



# Tendances technologiques

Lacs de données

Architecture d'entreprise, Direction générale du dirigeant principal  
de la technologie

Version 0.1

Date : 2019-07-12



## Table des matières

<b>Sommaire opérationnel</b> .....	<b>3</b>
<b>Sommaire technique</b> .....	<b>4</b>
<b>Utilisation par l'industrie</b> .....	<b>5</b>
<b>Utilisation par le gouvernement du Canada</b> .....	<b>6</b>
<b>Répercussions pour Services partagés Canada (SPC)</b> .....	<b>7</b>
Proposition de valeur .....	7
Défis .....	8
Considérations .....	10
<b>Références</b> .....	<b>13</b>

## Sommaire opérationnel

Dans un monde de plus en plus hyperconnecté, les entreprises ont du mal à faire face aux responsabilités de stockage, de gestion et de disponibilité rapide des données brutes. Pour décomposer davantage ces défis liés aux données :

- Les données se présentent dans de nombreuses structures différentes.
  - Non structurées
  - Semi-structurées
  - Structurées
- Les données proviennent de nombreuses sources disparates.
  - Applications d'entreprise
  - Fichiers bruts
  - Journaux d'exploitation et de sécurité
  - Transactions financières
  - Appareils et capteurs de réseau d'Internet des objets (IdO)
  - Sites Web
  - Recherche scientifique.
- Les sources de données sont souvent réparties géographiquement sur plusieurs sites.
  - Centres de données
  - Bureaux éloignés
  - Appareils mobiles

Afin de résoudre ces défis liés aux données, une nouvelle façon de gérer les données a été créée, ce qui a poussé les entreprises axées sur les données à inventer un nouveau mécanisme de stockage des données appelé lac de données.

Les lacs de données se définissent ainsi :

- Ils recueillent tout.
  - Un lac de données contient toutes les données, les sources brutes sur de longues périodes de temps ainsi que toutes les données traitées.
- Ils s'immergent n'importe où.
  - Un lac de données permet aux utilisateurs de plusieurs unités opérationnelles d'affiner, d'explorer et d'enrichir les données à leurs conditions.
- Ils offrent un accès flexible.
  - Un lac de données permet plusieurs modèles d'accès aux données par l'entremise d'une infrastructure partagée : par lots, interactifs, en ligne, de recherche, en mémoire et autres moteurs de traitement.

Les lacs de données constituent essentiellement une plateforme technologique pour la conservation des données. Leur valeur pour l'entreprise ne se réalise que par l'application au lac des compétences en sciences des données.

En résumé, les cas d'utilisation des lacs de données sont encore en cours de découverte. Les fournisseurs de services infonuagiques facilitent l'acquisition de lacs de données; aujourd'hui, les lacs de données sont principalement utilisés par les établissements de recherche, les services financiers, les télécommunications, les médias, le commerce de détail, la fabrication, les soins de santé, l'industrie pharmaceutique, l'industrie pétrolière et gazière, et les gouvernements.

## Sommaire technique

La mise en œuvre la plus populaire d'un lac de données se fait au moyen de la plateforme de source ouverte appelée Apache Hadoop. Apache Hadoop est une série d'utilitaires logiciels de source ouverte qui facilite l'utilisation d'un réseau de nombreux ordinateurs pour résoudre des problèmes impliquant des quantités massives de données et de calcul. Hadoop a été créé à l'origine par des chercheurs de Google; il s'agissait d'une méthode de stockage pour gérer l'indexation des sites Web sur Internet. À cette époque, il était appelé « système de fichiers Google ».

Un lac de données est un répertoire centralisé qui vous permet de stocker toutes vos données structurées et non structurées à toute échelle. Vous pouvez stocker vos données telles quelles, sans avoir à les structurer d'abord, et exécuter différents types d'analyses, depuis les tableaux de bord et les visualisations jusqu'aux traitements des mégadonnées, aux analyses en temps réel et à l'apprentissage automatique pour orienter de meilleures décisions.

Les données peuvent être intégrées au lac de données par traitement par lots ou traitement en temps réel des données en continu. De plus, les données elles-mêmes ne sont plus limitées par les décisions de schéma initiales et peuvent être exploitées plus librement par l'entreprise. Au-dessus de ce répertoire se trouve un ensemble de capacités qui permettent aux TI de fournir des données et analyses en tant que service (DaaS), selon un modèle offre-demande. Les TI jouent le rôle de fournisseur de données (fournisseur), tandis que les utilisateurs opérationnels (scientifiques des données, analystes opérationnels) sont des consommateurs.

Le modèle des données et analyses en tant que service permet aux utilisateurs de répondre librement à leurs besoins en matière de données et d'analyse. Les utilisateurs parcourent le catalogue de données du lac (un « Datapédia ») pour trouver et sélectionner les données disponibles et remplir un « panier » métaphorique (en fait un bac à sable analytique) de données avec lesquelles travailler. Une fois l'accès fourni, les utilisateurs peuvent utiliser les outils d'analyse de leur choix pour élaborer des modèles et obtenir des perspectives. Par la suite, les utilisateurs peuvent publier des modèles d'analyse ou placer les données raffinées ou transformées dans le lac de données pour les diffuser à l'ensemble de la communauté.

