesures intégrées : Built-in

Une fois une sonde Atlas connectée, elle lance automatiquement un ensemble de mesures prédéfinies, appelées *Built-in Measurements*. Les mesures personnalisées sont détaillées dans la section I.3.12. Les mesures peuvent être effectuées selon l'adressage IPv4 ou bien IPv6. Le choix du mode IPv4, IPv6 ou les deux, dépend de la capacité du réseau qui héberge la sonde Atlas.

8. — : Non disponible.

9. Exemple de `de-mai-as2857.anchors.atlas.ripe.net` avec la structure suivante : `pays-ville-ASN.anchors.atlas.ripe.net`.

10. Le matériel est au frais de l'hébergeur.

Il existe deux types de mesures : les mesures *One-Off*, ce sont les mesures qui s'exécutent une seule fois. Pour le deuxième type, ce sont les mesures qui s'exécutent périodiquement, à chaque intervalle de temps.

De base, les sondes Atlas assurent les mesures intégrées suivantes :

- Les informations sur la configuration du réseau dans lequel la sonde Atlas est déployée.
- L'historique de la disponibilité de la sonde Atlas.
- Les mesures du RTT (Round Trip Time) par traceroute.
- Les mesures ping vers un nombre de destinations prédéfinies.
- Les mesures traceroute vers un nombre de destinations prédéfinies.
- Les requêtes vers les instances des serveurs DNS (Domain Name System) racines.
- Les requêtes SSL/TLS (Secure Socket Layer/Transport Layer Security) vers un nombre de destinations prédéfinies.
- Les requêtes NTP (Network Time Protocol).

Chaque mesure a un identifiant ID unique. Cet identifiant indique le type de la mesure, s'il s'agit du ping, traceroute ou autres. Plus de détails sur la signification des identifiants des mesures sont disponibles sur RIPE Atlas ¹¹.

En plus des mesures intégrées, les sondes Atlas peuvent effectuer des mesures personnalisées. Ces mesures peuvent être lancées via l'interface web [3] ou bien via HTTP REST API. Toutefois, la planification des mesures personnalisées nécessite l'acquisition de ce qu'on appelle les "crédits" au sens RIPE Atlas.

I.3.11 Le système de crédits Atlas

Le système de crédits RIPE Atlas est une sorte de reconnaissance de la contribution des participants à ce projet. Un hébergeur d'une sonde Atlas reçoit un nombre de crédits en contrepartie de la durée pendant laquelle sa sonde reste connectée. D'autre part, il gagne d'autres crédits suivant les résultats de mesures générés par cette sonde. Les crédits gagnés peuvent être utilisés dans la création des mesures personnalisées, appelées *User Defined Measurements* (voir la section I.3.12). Les personnes ayant gagné des crédits peuvent les transférer vers une autre personne ayant besoin de ces crédits. Les crédits peuvent être obtenus via :

- L'hébergement d'une sonde Atlas ; à chaque utilisation d'une sonde, son hébergeur reçoit un nombre de crédits. La connexion d'une sonde Atlas au système durant une minute apporte 15 crédits.
- L'hébergement d'une ancre Atlas ¹².
- La recommandation à une personne d'héberger une sonde Atlas.
- En étant un sponsor du RIPE NCC. Le parrainage des sondes Atlas est disponible pour les organisations et les individus. Le sponsor reçoit le même nombre de crédits que les hébergeurs de ces sondes.

11. Source : <https://atlas.ripe.net/docs/built-in/>, consultée le 10/08/2018.

12. Les ancres Atlas sont décrites dans la section I.3.9.

- En étant un registre Internet régional (Local Internet Registry).
- La réception des crédits d'une autre personne via un transfert de crédits.

Le lancement des mesures personnalisées exploite les ressources de l'infrastructure RIPE Atlas d'une part, du réseau hôteur de la sonde d'autre part. Par conséquent, les mesures sont organisées afin d'éviter toute surcharge du système. Le coût d'une mesure dépend du type de la mesure et des options spécifiées. Le système calcule le nombre de crédits nécessaires pour effectuer une mesure donnée. Le nombre de crédits est déduit à chaque résultat reçu. Ci-dessous le coût unitaire des différents types de mesures.

Ping et ping6 :

$$\text{Coût unitaire} = N \times (\lfloor \frac{S}{1500} \rfloor + 1)$$

Où N est le nombre de paquets dans le train (par défaut 3) et S est la taille du paquet (par défaut : 48 octets).

DNS et DNS6 :

Coût unitaire pour UDP : 10 crédits/résultat
Coût unitaire pour TCP : 20 crédits/résultat

Traceroute et traceroute6 :

$$\text{Coût unitaire} = 10 \times N \times (\lfloor \frac{S}{1500} \rfloor + 1)$$

Où N est le nombre de paquets dans le train (par défaut 3) et S est la taille du paquet (par défaut : 40 octets).

SSLCert et SSLCert6 :

Coût unitaire = 10 crédits/résultat.

Exemple :

La planification d'une mesure ayant les caractéristiques suivantes nécessite 14,400 crédits.

La fréquence : deux fois par heure
La durée : deux jours (48 heures)
Le nombre de sondes : 5
Type de mesure : *traceroute*

Tel que :

$$\begin{aligned} 5 \times 2 \text{ mesures/heure} \times 48 &= 480 \text{ ligne résultat} \\ 30 \text{ credits/result} \times 480 \text{ results} &= 14,400 \text{ crédits} \end{aligned}$$

I.3.12 Les mesures personnalisées : User Defined measurement

En plus des mesures intégrées, par défaut, dans une sonde Atlas, il est possible de planifier des mesures personnalisées. Ce sont les mêmes types de mesures : ping, traceroute, HTTP Get, SSLCert, DNS, NTP et TLS. Cette planification coûte des crédits, en effet, il faut avoir assez de crédits pour lancer des mesures. L'interface web dédiée à la création d'une nouvelle mesure offre toutes les possibilités comme la personnalisation des éléments suivants :

- le type de la mesure ;
- la sélection des sondes Atlas réalisant la mesure ;
- la fréquence de la mesure et sa durée.

Chaque mesure est suivie via son état. Plusieurs états à distinguer : *specified*, *scheduled*, *ongoing*, *stopped*, *Forced to stop*, *no suitable probes* et enfin *failed*.

I.3.13 La sélection des sondes Atlas

La sélection des sondes Atlas pour effectuer une des mesures repose sur des critères suivants :

- numéro d'AS ;
- zone géographique via l'altitude et la longitude ;
- pays (ou zone géographique comme Europe) ;
- préfixe IP ;
- manuellement, avec les identifiants des sondes Atlas ;
- reprendre celles d'une mesure précédente.

Il existe une autre manière de regrouper les sondes : avec des étiquettes. Le système d'étiquettes sert comme indicateur des propriétés, des capacités, de la topologie du réseau et d'autres classifications des sondes. On distingue les étiquettes système et utilisateur. Chaque nom d'étiquette est lisible par un humain.

Les étiquettes utilisateurs sont associées à une sonde librement par son hébergeur. Les étiquettes système sont attribuées uniquement par l'équipe RIPE Atlas et sont mises à jour périodiquement, à priori chaque 4 heures. Des exemples d'étiquettes système sont disponibles sur RIPE Atlas ¹³.

I.3.14 Les sources de données Atlas

Les sondes Atlas génèrent trois types de données : leurs détails de connexions d'un jour donné, leurs résultats des mesures intégrées et personnalisées et les descriptions des mesures effectuées (meta-data).

Premièrement on trouve les données sur les sondes Atlas par jour. Les détails sur les sondes reprennent les informations des connexions, des réseaux et autres. A priori, les détails des

13. Source : <https://atlas.ripe.net/docs/probe-tags/>, consultée le 23/01/2018.

connexions des sondes sont disponibles pour la période du 13 mars 2014 jusqu'à ce jour¹⁴, un fichier JSON par jour. Les données de certains jours sont manquantes. La totalité des archives se trouve dans [5]. La taille d'une seule archive est entre 120 KB et 921 KB.

Deuxièmement, les résultats des mesures sont aussi archivés dans un serveur FTP. Seules les données des derniers 30 jours sont conservées en archives¹⁵. Les fichiers ont été nommés jusqu'au 15 mars 2018 de façon structurée comme suit :

\$TYPE-\$IPV-\$SUBTYPE-\$DATE.bz2

- \$TYPE peut être traceroute, ping, dns, ntp, http, sslcert.
- \$IPV version du protocole IP v4 ou v6.
- \$DATE date au format YEAR-MONTH-DAY. (exemple : 2017-06-13)
- \$SUBTYPE type de mesure builtin ou udm.

En considérant toutes les possibilités des types, la quantité de données générées quotidiennement est environ 25 Go¹⁶ et la taille des archives est entre 281M et 3.2G.

Depuis 15 mars 2018, les résultats des mesures sont regroupés différemment. 24 archives par jour, une seule archive pour chaque heure et type de mesure. L'archive ne distingue pas entre mesures IPv4 et IPv6, entre mesures intégrées et personnalisées. Il existe un attribut "af" qui distingue entre IPv4 et IPv6 et l'identifiant de la mesure pour distinguer les mesures intégrées et celles personnalisées (identifiant > 1, 000, 000).

Streaming API propose un service de récupération des résultats de mesures en temps réel, depuis les sondes publiques. Ainsi, elle fournit continuellement de nouveaux résultats en temps réel, obtenus par les sondes Atlas publiques, via une connexion de type HTTPS web-socket active tout le temps.

Troisièmement, on trouve des archives sauvegardées chaque semaine, elles décrivent les méta-datas des mesures. Une ligne objet JSON pour chaque mesure publique. Au moment de la consultation, la taille de chaque archive était entre 124 Mo et 1.5 Go. L'accès à ces données se fait de deux façons, via le téléchargement direct depuis un serveur FTP ou bien via streaming API. Les noms des archives sont bien structurés.

I.3.15 Les limitations du RIPE Atlas

De nombreux travaux ayant exploité les données générées par les sondes Atlas. Néanmoins, ce système connaît des bugs et des limitations. Les membres de la communauté RIPE Atlas s'engagent à remonter les bogues liées aux sondes Atlas. Tous les bogues sont répertoriés dans une rubrique dédiée [6].

RIPE Atlas connaît des limitations liées à la visualisation. Actuellement, RIPE Atlas supporte la visualisation des mesures de type ping ayant utilisé au maximum 20 sondes. Cette limitation concerne aussi le type traceroute, en effet, il est possible de visualiser seulement les mesures IPv6 built-in.

14. 15/08/2018.

15. Source : <https://data-store.ripe.net/datasets/atlas-daily-dumps/>, consultée le 05/04/2018.

16. Source : <https://ftp.ripe.net/ripe/atlas/data/README>, consultée le 26/03/2018.

Afin d'éviter la surcharge des sondes et de l'infrastructure, RIPE Atlas a limité le nombre de mesures périodiques de 10 à la fois et de 10 mesures de type one-off vers n'importe quelle cible à un moment donné. De plus, il n'est pas possible d'utiliser plus de 500 sondes par mesure.

Pour les mesures one-off (non périodiques), une sonde peut effectuer au plus 10 mesures en parallèle. RIPE Atlas limite aussi la fréquence des mesures personnalisées. Un hébergeur d'une sonde peut effectuer :

- Ping chaque 60 secondes (par défaut 240 secondes) ;
- Traceroute chaque 60 secondes (par défaut 900 secondes) ;
- SSL chaque 60 secondes (par défaut 900 secondes) ;
- DNS chaque 60 secondes (par défaut 240 secondes).

Dans le cas d'une déconnexion, la sonde continue à effectuer les mesures. Pour la version 1 et 2, la sonde est capable de sauvegarder les 6 dernières heures de données. Tandis qu'avec la version 3, une sonde est capable de sauvegarder les résultats de plusieurs mois. Une fois la sonde est connectée, elle envoie les données à l'infrastructure centrale.

Concernant la consommation des crédits par jour, RIPE Atlas limite cette consommation à 1,000,000 crédits.

I.3.16 Confiance aux données Atlas

De nombreux travaux ont exploité les données Atlas, cependant, peut-on faire confiance à la qualité de ces données ? les données sont-elles complètes ?

La question de la complétude des données est plus présente pour les mesures périodiques, celles qui se déroulent pendant une durée d et à un intervalle i . W. Shao et al. [33] ont traité les mesures manquantes. L'approche qu'ils ont adopté repose sur la corrélation entre l'absence de certaines mesures et les périodes durant lesquelles les sondes Atlas ont été déconnectées. Pour précision, RIPE Atlas maintient les détails des connexions/déconnexions des sondes Atlas. W. Shao et al. [33] ont étudié les mesures en provenance des sondes v3, effectuées entre le 01/06/2016 et le 01/07/2016 (UTC). Les auteurs ont combiné les informations relatives à la connexion/déconnexion des sondes et leurs mesures planifiées, leur approche se base sur l'attribut *timestamp* qui est présent dans chaque résultat de mesure et dans les états de connexions.

Nous avons discuté des limitations du RIPE Atlas en terme de mesures autorisées par jour. Cela n'empêche qu'il est possible qu'un nombre important de mesures soit effectué. De plus, plus d'un utilisateur peut s'intéresser à la même sonde Atlas. C'est la question traitée dans le travail de T. Holterbach et al. dans [21] : si les mesures lancées par les autres utilisateurs affectent les résultats obtenus par un autre utilisateur, et si c'est le cas, comment s'y entreprendre. Les expériences réalisées ont montré la présence de l'interférence entre les mesures à destination des sondes et cela de deux manières. Premièrement, les mesures depuis et à destination des sondes Atlas augmentent le temps reporté par la sonde et ils ont conclu que l'amélioration du CPU a permis de limiter les interférences sur le temps mesuré par les sondes Atlas. Deuxièmement, ils ont conclu que les mesures perdent la synchronisation avec l'infrastructure d'Atlas, pendant plus d'une heure, à cause de la charge concurrentielle que subit le système d'Atlas. Dans ce cas, l'amélioration du matériel ne peut pas résoudre le problème.

I.4 Projets existants de mesures d’Internet

Dans les sections précédentes, nous avons développé le projet RIPE Atlas comme étant une plateforme pour la collecte de données des réseaux informatiques. Toutefois, il existe d’autres projets similaires à RIPE Atlas. Les sections suivantes reprennent une liste non exhaustive des projets similaires à RIPE Atlas.

I.4.1 Test Traffic Measurement Service

Avant l’arrivée du RIPE Atlas, Le RIPE NCC (Réseaux IP Européens Network Coordination Centre) a assuré la mesure de la connectivité entre les réseaux via d’autres plateformes, comme la plateforme Test Traffic Measurement Service (TTM). Il s’agit d’un projet qui permet de mesurer la connectivité entre un nœud source et un nœud destination sur Internet. C’était une des manières pour suivre la connectivité entre le réseau source et le réseau destination.

L’idée était la mise en place d’un dispositif, test-box, qui génère du trafic. Ce dernier n’affecte pas l’infrastructure réseau en matière de bande passante. De plus, il n’a pas l’accès aux données du réseau dans lequel il est mis en place.

Ce service a été assuré et géré, pendant une période de 6 ans, par une équipe au sein du RIPE NCC. Les fonctionnalités assurées par ce service étaient de tester l’accessibilité à une destination via le *ping*, ainsi, les mesures effectuées étaient indépendantes des applications, elles dépendaient du réseau lui-même. RIPE NCC a arrêté la maintenance du TTM depuis le 1 juillet 2014 [29].

I.4.2 ProbeAPI

ProbeAPI [35] est une plateforme de mesure de l’état d’un réseau, cette plateforme couvre 170 pays et des milliers d’ISPs. *ProbeAPI* est utilisée par les développeurs, les administrateurs des réseaux et les chercheurs, ils peuvent lancer des mesures d’un réseau depuis différents réseaux.

Le logiciel *ProbeAPI* s’exécute dans plusieurs systèmes : dans des ordinateurs (Win32/64), Android via une installation dans les mobiles et les tablettes et dans des routeurs au sein du DD-WRT.

DD-WRT est un micrologiciel libre et gratuit, il est destiné aux routeurs sans fil et aux points d’accès. Il fonctionne avec un système d’exploitation Linux. Le rôle du DD-WRT est de remplacer le micrologiciel intégré aux routeurs par leurs fabricants. Ainsi, il est possible d’étendre des fonctionnalités du routeur en ajoutant d’autres fonctions supplémentaires.

ProbeAPI s’agit d’un logiciel qui tourne dans la machine de l’hébergeur. C’est pourquoi le suivi des réseaux dépend de la disponibilité de la machine qui le fait tourner. Cette dépendance affecte la disponibilité de la sonde logicielle, sa configuration et aussi les résultats de mesures.

Une étude comparative [36] entre les sondes Atlas et les sondes *ProbeAPI* est résumée dans le tableau I.3. En fin de cette étude, ils concluent qu’en comparant les résultats des mesures ICMP effectuées par les deux plateformes, des contrastes intéressantes ont été constatées. Les sondes Atlas ont montré un comportement stable lors de la réalisation des mesures, les résultats sont peu variables car les sondes sont indépendantes de l’utilisateur. Cependant, il était constaté qu’une forte variabilité au cours du temps pour les sondes logicielles (*ProbeAPI*), car elles dépendent fortement de l’hébergeur ; sa configuration réseau, sa disponibilité, etc.

Enfin, la force des sondes logicielles comme ProbeAPI réside dans sa capacité à effectuer des mesures depuis la couche application, la plus proche de l'utilisateur. L'exemple de l'évaluation du Time To First Byte et le taux de transfert dans deux pays.

« *Le **Time to First Byte (TTFB)** est le temps de chargement du premier octet, c'est la mesure qui nous permet d'évaluer la vitesse d'accès à un serveur. Plus la mesure est basse et plus le serveur commencera à servir les ressources rapidement.* »^a

a. Source : <https://www.skyminds.net/calculer-le-time-to-first-byte-ttfb-dun-serveur/>, consultée le 10/08/2018.

| RIPE ATLAS | PROBEAPI |
|---|---|
| Matériel homogène a un comportement prévisible | Matériel hétérogène a un comportement imprévisible |
| Connexions stables vu l'indépendance du software utilisateur | Connexions instables vu la dépendance du software utilisateur |
| Indépendance de l'OS et ses limitations ou vulnérabilités | Liaison à l'OS et ses limitations ou vulnérabilités, cependant utile pour les mesures au niveau application |
| La distribution des sondes est coûteuse, difficile de couvrir certaines régions | Mise en place du logiciel est rapide et moins chère, avec facilité de couvrir plusieurs régions |
| Les mesures HTTP se limitent aux ancres pour des raisons de sécurité | HttpGet, DNS et page-load sont disponibles via des librairies Mozilla et chromium, et ce pour toutes les destinations |

TABLE I.3 – Comparaison entre sondes Atlas et ProbeAPI

Malgré le niveau de couverture assuré par ProbeAPI, cependant ces sondes se connectent et se déconnectent fréquemment, ce qui montre une forte volatilité. Cette volatilité est liée à la dépendance des sondes ProbeAPI de leur hébergeur ; tant qu'il est connecté, la sonde ProbeAPI est prête pour effectuer les mesures. Toutefois, si l'hébergeur est déconnecté, la sonde ProbeAPI ne peut pas effectuer des mesures, d'où le basculement fréquent entre les deux états : connectée et déconnectée.

I.4.3 Archipelago

Archipelago (Ark) [1] est l'infrastructure de mesures actives du CAIDA [2]. Elle est au service des chercheurs en réseau depuis 2007. L'objectif de ce projet est de couvrir un maximum de régions afin de collecter un maximum de mesures. Ensuite, produire des visualisations qui améliorent la vue globale de l'Internet. Pour précision, c'est un Raspberry Pi 2nd gen.

I.4.4 DIMES

DIMES [34] est un logiciel qui devrait être installé dans une machine. Une fois installé, il fonctionne de sorte que la consommation d'énergie soit minimale et qu'il n'existe aucun impact sur les performances de la machine ou sur la connexion. L'objectif de *DIMES* est de collecter un maximum de données afin d'explorer la topologie d'Internet.

I.4.5 SamKnows

SamKnows [7] est une plateforme globale des performances d'Internet, elle regroupe les ISPs, ingénieurs, universitaires, codeurs et des organismes de régulation. Son objectif est d'évaluer les performances du haut débit des utilisateurs finaux et de trouver les problèmes avant que les clients ne commencent à se plaindre.

I.5 Quelques cas d'utilisation des données collectées par les sondes Atlas

Plusieurs travaux ont exploités les données collectées par les sondes Atlas. Ces travaux peuvent être classés de plusieurs manières, par exemple par thème, par type de mesures utilisé, etc. Nous distinguons les travaux ayant exploité les données collectées par les sondes Atlas à travers les mesures *built-in* ou bien ceux ayant utilisé les données des mesures personnalisées. Pour les premiers, ils permettent d'exploiter au mieux ces données sans surcharger le système RIPE Atlas, car ces données sont collectées, par défaut, quotidiennement. Tandis que les autres peuvent introduire une charge supplémentaire sur ces sondes. D'autre part, certains travaux ont manipulés aux données par type de mesure : traceroute, ping, HTTP, etc. Nous allons présenter brièvement quelques travaux par thème.

