Le *cloud* au service secret de sa Majesté
Enjeux et perspectives des *clouds* pour les services de renseignement

Marina de CASTRO
Attachée d'administration
Ministère de l'Intérieur
Advisor du CyberCercle

Au second semestre de l'année 2021, les services de renseignement anglais annonçaient la signature d'un contrat avec *Amazon Web Service* pour l'hébergement de données, en particulier celles classifiées. L'accord est d'une importance majeure puisqu'il concerne trois pôles stratégiques du renseignement britannique. Ainsi, ce sont les données du *Government Communications Headquarters* (GCHQ), du *Security Service* (MI5) et du *Secret Intelligence Service* (MI6) qui seront stockées sur un *cloud* unique hébergé par *Amazon*. Cette solution vise à leur permettre des facilités de stockage, de recherche et de partage de l'information. Ces trois entités, dont les activités englobent la sécurité intérieure comme extérieure (MI5 et MI6) mais aussi la sécurité des systèmes d'information (GCHQ) s'en remettent donc à une entreprise américaine pour garantir la sauvegarde de leurs données confidentielles. Si le montant de la tractation n'a pas été communiqué, il est intéressant de noter qu'*Amazon Cloud Service* compte plus d'un million de clients au cœur de métiers divers tels que le spatial, la téléphonie, l'enseignement universitaire ou encore l'industrie agroalimentaire.

Le Royaume-Uni ne fait pas figure de pionnier en confiant l'hébergement de ses données classifiées à *Amazon*. En effet, en 2014, la CIA américaine avait également recours au service de ce géant du Web. En revanche, ce qui est une première dans le cas britannique c'est l'emploi des services d'une entreprise étrangère pour traiter des données aussi sensibles que celles des agences de renseignement de sa Majesté.

En dépit de la « relation spéciale » anglo-américaine qui peut exister, le recours aux *clouds* d'industriels étrangers pour l'hébergement de ce type de données pose nécessairement des questions de souveraineté. Quels sont les enjeux que ces *clouds* revêtent pour les services de renseignement ? Les risques encourus sont-ils si compromettants et préjudiciables pour les intérêts vitaux de ces nations ? Quels nouveaux champs ces technologies de stockage ouvrent-elles pour la veille, la collecte, l'analyse et la diffusion du renseignement ?

Des enjeux multiples en termes de réponse opérationnelle

Pour les services de renseignement, le recours aux *clouds* pour l'hébergement de données revêt deux objectifs principaux et complémentaires. Tout d'abord, le regroupement et la centralisation de toutes ces données sensibles faciliteront indéniablement l'usage de technologies de type Big Data pour le traitement et la primo-analyse de ces dernières. Ensuite, cela pavera la voie aux innovations liées à l'intelligence artificielle en matière de production et de gestion du renseignement multi-capteurs[1].

En effet, l'usage des *clouds* devrait permettre aux Agences de développer des lacs de données plus connus sous l'appellation anglo-saxonne *data lakes*. Il s'agit d'un référentiel qui permet de regrouper des données structurées c'est-à-dire provenant de bases de données relationnelles, des données semi-structurées à l'exemple de fichiers CSV ou XML et des données non-structurées comme des e-mails ou des documents PDF. Les *data lakes* peuvent aussi contenir des données binaires correspondant à des images, des vidéos ou des audios. Cette flexibilité est un avantage. Sa finalité est d'intégrer des données brutes dans un ensemble cohérent, de les transformer et de les rendre exploitables à des fins d'analyse. De telles solutions accélèrent l'intégration et offrent un contenu homogène à exploiter même si la nature des données d'entrée est aussi diverse que variée. Alors qu'hier il fallait plusieurs semaines ou plusieurs mois pour exploiter des données brutes et disparates, aujourd'hui il ne faut plus que quelques minutes. Rendant ainsi l'analyse plus aisée et plus rapide, cet environnement accroît l'opérationnalité de la donnée qui n'est plus uniquement utilisée dans l'objectif de rendre compte. Elle devient ainsi un véritable outil de pilotage, une force motrice qui intervient dès l'orientation du cycle du renseignement. A ce titre, elle favorise le *machine learning*[2] et permet aux analystes et *data scientists* de produire des modèles prédictifs. Ces derniers concourent directement à la planification des opérations et alimentent la fonction stratégique « connaissance et anticipation ».