Bien que l'approvisionnement d'un bac à sable analytique constitue une utilisation principale, le lac de données a aussi d'autres applications. Par exemple, il peut être

utilisé pour l'ingestion de données brutes, la conservation des données et l'application extraire, transformer, charger. Ces données peuvent ensuite être chargées dans un entrepôt de données d'entreprise. Pour tirer parti de la flexibilité offerte par le lac de données, les organisations doivent personnaliser et configurer le lac de données en fonction de leurs besoins et domaines particuliers.

## Utilisation par l'industrie

Les lacs de données sont utilisés de diverses façons dans l'industrie :

- **Ingestion de sources de données semi-structurées et non structurées (aussi appelées mégadonnées)** telles que les relevés d'équipement, les données de télémétrie, les journaux, les données en continu, etc. Un lac de données est une excellente solution pour stocker des données de type IdO (Internet des objets) qui sont traditionnellement plus difficiles à stocker, et peut appuyer une analyse en temps quasi réel. Au besoin, vous pouvez également ajouter des données structurées (c.-à-d. extraites d'une source de données relationnelle) à un lac de données si votre objectif est d'obtenir un répertoire unique de toutes les données qui seront disponibles par l'entremise du lac.
- **Analyse expérimentale** des données avant que leur valeur ou leur finalité n'ait été pleinement définie. La souplesse est importante pour toute entreprise de nos jours, de sorte qu'un lac de données peut jouer un rôle important dans des situations de type « preuve de valeur » en raison de l'approche « extraire, transformer, charger » mentionnée ci-dessus.
- **Soutien analytique avancé.** Un lac de données est utile aux scientifiques et analystes des données pour fournir des données et expérimenter avec ces données.
- **Stockage des données d'archives et historiques.** Parfois, les données sont rarement utilisées, mais il faut qu'elles soient disponibles pour analyse. Une stratégie de lac de données peut être très précieuse pour appuyer une stratégie d'archivage active.
- Capacités de **traitement réparties** associées à un entrepôt de données logiques.

### Comment la Banque TD a rendu son lac de données plus utilisable

<https://www.datanami.com/2017/10/03/td-bank-made-data-lake-usable/>

La Banque Toronto-Dominion (Banque TD) est l'une des plus grandes banques d'Amérique du Nord, avec 85 000 employés, plus de 2 400 sites répartis entre le Canada et les États-Unis et des actifs avoisinant le billion de dollars. En 2014, l'entreprise a décidé de normaliser l'entreposage de ses données pour diverses fonctions d'informatique décisionnelle et de rapports réglementaires. L'entreprise a acheté une répartition Hadoop et s'est lancée dans la construction d'une importante grappe qui pourrait fonctionner comme un lac centralisé pour stocker les données provenant de plusieurs services.

## Utilisation par le gouvernement du Canada

En 2019, le Secrétariat du Conseil du Trésor du Canada (SCT), en partenariat avec Services partagés Canada et d'autres ministères, a trouvé un responsable opérationnel pour élaborer une stratégie de service en matière de lac de données (un répertoire de données brutes) afin que le gouvernement du Canada puisse tirer parti des mégadonnées et des innovations du marché pour favoriser de meilleures analyses et promouvoir la mise en commun horizontale des données.<sup>i</sup>

Les mégadonnées sont utilisées pour le stockage et le traitement des données et des informations contenues dans des ensembles de données qui sont si importants ou complexes que les applications traditionnelles de traitement des données ne peuvent en effectuer l'analyse. Les mégadonnées peuvent rendre accessibles des quantités presque illimitées d'informations, ce qui améliore la prise de décisions fondées sur les données et élargit les initiatives de données ouvertes. L'informatique décisionnelle comprend la création, l'agrégation, l'analyse et la visualisation des données dans le but d'éclairer et de faciliter la gestion et la stratégie opérationnelles. Le SCT, en collaboration avec les ministères, dirigera l'élaboration des besoins liés à une plateforme d'analyse d'entreprise<sup>ii</sup>.

L'élaboration des lacs de données au sein du gouvernement du Canada est une initiative plus récente. Cette situation s'explique principalement par le fait que le gouvernement du Canada concentre ses ressources sur la mise en œuvre d'initiatives infonuagiques. Toutefois, certains ministères du gouvernement du Canada s'emploient à élaborer des environnements de lacs de données en tandem avec des initiatives infonuagiques.

En particulier, Emploi et Développement social Canada (EDSC) prépare l'installation de plusieurs lacs de données afin de permettre la création d'un écosystème de lacs de données, d'une analyse de données et d'un ensemble d'outils d'apprentissage automatique. Cela permettra à EDSC de communiquer l'information horizontalement, de façon efficace et sécuritaire, tout en offrant un large éventail de capacités en matière d'analyse de données. EDSC vise à tenir à jour les capacités actuelles en matière de données et d'analyse tout en explorant de nouvelles possibilités pour réduire les lacunes et faire évoluer continuellement nos services afin de répondre aux besoins des clients<sup>iii</sup>.