I.5.1 Détection des coupures d'Internet

Les données collectées par les sondes Atlas ont permis de valider certaines coupures d'Internet. Par exemple, la coupure concernant le point d'échange AMS-IX (Amsterdam Internet Exchange). En 2015, R. Kisteleki et al. [22] ont évalué l'état des pings en provenance des sondes Atlas à destination de trois ancres Atlas qui se trouvent dans AMS-IX. En effet, peu de pings ont réussi à atteindre leurs destinations, cependant, certains pings n'ont pas réussi à le faire. Ils ont conclu qu'il existe un problème du réseau, et le problème concerne les ancres plutôt que les sondes ayant lancé le ping. De même pour DNS, ils ont constaté l'absence des données DNS sensées être collectées par les ancres Atlas à destination du K-root pendant la période de la coupure.

I.5.2 Aide à la prise de décision

L'utilisation des sondes Atlas n'est pas limitée au domaine de recherche seulement, elle a permis aussi d'aider à la prise de décision pour certaines implantations et pour la mise en place des équipements comme les routeurs, les data-centers, les IXPs, etc.

Les ingénieurs de *Wikimedia Foundation* et du RIPE NCC ont collaboré dans un projet [10] pour étudier la latence vers les sites du Wikimedia. L'idée était d'exploiter la distribution des sondes Atlas dans le monde en vue de mesurer la latence vers les sites du Wikimedia. L'étude de la latence va permettre d'améliorer l'expérience des utilisateurs vers ces sites en réduisant la latence. Comme Wikimedia avait l'intention d'étendre son réseau de datacenters, ils ont profité des résultats de cette étude pour choisir les futurs emplacements de leurs data-centers.

Un groupe de chercheurs africains a évalué le routage inter-domaine afin d'étudier les emplacements adéquats pour la mise en place d'un IXP [30]. Après avoir analysé les données des mesures collectées par les sondes Atlas, ils ont constaté que le trafic de et à destination de l'Afrique quitte le continent vers les États-Unis ou bien l'Europe pour revenir en Afrique, d'où l'intérêt d'investir dans la mise en place des IXPs dans ce continent.

I.5.3 Le suivi des censures

En 2014, des chercheurs ont examiné les incidents de type content-blocking en Turquie et en Russie tout en prenant en considération le respect de l'aspect éthique des données. Ils ont aussi élaboré un aperçu comparatif des différents outils permettant de mesurer les réseaux [13]. C. Anderson et al. ont repris deux cas d'études où une censure a été appliquée : la Turquie et la Russie. L'idée de C. Anderson et al. est de créer des méthodes pour analyser ces censures en se basant sur les données collectées par les sondes Atlas.

Il existe plusieurs pratiques pour appliquer la censure. Ces pratiques dépendent des objectifs de cette censure ; bloquer un site web, rediriger le trafic, filtrer l'accès à travers des mots clés, etc.

En mars 2014, des utilisateurs turcs ont été interdits d'accéder au réseau social *Twitter*. Ce filtrage a été fait en utilisant *DNS Tampering* et *IP Blocking*. Comme ces deux pratiques sont évaluables avec les sondes Atlas, ils ont planifié des mesures vers plusieurs destinations et depuis un nombre de sondes. Ces mesures sont reprises en détail dans le tableau I.4.

| Cible | Type | Sondes | Fréquence (s) | Crédits |
|---------|---------|--------|---------------|---------|
| Twitter | SSL | 10 | 3,600 | 2,400 |
| YouTube | SSL | 10 | 3,600 | 2,400 |
| Tor | SSL | 10 | 3,600 | 2,400 |
| Twitter | DNS (U) | 10 | 3,600 | 2,400 |
| YouTube | DNS (U) | 10 | 3,600 | 2,400 |
| Twitter | Tracert | 10 | 3,600 | 7,200 |

TABLE I.4 – Les détails des mesures effectuées dans le travail de C. Anderson [13]

L'analyse de données obtenues a permis de détecter six changements concernant les décisions du filtrage. Plus de détails se trouvent dans [13].

Quant à la Russie, les autorités ont décidé de mettre le blog d'*Alexei Navalny* sur *LiveJournal* dans la liste noire. En même temps, certains médias indépendants ont été aussi filtrés, l'exemple du site *grani.ru*. Pour le site *Grani*, les sondes Atlas ont reçu des réponses DNS aberrantes, d'où l'impossibilité de joindre *grani.ru*. Cependant, le filtrage du site *navalny.livejournal.com* a pris une autre forme, c'était une redirection d'adresse IP. La réponse d'une requête vers ce site donne 208.93.0.190 au lieu de 208.93.0.150. Ces deux adresses sont inclut dans le préfixe 208.93.0.0/22 géré par *LiveJournal Inc*. 208.93.0.190 correspond au contenu non-blacklisted, alors que 208.93.0.150 correspond au contenu correct.

I.5.4 Le suivi des performances d'un réseau

Les ancres Atlas

Les ancres Atlas ont des capacités avancées que les sondes Atlas. Les ancres servent comme cibles aux mesures des sondes. De plus, elles sont capables de fournir des détails sur l'état du réseau dans lequel elles sont déployées. S. Gasmi, un hébergeur d'une ancre Atlas, a développé un outil disponible au public¹⁷. A partir des données collectées par les ancres Atlas, cet outil permet d'analyser la qualité de la connectivité d'un réseau (ou d'un AS) et permet de suivre les changements relatifs à la topologie des réseaux.

Par exemple, il a constaté que la vérification du BGP Prepending et des communautés BGP peut être faite en considérant les éléments suivants : adresse IP source, AS source, pays, le RTT

17. Source : <http://ripeanchor.sdv.fr/>, consultée le 08/08/2018.

du ping et les chemins du traceroute. En particulier, S. Gasmi a évalué deux corrélations. Dans un premier temps, il a visualisé la corrélation entre l'AS path et Round Trip Time (RTT). Il a regroupé des sondes par pays, ensuite, il a calculé, par ce pays, la moyenne du nombre de sauts et la moyenne du RTT des requêtes à destination de l'ancre depuis ces sondes Atlas. La figure I.8 reprend les résultats obtenus. Aucun renseignement sur la période des données. Pour les sondes en provenance de la France, le nombre de sauts et le RTT entre les sondes déployées en France sont faibles car l'ancre (la cible) se trouve aussi en France.

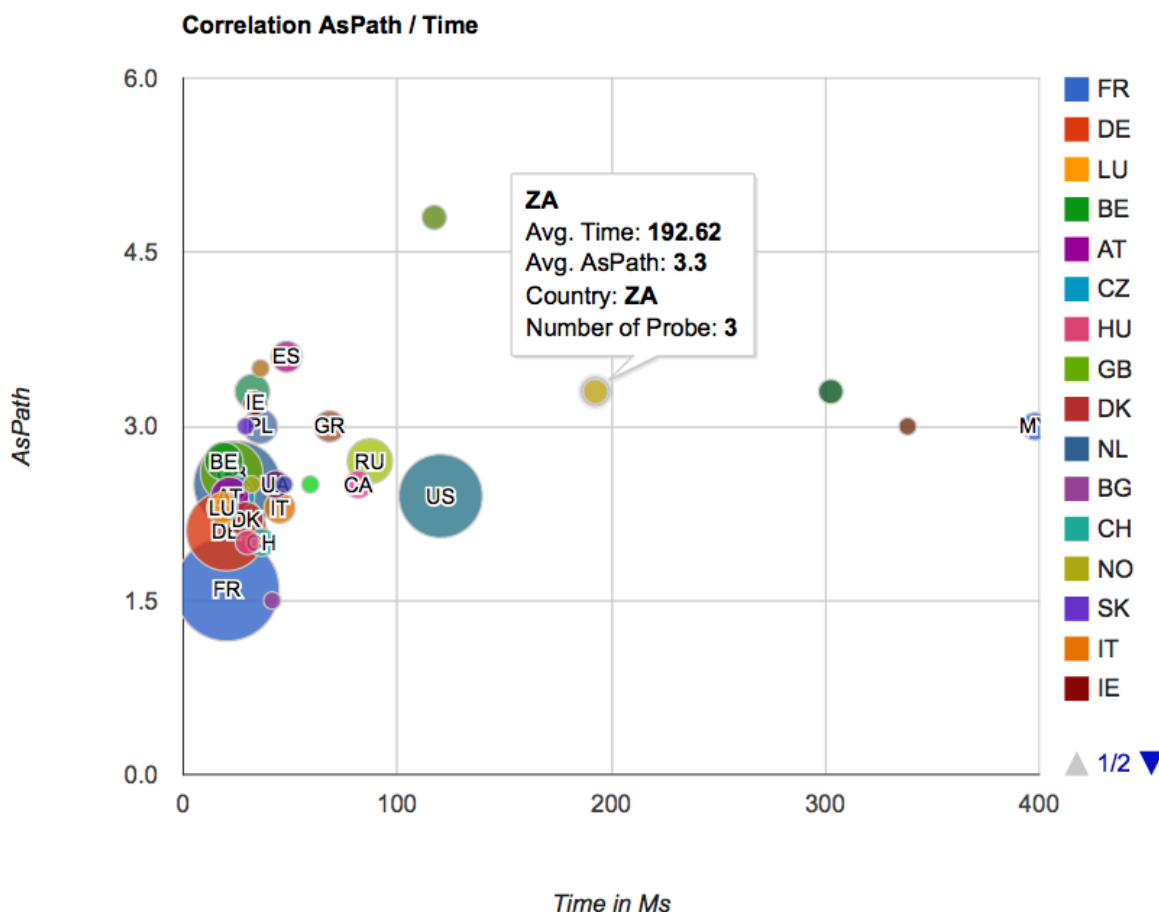


FIGURE I.8 – La corrélation entre la moyenne des AS paths et la moyenne des RTTs [17]

Ensuite, S. Gasmi a mesuré le RTT entre des sondes Atlas dans le monde et son ancre, il a aussi visualisé le nombre de sauts parcourus entre des sondes Atlas à travers le monde et son ancre. Ces deux visualisations permettent d'avoir une idée sur la latence entre certains pays et le pays de l'ancre en question. Plus de détails sur l'approche sont disponibles dans [17].

La vérification de la cohérence du Traceroute

Les chemins parcourus par traceroute pour aller d'une source s vers une destination d changent au cours du temps pour plusieurs raisons. Par exemple, suite à un changement BGP, à une répartition des charges, à des pannes des routeurs, à des pannes des liens physiques, etc.

Traceroute Consistency Check peut reprendre les chemins obtenus via traceroute au cours du temps. L'objectif est de suivre les nœuds apparaissant dans le chemin allant de s à d aux instants t , $t + 1$, $t + 2$, etc, et cela afin de voir les nœuds traversés plus fréquemment au cours du temps. Le chemin est mis à jour via Atlas streaming API.

L'outil proposé dessine les chemins traceroute comme étant un graphe dirigé, chaque nœud est coloré suivant sa cohérence. Le code source du projet est disponible sur GitHub [14]. La figure I.9 présente un exemple de la visualisation proposée. Ce résultat concerne la mesure 1663314¹⁸. Ce sont des traceroutes à destination de l'adresse 213.171.160.1, effectués entre le 02/05/2014 13 : 00 et le 03/05/2014 15 : 00.

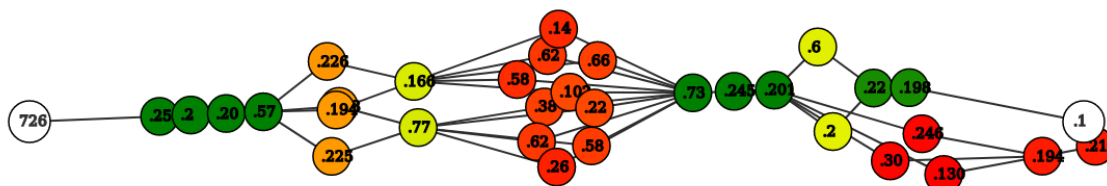


FIGURE I.9 – Visualisation des changements des chemins traceroute [14]

BGP+traceroute

C'est une combinaison des données BGP (RIPE RIS) et traceroute (RIPE Atlas). L'objectif de ce projet est de partir d'un AS path pour enfin géolocaliser les ASs. L'idée est de prendre un AS path des données RIPE RIS, puis, récupérer le préfixe (bloc d'adresses IP) annoncé via cet AS path, ensuite, lancer un traceroute vers une des adresses du bloc. Enfin, géolocaliser les ASs via les données du traceroute. Le code source et la présentation de ce projet sont disponibles GitHub [20, 19].

BGP Atlas Monitor (BAM)

Le projet *BAM* vise la visualisation, en temps réel, des informations utiles pour les opérateurs des réseaux. Par exemple, BAM montre la visibilité des préfixes obtenus par RIPE RIS. De plus, il est possible de voir le délai du *ping* obtenu via les sondes Atlas. Le code source est disponible sur GitHub [18]. En fournissant un ASN (identifiant d'un AS), *BAM* récupère les préfixes IPv4 et IPv6 et leur visibilité et il montre aussi les sondes dans cet AS. L'outil offre les fonctionnalités suivantes :

- les préfixes annoncés par un ASN ;
- la visibilité d'un ASN ;
- la visibilité d'un préfixe ;
- la liste des sondes par AS ;
- les objets route des préfixes.

18. Source : <https://atlas.ripe.net/measurements/1663314/>, consultée le 05/08/2018.

Prédiction des routeurs provoquant la perte des paquets

Dans l'étude [16], R. Fontugne et al. ont modélisé le comportement des routeurs, ils ont développé un modèle qui permet d'estimer l'endroit de la perte des paquets. A partir des trace-routes passant par un routeur r à une destination d , ils ont construit un modèle de forwarding pour ce routeur. Ce modèle reprend les prochains sauts (routeurs) et la fréquence de passage par ces derniers. Si le routeur r change le prochain saut qui a eu "l'habitude" de traverser pour atteindre d , alors cela pour être un indicateur de l'origine de la perte des paquets.

I.5.5 Le suivi des détours dans un trafic local

Dans leur travail [12], E. Aben et al. avaient l'objectif de voir comment les mesures du RIPE Atlas peuvent fournir un aperçu sur le chemin du trafic local à un pays. Précisément si ce trafic traverse un autre pays en revenant au pays du départ. Ce qui pourrait aider à améliorer les performances et l'efficacité des IXPs. L'objectif est d'analyser les chemins identifiés dans le trafic d'Internet entre les sondes Atlas dans un pays donné et essayer d'identifier si le trafic traverse les IXPs.

France-IX est un point d'échange Internet (IXP) français créé en juin 2010. Afin d'apprendre la topologie de routage, un RIS route collector (RRC21) a été installé au sein du France-IX. Actuellement, la France compte 755 sondes Atlas et 9 ancrés. Une ancre sur les 9 est installée au sein de France-IX.

Une des questions posées c'est de savoir si le trafic local de la France reste localement. Les sondes Atlas ne permettent pas de mesurer le trafic entre deux points, alors qu'elles permettent de calculer le chemin entre deux points, adresses IP, ce qui permet d'inférer les sauts par lesquels il passe le trafic. Le travail [11] s'intéresse au trafic depuis et vers une sonde en France en se basant sur l'étude [12].

Les résultats obtenus de l'analyse des détours peuvent être intéressants pour les opérateurs des réseaux afin d'améliorer leurs services, ainsi intéressants pour les IXPs tels qu'ils peuvent proposer des services de peering dans les endroits où il le faut.

I.5.6 Visualisation : indicateurs et dashboard

L'objectif de certains travaux est d'exploiter les données collectées par les sondes Atlas pour concevoir des tableaux des indicateurs. Par exemple, à partir des données de connexion/déconnexion des sondes Atlas, on peut visualiser les sondes connectées, déconnectées, abandonnées. Un autre projet avait comme objectif la reconstruction d'un graphe reprenant les routeurs (nœuds) impliqués dans certains traceroutes, ainsi, identifier les nœuds les plus traversés. D'autres travaux ont repris les détails de la latence, essentiellement, sont les valeurs des RTTs dans les pings et les traceroutes qui permettent de visualiser ce type d'information.

La liste des travaux basés sur le projet RIPE Atlas est très longue. Nous avons essayé d'énumérer quelques projets, les classer par thèmes, toutefois, ce n'est pas un classement unique, tel qu'on peut retrouver un travail dans plus d'une catégorie, ou bien les classer par un autre classement.

I.6 Conclusion

Dans ce chapitre, nous avons présenté les sondes Atlas et leur fonctionnement, ainsi que quelques travaux ayant impliqué des données collectées par ces sondes. Ces travaux ont visé différents thèmes, tels que la prise de décision, le suivi des censures, la conception des tableaux

de visualisation, etc. Dans le reste de ce document, le choix a été porté sur un sujet lié aux performances des réseaux informatiques. Il s'agit d'un outil de détection des anomalies dans les délais des liens dans les réseaux informatiques, en se basant sur les mesures de traceroute. Une présentation détaillée de l'algorithme de détection est donnée dans le chapitre II.

Chapitre II

La détection des anomalies dans les délais d'un lien

Dans le présent chapitre, nous présentons l'outil de détection des anomalies dans les délais d'un lien. Cet outil a été conçu dans le cadre du travail de Fontugne et al. [16]. Nous avons choisi de réutiliser ce travail qui se base sur les traceroutes collectés par les sondes Atlas, et ce en vue de l'évaluer avec quelques technologies Big Data.

II.1 Présentation générale

Dans leur travail [16], Fontugne et al. ont exploité la distribution répandue des sondes Atlas dans le monde afin d'étudier un des problèmes relatifs aux performances des réseaux informatiques.

Le travail de Fontugne et al. [16] se base principalement sur une des mesures effectuées par les sondes Atlas : la requête traceroute, et ce dans le but d'analyser le délais d'un lien topologique sur les réseaux informatiques. L'idée de ce travail est de collecter les résultats des requêtes traceroute effectuées par les sondes Atlas, d'en déterminer une valeur de référence du délai du lien en question sur base de l'historique, et ensuite de comparer la référence avec la valeur courante. Cette référence est mise à jour au fur et au mesure de l'analyse de nouveaux traceroutes. La comparaison de la référence avec la valeur courante ne se déclenche qu'à partir d'avoir une référence assez représentable de l'état réel du lien. A savoir que le déclenchement de cette comparaison est personnalisable.

En pratique, il est difficile d'avoir une idée globale et exacte sur la topologie d'Internet. Toutefois, les opérateurs des réseaux informatiques disposent d'un aperçu de l'état des entités qui forment leurs réseaux, les relations entre ces entités ainsi que les éventuels problèmes. Avec la distribution abondante des sondes Atlas dans le monde en terme de type d'adressage : sondes Atlas supportant seulement l'adressage IPv4, d'autres qui supportent en plus l'adressage IPv6, en terme de la diversité géographique, la diversité en terme d'ASs hébergeant les sondes Atlas, etc, il était possible d'aborder les délais dans les réseaux informatiques à travers de nouvelles approches, reposées sur des fondements statistiques. Parmi les points forts de l'analyse menée par Fontugne et al., c'est la possibilité de valider les méthodes proposées avec des événements marquants sur Internet.

Le travail de Fontugne et al. [16] reprend trois méthodes basées sur les traceroutes collectées par les sondes Atlas :

1. la détection des changements des délais que subissent les liens intermédiaires dans les

traceroutes ;

2. la conception d'un modèle de forwarding pour un routeur donné. Ce modèle prédit l'acheminement du trafic afin d'identifier les routeurs et les liens en panne dans le cas d'un problème de perte de paquets ;
3. la création d'un score par Système Autonome afin d'évaluer l'état de ce dernier.

Dans la suite de ce travail, nous reprenons seulement la première méthode. Il s'agit d'étudier le délai d'un lien topologique ; c'est le délai entre deux routeurs adjacents sur Internet. Nous aurions aimé avoir suffisamment de temps pour pouvoir aborder les autres méthodes.

II.2 Pourquoi analyser les délais des liens réseaux

Il existe un bon nombre de sujets à traiter en exploitant les données collectées par les sondes Atlas. Quelques exemples de sujets ont été déjà présentés dans la section I.5. Notre choix de reprendre le travail de Fontugne et al. a été fait après avoir parcouru un ensemble de travaux basés sur le projet RIPE Atlas.

Pour certains travaux, il n'était pas possible de les reprendre suite au manque de détails et de l'accès à certaines ressources réseaux utilisées dans ces travaux¹. Par exemple, afin d'étudier la censure dans un pays donné, il faut être au courant des spécificités de ce pays, les circonstances politiques et sociales, les opposants, etc. Un autre exemple concerne le sujet de la détection des détours du trafic local dans un pays. Anant Shah et al. ont travaillé sur la détection des détours [32] où ils ont présenté les contraintes relatives à ce sujet : la difficulté d'avoir une géolocalisation exacte d'une adresse IP d'une part et l'absence des informations de peering entre les ASs d'autre part. Ces dernières peuvent changer complètement les conclusions finales.