De plus, l'avènement des réseaux sociaux et de l'Internet des objets[3] (IoT) mettent en exergue le traitement d'importants volumes données et leur collecte massive par des procédés tels que le *scrapping*[4] ou le *crawling*[5], par exemple. Ainsi approvisionnés, les *data lakes* permettraient de systématiser la pratique du *machine learning* à des fins opérationnelles. En février 2020, Jérémy Fleming, le directeur du GCHQ, expliquait que ces moyens devenaient une priorité pour l'espionnage britannique à deux égards : la lutte contre le terrorisme et la lutte contre la manipulation de

1 Le renseignement multi-capteurs désigne une forme de renseignement qui agrège des informations émanant d'origines de collecte différentes : électromagnétique, imagerie, humaine ou encore sources ouvertes.

2 Le *machine learning* est une méthode de programmation qui se base sur les statistiques et les probabilités pour permettre aux ordinateurs d'apprendre « par eux-mêmes » en s'entraînant sur de larges jeux de données (*datasets*).

3 L'Internet des objets désigne le flux de données émis par tous les objets connectés : de la montre au réfrigérateur en passant par l'assistant vocal.

4 Le *scrapping* est une technique de collecte automatisée de données sur des sites web. Généralement des scripts ou des bots extraient le contenu de pages web pour alimenter des bases de données ou des outils de veille et d'analyse.

5 Le *crawling* désigne l'exploration de sites webs par des robots à des fins d'indexation de contenus. Les bots scannent les pages webs et récupèrent leur codes sources pour les indexer. Googlebot est l'outil de Google pour cet usage.

l'information. En effet, sur de grands volumes de données, le *machine learning* et à terme l'intelligence artificielle permettent d'établir des corrélations invisibles à l'œil nu. Cette méthode est également très utilisée en matière de reconnaissance vocale, de transcription et de traduction des conversations et signaux interceptés. Enfin, la croissance des préoccupations autour des champs immatériels constitue un enjeu déterminant à l'égard de l'usage de *clouds* par certains services de renseignements.

Des risques assumés en matière de souveraineté

De manière générale, si le stockage des données est un enjeu de souveraineté, il l'est encore plus lorsqu'il s'agit de données sensibles et classifiées. Ainsi, ce recours aux services d'hébergement de l'américain *Amazon Web Services* (AWS) soulève de nombreuses interrogations au Royaume-Uni, mais également au sein de viviers d'experts internationaux. Une nouvelle fois, ces pratiques relancent le débat au sujet de la souveraineté de la Grande-Bretagne sur ses institutions et ses industries stratégiques. Alors que le gouvernement britannique a exclu le chinois Huawei de son réseau télécom 5G, il confie les données de ses services de renseignement à un géant des États-Unis.

En 2013, la CIA signait déjà un accord avec AWS. Toutefois, la question de la souveraineté était moins palpable puisqu'une agence de renseignement américaine contractualisait avec une entreprise dont le siège est dans l'état de Washington. Les notions d'influence ou d'ingérence étrangère n'avaient alors pas lieu d'être. *Amazon* se veut rassurant et précise que l'intégralité des données sera conservée au Royaume-Uni et que la multinationale n'aura aucun accès aux informations contenues dans son *cloud*. Son directeur de la sécurité, Stephen Schmidt, rappelait la philosophie de l'entreprise de « *sécurité par l'obscurité* ». Les employés qui travaillent dans ces segments disposent d'habilitations particulières et ne sont informés que du strict nécessaire à l'exercice de leur travail. Les standards de sécurité d'AWS demeurent très rigoureux.