Le travail sur lequel nous nous basons [16] pour l'évaluation de quelques technologies Big Data s'inscrit dans les travaux traitant les performances des réseaux. Dans la suite de ce document, ce travail est appelé travail de référence. Ce travail a été choisi comme référence pour plusieurs raisons :

Les données utilisées Les auteurs de ce travail ont exploité des données déjà présentes dans la base de données du RIPE Atlas. Ainsi, il n'y a pas besoin de lancer des mesures qui nécessitent la possession d'assez de crédits². De plus, les mesures intégrées montrent plus de stabilité par rapport aux mesures personnalisées. Les destinations des mesures intégrées sont prédéfinies, généralement ce sont des instances des serveurs DNS et des serveurs gérés par RIPE NCC. Alors que, les mesures personnalisées peuvent concerner des destinations moins stables en terme de disponibilité.

La clarté du travail La communauté RIPE Atlas est active en nombre de travaux publiés. Dans certains cas, ces travaux reprennent seulement les résultats. Pour certains, la méthodologie a été bien détaillée. Pour d'autres, la méthodologie adoptée était très brève. En ce qui concerne le travail choisi, la méthodologie est bien détaillée.

1. Par exemple, les informations du peering entre les Systèmes Autonomes, la géolocalisation des adresses IP, etc.

2. Voir la section I.3.11.

La disponibilité du code source La détection des anomalies proposée a été mise en pratique à travers un outil. Le code source de cet outil est disponible sur GitHub [15]. L'accès au code source de l'outil nous a permis de bien comprendre le processus de la détection.

La possibilité de la validation des résultats Comme il est cité dans le travail [16], les auteurs ont démontré la cohérence de l'outil de détection avec des événements réels comme les attaques DDOS.

Une attaque de type Distributed Denial of Service (**DDOS**) vise la disponibilité d'un serveur en surchargeant ce dernier avec un trafic depuis différentes sources, d'où le terme *Distributed*.

II.3 L'étude des délais des liens

II.3.1 Les données utilisées dans l'analyse des délais

La méthode conçue pour la détection des changements des délais se base sur des fondements statistiques. Ces derniers sont capables de montrer leurs performances si la taille des échantillons³ considérés est grande. Afin de surveiller un grand nombre de liens sur Internet, il faut avoir un grand nombre de sondes Atlas avec une certaine diversité. Les deux catégories de traceroutes utilisés sont :

- *builtin* : ce sont les traceroutes effectués par toutes les sondes Atlas vers les instances des 13 serveurs DNS racines. Les traceroutes sont effectués chaque 30 minutes. En pratique, certains serveurs racines DNS déploient l'anycast. Au moment de la réalisation du travail de référence⁴, c'étaient des traceroutes vers 500 instances des serveurs DNS racines ;

DNS Anycast est une solution utilisée pour accélérer le fonctionnement des serveurs DNS. Les serveurs DNS adoptant cette approche fournissent des temps de réponse plus courts, et ce partout dans le monde. Les requêtes en provenance de l'utilisateur sont redirigées vers un nœud adéquat suivant un algorithme prédéfini.

- *anchoring* : ce sont les traceroutes effectués par environ 400 sondes Atlas à destination de 189 serveurs⁵ et ce chaque 15 minutes.

En ce qui concerne le nombre de traceroutes analysés, le tableau II.1 reprend plus de détails ; le nombre de traceroutes ainsi que le nombre de sondes impliquées dans ces traceroutes, et ce pour les deux types d'adressages (IPv4 et IPv6).

| | Nombre de traceroutes | Nombre de sondes |
|------|-----------------------|------------------|
| IPv4 | 2.8 milliards | 11, 538 |
| IPv6 | 1.2 milliards | 4, 30 |

TABLE II.1 – Récapitulatif des traceroutes utilisés dans le travail de référence

3. L'échantillon de la métrique qui caractérise un lien : RTT différentiel.

4. Année 2017.

5. Sondes Atlas ayant des fonctionnalités avancées.

Pour précision, l'étude des délais des liens ne concerne pas les adresses privées, ainsi, le suivi des délais ne concerne pas les réseaux privés. De plus, ce suivi se base sur les requêtes de type traceroute, et traceroute reprend une partie de la topologie de l'Internet. En effet, les liens considérés sont ceux topologiques et ne sont pas les liens physiques.

II.3.2 Le principe de la détection des changements des délais

Le processus de la détection des anomalies dans les délais d'un lien repose sur une métrique caractérisant un lien dans un réseau informatique : le RTT différentiel.

Définition du RTT différentiel

Avant de définir le RTT différentiel, soit la définition du RTT :

ICMP (Internet Control Message Protocol) est un protocole utilisé pour véhiculer des messages de contrôle sur Internet.

RTT est obtenu en calculant la différence entre le timestamp associé à l'envoi du paquet vers une destination et le timestamp associé à la réception de la réponse ICMP de cette destination. C'est une métrique pour évaluer les performances d'un réseau en matière de temps de réponse. Les mesures du RTT sont fournies par les utilitaires traceroute et ping. En ce qui concerne traceroute, il fournit les sauts impliqués dans le chemin de forwarding, c'est le chemin parcouru par le trafic entre la source et la destination. Le RTT inclut le temps de transmission, du quering et du traitement.

La figure II.1 (a) illustre le RTT entre la sonde P et deux routeurs B et C. Le RTT différentiel du lien entre B et C adjacents, noté Δ_{PBC} , est la différence du RTT entre la sonde P et B (bleu) d'une part, et du RTT entre la sonde P et C (rouge) d'autre part, comme est illustré dans la figure II.1 (b).

$$\begin{aligned}
 \Delta_{PBC} &= RTT_{PC} - RTT_{PB} \\
 &= \delta_{PA} + \delta_{AB} + \delta_{BC} + \delta_{CD} + \delta_{DA} + \delta_{AP} - \delta_{PA} - \delta_{AB} - \delta_{BA} - \delta_{AP} \\
 &= \delta_{BC} + \delta_{CD} + \delta_{DA} - \delta_{BA} \\
 &= \delta_{BC} + \varepsilon_{PBC}
 \end{aligned}$$

où δ_{BC} est le délai du lien BC et ε_{PBC} est la différence entre les deux chemins de retour (B vers P et C vers P).

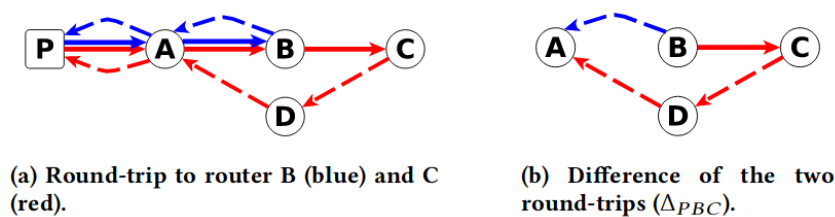


FIGURE II.1 – (a) Le RTT entre la sonde P et les routeurs B et C. (b) La différence entre les chemins de retour depuis les routeurs B et C vers la sonde P. Source : [16]

Le principe de la détection des changements des délais L'évolution du délai d'un lien est déduit de l'évolution de son RTT différentiel. Reprenons d'abord la formule du RTT différentiel du lien BC : $\delta_{BC} + \varepsilon_{PBC}$.

La valeur de δ_{BC} dépend de l'état des deux routeurs B et C et ne dépend pas de la sonde P. Tandis que la valeur de ε_{PBC} dépend de la sonde P ; la destination des deux chemin de retours.

Supposons qu'on dispose d'un nombre n de sondes Atlas $P_i, i \in [1, n]$, telles que toutes les sondes ont un chemin de retour ε_{P_iBC} différent depuis B et depuis C. Les n RTTs différentiels Δ_{P_iBC} du lien BC pour chacune des sondes P_i partagent la même composante δ_{BC} .

$$\Delta_{P_1BC} = \delta_{BC} + \varepsilon_{P_1BC}$$

$$\Delta_{P_2BC} = \delta_{BC} + \varepsilon_{P_2BC}$$

.....

$$\Delta_{P_iBC} = \delta_{BC} + \varepsilon_{P_iBC}$$

$$\Delta_{P_nBC} = \delta_{BC} + \varepsilon_{P_nBC}$$

Les valeurs de ε_{P_iBC} sont indépendantes. L'indépendance de ces valeurs implique que la distribution Δ_{P_iBC} est estimé stable au cours du temps si δ_{BC} est constant. Toutefois, un changement significatif de la valeur de δ_{BC} influence les valeurs des RTTs différentiels, ainsi, la distribution des RTTs différentiels Δ_{P_iBC} change. D'où l'idée de l'évolution du délai d'un lien est déduit de l'évolution de son RTT différentiel.

La détection des anomalies dans les délais repose sur le théorème central limite (TCL). Ce théorème annonce que si on a une suite de variables aléatoires X_i indépendantes ayant la même espérance μ et la même variance σ^2 , la moyenne de ces variables aléatoires est une variable aléatoire qui suit une loi normale.

II.4 L'étude des délais des liens en pratique : l'évolution du RTT différentiel des liens

L'évolution des RTTs différentiels est une des applications du principe décrit dans II.3.2. Pour un lien donné, le suivi s'étale sur plusieurs périodes. En plus du suivi du RTT différentiel, il y a aussi l'identification d'éventuelles anomalies pour ce lien.

II.4.1 Les étapes principales de détection

Les étapes du processus de détection peuvent être résumées dans les éléments suivants :

1. Calcul des RTTs différentiels de chaque lien identifié dans les traceroutes.
2. Caractérisation des liens avec le théorème central limite (CLT).
3. Comparaison de l'état de chaque lien avec sa référence et l'identification des anomalies.
4. Mise à jour de la référence du lien.

II.4.2 Description des paramètres de l'analyse des délais

La détection des changements des délais nécessite l'ajustement d'un nombre de paramètres. Ces paramètres ont des valeurs par défaut définies dans le travail de référence. Ci-dessous les paramètres à ajuster afin d'obtenir l'évolution des RTTs différentiels d'un lien ainsi que les éventuelles anomalies.

traceroutes : ce sont l'ensemble des résultats de requêtes traceroute. La détection des anomalies est effectuée suivant le nombre de traceroutes analysés.

start : c'est la date de début de l'analyse. Seuls les traceroutes effectués par les sondes Atlas à partir de cette date qui sont analysés.

end : c'est la date marquant la fin de l'analyse. Comme le paramètre *start*, c'est la date maximales des traceroutes effectués par les sondes Atlas à considérer.

timeWindow : ce paramètre est exprimé en secondes. Il correspond à la durée des périodes de l'analyse. Ces périodes ont la même durée. Autrement dit, la durée entre *start* et *end* est divisée sur cette même taille : *timeWindow*, c'est qui est illustré dans la Figure II.2.

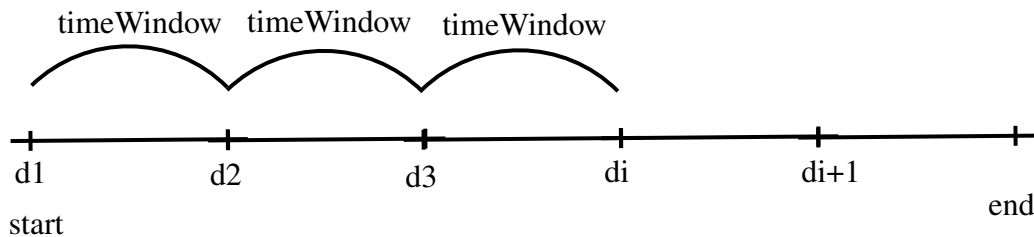


FIGURE II.2 – Illustration des périodes de l'analyse entre la date de début et la date de fin

minSeen : ce paramètre est un entier, il indique le déclenchement de la comparaison du RTT différentiel de référence d'un lien avec sa valeur courante. Par exemple, si $minSeen = 24$ et $timeWindow = 3600$ (1 heure), cela implique que la référence de tout lien est construite des valeurs obtenues tout au long la durée $24 * timeWindow$, donc après avoir analysé les traceroutes d'une journée. A la 25^{ème} *timeWindow*, la détection des anomalies s'effectue.

alpha : noté α , $\alpha \in [0, 1]$, c'est le paramètre de la moyenne mobile exponentielle calculée. « *Les méthodes de lissage exponentiel sont un ensemble de techniques empiriques de prévision qui accordent plus ou moins d'importance aux valeurs du passé d'une série temporelle.* ⁶ »

La moyenne mobile exponentielle est utilisée pour calculer la médiane des RTTs différentiels de référence durant une période d_k . Pour calculer la valeur de la médiane des RTTs différentiels de référence \bar{m}_t courant la période t et pour le lien l , soit :

$$\bar{m}_t = \alpha m_t + (1 - \alpha) \bar{m}_{t-1}$$

6. Source : <https://perso.math.univ-toulouse.fr/lagnoux/files/2013/12/Chap6.pdf>, consultée le 30/09/2018.

m_t la médiane des RTTs différentiels observée pour l durant le *timeWindow* t .

\overline{m}_{t-1} la médiane des RTTs différentiels de référence durant le *timeWindow* $t - 1$.

Pour précision, $\{m_t\}$ et $\{\overline{m}_t\}$ désignent deux ensembles différents. Le premier est l'ensemble de la médiane des RTTs différentiels de chaque période d_k . Or, le deuxième est l'ensemble de la médiane des RTTs différentiels construite en utilisant la méthode de la moyenne mobile exponentielle. Le calcul de cette dernière prend en compte les médianes des RTTs différentiels précédentes ainsi que la médiane des RTTs différentiels courante. La participation de ces dernières dans le calcul de la référence est dirigé par le paramètre α .

Le paramètre α contrôle l'importance des mesures précédentes par rapport aux mesures récentes. De ce fait, «plus α est proche de 1 plus les observations récentes influent sur la prévision, à l'inverse un α proche de 0 conduit à une prévision très stable prenant en compte un passé lointain.»[9]. Dans le travail [16], le paramètre α prend par défaut 0.01 comme valeur.

confInterval : l'intervalle de confiance est défini comme représentant les valeurs probables que peut prendre une moyenne, si l'on accepte une marge d'erreur définie à l'avance à travers la valeur de la marge d'erreur. Il existe plusieurs méthodes pour calculer l'intervalle de confiance d'une proportion. Parmi les critères impliqués sur le choix de la méthode de calcul, on note le nombre total d'expériences.

le RTT différentiel courant et celui de référence d'un lien donné sont représentés, en pratique, par des intervalles de confiance. le calcul des intervalles de confiance est approché par le score de Wilson. Le score de wilson a été adopté car il donne des résultats même si la distribution pour laquelle l'intervalle de confiance est calculé est petite. Pour chaque lien et pour chaque période d_k , deux intervalles de confiances sont calculés, pour ensuite évaluer le chevauchement entre ces deux intervalles. L'outil de détection utilise une marge d'erreur égale à 0.05.

II.4.3 L'évolution du RTT différentiel d'un lien et la détection des anomalies

Nous avons décrit les étapes principales de détection dans la section II.4.1. Nous présentons dans ce qui suit l'algorithme de détection en détail. Les paramètres d'entrée de cet algorithme sont détaillés dans la section II.4.2. L'algorithme fournit en résultat l'évolution du RTT différentiel de chaque lien, tout au long de la période de l'analyse, ainsi que son RTT différentiel de référence. Dans le cas d'anomalies détectées, celles-ci sont identifiées tout en précisant la date pendant laquelle l'anomalie a été détecté ainsi que la valeur du RTT différentiel ayant engendré cette anomalie.

Un traceroute est un ensemble de sauts H auxquels sont joints l'identifiant de la sonde ayant effectué la requête traceroute et la destination de la requête. Chaque saut est décrit par un ensemble de signaux S . Chaque signal décrit le routeur ayant émis une réponse à la sonde parmi les routeurs traversés avant d'atteindre la destination finale. Pour le saut h_i , on note trois⁷ signaux $s_{i,j}$, $j \in [1, 3]$ dont le routeur émettant le signal est $from_{i,j}$ et le RTT est égal à $rtt_{i,j}$. C'est ce qu'illustre la Figure II.3.

7. Le nombre trois peut varier dans certains cas. L'outil de détection conçu est adapté à tout nombre de signaux.

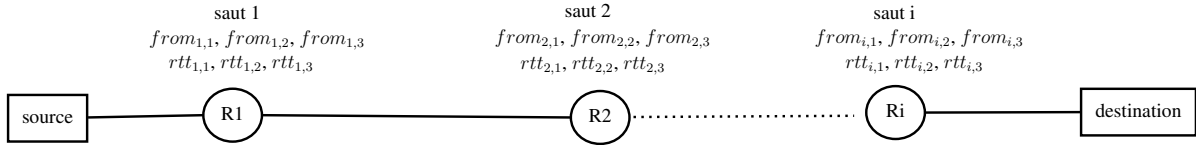


FIGURE II.3 – Illustration des sauts d'un traceroute avec leurs informations

II.4.4 Processus de détection des anomalies

Soient d_1, d_2, \dots, d_N les k périodes entre *start* et *end* avec $k \in [1, N]$. La différence entre le début de la période d_{k+1} et le début de la période d_k est égale à *timeWindow* pour tout $k + 1 \leq N$.

Le processus de la détection des anomalies passe par plusieurs étapes. D'abord on regroupe les traceroutes à analyser par période d_k (étape 1). Ensuite, on prépare les traceroutes de toute période d_k en y appliquant un nombre d'opérations (étapes 2 à 6). A la fin de la préparation des traceroutes de toutes les périodes, on ne considère que les données qui concernent le lien à suivre. Ce sont les données qui servent à la comparaison des délais d'un lien avec les valeurs de référence (étape 7). Les étapes de détection sont les suivantes :

1. Regroupement des traceroutes par période d_k . En effet, chaque période d_k est associée à un ensemble de traceroutes ayant été effectués entre d_k et $d_{k+1} - 1$, avec $k \in [1, N]$, $k + 1 \leq N$ et d_k est exprimé en secondes (Unix Timestamp). Les étapes (2 à 6) concernent les traceroutes de chaque d_k .

2. Vérification de la validité de chaque traceroute. Les traceroutes sont filtrés en prenant en considération les points suivants :

- élimination des traceroutes échoués complètement ;
- élimination des signaux contenant une adresse IP privée ;
- élimination des signaux qui ne contiennent pas un RTT ou qui contiennent un RTT négatif ;
- élimination des signaux échoués.

Il existe deux sortes d'échecs dans un traceroute : échec complet et échec partiel. Dans le premier, la sonde Atlas ne réussit pas à atteindre la destination, dans ce cas, la liste des sauts est vide. Dans le deuxième cas, l'échec peut concerner un ou plusieurs saut, ou bien il peut concerner un, deux ou trois signaux d'un saut. A la fin de cette étape, nous obtenons une liste de traceroutes.

3. Calcul de la médiane des RTTs par saut Pour tout saut d'un traceroute, on calcule la médiane des RTTs par $from_{i,j}$. Soit le saut $h_i = \{s_{i,j} | s_{i,j} \in S\}$ La médiane des RTTs d'un saut h_i est :

$mediane_rtt(h_i) = \{median(\{rtt_{i,j}\})\}$ pour tout $rtt_{i,j}$ ayant la même $from_{i,j}$. Autrement dit, le nouveau saut du traceroute est reconstruit en regroupant les signaux par adresse IP ($from_{i,j}$) et ensuite en calculant la médiane de leurs RTTs ($rtt_{i,j}$). Par exemple, si on a un saut dont les trois signaux sont en provenance du même routeur, le nouveau saut est présenté par ce routeur avec un RTT obtenu en calculant la médiane des trois RTTs.

4. Inférence des liens topologiques par traceroute. Un lien est formé par chaque paire de routeurs consécutifs dans un traceroute. De manière générale, la Figure II.4 illustre la constitution des liens possibles dans un traceroute. Soient $R_{a,i}$, avec $i \in 1, 2, \dots, N$, l'ensemble de routeurs pour le saut a et $R_{b,j}$, avec $j \in \{1, 2, \dots, M\}$, l'ensemble de routeurs pour le saut b , avec N et M deux entiers. Les liens construits sont ceux partant de tout $R_{a,i}$ vers tout $R_{b,j}$, où a et b sont deux sauts consécutifs. A l'issue de cette étape, pour tout traceroute, on obtient la liste des liens possibles tout en reprenant des informations générales de la requête traceroute.

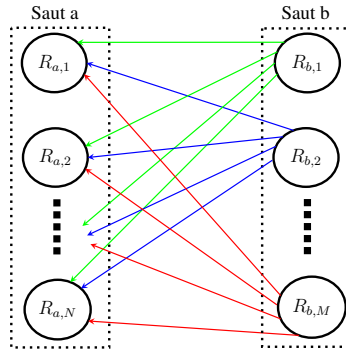


FIGURE II.4 – Inférence des liens possibles entre les routeurs des deux sauts consécutifs $R_{a,i}$ et $R_{b,j}$

5. Caractérisation des liens avec leurs RTTs différentiels. A cette étape, on calcule le RTT différentiel de chaque lien, au sein d'un traceroute donné, en calculant la différence entre les RTTs des deux routeurs impliqués dans le lien en question. En plus du RTT différentiel, on note aussi la sonde Atlas ayant effectué la requête traceroute où le lien a été identifié.

6. Fusion des informations d'un lien Etant donné qu'un lien (IP1, IP2) peut être identifié plusieurs fois pendant une même période d_k d'une part, et le lien (IP2, IP1) est similaire au lien (IP1, IP2) d'autre part, la fusion permet de construire une nouvelle distribution des RTTs différentiels caractérisant le lien (IP1, IP2) qui reprend les RTTs différentiels du lien (IP1, IP2) ainsi que ceux du lien (IP2, IP1).

Dans le travail de référence, la similarité de deux liens implique que les RTTs différentiels qui caractérisent un lien (IP2, IP1) caractérisent aussi le lien (IP1, IP2); ces deux liens partagent les mêmes routeurs.

A la fin de l'étape 6, tous les traceroutes sont préparés. A présent, l'objectif est d'identifier les dates pendant lesquelles des anomalies ont été détectées. Pour ce faire, l'idée du travail de référence est de conserver, pour un lien donné, une référence du RTT différentiel médian. Cette référence est d'abord comparée avec la médiane courante du RTT différentiel, puis, mise à jour au fur et à mesure tout au long de la période d'analyse.

7. Calcul de la médiane des RTTs différentiels et l'intervalle de confiance courant du lien analysé. Pour un lien donné, on calcule la médiane des RTTs différentiels d'une d_k , ensuite on calcule les deux bornes de l'intervalle de confiance courant pour ce d_k .

8. Mise à jour de la médiane et de l'intervalle de référence du lien analysé. La médiane des RTTs différentiels de référence ainsi que l'intervalle de confiance de référence sont mise à jour tant que la période déclenchant la détection des anomalies n'est pas atteinte. La mise à jour de la référence : médiane et les bornes de l'intervalle de confiance se base sur la moyenne

mobile exponentielle . Cette méthode prend en compte les valeurs antérieures ainsi que celles récentes en étant dirigée par le paramètre *confInterval* décrit dans la section II.4.2.

9. Comparaison des intervalles de confiance La comparaison de l'état courant du lien avec celui de référence est effectuée en analysant le chevauchement d'intervalles de confiance courant et de référence. Le délai d'un lien est jugé normal si son intervalle de confiance courant est inclus dans l'intervalle de confiance de référence. C'est le cas 1 dans la Figure II.5, *referenceLow* et *referenceHight* sont les bornes de l'intervalle de confiance de référence. *currentLow* et *currentHight* sont les bornes de l'intervalle de confiance courant.

D'après le travail de référence, on distingue quatre cas possibles illustrés dans la Figure II.5 :

Cas 1 : le délai du lien est normal.

Cas 2 : le délai du lien est anormal.

Cas 3 : le délai du lien est anormal.

Cas 4 : le délai du lien est anormal.

Dans le cas où le délai est jugé anormal, on introduit ce qu'on appelle la *déviaton*. Cette métrique caractérise l'anomalie détectée. Elle est calculée différemment dans le cas où le délai est anormal.

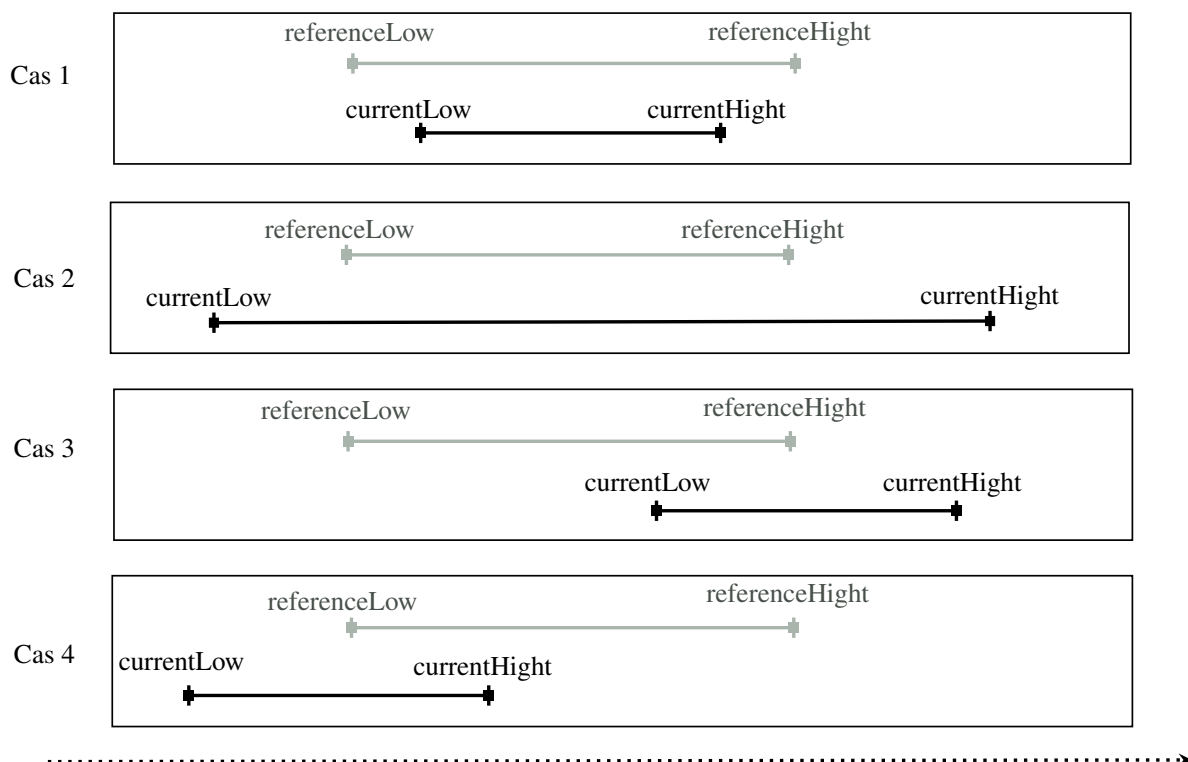


FIGURE II.5 – La comparaison des deux intervalles de confiance : courant et référence

Ces étapes sont reprises dans la Figure II.6. [!]

II.4.5 Vue globale des étapes de la détection des anomalies à travers l'évolution des RTTs différentiels

La figure II.6 présente la succession des étapes de la détection des anomalies dans les délais d'un lien donné.

II.5 Conclusion

Nous avons abordé les différentes étapes de la création de l'évolution du RTT différentiel des liens identifiés dans les traceroutes analysés. Cette évolution présente aussi les anomalies si sont présentes. C'est une évolution des RTTs différentiels d'un lien mais aussi un indicateur sur les délais des liens. Les approches adoptées dans la création de l'outil de détection suppose qu'on dispose d'assez de traceroutes pour assurer les performances des méthodes conçues. Etant donnée que les sondes Atlas sont nombreuses, la quantité de données générées est très grande, d'où la nécessité d'utiliser des technologies adoptées à la quantité importante de données.

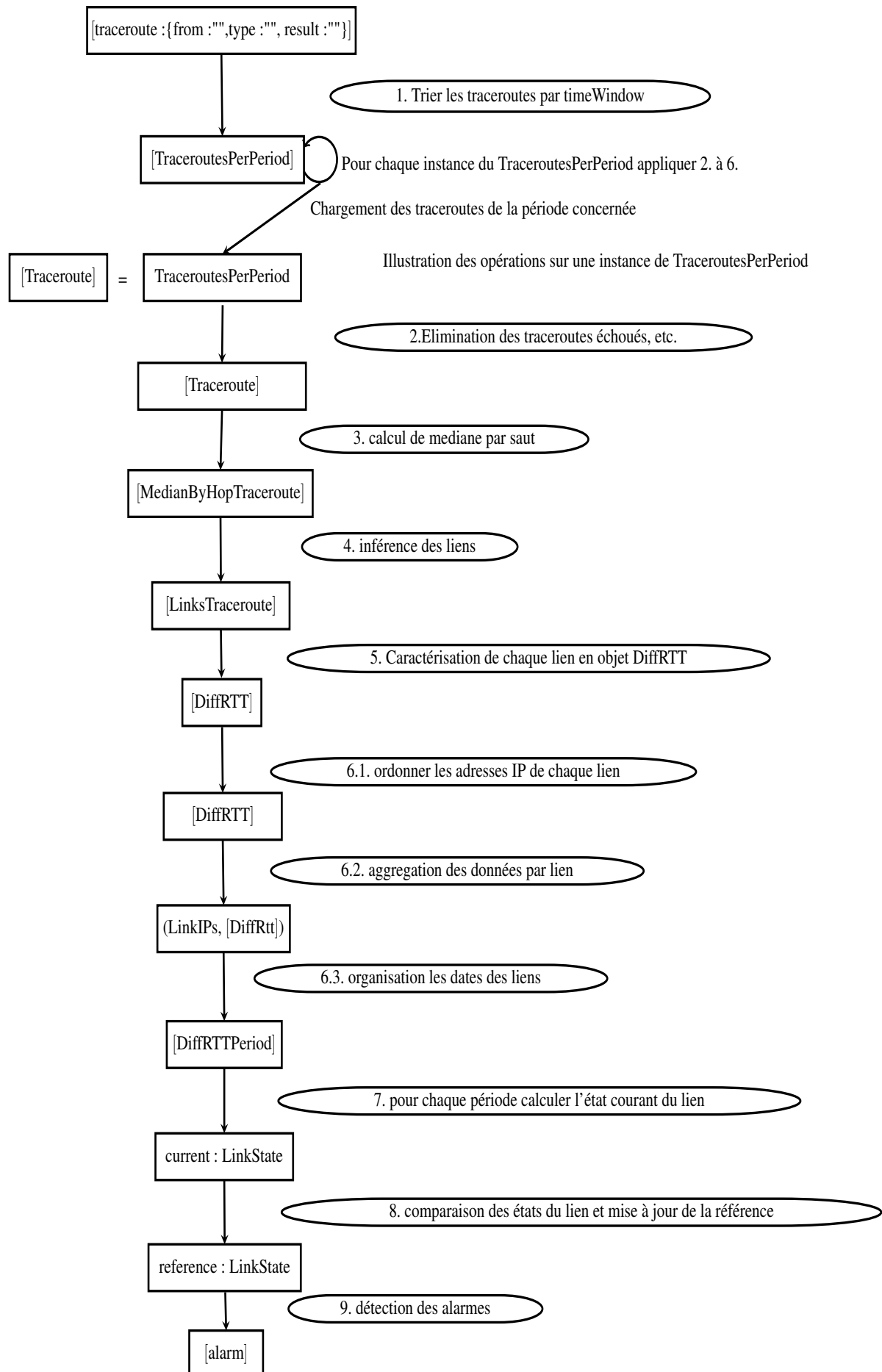


FIGURE II.6 – Le processus de la détection des anomalies dans les délais des liens

Chapitre III

Introduction au Big Data

III.1 Introduction

Big Data est un terme associé aux données massives, rapidement générées, ayant une grande diversité où les outils traditionnels sont incapables de gérer ces données. La complexité du Big Data est dû au fait que tout type de données peut être utilisé, et ce en vue de livrer la bonne information à la bonne personne et au bon moment dans le but d'aider à prendre les bonnes décisions. Dans ce chapitre, nous présentons brièvement un des processus d'analyse de données. Ensuite, nous décrivons quelques concepts impliqués dans un processus d'analyse des données massives. Enfin, nous parcourons un ensemble de technologies utilisées dans l'analyse des données à grande échelle.

III.2 Processus d'analyse de données massives

Les étapes d'un processus d'analyse de données, à grande échelle ou non, sont différentes selon le processus adopté. Ce dernier peut impliquer plusieurs besoins et concepts. En particulier, on note le besoin au stockage de données massives, au traitement de données massives et le besoin à la visualisation des résultats des traitements appliqués. Le choix du traitement à appliquer sur les données est dirigé par les objectifs de l'analyse menée. Par exemple, une analyse peut envisager la création d'un système de recommandations dans les sites Internet, la conception d'un outil de prédiction basé sur les données historiques, un système de suivi en temps réel, etc. Ces applications appuient sur des algorithmes basés sur les mathématiques, en particulier les mathématiques statistiques et probabilistes. Afin d'assurer l'efficacité de l'analyse de données massives, la coopération de plusieurs ressources accélère considérablement les étapes de l'analyse, c'est ce que l'informatique distribuée a apporté.

Exemple d'un processus d'analyse de données

Le processus d'analyse de données passe généralement par des étapes classiques. A ces étapes peuvent s'ajouter d'autres étapes intermédiaires ou supplémentaires. Pour précision, non seulement ces étapes peuvent s'appliquer dans le cas où les données sont volumineuses, mais aussi dans le cas des données manipulées par des outils traditionnels comme les bases de données relationnelles. IBM Knowledge Center¹ a résumé le processus de l'analyse de données

1. Source : https://www.ibm.com/support/knowledgecenter/fr/SSEPGG_9.5.0/com.ibm.im.easy.doc/c_dm_process.html, consultée le 06/08/2018.

dans les étapes suivantes : définition de problème, exploration de données, préparation de données, modélisation, évaluation et enfin l'étape de déploiement.

Définition de problème C'est le point d'entrée vers tout projet d'analyse de données. Les objectifs doivent être clairs. A la fin de cette étape, on connaît les objectifs de l'analyse, mais pas forcément comment cette analyse va être menée en terme de technologies utilisées, les méthodes statistiques, etc.

Exploration de données Ce que caractérise cette étape, c'est la découverte de la nature des données, les sources des données, la qualité des données, etc.

Préparation de données C'est durant cette phase que se passent les opérations d'ETL (Extract-transform-load), ce sont les opérations de chargement, nettoyage et de la transformation des données. Durant cette étape aussi que des tentatives sont faites pour créer le schéma de données. Dans certains cas, cela amène à chercher des sources complémentaires de données ou bien de nouvelles transformations à appliquer sur les données. Tel qu'un schéma comporte les différentes tables, les attributs pertinents et autres.

Modélisation A l'issue de cette phase, un modèle de qualité est créé. Un modèle qui doit répondre aux objectifs précédemment établis. Cette étape implique des principes en statistiques, probabilités, *machine learning*, etc.

Evaluation C'est à cette étape qu'on valide le modèle créé à l'étape de Modélisation. L'évaluation et la validation du modèle prend en compte les objectifs prédéfinis ; si le modèle répond aux attentes précédemment exprimées.

Déploiement C'est l'utilisation des outils existants ou la mise place d'une solution qui convient aux besoins spécifiques en terme de visualisation des résultats de l'analyse.

III.3 Quelques concepts associés au Big Data

Etant donné que le domaine du Big Data implique plusieurs concepts, nous présentons une liste de concepts non exhaustive.

III.3.1 Définition du Big Data : Volume, Vitesse, Variété et Véracité

IBM définit le Big Data suivant les quatre dimensions suivants : volume, variété, vitesse et véracité.

Volume de données La quantité de données manipulées par les outils traditionnels de la gestion des données est de l'ordre de Gigaoctets (GB) et de Téraoctets (TB). Toutefois, le Big Data est mesuré en Pétaoctets (PB), Exaoctets (EB), voire plus. Une des premières applications du Big Data est la recherche dans Word-Wide Web (WWW). Selon une étude ([31], 2013) de l'International Data Corporation (IDC), le volume de données va atteindre 40 Zettaoctets² par entreprise en 2020 .

Vitesse de données le Big Data est généré par des milliards d'appareils, les données générées sont communiquées avec la vitesse de la lumière via l'Internet. L'augmentation de la vitesse de l'Internet est une des raisons ayant contribué à l'augmentation de la vitesse de la génération de données. Par exemple, Walmart (international discount retail chain) génère environ

2. 1 ZB = 1000000000000 GB

2.5 Pétaoctets de données chaque heure via les transactions de ses consommateurs³.

Variété de données Le Big Data inclut toutes les formes des données, des fonctions diversifiées des données et des sources variées des données.

Le premier aspect de la variété des données massives est la **forme** de celles-ci. Les données manipulées incluent du texte, des graphes, des cartes, des vidéos, des photos, etc.

Le deuxième aspect de la variété des données massives concerne les **fonctions** assurées par ces données. Des données sont issues des conversations humaines, d'autres des transactions des consommateurs, ou bien des données archivées, etc.

Les **sources** du Big Data est le troisième aspect de variété. Des données sont en provenance des téléphones mobiles, des tablettes ou des ordinateurs portables, des fichiers journaux, du réseau de capteurs, etc. Les sources du Big Data peuvent être classées en trois grandes catégories : communications *human to human* comme les conversations échangées dans les réseaux sociaux, communications *human to machine* comme l'accès des utilisateurs aux données dans le web et enfin communications *machine to machine* comme les données issues de la communication entre les capteurs dans un réseau de capteurs.

Véracité de données La véracité concerne la crédibilité et la qualité de données. La mauvaise qualité de données est due à de nombreuses raisons, telles que les pannes techniques comme le dysfonctionnement des appareils comme les capteurs, les erreurs humaines, etc. De plus, les données peuvent être intentionnellement erronées pour des raisons de concurrence ou des raisons stratégiques.

III.3.2 L'architecture standard du Big Data

L'architecture standard du Big Data présentée dans ce qui suit est celle proposée dans [28]. Cette architecture est composée des couches suivantes : *Data sources*, *Data Ingest*, *Batch processing*, *Stream Processing*, *Data organizing*, *Infrastructure*, *Distributed File System* et enfin la couche *Data consumption*. La Figure III.1 reprend l'organisation des différentes couches de cette architecture.

Dans un premier temps, les données sont accueillies via la couche *ingest system* par diverses sources de données. Ensuite, les données sont traitées dans deux modes différents : *stream processing* et *batch processing*. Les résultats de ce traitement peuvent être envoyées vers les bases de données NoSQL (III.3.3) pour une utilisation ultérieure, ou bien utiliser ces résultats comme entrées pour d'autres applications. Une solution Big Data comprend typiquement ces couches logiques. Chaque couche peut être représentée par une ou plusieurs technologies disponibles. Reprenons chaque couche logique :

Data sources layer Le choix des sources de données pour une application donnée dépend des objectifs qui dirigent l'analyse en question. Les sources avec leurs différents aspects sont détaillées dans la section III.3.1.

Data Ingest layer Cette couche permet de récupérer les données depuis les différentes sources de données. Les données sont accueillies à travers des points d'entrées multiples. Ces points sont capables de recevoir ces données ayant une vitesse variable ainsi qu'une

3. Source : <https://www.bernardmarr.com/default.asp?contentID=690>, consultée le 30/06/2018.

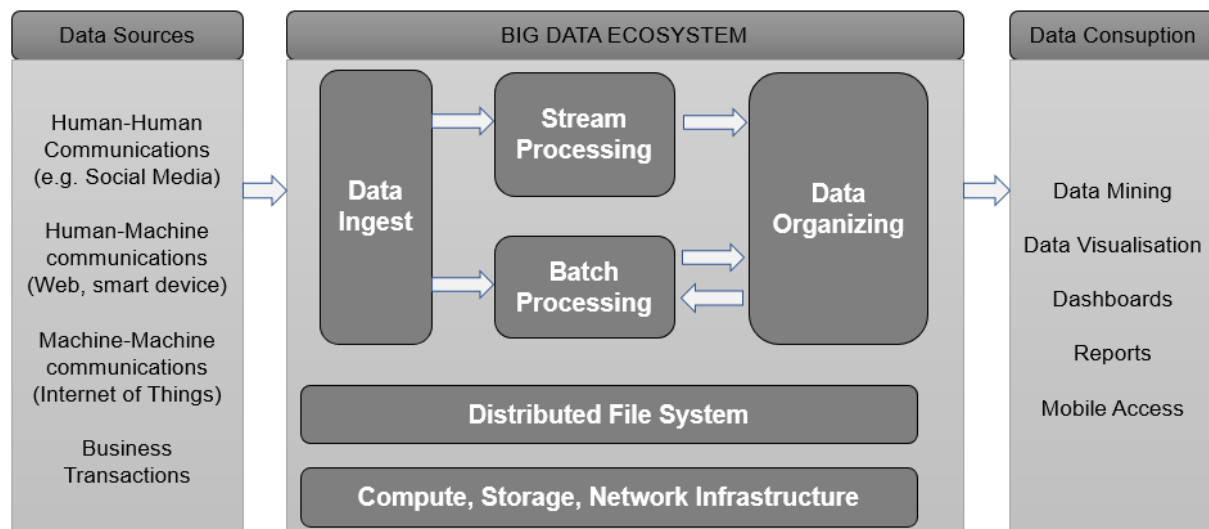


FIGURE III.1 – Architecture standard du Big Data [28]

quantité aussi variable. Après avoir traversé la couche *Data Ingest*, les données sont envoyées au *batch processing system*, au *realtime processing system*, ou bien à un système de stockage particulier.

Batch processing layer Les données reçues sur cette couche sont celles en provenance du *Data Ingest* ou bien d'une des bases de données NoSQL. Ces données sont ensuite traitées, par exemple, en utilisant les techniques de la programmation parallèle en vue de fournir les résultats souhaités dans un temps raisonnable. La présente couche doit avoir connaissance des sources de données, les types de données, les algorithmes qui vont travailler sur ces données et enfin les résultats souhaités. Les résultats des traitements peuvent être utilisés par une des applications ou bien sauvegarder ces données dans une des bases de données adaptées.