Si la problématique de la souveraineté peut se poser en termes techniques, elle doit aussi se poser en matière géopolitique. Elle devient d'autant plus capitale dans un contexte international marqué par une course effrénée à la technologie et à l'innovation. En 2020, Richard Moore, le patron de la sécurité extérieure britannique, exprimait son inquiétude d'être distancé par la Chine et la Russie. Il expliquait à l'antenne de la BBC que ces nations « *mettent de l'argent et de l'ambition dans l'intelligence artificielle et l'informatique quantique […] car ils savent que la maîtrise de ces technologies leur donnera un avantage compétitif* ». Il affirmait enfin que « *nous pourrions connaître plus de progrès technologiques les dix prochaines années qu'au cours du siècle dernier, avec un impact en termes de perturbations égal à celui de la révolution industrielle* ». Les craintes qu'il a ainsi exprimées sont partagées par ses homologues d'autres pays européens. La menace chinoise est particulièrement étudiée puisque l'empire du Milieu pourrait dominer d'ici quelques années de nombreuses technologies de pointe parmi lesquelles l'intelligence artificielle. La collecte, le traitement et le stockage de la donnée pourraient devenir des sources de tension, voire de conflits, dans les décennies à venir. Cette méfiance fait écho aux accusations portées par le Royaume-Uni à l'égard de l'équipementier Huawei ou encore aux campagnes de désinformation en ligne fomentées par le gouvernement chinois, notamment dans le contexte de la pandémie de Covid-19.

Le contrat conclu entre les agences de renseignement britanniques et *Amazon* peut, de prime abord, interroger sur la souveraineté de la donnée. Toutefois, le risque semble être connu et assumé. Il s'agirait davantage d'un choc culturel d'ampleur à absorber. Dans des milieux où le cloisonnement était une règle d'or, la technologie impose de désensiler l'information pour en tirer le meilleur profit. L'impression de décloisonnement que ce recours aux *clouds* peut induire ne contrevient pourtant pas aux traditionnels « besoins d'en connaître ».

Eu égard à la fulgurance de l'information et à l'accélération du cycle d'innovation digitale, les entreprises nationales britanniques n'ont pas été en mesure de fournir des capacités de stockage de données dans le *cloud* satisfaisant au besoin opérationnel. Contrairement à Q dans James Bond, un tel niveau d'expertise technique ne s'obtient pas en un claquement de doigts. Pour Londres, l'étroite coopération avec les acteurs de la Tech' et les GAFAM a été incontournable.

A Paris, le gouvernement préfère soutenir des projets de *clouds* souverains à l'exemple de « Bleu », le *cloud* de confiance porté par Orange et Capgemini. Par ailleurs, en octobre 2021 Google Cloud et Thalès (leader français des hautes technologies sur le marché de l'aérospatial et de la défense) annonçaient le développement conjoint de projets répondant aux exigences du label « *cloud* de confiance ». Avec le recours massif aux *clouds* par les acteurs privés et publics, les institutions et les services de renseignement prennent aujourd'hui la mesure du gain de performance qu'ils procurent, mais également des nouveaux risques opérationnels qui les accompagnent, notamment en termes de confidentialité, de gouvernance et de souveraineté de la donnée.

Sources

BABUTA Alexander, JANJEVA Ardi et OSWALD Marion « Artificial *intelligence and UK national security : policy considerations* », Rusi.org, 27 avril 2020

BLANJEAN Romain, « P*ourquoi investir dans le cloud souverain quand on est déjà sur le cloud public ?* », Zdnet.fr, 30 novembre 2021

CORERA Gordon, « *UK spies will need artificial intelligence* », BBC.com, 27 avril 2020

DELUZARCHE Céline, « *Les services secrets britanniques font appel à Amazon pour stocker leurs documents secrets défense* », Korii.slate.fr, 28 octobre 2021

FILIPONE Dominique, « *Les services secrets britanniques misent sur le cloud AWS* », Le monde informatique, 27 octobre 2021

KONKEL Frank, « *The details about CIA's deal with Amazon* », Theatlantic.com, 17 juillet 2014

LIÈVRE Florence, « Capgemini *et Orange annoncent le projet de créer « Bleu » une société qui fournira un cloud de confiance en France* », Orange.com, 27 mai 2021

TOUSSAINT Léo, « *AWS : une région secrète pour les services de renseignement américains* », Siecledigital.fr, 22 novembre 2017

VONINTSOA, « Amazon *signe un contrat cloud avec les agences d'espionnage britannique* », intelligence-artificielle.com, 15 novembre 2021

« Data *lake : la solution reine du Big Data* », Journaldunet.fr, 15 mai 2018