Stream Processing layer Cette couche approvisionne les données directement d'une des entrées du *Data Ingest layer*; c'est ce que différencie cette couche de la couche *Batch processing layer*. En revanche, *Stream Processing* est similaire à la couche *Batch processing* en matière des techniques de la programmation parallèle utilisées ainsi que la nécessité d'avoir les détails sur les sources des données, les types de données et les résultats souhaités.

Data organizing layer Le rôle de cette couche est d'organiser les données afin de faciliter l'accès à ces dernières. Ce sont les données obtenues de la part de la couche *Stream Processing* ainsi que la couche *Batch processing*. Cette couche est représentée par les bases de données NoSQL.

Infrastructure layer Cette composante est responsable de la gestion des ressources de stockage, les ressources du calcul et la gestion de la communication. Par exemple, les fonctionnalités de cette couche sont fournies à travers le cloud computing.

Distributed File System layer Cette couche permet de stocker une grande quantité de données, de sorte que ces données soient rapidement et facilement accessibles à toute les couches qui forment un système du Big Data. C'est ce que assure, par exemple, Hadoop Distributed File System (HDFS).

Data consumption layer Cette dernière couche utilise les résultats obtenus par les couches de l'analyse. Les résultats fournis peuvent être exprimés avec des rapports, des dashboards, des visualisations, un moteur de recommandation ou tout autre format.

III.3.3 Les bases de données NoSQL (Not Only SQL)

Introduction

Au cours de ces dernières années, on constate une révolution dans le stockage de données non structurées ayant une taille importante. De plus, les objets à sauvegarder sont complexes ; ils sont issus de sources hétérogènes. Cette complexité a mis en question les performances des bases de données relationnelles.

Le terme NoSQL est apparu pour la première fois en 1998. Carlo Strozzi a parlé des bases de données relationnelles qui n'utilisent pas le SQL comme langage d'interrogation des tables. Des années plus tard, des solutions open source basées sur ce concept ont vu le jour.

Les bases de données relationnelles sont conçues pour gérer les données structurées, optimisées pour offrir la précision et la cohérence de données. De plus, elles sont utilisées par la majorité des entreprises pour plusieurs raisons comme la facilité d'utilisation, la disponibilité de plusieurs produits et développeurs, etc. Ces dernières années, avec l'augmentation exponentielle de la quantité de données générées par certaines entreprises, ces dernières ont constaté l'insuffisance des Systèmes de Gestion de Bases de Données Relationnelles (SGBDR) pour répondre à leurs besoins.

Les bases de données NoSQL sont conçues pour gérer des volumes de données importants. Le flux ainsi que la structure de ces données sont imprévisibles. C'est pourquoi les bases de données relationnelles ne sont pas convenables. L'idée de des bases de données NoSQL, c'est d'abord assurer la capacité de stocker des données à grande échelle dont leur quantité évolue rapidement, voire exponentiellement. En deuxième lieu, les données stockées doivent être interrogées avec efficacité. Les données stockées dans les bases de données NoSQL n'obéissent pas à un modèle prédéfini comme le cas des bases de données relationnelles. Cette flexibilité est une des caractéristiques des bases de données NoSQL.

Types de base de données NoSQL

Il existe quatre catégories distinctes de bases de données NoSQL. Chaque catégorie répond à des besoins particuliers. Notamment, on distingue les bases de données clé-valeur, document, graphe et colonne.

Clé-valeur Une base de données de type clé-valeur repose sur le paradigme clé-valeur ; chaque donnée, que ce soit un nombre, du texte ou tout autre type est associé à une clé unique. Cette clé est le seul moyen d'accéder aux données stockées. Dans les bases de données NoSQL de type clé-valeur, les enregistrements n'adhèrent pas à une structure prédéfinie. Par exemple, on peut avoir le premier enregistrement de type entier et le deuxième enregistrement de type texte. Cela assure une forte évolutivité grâce à l'absence d'une structure ou de typage. La Figure III.2 reprend un exemple d'une base de données NoSQL de type clé-valeur.

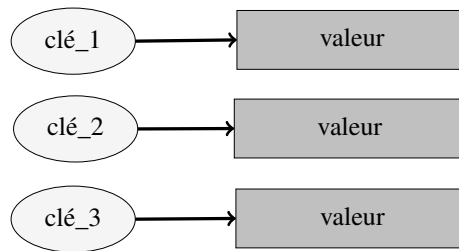


FIGURE III.2 – Illustration d’une base de données NoSQL de type clé-valeur

Document Une base de données NoSQL de type document permet de stocker les données en reposant sur le paradigme clé-valeur. Toutefois, les valeurs stockées sont complexes, il s’agit de documents de type JSON, XML, etc. L’accès aux données d’un enregistrement peut se faire de manière hiérarchique. La possibilité de stocker des objets complexes et hétérogènes est un des points forts des bases de données NoSQL de type document. Un exemple est fourni dans la Figure III.3.

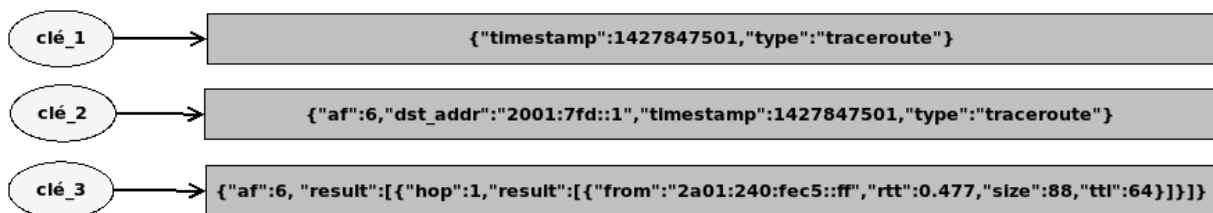


FIGURE III.3 – Illustration d’une base de données NoSQL de type document

Colonnes Dans les bases de données traditionnelles, les données sont stockées sur des lignes. Dans le cas d’une base NoSQL orientée colonne, les données sont stockées par colonne. L’interrogation de ce type de bases travaille sur une colonne particulière sans devoir passer par les autres colonnes comme dans les bases de données relationnelles classiques. Une base de données de type colonne, illustrée dans la Figure III.4, est adaptée pour les requêtes analytiques comme les requêtes d’agrégation (moyennes, maximum, etc).

| clé | prob_id | clé | af | clé | msm_id |
|-----|---------|-----|----|-----|--------|
| 1 | 233 | 1 | 4 | 1 | 5001 |
| 4 | 10 | 2 | 6 | 2 | 5006 |
| | | 3 | 4 | 3 | 16345 |
| | | 4 | 4 | 4 | 6026 |

FIGURE III.4 – Illustration d’une base de données NoSQL de type colonne

Graphe Dans une base de données de type graphe, les données stockées sont les nœuds, les liens et les propriétés sur les nœuds et sur les liens. Un exemple de base de données NoSQL de type graphe est le réseau social ; chaque entité représente une personne et les relations entre ces personnes peuvent prendre plusieurs formes. Comme il est illustré dans la Figure III.5.

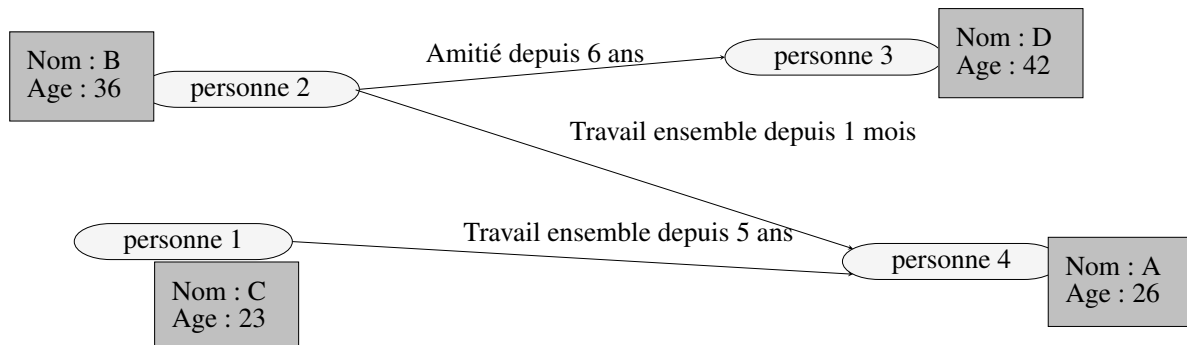


FIGURE III.5 – Illustration d’une base de données NoSQL de type graphe

Il existe plusieurs implémentations des quatre types de bases de données NoSQL. Chaque implémentation favorise un ou plus des éléments suivants : la disponibilité des données, la cohérence des données et la tolérance au partitionnement. C’est ce qu’explique le théorème CAP.

Big Data et le théorème CAP :

Dans le but d’assurer un traitement rapide de données à grande échelle, ces dernières sont réparties sur un nombre de machines (ou nœuds). Le théorème CAP annonce que dans le cadre d’un système distribué où les données sont réparties sur plusieurs machines, une base de données ne peut pas garantir les trois attributs suivants : *Consistency*, *Availability*, et *Partition Tolerance* en même temps.

Consistency (ou intégrité) Chaque donnée a un seul état visible depuis l’extérieur. Par exemple, les différents serveurs hébergeant la base de données voient tous les mêmes données. C’est pourquoi une lecture faite après une écriture doit renvoyer la donnée précédemment écrite.

Availability (ou disponibilité) Une base de données doit toujours fournir une réponse à une requête d’un client.

Partition tolerance (ou la tolérance au partitionnement) Une coupure du réseau entre deux nœuds ou l’indisponibilité d’un de ces nœuds ne devrait pas affecter le bon fonctionnement du système. Tout de même, ce dernier doit répondre à la demande d’un client.

Les trois attributs du théorème CAP s’opposent entre eux. On distingue les trois scénarios possibles :

- Le couple **CA** : les SGBDR adoptent les deux attributs C et A, qui sont une forte cohérence et disponibilité. Cependant, l’attribut partitionnement réseau n’est pas toujours pris en compte.
- Le couple **CP** : les implémentations du C et du P assurent la tolérance aux pannes en distribuant les données sur plusieurs serveurs. Malgré cette réplication, ces implémentations assurent la cohérence des données même en présence de mises à jour concurrentielles.

- Le couple **AP** : les implémentations du A et du P assurent un temps de réponse rapide et une réplication des données. Cependant, les mises à jour étant asynchrones, la garantie que la version d'une donnée soit bonne, ne peut pas être assurée.

La Figure III.6 présente des implémentations des différents types de bases de données NoSQL pour chaque couple CA, CP et AP.

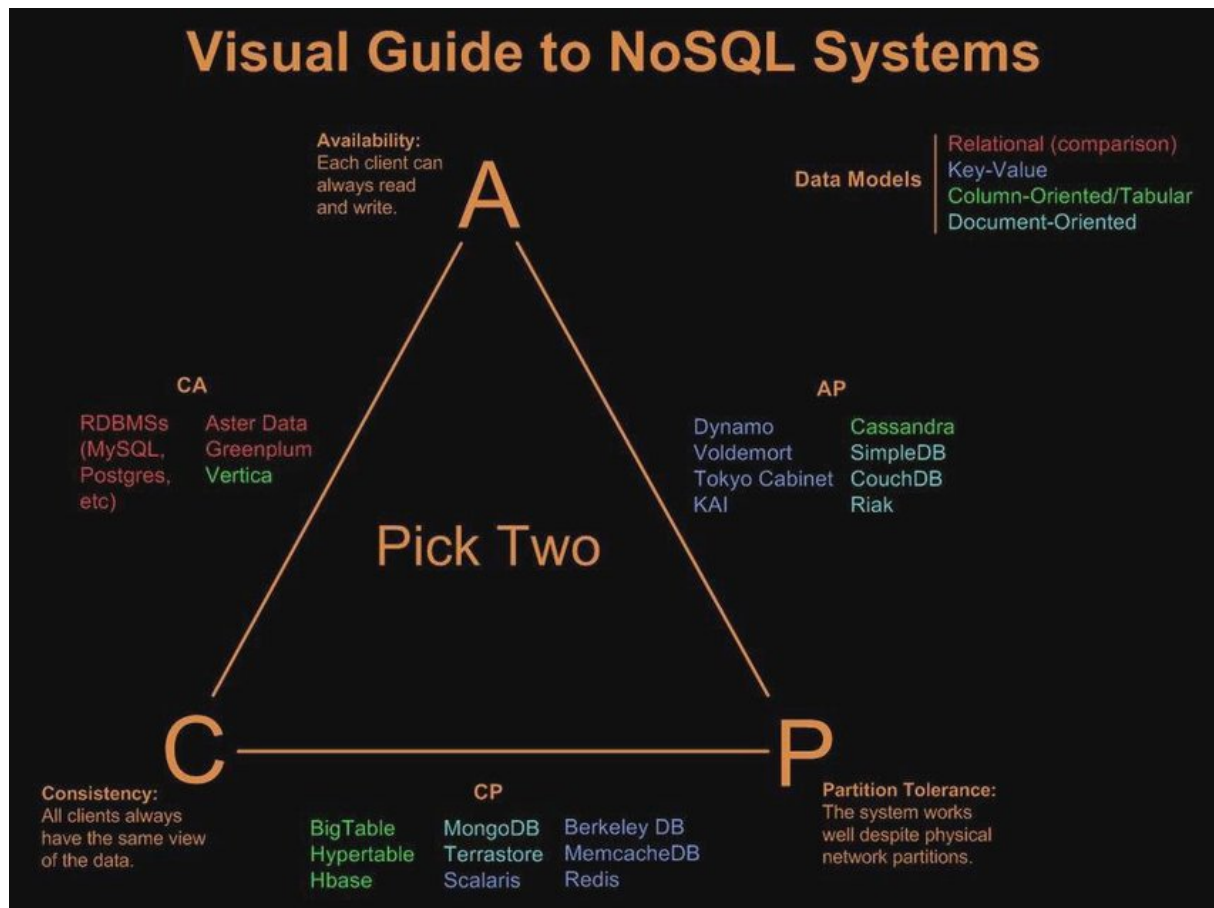


FIGURE III.6 – Bases de données NoSQL suivant le théorème de CAP

Source : <https://payberah.github.io/files/download/p2p/nosql.pdf>, consultée le 05/08/2018.

Le choix d'une base de données relationnelle ou NoSQL dépend des besoins des entreprises. En terme de tendances, la Figure III.7 reprend un classement des SGBDs au 1 août 2018. La suite de la liste ainsi que la méthode qui dirige ce classement sont disponibles sur le site web *DB-Engines Ranking*⁴. Parmi les critères du classement, on trouve le nombre de références du SGBD sur les sites Internet.

4. Source : <https://db-engines.com/>, consultée le 01/08/2018.

343 systems in ranking, August 2018

| Rank | | | DBMS | Database Model | Score | | |
|----------|----------|----------|------------------------|-------------------|----------|----------|----------|
| Aug 2018 | Jul 2018 | Aug 2017 | | | Aug 2018 | Jul 2018 | Aug 2017 |
| 1. | 1. | 1. | Oracle + | Relational DBMS | 1312.02 | +34.24 | -55.85 |
| 2. | 2. | 2. | MySQL + | Relational DBMS | 1206.81 | +10.74 | -133.49 |
| 3. | 3. | 3. | Microsoft SQL Server + | Relational DBMS | 1072.65 | +19.24 | -152.82 |
| 4. | 4. | 4. | PostgreSQL + | Relational DBMS | 417.50 | +11.69 | +47.74 |
| 5. | 5. | 5. | MongoDB + | Document store | 350.98 | +0.65 | +20.48 |
| 6. | 6. | 6. | DB2 + | Relational DBMS | 181.84 | -4.36 | -15.62 |
| 7. | 7. | ↑ 9. | Redis + | Key-value store | 138.58 | -1.34 | +16.68 |
| 8. | 8. | ↑ 10. | Elasticsearch + | Search engine | 138.12 | +1.90 | +20.47 |
| 9. | 9. | ↓ 7. | Microsoft Access | Relational DBMS | 129.10 | -3.48 | +2.07 |
| 10. | 10. | ↓ 8. | Cassandra + | Wide column store | 119.58 | -1.48 | -7.14 |
| 11. | 11. | 11. | SQLite + | Relational DBMS | 113.73 | -1.55 | +2.88 |
| 12. | 12. | 12. | Teradata + | Relational DBMS | 77.41 | -0.82 | -1.83 |
| 13. | 13. | ↑ 16. | Splunk | Search engine | 70.49 | +1.26 | +9.03 |
| 14. | 14. | ↑ 18. | MariaDB + | Relational DBMS | 68.29 | +0.78 | +13.60 |
| 15. | ↑ 16. | ↓ 13. | Solr | Search engine | 61.90 | +0.38 | -5.06 |
| 16. | ↓ 15. | ↓ 14. | SAP Adaptive Server + | Relational DBMS | 60.44 | -1.68 | -6.48 |
| 17. | 17. | ↓ 15. | HBase + | Wide column store | 58.80 | -1.97 | -4.72 |
| 18. | 18. | ↑ 20. | Hive + | Relational DBMS | 57.94 | +0.32 | +10.64 |
| 19. | 19. | ↓ 17. | FileMaker | Relational DBMS | 56.05 | -0.33 | -3.60 |
| 20. | 20. | ↓ 19. | SAP HANA + | Relational DBMS | 51.93 | +0.33 | +3.96 |

FIGURE III.7 – Un classement des SGBDs sur *DB-Engines Ranking* du 1 août 2018

Source : <https://db-engines.com/en/ranking>, consultée le 01/08/2018.

III.3.4 Schema on Write VS Schema on Read

Lors du chargement des données depuis leurs sources de stockage, on distingue deux approches : *Schema on Write* et *Schema on Read*.

Dans la première, il faut définir les colonnes, le format de données, les types, etc. La lecture des données est rapide et moins coûteuse étant donné l'effort entrepris pour définir la structure. C'est le cas des bases de données relationnelles.

Dans la deuxième, les données sont chargées telles qu'elles sont, sans transformations ou changements. L'interprétation de ces données se fait lors de la lecture, et cela dépend des besoins pour lesquels les données sont analysées. Ainsi, les mêmes données peuvent être lues de différentes manières. Par exemple, l'action de lire les données d'une colonne, qu'elles soient de type entier ou bien chaîne de caractère d'un fichier CSV est la même, mais le type de la donnée qui diffère. C'est l'approche utilisée par Amazon Athena (voir III.4.3). Les Figures III.8 et III.9 illustrent la différence entre ces deux approches.

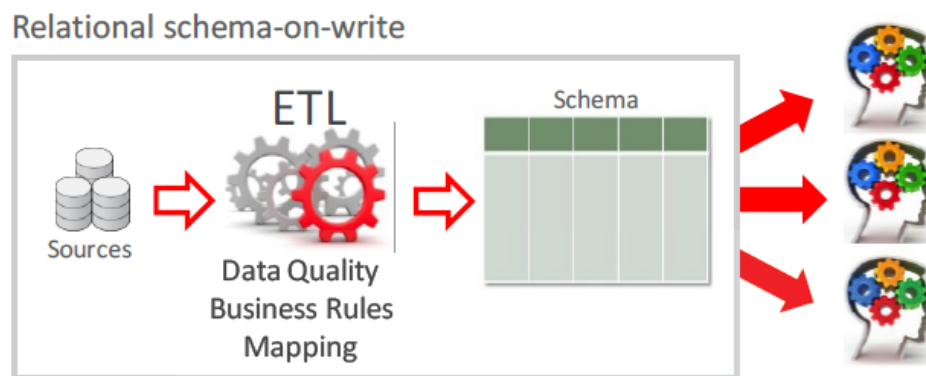


FIGURE III.8 – Schema on Write (SGBDR)

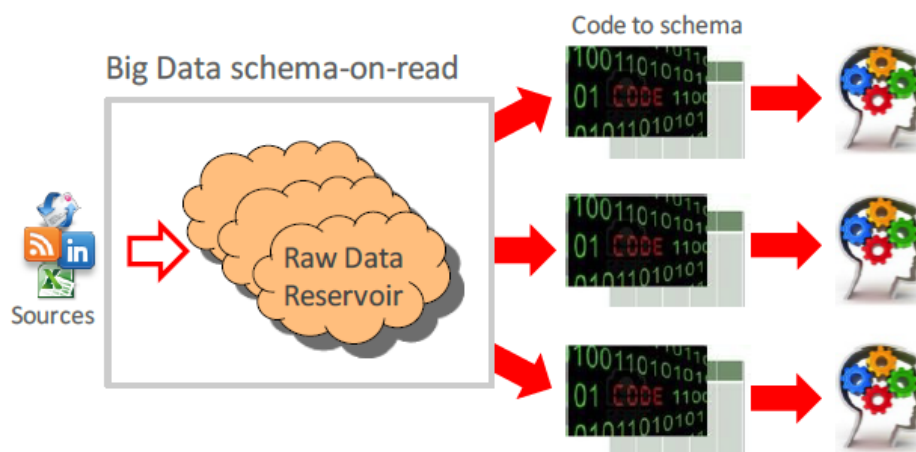


FIGURE III.9 – Schema on Read (Big Data)

Source : <https://blogs.oracle.com/datawarehousing/big-data-sql-quick-start-schema-on-read-and-schema-on-write-part11>, consultée le 05/08/2018.

La meilleure approche dépend des besoins de l'analyse. La première approche est meilleure en performances, en revanche, la deuxième est tolérante aux erreurs humaines.

III.3.5 L'informatique distribuée et l'analyse de données massives

Il existe deux stratégies pour appliquer des traitements sur un grand ensemble de données :

- Par distribution des traitements (*scaling* des traitements) : les traitements sont distribués sur un nombre de nœuds important. De ce fait, les données sont amenées jusqu'à ces nœuds ;
- Par distribution des données (*scaling* des données) : les données sont distribuées sur un nombre important de nœuds. D'ailleurs cela permet de stocker un maximum de données.

Il s'agit d'amener les traitements aux machines sur lesquelles les données sont stockées. Du fait que le stockage de données est réparti sur plusieurs machines, il est possible de traiter des données très volumineuses en un temps optimal. La première mise en œuvre de cette approche est le schéma Map-Reduce.

MapReduce est un patron d'architecture de développement informatique, inventé par Google, dans lequel sont effectués des calculs parallèles, et souvent distribués, de données potentiellement très volumineuses, typiquement supérieures en taille à 1 téraoctet^a.

a. Source : <https://fr.wikipedia.org/wiki/MapReduce>, consultée le 20/12/2018.

III.4 Parcours de quelques technologies du Big Data

La liste des technologies dans le domaine du Big Data est en expansion continue pour répondre au mieux aux besoins de l'analyse de données massives. C'est pourquoi nous allons parcourir une liste non exhaustive des technologies liées au Big Data. En particulier, ce sont les technologies expérimentées pour analyser les traceroutes en provenance du RIPE Atlas.

III.4.1 MongoDB

MongoDB⁵ est une base de données NoSQL de type Document⁶. MongoDB est classé parmi les SGBDs adoptant le couple CP (Consistency et Partition Tolerance) dans le théorème CAP⁷. Une base de données créée dans MongoDB est un ensemble de collections. Une collection dans MongoDB est équivalente à une table dans un SGBDR.

En 2016, MongoDB devient disponible en mode cloud sous le nom MongoDB Atlas⁸. Il est distribué à travers les trois fournisseurs du cloud : Amazon Web Services (AWS), Google Cloud Platform et Microsoft Azure. En terme de tarifs, plusieurs formules sont proposées⁹, y inclut l'offre gratuite pour expérimenter MongoDB Atlas. Les frais d'utilisation du service MongoDB Atlas dépend du stockage, de la RAM allouées et des options choisies. Les documents sont stockés dans MongoDB sous format BSON.

BSON (ou Binary JSON) est un format utilisé pour stocker et transférer les données dans la base de données MongoDB. BSON facilite la représentation des structures de données simples et des tableaux associatifs^a.

a. Source : <https://fr.wikipedia.org/wiki/BSON>, consultée le 02/08/2018.

III.4.2 Amazon DynamoDB

5. Source : <https://www.mongodb.com/>, consultée le 02/08/2018.

6. Une base de données NoSQL de type document est décrite dans la section III.3.3.

7. Le théorème CAP est décrit dans la section III.3.3

8. Source : <https://www.mongodb.com/cloud/atlas>, consultée le 02/08/2018.

9. Source : <https://www.mongodb.com/cloud/atlas/pricing>, consultée le 02/08/2018.

Amazon DynamoDB¹⁰ est une base de données NoSQL de type clé-valeur distribuée, gérée par les services d'Amazon. Elle est capable de stocker un volume important de données limité par la capacité de l'infrastructure d'AWS. Amazon DynamoDB est un service simple et facile à utiliser, il ne nécessite aucune configuration préalable.

Amazon DynamoDB est une base de données évolutive de façon abstraite pour l'utilisateur final. Elle offre des performances constantes à une échelle essentiellement infinie, limitée uniquement par la taille physique du cloud AWS. Elle est flexible. Aucun schéma n'est requis pour stocker les données. Les frais d'utilisation de ce service dépendent de trois éléments¹¹ :

- la quantité de données stockées : DynamoDB est facturé par Go d'espace disque utilisé (0, 250 USD par Go par mois) ;
- la capacité en lecture par seconde (0, 470 USD par unité de capacité d'écriture par mois) ;
- la capacité en écriture par seconde (0, 090 USD par unité de capacité de lecture par mois) ;

III.4.3 Amazon S3, Amazon Glue et Amazon Athena

La combinaison d'Amazon S3, Amazon Glue et Amazon Athena permet de créer un environnement Big Data capable d'assurer le stockage de données, le chargement de données et l'interrogation de données.

Amazon S3¹² est un service de stockage d'objets dans le cloud. Il est conçu pour stocker et récupérer toute quantité de données. Il peut assurer 99, 999999999 % de durabilité. La sécurité et l'accès aux données sont assurés. Il existe plusieurs classes de stockage qui répondent aux différents besoins.

Les fichiers des données sont organisés dans ce qu'on appelle un compartiment, c'est une simulation de dossier dans un système d'exploitation. A l'intérieur d'un compartiment, il est possible de créer des compartiments imbriqués. C'est une simulation d'arborescence de dossiers car physiquement cet arborescence n'existe pas. En ce qui concerne les frais du service AWS S3, le tableau III.1 décrit les tarifs de la formule standard.

| Région | UE (Irlande) |
|---------------------------------------|-------------------|
| Première tranche de 50 To/mois | 0, 023 USD par Go |
| 450 To suivants/mois | 0, 022 USD par Go |
| Plus de 500 To/mois | 0, 021 USD par Go |

TABLE III.1 – Les tarifs du AWS S3 (formule Stockage standard S3)

Source : <https://aws.amazon.com/fr/s3/pricing/>, consultée le 05/08/2018.

Amazon Glue¹³ est un service d'extraction, de transformation et de chargement. L'objectif de ce service est de découvrir les données, les transformer et les rendre accessibles à la recherche et à l'interrogation. Amazon Glue est utile pour la construction des entrepôts de données ; il découvre les métadonnées relatives aux magasins de données et les rendre accessibles dans un

10. Source : <https://aws.amazon.com/fr/dynamodb/>, consultée le 02/05/2018.

11. Source : <https://aws.amazon.com/fr/dynamodb/pricing/>, consultée le 02/05/2018.

12. Source : <https://aws.amazon.com/fr/s3/>, consultée le 06/07/2018.

13. Source : <https://aws.amazon.com/fr/glue/>, consultée le 06/07/2018.

catalogue central. En prenant en entrée les données présentes dans un compartiment dans Amazon S3, Amazon Glue découvre le schéma de ces données. Il dispose de plusieurs classificateurs intégrés pour la découverte des données. Par exemple un classificateur pour trouver le schéma de données en format JSON, XML, etc. Si les classificateurs intégrés ne répondent pas aux besoins particuliers, il est possible de créer des classificateurs personnalisables.

Les frais de ce service dépendent du temps écoulé lors de l'analyse des données par les robots d'analyse durant la découverte du schéma. A ces frais, ils s'ajoutent les frais du catalogue de données qui va être peuplé par les résultats fournis par les robots de l'analyse. Par exemple, on paye 0,44 USD par heure par DPU¹⁴, il est facturé à la seconde avec un minimum de 10 minutes par robot d'analyse exécuté. Plus de détails sont disponibles sur Amazon Glue¹⁵.

Amazon Athena¹⁶ est un service de requêtes interactif. Il permet d'interroger les données présentes dans Amazon S3 avec des requêtes SQL plus avancées. Le service Amazon Athena est considéré comme *serverless*. Amazon Athena utilise l'approche *schema-on-read* (voir la section III.3.4) afin de projeter le schéma donné en entrée sur les données au moment de l'exécution de la requête SQL demandée. Le schéma sur lequel les données peuvent être projetées peut être créé manuellement ou bien utiliser le catalogue créé dans Amazon Glue. Le service Amazon Athena est facturé suivant la quantité de données analysée. Précisément, 5 USD par To de données analysées.

Une *requête est interactive* si on peut obtenir immédiatement une réponse à la requête. Dans le cas échéant, les résultats sont obtenus dans le cadre d'un code source pour un des langages de programmation, souvent à travers une API.

Serverless peut être décomposé en *server* et *less*. Un outil est *serverless* quand l'utilisateur final de cet outil peut l'utiliser sans se soucier de toute configuration ou gestion des serveurs derrière ce service. C'est un mécanisme présent beaucoup sur le cloud.

L'exécution des requêtes SQL est effectuée par le moteur de requêtes SQL Presto. pour les instruction DDL, elles sont effectuées par *Hive Data Definition Language*¹⁷. Les requêtes DDL incluent la création, la suppression et la mise à jour de la structure de la table dans le cas d'une base de données relationnelles, d'une collection, d'une vue, etc.

14. DPU : unité de traitement des données.

15. Source : <https://aws.amazon.com/fr/glue/pricing/>, consultée le 05/08/2018.

16. Source : <https://aws.amazon.com/fr/athena/>, consultée le 06/07/2018.

17. Source : <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>, consultée le 05/08/2018.

Presto^a est un moteur de requêtes SQL open source destiné au Big Data. Il permet d'exécuter des requêtes analytiques interactives sur des données de taille importante ; jusqu'à des Pétaoctets de données.

Presto interroge les données où elles sont hébergées. Ce qui inclut les bases de données relationnelles, Amazon S3 et autres dépôts propriétaires. De plus, une même requête SQL peut combiner plusieurs sources de données. C'est intéressant pour les organisations ayant plusieurs sources de données. Il fournit les résultats en quelques secondes, voire quelques minutes. Il supporte les types de données complexes comme les objets JSON, un tableau d'éléments, etc. Il supporte aussi des opérations complexes sur les données.

a. Source : <http://prestodb.io/>, consultée le 01/08/2018.

Hive Data definition language (DDL) est un sous-ensemble de déclarations qui décrivent la structure de données dans Apache Hive. Principalement, ce sont les instructions de création, suppression et de mise à jour de la structure des objets comme les bases de données, les tables, les vues et autres.

III.4.4 Apache Spark

Apache Spark¹⁸ est un framework de calcul distribué. C'est un ensemble de composantes conçues pour assurer la rapidité, la facilité d'utilisation ainsi que la flexibilité dans l'analyse des données à grande échelle. Plusieurs APIs sont disponibles pour interagir avec Spark et appliquer les transformations sur les données à analyser.

Core Concepts et architecture de Spark

Spark Clusters et Resource Management System Spark est un système distribué conçu pour traiter les données massives rapidement et avec efficacité. Ce système est déployé sur un ensemble de machines, qu'on appelle *Spark cluster*. La taille du cluster en nombre de machines est variable, il existe un cluster avec peu de machines mais aussi un cluster avec des milliers de machines. En vue de gérer efficacement les machines d'un cluster, les entreprises recourent à un système de gestion de ressources tel que Apache YARN¹⁹ ou Apache Mesos²⁰. Les deux composantes les plus importantes dans un système de gestion de ressources sont : le *cluster manager* et le *worker*.

Le *cluster manager* a une vue globale de l'emplacement des *workers* ; la mémoire qu'ils ont et le nombre de cœurs CPU dont chaque worker dispose. Le rôle du *cluster manager* est d'orchestrer le travail en le désignant à chaque worker. Tandis que le rôle d'un *worker* est de fournir les informations utiles pour le *cluster manager* ainsi que la réalisation du travail y assigné. La Figure III.10 montre l'interaction entre une application spark, le cluster manager et les *workers*.

18. Source : <https://spark.apache.org/>, consultée le 14/12/2018

19. Description dans <https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html>, consulté le 09/12/2018.

20. Site officiel <https://mesos.apache.org/>, consulté le 09/12/2018.

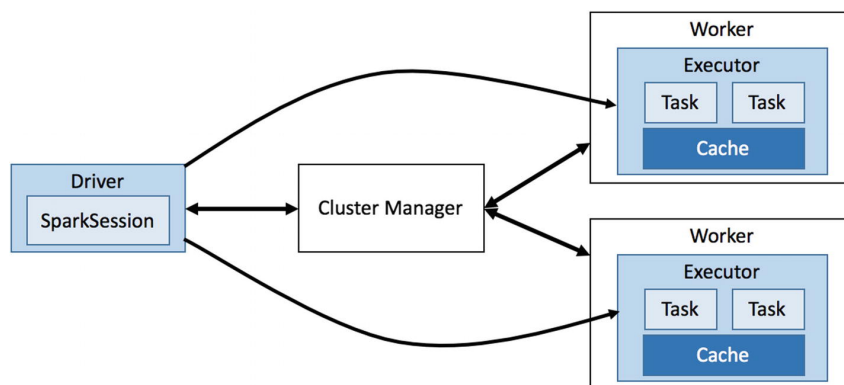


FIGURE III.10 – Interaction entre une application Spark et le cluster manager. Source : [25]

Application Spark Une application Spark consiste en deux parties. La première partie concerne la logique décrivant les traitements à appliquer sur les données. Cette logique est décrite en utilisant les APIs²¹ disponibles. La deuxième partie est appelée le *driver*, c'est le coordinateur principal d'une application Spark. Le driver interagit avec le cluster manager afin de trouver les machines sur lesquelles le traitement de données doit être réalisé. Ainsi, pour chacune de ces machines, le driver Spark lance le processus *executor* en passant par le cluster manager. Un autre rôle du driver Spark est de gérer et de distribuer les tâches Spark en provenance de l'application Spark sur chaque executor. Pour précision, dans la Figure III.10, la classe *SparkSession* est le point d'entrée vers une application Spark.

Spark driver et executor Chaque Spark *executor* est alloué exclusivement à une application Spark spécifique et la durée de vie d'un *executor* est celle de l'application Spark.

Spark utilise l'architecture master-slave. Spark *driver* est le master et Spark *executor* est le slave. De ce fait, une application Spark n'a qu'un seul Spark driver et plusieurs Spark *executors*. Chaque Spark *executor* s'occupe d'un traitement sur une partie de la totalité des données à analyser. De cette manière, Spark est capable de traiter les données de façon parallèle. La Figure III.11 illustre un exemple d'un cluster. Ce dernier est composé de trois *executors*.

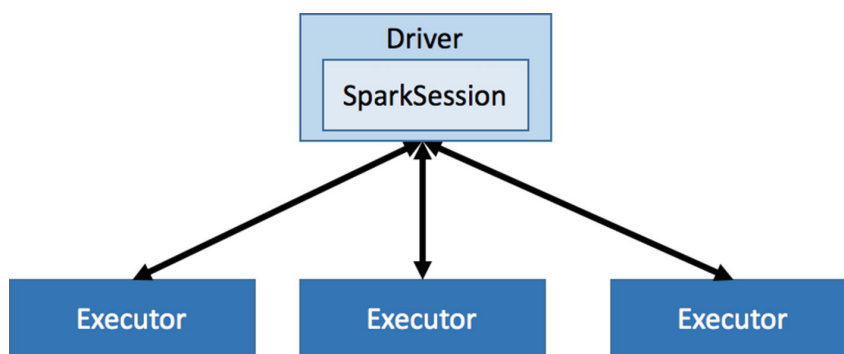


FIGURE III.11 – Un exemple d'un cluster formé de trois executors. Source : [26]

Spark Unified Stack Spark offre ce qu'on appelle Spark *Stack*. C'est un ensemble de composantes construites dessus la composante Spark Core. Ces composantes sont conçues pour répondre à des besoins spécifiques :

21. API en Java, Scala, Python ou R.

- Spark SQL est conçu pour le traitement interactive ;
- Spark Streaming est utilisé pour les traitements en temps réel ;
- GraphX est destiné au traitement de graphe ;
- MLib est conçu pour machine learning ;
- SparkR est consacré au traitement lié au machine learning en utilisant R.

Spark Core est la base du moteur Spark pour le traitement distribué de données. On distingue deux parties formant Spark Core. Premièrement, la partie concernant l'infrastructure distribué du calcul. Cette dernière est responsable de la distribution, la coordination et de la planification des tâches sur les différentes machines formant le cluster. De plus, cette partie gère l'échec d'un traitement donné et le transfert de données entre les machines. Le deuxième élément formant Spark Core est appelé RDD (Resilient Distributed Dataset). Un RDD est une collection partitionnée d'objets, tolérante aux pannes et en lecture seule. La Figure III.12 présente les différentes entités du Spark Unified Stack avec Spark Core.

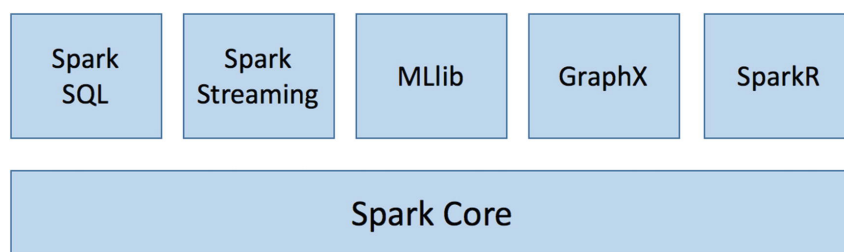


FIGURE III.12 – Spark Unified Stack. Source : [27]

Resilient Distributed Datasets

Spark dispose d'une abstraction notée Resilient Distributed Datasets (ou RDDs). Un RDD est une collection d'objets immuable. Ces objets sont répartis sur les nœuds du cluster afin d'être traités en parallèle. Un RDD peut être persisté dans la mémoire pour éventuelle réutilisation. D'ailleurs de cette façon, on constate l'amélioration des performances. En outre, un RDD peut être persisté sur un disque.

Les RDDs supportent deux types d'opérations sur les objets stockés : les transformations et les actions. Une transformation appliquée sur un RDD crée un nouveau RDD, par exemple la transformation *filter* retourne un RDD ayant vérifié la condition donnée en entrée. Toutefois, une action appliquée sur un RDD retourne une seule valeur, par exemple l'action *count* calcule le nombre d'objets d'un RDD. La Figure III.13 illustre un flux de données avec l'utilisation de Spark.

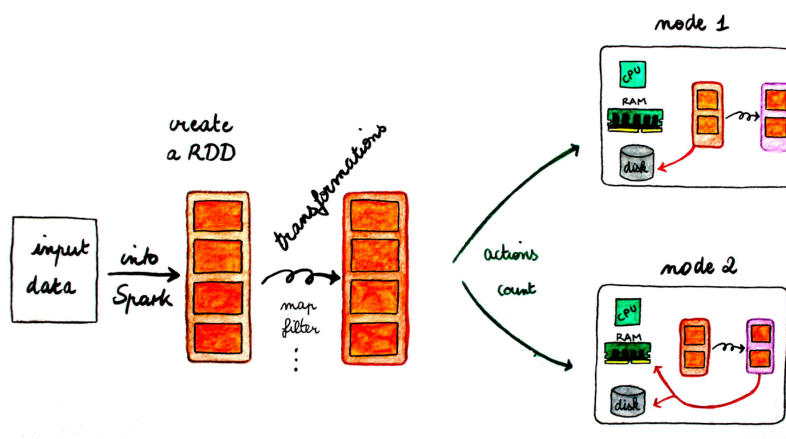


FIGURE III.13 – Exemple d'un flux de données avec Spark

Source : <https://www.duchess-france.org/starting-with-spark-in-practice/>, consultée le 15/12/2018.

input data sont les données à analyser en utilisant Spark. Ces données sont récupérées depuis des sources extérieures vers Spark. Ce dernier crée un RDD basé sur ces données. Un RDD est représenté par le rectangle en orange, les morceaux en orange dans le rectangle représentent les partitions d'un RDD. Il existe plusieurs transformations à appliquer sur un RDD, avec la possibilité d'enchaîner plusieurs transformations. Comme une transformation est à la base *lazy*, les partitions sont partagées sur les nœuds du cluster qu'à suite à l'appel d'une action. Une fois une partition est localisé sur un nœud donné, les transformations ainsi que les actions peuvent s'enchaîner.

En cas de perte de partition pour une raison ou une autre, Spark est capable de reproduire automatiquement la partition en question. Cette fonctionnalité est assurée via le DAG (Direct Acyclic Graph). Dans ce graphe, Spark enregistre toutes les opérations appliquées sur un RDD.

III.5 Conclusion

Dans ce chapitre, nous avons décrit brièvement quelques technologies du Big Data, car la liste de toutes les technologies est très longue. Afin de découvrir ces technologies en pratique, nous allons aborder dans le chapitre IV l'utilisation de ces technologies dans le cas de l'analyse des délais d'un lien décrite dans la chapitre II.

Chapitre IV

Application de quelques technologies Big Data sur l'analyse des traceroutes

IV.1 Introduction (to adapt)

Ce chapitre reprend un ensemble de technologies destinées à la manipulation des données massives. Ce sont les technologies que nous avons expérimenté pour analyser les traceroutes disponibles dans le dépôt de RIPE Atlas. On ne peut pas comparer ces technologies entre elles car elles ne se trouvent pas dans la même catégorie ; quelques technologies n'assurent que le stockage, une autre technologie gère l'analyse ainsi que le stockage. En revanche, ce sont des technologies permettant d'analyser les données massives telles qu'elles représentent une ou plusieurs étapes d'un processus d'analyse de données (voir un exemple du processus d'analyse de données dans [III.2](#)). L'évaluation des performances de chacune des technologies est faite sur une machine ayant les caractéristiques reprises dans le tableau[!].

IV.2 Critères d'évaluation des technologies du Big Data

Les critères d'évaluation d'une technologies par rapport à une autre dépend de plusieurs entrées. En ce qui concerne les critères sur lesquels nous allons évaluer les différents technologies du Big Data sont les suivants :

Frais d'utilisation d'un technologie.

Temps d'exécution nécessaire pour fournir les résultats finaux.

Flexibilité du schéma de données

L'évolutivité de l'environnement Big Data mis en place pour les nouvelles données.

Configuration de l'environnement

Limitations en stockage de données

Analyse temps réel

Networking et bande passante

IV.3 MongoDB

Application sur MongoDB

MongoDB est la technologie Big Data utilisé par R. Fontugne dans l'implémentation de l'outil de détection. Les traceroutes sont organisés, dans MongoDB, dans des collections. Chaque collection stocke les traceroutes effectués lors de la journée *YYYY_MM_DD* et en adressage *V*. Dans le cas de l'adressage IPv4, *V* ne prend aucune valeur, or, *V* prend la valeur 6 s'il s'agit de l'adressage IPv6. La nomination structurée des collections permet de ne récupérer que les traceroutes concernés par une analyse lancée. Ainsi, le nom d'une collection est structuré comme suit : *tracerouteV_YYYY_MM_DD*.

MongoDB est une technologie conçue pour assurer le stockage de données dans un processus d'analyse de données. En particulier, MongoDB supporte les données non structurées. Par exemple, dans certains cas, les traceroutes planifiés ne réussissent pas à atteindre une destination, dans ce cas le contenu d'un traceroute est différent de celui réussi. De plus, MongoDB est adapté non seulement aux projets visant la lecture de données massives mais aussi aux projets où on envisage la mise à jour d'un objet (modification ou suppression). Généralement l'analyse de données à grande échelle se limitent qu'au mode lecture.

Malgré la convenance de MongoDB aux données non structurées et massives, l'utilisation de telle base de données, en version locale, nécessite l'ajustement de la machine locale où MongoDB tourne.

Les limitations de MongoDB L'implémentation proposée de l'outil de détection utilise la version locale de la base de données MongoDB pour le stockage des données. La quantité de données dont MongoDB peut stocker dépend de l'espace mémoire de stockage disponible dans la machine dans laquelle MongoDB est installé. De plus, les performances d'une détection lancée concernant une période donnée dépendent de la RAM de la machine en question. Pour conclure, l'utilisation de la version locale d MongoDB pour analyser les traceroutes à travers l'outil de détection dépend typiquement de la machine locale.

IV.4 Amazon DynamoDB

Application sur les traceroutes

L'évolutivité est une des caractéristiques attirantes des services web d'Amazon. En particulier, c'est le cas d'Amazon DynamoDB. Une implémentation basée sur Amazon DynamoDB n'a pas à se soucier de la capacité de stockage de données. Toutefois, au moment de la récupération et de la manipulation de ces données, il faut ajuster les ressources pour pouvoir récupérer et traiter une quantité importante de données. A titre indicatif, une heure de traceroutes fait en moyen 620 MB en format compressé, ce que revient à environ 9 GB en format texte.

IV.5 Amazon S3, Amazon Glue et Amazon Athena

Application sur les traceroutes

Le processus de l'analyse combinant les trois services d'Amazon (Amazon S3, Amazon Glue et Amazon Athena) est illustré par la Figure [IV.1](#)¹.

1. Amazon Redshift et Amazon Quicksight ne sont pas utilisés ici.

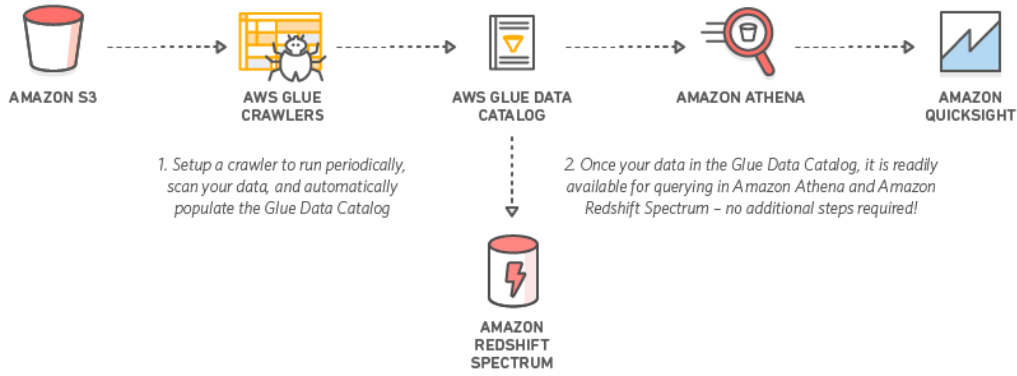


FIGURE IV.1

Source : https://docs.aws.amazon.com/fr_fr/athena/latest/ug/glue-best-practices.html, consultée le 16/12/2018.

Afin d'utiliser Amazon Athena, nous avons besoin du schéma des données. Il s'agit d'une table comme les tables dans un SGBDR. Pour ce faire, nous avons lancé avec Amazon Glue la détection automatique du schéma d'un ensemble de traceroutes enregistrés dans un fichier faisant 500 MB. Toutefois, la détection a échoué. Autrement dit, Amazon Glue n'a pas pu inférer le schéma d'une seule table capable de lire tout traceroute dans ce fichier. L'échec de l'inférence est dû au fait que le fichier contient des traceroutes différents en terme de structure, car la structure dépend du firmware de la sonde ayant effectué le traceroute. Pour finir, le schéma des traceroutes a été créé manuellement (voir la section A.1 dans l'annexe A). La table a été créée en utilisant le partitionnement des données dans un compartiment S3. Plus de détails sur le partitionnement sont données dans la section A.2 dans l'annexe A.

Une fois les fichiers de données sont synchronisés vers le compartiment AWS S3 et le schéma de données est créé, on passe à l'interrogation de données en utilisant les requêtes SQL basées sur Presto.

Pour intégrer Amazon Athena dans l'outil de détection, on distingue deux possibilités. La première possibilité n'utilise Athena que pour récupérer, dans la machine locale, les traceroutes vérifiés en terme de validité (étape 1 et 2 dans II.4.4). Les traitements qui suivent (étapes à partir de 3) sont effectués dans la machine locale. Dans ce cas, l'utilisation des technologies du Big Data est limité qu'au niveau stockage de données massives. Tandis que la deuxième possibilité vise la maximisation des traitements au sein de l'infrastructure d'Athena. De ce fait, la machine locale n'a qu'à recevoir les derniers résultats de la détection. Pour la deuxième possibilité, les données doivent être manipulées de sorte à maximiser, au niveau d'Amazon Athena, les traitements relatives aux étapes décrites dans la section II.4.4.

Ainsi le défi est de trouver la requête ou bien l'ensemble de requêtes SQL à exécuter sur Athena en vue d'avoir l'évolution du RTT différentiel d'un lien donné.

Supposons qu'il existe une requête SQL capable de trouver les liens possibles avec leur RTT différentiel. A l'étape 4 dans II.4.4, on construit la distribution des RTTs différentiels pour tout lien l identifié dans les traceroutes de la période d_i . Cette distribution est mise à jour à chaque l identifié dans un des traceroutes de la période d_i . Soient $T_i = \{t_{i,j}\}$ l'ensemble de traceroutes effectués durant d_i , $j \in [1, R]$ et R est le nombre de traceroutes durant d_i . On peut décrire le parcours des traceroutes brièvement dans le pseudo-code suivant, sachant que les détails ne sont données, l'objectif est d'illustrer la convenance d'Athena au traitement souhaité :

Algorithm 1 Une partie de l'étape 4 du processus de la détection des anomalies des délais

```

1: for all  $t_{i,j} \in T_i$  do
2:    $links \leftarrow \text{getLinksFromTraceroute}(t_j)$ 
3:   for all  $l \in links$  do
4:      $\text{updateLinkRttDistribution}(l)$ 
5:   end for
6: end for

```

Avec :

- $\text{getLinksFromTraceroute}(t_j)$ énumère tous les liens possibles dans le traceroute t_j .
- $\text{updateLinkRttDistribution}(l)$ ajoute le RTT différentiel du lien l , précédemment calculé, à la distribution des RTTs différentiels courante de ce lien.

Le service Athena est conçu pour la lecture de données, toute mise à jour de données n'est pas possible avec ce service. C'est pourquoi la distribution des RTTs différentiels de chaque lien identifié doit être sauvegardée dans un endroit accessible en lecture et en écriture, par exemple dans un compartiment AWS S3. Que ce soit un fichier reprenant la distribution des RTTs différentiel par un seul lien ou bien un fichier pour tous les liens, à la ligne 4 du pseudo-code 1, un fichier doit être lu et mise à jour avec de nouvelle valeur. Pour une période d_i d'une heure, le nombre de traceroutes est de l'ordre de milliers. Chaque traceroute peut inclure L liens. Dans ce cas, le nombre de mise à jour de la distribution des RTTs différentiels est $R \times L$. Avec une autre technologie qui travaille en mémoire, les résultats sont donnés plus rapidement.

Evaluation des performances d'Amazon S3, Amazon Glue et Amazon Athena

IV.6 Spark Apache avec Scala

Les fonctionnalités de Spark sont accessibles avec les APIs en Scala, Java et Python. Nous avons choisi l'utilisation de l'API en Scala parce que Scala est le langage natif de Spark. De plus, Scala est interopérable avec Java. En pratique, on travaille sur les collections d'objets, sur lesquelles on applique des fonctions. Dans une implémentation en Spark, nous avons besoin des classes modélisant les objets tout au long de l'analyse de données. De plus, nous avons besoin de définir les fonctions à appliquer tout au long des étapes de cette analyse.

IV.6.1 Application sur traceroutes

Comme complément aux étapes décrites dans II.4.4, on présente les différentes classes permettant de modéliser les données tout au long du processus de l'analyse. La définition de ces classes est liée au langage *Scala*.

Soient les classes suivantes utilisées :

La classe Signal modélise un signal². Ainsi, *from* est l'adresse IP du routeur émettant ce signal, *rtt* est le Round Trip Time entre la sonde Atlas et ce routeur et enfin *x* est un indicateur de l'échec du signal.

2. Un signal dans le contexte d'un traceroute.

Listing IV.1 – La classe Signal en Scala

```
case class Signal(
    rtt : Option[Double],
    x : Option[String],
    from : Option[String])
```

La classe Hop modélise un saut dans un traceroute. On caractérise un saut par son identifiant noté *hop*. Celui-ci prend comme valeur un entier commençant à 1 et la liste des signaux relatifs à ce saut notée par *result*. Généralement un saut est représenté par 3 signaux.

Listing IV.2 – La classe Hop en Scala

```
case class Hop(
    var result : Seq[Signal],
    hop : Int)
```

La classe Traceroutes modélise le résultat d'une requête traceroute effectuée par une sonde Atlas. Cette modélisation se limite aux données qui nous intéressent dans la présente analyse.

dst_name représente l'adresse IP de la destination de la requête traceroute, *from* est l'adresse IP de la sonde, *prb_id* est l'identifiant de la sonde, *msm_id* est l'identifiant de mesure, *timestamp* est le temps auquel la requête traceroute a été effectuée et enfin on trouve la liste des sauts qui représentent les routeurs traversés par le trafic entre la source et la destination.

Listing IV.3 – La classe Traceroute en Scala

```
case class Traceroute(
    dst_name : String,
    from : String,
    prb_id : BigInt,
    msm_id : BigInt,
    timestamp : BigInt,
    result : Seq[Hop])
```

La classe TraceroutesPerPeriod permet de présenter les traceroutes après les avoir trié suivant la période pendant laquelle ils ont été effectués. *timeWindow* est le temps unix marquant le début de la période³ et *traceroutes* est la liste des traceroutes effectués pendant cette période.

A l'étape 2, l'objectif était d'agréger les signaux par routeur source et ensuite calculer la médiane des RTTs par ce routeur. Par conséquent, un traceroute est présenté différemment, ce qui est illustré par la classe *MedianByHopTraceroute*.

La classe PreparedSignal est une agrégation de tous les signaux, d'un saut donné, par le routeur *from*, la médiane des RTTs calculée est présentée par *medianRtt*.

Listing IV.4 – La classe PreparedSignal en Scala

```
case class PreparedSignal(
    medianRtt : Double,
    from : String)
```

3. Pour précision, la fin de la période peut être inférée en prenant deux débuts de deux périodes car la durée d'une période est fixe tout au long de l'analyse.

La classe PreparedHop modélise un saut après avoir agrégé ses signaux.

Listing IV.5 – La classe PreparedHop en Scala

```
case class PreparedHop(
  var result : Seq[ PreparedSignal ],
  hop :      Int )
```

La classe MedianByHopTraceroute modélise un traceroute après avoir agrégé ses sauts. Par rapport au traceroute d'avant l'agrégation, seule la liste des sauts a subi un changement.

Listing IV.6 – La classe MedianByHopTraceroute en Scala

```
case class MedianByHopTraceroute(
  dst_name : String ,
  from :    String ,
  prb_id :  BigInt ,
  msm_id :  BigInt ,
  timestamp : BigInt ,
  result :  Seq[ PreparedHop ])
```

La classe Link modélise un lien topologique. Ce dernier est défini par deux adresses IP *ip1* et *ip2* et par son RTT différentiel calculé *rttDiff*.

Listing IV.7 – La classe Link en Scala

```
case class Link(
  ip1 : String ,
  ip2 : String ,
  rttDiff : Double)
```

La classe LinksTraceroute permet de modéliser un traceroute après avoir inféré tous les liens de ce dernier. Ainsi, la liste des sauts est remplacée par la liste des liens (*links*).

Listing IV.8 – La classe LinksTraceroute en Scala

```
case class LinksTraceroute(
  dst_name : String ,
  from :    String ,
  prb_id :  BigInt ,
  msm_id :  BigInt ,
  timestamp : BigInt ,
  links :   Seq[ Link ])
```

A l'étape 5, l'objectif était de passer d'un traceroute à une liste de liens caractérisés par les informations générales sur la sonde Atlas, la mesure Atlas, etc. Chaque élément de cette liste est représenté par la classe *DiffRtt*, où *LinkIPs* représente les deux adresses IP d'un lien donné.

La classe LinkIPs permet représenter un lien par seulement ses deux adresses IP *ip1* et *ip2*.

Listing IV.9 – La classe LinkIPs en Scala

```
case class LinkIPs(
```

```
ip1 : String ,
ip2 : String )
```

La classe DiffRtt est une représentation plus détaillée d'un lien, en plus de son RTT différentiel, on ajoute d'autres informations. Les adresses IP d'un lien sont modélisées par la classe *LinkIPs*.

Listing IV.10 – La classe DiffRtt en Scala

```
case class DiffRtt(
  rtt : Double ,
  var link : LinkIPs ,
  probe : BigInt )
```

A l'étape 6.3, on souhaite normaliser les dates de chaque lien; peu importe le moment pendant lequel le traceroute a été effectué durant une période d_i , on note seulement le début de cette période. Ainsi, la classe *DiffRTTPeriod* reprend un *lien* donné, les différentes sondes Atlas ayant identifié ce lien (*probes*), les RTTs différentiels de ce lien tout au long de la période et enfin les dates associées à chaque RTT différentiel.

La classe DiffRTTPeriod

Listing IV.11 – La classe DiffRTTPeriod en Scala

```
case class DiffRTTPeriod(
  link : LinkIPs ,
  probes : Seq[ BigInt ] ,
  rtt : Seq[ Double ] ,
  var dates : Seq[ Int ] )
```

A la fin des opérations de l'étape 6, on reprend pour chaque période, pour un lien donné, les RTTs différentiels ainsi que leurs dates. Ensuite, on construit les bornes de l'intervalle de confiance courants pour ce lien et les bornes de l'intervalle de confiance de référence, et ce afin de comparer ces deux intervalles en vue d'inférer les anomalies possibles du délais de ce lien.

La classe LinkState permet de modéliser les intervalles de confiance d'un lien pendant une période d_i donnée. *valueLow* est la borne inférieure de l'intervalle de confiance, *valueHi* est la borne supérieure de l'intervalle de confiance, *valueMedian* est la médiane des RTTs différentiels et enfin *valueMean* est la moyenne des RTTs différentiels. Pour précision, les données concernant l'état d'un lien sont sous forme d'une liste. L'idée est de garder l'historique de ces valeurs durant toute la période de l'analyse. Cette historique est exploitée pour tracer l'évolution du RTT différentiel du lien. Cependant, la comparaison utilise les valeurs du dernier état du lien.

Listing IV.12 – La classe LinkState en Scala

```
case class LinkState(
  var valueMedian : Seq[ Double ] ,
  var valueHi : Seq[ Double ] ,
  var valueLow : Seq[ Double ] ,
  var valueMean : Seq[ Double ] )
```

| | |
|----------|--|
| Type | OpenVZ container |
| RAM (MB) | 32768 |
| CPU | 64 (The logical CPU number of a CPU as used by the Linux kernel) |

TABLE IV.1 – Caractéristiques de la machine de test

| Durée (début - fin) | Taille (MB) | Nb traceroutes | Temps |
|---------------------|-------------|----------------|-------|
| 5 min | | | |
| 15 min | | | |
| 30 min | | | |
| 1h min | | | |
| 2h min | | | |
| 6h min | | | |
| 12h min | | | |
| 1 jour | | | |
| 2 jour | | | |
| 3 jour | | | |
| 4 jour | | | |

This is where authors provide additional information about the data, including whatever notes are needed.

TABLE IV.2

IV.7 Comparatif des performances

Les sections suivantes décrivent les performances obtenues en terme de temps écoulé durant à la fin d'une analyse. Nous allons aborder les performances pour les trois implémentations : MongoDB, Les trois services d'Amazon et Spark Apache.

IV.7.1 Caractéristique de l'environnement de test

Le tableau IV.1 présente les caractéristique de la machine sur laquelle nous avons effectué les différents tests.

IV.7.2 Spark Apache

IV.8 Récapitulatif des technologies Big Data

Le Tableau x reprend les temps d'exécution, exprimés en micro-secondes CPU, sur les différents ensembles de données. Ces tests ont été effectués sur une machine ayant les caractéristiques suivantes : Dell Dual Core, 2.66 GHz, 2 Gb RAM, système SuSE Linux 10.0 (kernel 2.4.2), java 1.5.0, etc. Pour calculer le temps CPU, la classe ThreadMXBean a été utilisée.

| | | | | | |
|---------------|---------------------------------|----------|-------------|---------------|--------------|
| MongoDB | DynamoDB | AWS (S3) | AWS (Glue) | AWS (Athena) | Apache Spark |
| Technologies | | | | | |
| Stockage | | | | | |
| traitement | | | | | |
| visualisation | | | | | |
| | flexible (schema-less database) | | | | |

TABLE IV.3 – Récapitulatif des technologies Big Data expérimentées dans l’analyse des délais

IV.8.1 Notes sur les données

- Champs manquants : timestamp

IV.9 Conclusion

ma conclusion.

Conclusion

Les données en provenance du RIPE Atlas analysées dans ce travail passent à l'échelle dès qu'on considère plusieurs heures d'échantillons. A travers ce travail, nous avons évalué des technologies Big Data pour analyser des échantillons de données.

Aujourd'hui, il existe de nombreuses technologies Big Data. Le choix d'une technologie ou une autre dépend de plusieurs facteurs. Dans un premier temps, nous avons utilisé deux technologies conçues pour le stockage de données à grande échelle : MongoDB et DynamoDB. Ensuite, nous avons expérimenté trois services d'Amazon, le premier pour le stockage de données, le deuxième pour la découverte de données et le troisième service permet est consacré pour l'interrogation de données. Enfin, nous avons utilisé un framework qui gère le traitement distribué de données dans un cluster de machines.

Nous avons à disposition une vérité de données, à titre indicatif, des années de mesures effectuées par les sondes Atlas. Notre premier objectif de l'évaluation est d'évaluer les performances des choix technique. C'est pourquoi nous n'avons pas défini des critères pour choisir l'ensemble de données. Nous avons évalué les performances des technologies en terme de temps écoulé tout au long de l'analyse de différents échantillons. Les performances de MongoDB dépendent de ... Toutefois, les performances des trois services d'Amazon dépendent de Tandis que Apache Spark dépend de

Si nous aurions plus de temps, nous aurions aimé évalué les performances de l'outil de détection conçu par Fontugne[16] en terme de précision dans la détection d'anomalies dans les délais des liens.

Annexe A

Amazon Athena

A.1 Création de la table traceroutes

La création d'une table reprend plusieurs parties :

- les colonnes de la table avec le type correspondant (int, string, array pour définir une liste, struct pour définir un objet) ;
- LOCATION : c'est l'endroit où les données sont stockées dans Amazon S3, il faut préciser le chemin vers le compartiment de données ;
- ROW FORMAT SERDE : elle définit la manière dont chaque ligne d'un fichier de données est sérialisée/désérialisée par Amazon Athena ;
- PARTITIONED BY : elle définit la manière dont les données sont organisées dans le compartiment de données ;
- WITH serdeproperties : elle définit les options de la sérialisation/désérialisation.

```
CREATE EXTERNAL TABLE traceroutes_api (  
    af int ,  
    bundle int ,  
    dst_addr string ,  
    dst_name string ,  
    fw int ,  
    endtime int ,  
    'from' string ,  
    group_id int ,  
    lts int ,  
    msm_id int ,  
    msm_name string ,  
    paris_id int ,  
    prb_id int ,  
    proto string ,  
    size int ,  
    src_addr string ,  
    'timestamp' int ,  
    ttr float ,  
    type string ,  
    result array< struct< hop:int , error:string , result:array<  
        struct<x:string , err:string , 'from':string , ittl:int , edst:string ,  
            late:int , mtu:int , rtt:float , size:int , ttl:int , flags:string ,  
            dstoptsize:int , hbhoptsize:int , icmpext:
```

```

        struct<version:int, rfc4884:int, obj:array<
            struct<class:int, type:int, mpls:array<struct<exp:
                int, label:int, s:int, ttl:int>>>>>>>>>>>>
        )
PARTITIONED BY (
    af_string,
    type_string,
    msm_string,
    year_string,
    month_string,
    day_string,
    hour_string
)
ROW FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe'
WITH serdeproperties ('paths'='af,bundle,dst_addr,dst_name,fw,endtime,
    from,lts,msm_id,paris_id,prb_id,proto,size,src_addr,timestamp,
    type,fw,msm_name')
LOCATION 's3://ripeatlasdata/traceroute/source=api/'

```

A.2 Partitionnement de données dans un compartiment AWS S3

A.2.1 Présentation du partitionnement

Le partitionnement de données présentes dans un compartiment Amazon S3 permet de limiter la quantité de données à analyser par une requête Amazon Athena. Le partitionnement améliore les performances d'Amazon Athena. D'une part, on obtient une réponse rapidement, d'autre part, on réduit les coûts engendrés suite à l'utilisation du service car on est facturé selon la quantité de données analysées.

Les partitions créées jouent un rôle similaire à celui d'une colonne durant l'interrogation d'une table dans Athena.

Prenons un exemple, nous avons des traceroutes ayant comme adressage IP la version 4 et d'autres traceroutes ont l'adressage IPv6 :

```

s3://ripeatlasdata/traceroute/
                                type=4/
                                type=6/

```

Sans l'utilisation du partitionnement et si on souhaite récupérer que les traceroutes ayant comme adressage IPv4, toutes les données (type = 4 et type = 6) sont analysées. Toutefois, en partitionnant les données suivant le type d'adressage, seuls les fichiers dans le dossier type = 4 qui sont analysés. Par conséquent, le partitionnement permet de réduire les coûts d'utilisation du service Amazon Athena, surtout si la quantité de données est très importante.

A.2.2 Application du partitionnement sur les traceroutes Atlas

Les traceroutes en provenance des sondes Atlas sont organisés comme est illustré dans la Figure :

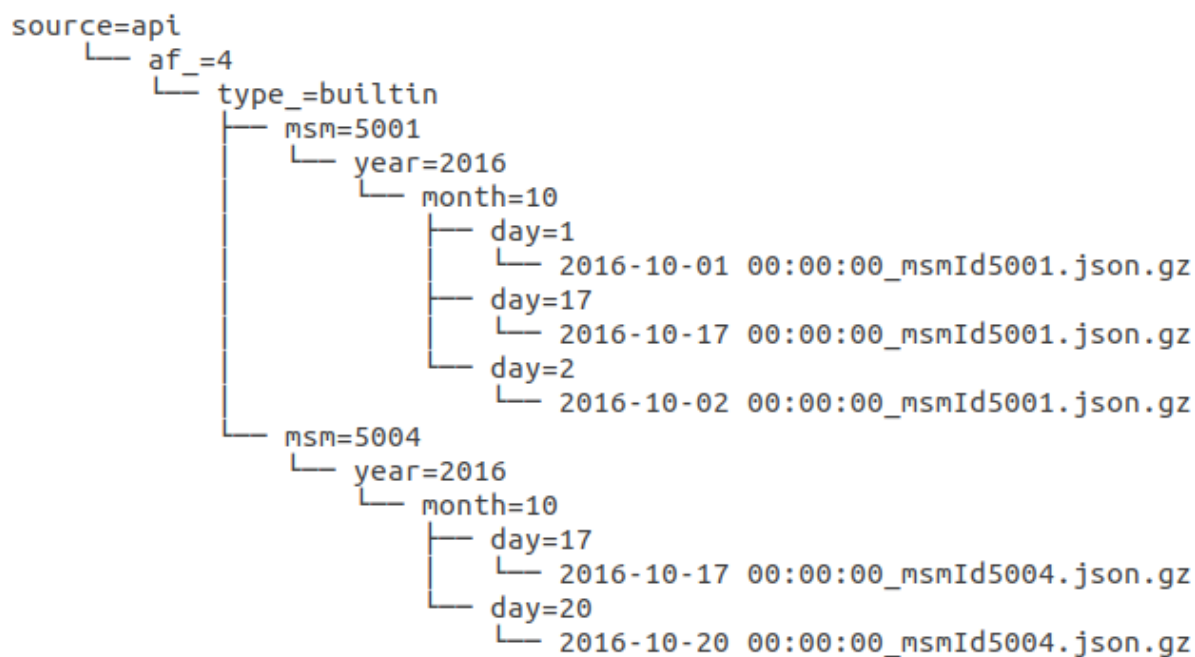


FIGURE A.1 – L'organisation des traceroutes dans un compartiment Amazon S3

A.2.3 L'interrogation de données sur Amazon Athena

L'interrogation de données présentes sur Amazon S3 via Amazon Athena est effectuée avec une requête SQL. On peut décomposer une requête SQL en trois parties principales :

Les données sollicitées Ça peut être une colonne ou bien une colonne transformée suite à l'application d'une ou de plusieurs fonctions de Presto.

Les partitions de données concernées La requête SQL est paramétrée de sorte à limiter les données à analyser sur Amazon S3.

Les paramètres appliqués sur les colonnes La requête SQL doit filtrer les données suivant les conditions sur les colonnes de la table.

En pratique, nous avons créé la requête SQL reprise dans la Figure A.2.

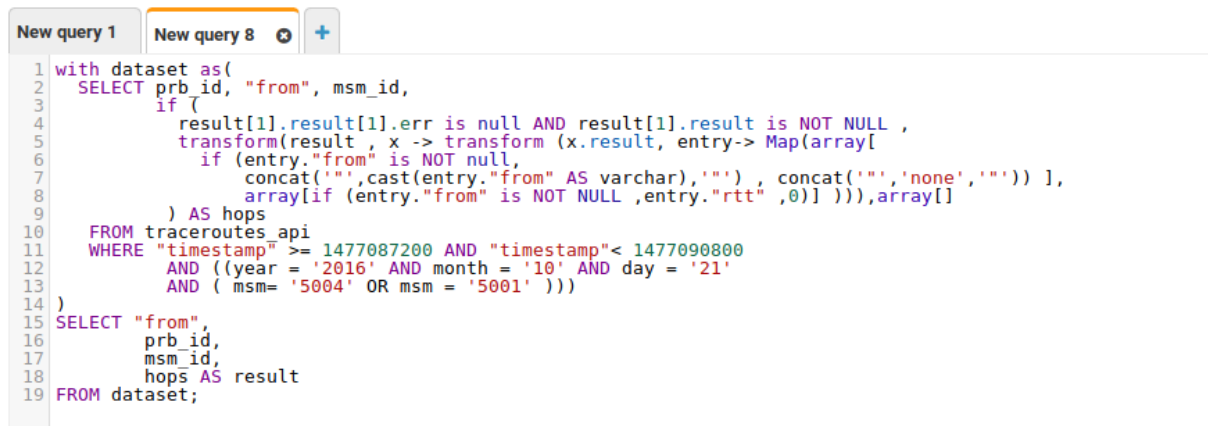
Pour les données sollicitées, ce sont les trois colonnes *prb_id*, *from*, *msm_id* (ligne 2) et la liste de saut (*hops*) obtenue après quelques vérifications (lignes 3 à 9). A la ligne 10, c'est la table créée pour les traceroutes (voir les détails de la table dans A.1). Les traceroutes à analyser sont ceux obtenus en vérifiant les conditions dans la ligne 12 et 13. C'est à dire, Athena va regarder les traceroutes qui se trouvent dans les sous dossiers year=2016

month=10

day=21 qui se trouvent aussi dans les deux dossiers msm=5001 et msm=5004. Ce que revient à chercher les traceroutes dans les deux endroits suivants :

```
s3://ripeatlasdata/traceroute/source=api/af_=4/type_=builtin/msm=5001/year=2016/month=10/day=21
```

```
s3://ripeatlasdata/traceroute/source=api/af_=4/type_=builtin/msm=5004/year=2016/month=10/day=21
```

The image shows a screenshot of the Amazon Athena console. At the top, there are tabs for 'New query 1' and 'New query 8'. The 'New query 1' tab is active, displaying a SQL query. The query is a complex SELECT statement with a CTE named 'dataset'. It filters data from the 'traceroutes_api' table based on a timestamp range and specific month, day, and msm values. The query uses a transform function to process the 'result' column, handling nulls and concatenating values. The final result is a SELECT statement that joins the 'dataset' CTE with the 'hops' table to produce the final output.

```
1 with dataset as(  
2   SELECT prb_id, "from", msm_id,  
3   if (  
4     result[1].result[1].err is null AND result[1].result is NOT NULL ,  
5     transform(result , x -> transform (x.result, entry-> Map(array[  
6       if (entry."from" is NOT null,  
7         concat('"' ,cast(entry."from" AS varchar),'"' ) , concat('"' , 'none' , '"') ) ],  
8       array[if (entry."from" is NOT NULL ,entry."rtt" ,0)] ))) ,array[]  
9   ) AS hops  
10  FROM traceroutes_api  
11  WHERE "timestamp" >= 1477087200 AND "timestamp"< 1477090800  
12     AND ((year = '2016' AND month = '10' AND day = '21'  
13     AND ( msm= '5004' OR msm = '5001' )))  
14 )  
15 SELECT "from",  
16        prb_id,  
17        msm_id,  
18        hops AS result  
19 FROM dataset;
```

FIGURE A.2 – Une exemple d’une requête SQL sur Amazon Athena

Table des figures

| | | |
|--------|---|----|
| I.1 | Génération 1 | 6 |
| I.2 | Génération 2 | 6 |
| I.3 | Génération 3 | 6 |
| I.4 | Les trois générations des sondes Atlas | 6 |
| I.5 | La connexion des sondes Atlas à l'infrastructure RIPE Atlas [23] | 7 |
| I.6 | L'architecture du système RIPE Atlas | 8 |
| I.7 | Les étapes d'établissement d'une connexion entre la sonde Atlas et l'architecture RIPE Atlas | 10 |
| I.8 | La corrélation entre la moyenne des AS paths et la moyenne des RTTs [17] | 22 |
| I.9 | Visualisation des changements des chemins traceroute [14] | 23 |
| II.1 | (a) Le RTT entre la sonde P et les routeurs B et C. (b) La différence entre les chemins de retour depuis les routeurs B et C vers la sonde P. Source : [16] | 29 |
| II.2 | Illustration des périodes de l'analyse entre la date de début et la date de fin | 31 |
| II.3 | Illustration des sauts d'un traceroute avec leurs informations | 33 |
| II.4 | Inférence des liens possibles entre les routeurs des deux sauts consécutifs $R_{a,i}$ et $R_{b,j}$ | 34 |
| II.5 | La comparaison des deux intervalles de confiance : courant et référence | 35 |
| II.6 | Le processus de la détection des anomalies dans les délais des liens | 37 |
| III.1 | Architecture standard du Big Data [28] | 41 |
| III.2 | Illustration d'une base de données NoSQL de type clé-valeur | 43 |
| III.3 | Illustration d'une base de données NoSQL de type document | 43 |
| III.4 | Illustration d'une base de données NoSQL de type colonne | 43 |
| III.5 | Illustration d'une base de données NoSQL de type graphe | 44 |
| III.6 | Bases de données NoSQL suivant le théorème de CAP | 45 |
| III.7 | Un classement des SGBDs sur <i>DB-Engines Ranking</i> du 1 août 2018 | 46 |
| III.8 | Schema on Write (SGBDR) | 47 |
| III.9 | Schema on Read (Big Data) | 47 |
| III.10 | Interaction entre une application Spark et le cluster manager. Source : [25] | 52 |
| III.11 | Un exemple d'un cluster formé de trois executors. Source : [26] | 52 |
| III.12 | Spark Unified Stack. Source : [27] | 53 |
| III.13 | Exemple d'un flux de données avec Spark | 54 |
| IV.1 | | 57 |
| A.1 | L'organisation des traceroutes dans un compartiment Amazon S3 | 68 |
| A.2 | Une exemple d'une requête SQL sur Amazon Athena | 69 |

Liste des tableaux

| | | |
|-------|---|----|
| I.1 | Les caractéristiques du matériel des trois générations des sondes Atlas | 6 |
| I.2 | Comparaison entre sondes et ancres RIPE Atlas | 12 |
| I.3 | Comparaison entre sondes Atlas et ProbeAPI | 19 |
| I.4 | Les détails des mesures effectuées dans le travail de C. Anderson [13] | 21 |
| II.1 | Récapitulatif des traceroutes utilisés dans le travail de référence | 28 |
| III.1 | Les tarifs du AWS S3 (formule Stockage standard S3) | 49 |
| IV.1 | Caractéristiques de la machine de test | 62 |
| IV.2 | | 62 |
| IV.3 | Récapitulatif des technologies Big Data expérimentées dans l'analyse des délais | 63 |

Listings

| | |
|---|----|
| IV.1 La classe Signal en Scala | 59 |
| IV.2 La classe Hop en Scala | 59 |
| IV.3 La classe Traceroute en Scala | 59 |
| IV.4 La classe PreparedSignal en Scala | 59 |
| IV.5 La classe PreparedHop en Scala | 60 |
| IV.6 La classe MedianByHopTraceroute en Scala | 60 |
| IV.7 La classe Link en Scala | 60 |
| IV.8 La classe LinksTraceroute en Scala | 60 |
| IV.9 La classe LinkIPs en Scala | 60 |
| IV.10 La classe DiffRtt en Scala | 61 |
| IV.11 La classe DiffRTTPeriod en Scala | 61 |
| IV.12 La classe LinkState en Scala | 61 |

Bibliographie

- [1] Archipelago (Ark) Measurement Infrastructure. URL : <http://www.caida.org/projects/ark/>. (consulté le 19/01/2018).
- [2] Center for Applied Internet Data Analysis (CAIDA). URL : <http://www.caida.org/>. (consulté le 06/04/2018).
- [3] Create a New Measurement - RIPE Atlas. URL : <https://atlas.ripe.net/measurements/form/>. consulté le 05/08/2018.
- [4] Le dépôt des données RIPE Atlas. URL : <https://data-store.ripe.net/datasets/atlas-daily-dumps/>. (consulté le 26/07/2018).
- [5] Les archives des détails des sondes atlas. URL : <https://ftp.ripe.net/ripe/atlas/probes/archive/>. (consulté le 28/01/2018).
- [6] RIPE Atlas - Known Bugs and Limitations. URL : <https://atlas.ripe.net/docs/bugs/>. (consulté le 05/04/2018).
- [7] Samknows. URL : <https://www.samknows.com/global-platform>. (consulté le 23/01/2018).
- [8] Xport pro. URL : <https://www.lantronix.com/products/xport-pro/>. (consulté le 08/08/2018/).
- [9] Lissages exponentiels. URL : https://www.math.u-psud.fr/~goude/Materials/time_series/cours3_lissage_expo.pdf, 2017. (consulté le 30/09/2018).
- [10] ABEN, E. How RIPE Atlas Helped Wikipedia Users. URL : <https://labs.ripe.net/Members/emileaben/how-ripe-atlas-helped-wikipedia-users>, 2014. (consulté le 18/08/2018).
- [11] ABEN, E. Looking at France-IX with RIPE Atlas and RIS. URL : <https://labs.ripe.net/Members/emileaben/looking-at-france-ix-with-ripe-atlas-and-ris>, 2015. (consulté le 08/08/2018).
- [12] ABEN, E. Measuring Countries and IXPs with RIPE Atlas. URL : <https://labs.ripe.net/Members/emileaben/measuring-ixps-with-ripe-atlas>, 2015. (consulté le 08/08/2018).
- [13] ANDERSON, C., WINTER, P., AND ROYA. Global network interference detection over the RIPE atlas network. In *4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14)* (San Diego, CA, 2014), USENIX Association.

- [14] DONATO, V. D. Traceroute Consistency Check. URL : <https://github.com/vdidonato/Traceroute-consistency-check>, 2015. (consulté le 08/08/2018).
- [15] FONTUGNE, R. InternetHealthReport - tartiflette. URL : <https://github.com/InternetHealthReport/tartiflette/tree/master/dataManipulation>.
- [16] FONTUGNE, R., ABEN, E., PELSSER, C., AND BUSH, R. Pinpointing delay and forwarding anomalies using large-scale traceroute measurements. *CoRR abs/1605.04784* (2016).
- [17] GASMI, S. Visualising RIPE Atlas Anchor Measurements. URL : https://labs.ripe.net/Members/salim_gasmi/visualising-ripe-atlas-anchor-measurements, 2015. (consulté le 08/08/2018).
- [18] GUILLAUME VALADON, FRANCOIS CONTAT, M. H., AND HOLTERBACH, T. BGP Atlas Monito (BAM). URL : <https://github.com/guedou/bam>, 2015. (consulté le 08/08/2018).
- [19] HEROLD, J. Bgp + traceroute presentation. URL : <https://labs.ripe.net/Members/becha/ripe-atlas-hackathon-presentations/bgp-traceroute>, 2015. (consulté le 08/08/2018).
- [20] HEROLD, J. BGP + Traceroute using RIPE NCC Atlas. URL : <https://github.com/wires/bgp-traceroutes>, 2015. consulté le 08/08/2018.
- [21] HOLTERBACH, T., PELSSER, C., BUSH, R., AND VANBEVER, L. Quantifying interference between measurements on the ripe atlas platform. In *Proceedings of the 2015 Internet Measurement Conference* (New York, NY, USA, 2015), IMC '15, ACM, pp. 437–443.
- [22] KISTELEKI, R. The AMS-IX Outage as Seen with RIPE Atlas. URL : <https://labs.ripe.net/Members/kistel/the-ams-ix-outage-as-seen-with-ripe-atlas>, 2015. (consulté le 23/01/2018).
- [23] KISTELEKI, R. RIPE Atlas Architecture - how we manage our probes. URL : <https://labs.ripe.net/Members/kistel/ripe-atlas-architecture-how-we-manage-our-probes>, 2017. consulté le (08/08/2018).
- [24] KISTELEKI, R. RIPE Atlas probes as IoT devices. URL : <https://labs.ripe.net/Members/kistel/ripe-atlas-probes-as-iot-devices>, 2017. (consulté le 21/12/2017).
- [25] LUU, H. *Beginning Apache Spark 2 - 2018*, 1 ed. Apress, 2018, p. 4.
- [26] LUU, H. *Beginning Apache Spark 2 - 2018*, 1 ed. Apress, 2018, p. 6.
- [27] LUU, H. *Beginning Apache Spark 2 - 2018*, 1 ed. Apress, 2018, p. 7.
- [28] MAHESHWARI, A. *Big Data*. 2017.
- [29] RIPE NCC. Test Traffic Measurement Service (TTM). URL : <https://www.ripe.net/analyse/archived-projects/ttm>. (consulté le 14/01/2018).

- [30] RODERICK, F. On the Diversity of Interdomain Routing in Africa. URL : https://labs.ripe.net/Members/fanou_roderick/on-the-diversity-of-interdomain-routing-in-africa, 2015. (consulté le 11/01/2018).
- [31] SAGIROGLU, S., AND SINANC, D. Big data : A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (May 2013), pp. 42–47.
- [32] SHAH, A., FONTUGNE, R., AND PAPADOPOULOS, C. Towards characterizing international routing detours. In *Proceedings of the 12th Asian Internet Engineering Conference (AINTEC)* (Bangkok, Thailand, Nov. 2016), ACM, p. to appear.
- [33] SHAO, W., ROUGIER, J., DEVIENNE, F., AND VISTE, M. Missing measurements on RIPE atlas. *CoRR abs/1701.00938* (2017).
- [34] SHAVITT, Y., AND SHIR, E. Dimes : Let the internet measure itself. *SIGCOMM Comput. Commun. Rev.* 35, 5 (Oct. 2005), 71–74.
- [35] SPEED CHECKER. Global Internet testing - ProbeAPI. URL : <http://probeapi.speedchecker.xyz/>. (consulté le 06/08/2018).
- [36] VARAS, C. A Practical Comparison Between RIPE Atlas and ProbeAPI. URL : https://labs.ripe.net/Members/cristian_varas/a-practical-comparison-between-ripe-atlas-and-probeapi, 2016. (consulté le 19/01/2018).