# Répercussions pour Services partagés Canada (SPC)

## Proposition de valeur

Il y a trois propositions de valeur communes pour la poursuite des lacs de données. 1) Ils peuvent fournir un moyen facile et accessible d'obtenir des données plus rapidement; 2) Ils peuvent créer un point d'entrée des données unique pour aider à relier et à fusionner les silos d'information au sein d'une organisation; et 3) Ils peuvent fournir un environnement expérimental pour les scientifiques des données expérimentés afin de permettre de nouvelles perspectives d'analyse.

Les lacs de données peuvent fournir des données aux consommateurs plus rapidement en offrant des données sous une forme plus brute et plus facilement accessible. Les données sont stockées dans leur format d'origine avec peu ou pas de traitement, elles sont optimisées afin d'être stockées en grandes quantités dans leur format d'origine. En permettant aux données de rester dans leur format d'origine, un flux de données beaucoup plus rapide est disponible pour des requêtes et des analyses illimitées. Un lac de données peut aider les consommateurs de données à contourner les applications strictes de récupération des données et de données structurées, comme un entrepôt de données ou un mini-entrepôt de données. Cela a pour effet d'améliorer la flexibilité des données d'une entreprise. Certaines entreprises ont en fait utilisé les lacs de données pour remplacer les environnements d'entrepôt existants, où la mise en œuvre d'un nouvel entrepôt de données est plus coûteuse. Un lac de données peut contenir des données non raffinées, ce qui est utile lorsqu'une structure de données opérationnelles est inconnue ou lorsqu'un consommateur de données a besoin d'accéder rapidement à ces données.

Un lac de données n'est pas une source unique de vérité. Un lac de données est un emplacement centralisé dans lequel convergent et sont stockées les données provenant de toutes les sources de données, quel que soit leur formatage. En tant que point unique pour l'entrée de données, les sections d'une entreprise peuvent mettre en commun leurs informations dans le lac de données et augmenter la communication des renseignements avec d'autres parties de l'organisation. De cette façon, tous les membres de l'organisation ont accès aux données. Un lac de données peut augmenter la mise en commun horizontale des données au sein d'une organisation en créant ce point d'entrée de données unique. L'utilisation de plusieurs outils de stockage et de traitement permet aux analystes d'extraire rapidement la valeur des données afin d'éclairer les décisions opérationnelles clés.

Un lac de données est optimisé pour l'exploration et constitue un environnement expérimental permettant aux scientifiques des données expérimentés de découvrir de nouvelles perspectives à partir des données. Les analystes peuvent superposer le contexte aux données pour en extraire la valeur. Toutes les organisations veulent accroître la souplesse analytique et opérationnelle. L'approche architecturale du lac

de données peut stocker de grands volumes de données, ce qui peut permettre aux équipes transversales de mettre leurs données en commun dans un emplacement centralisé en complétant leurs systèmes d'enregistrement par des systèmes de perspective. Ce sont les lacs de données qui présentent le plus d'avantages potentiels pour les scientifiques des données expérimentés et compétents. Les données structurées, non structurées et semi-structurées, provenant généralement du même ensemble de données, peuvent contenir des perspectives commerciales, prédictives et prescriptives auparavant impossibles à obtenir à partir d'une plateforme structurée, comme on l'observe dans les entrepôts et mini-entrepôts de données.

## Défis

Bien que la technologie des lacs de données présente de nombreux avantages pour les organisations qui traitent de grandes quantités de données, elle comporte ses propres défis. Par exemple :

### **Gouvernance des données et problèmes sémantiques**

Le plus grand défi pour les lacs de données est de résoudre les différents besoins en matière de gouvernance des données dans une seule plateforme de données centralisée. Les lacs de données échouent surtout lorsqu'ils manquent de gouvernance, d'utilisateurs autodisciplinés et d'un flux de données rationnel. Souvent, les mises en œuvre de lacs de données se concentrent sur le stockage des données plutôt que sur leur gestion. Les lacs de données ne sont pas optimisés pour l'application ou la cohérence sémantiques. Ils sont conçus pour la flexibilité sémantique, pour permettre à toute personne de fournir un contexte aux données si elle a les compétences pour le faire.

Mettre des données au même endroit ne supprime pas leur ambiguïté ou leur signification. Les lacs de données constituent un environnement de modèle de stockage sans contrainte et « sans compromis », sans les garanties de gouvernance des données communes aux entrepôts ou aux mini-entrepôts de données. Des métadonnées appropriées sont essentielles pour un lac de données; sans elles, le lac de données ne fonctionnera pas comme prévu. Il est souhaitable de considérer les métadonnées comme le détecteur de poissons dans le lac de données.

### **Manque de qualité des données et de confiance dans les données**

La qualité des données et la confiance dans les données sont un enjeu permanent pour de nombreuses organisations. Bien que les outils de découverte des données puissent appliquer l'apprentissage automatique à des ensembles de données connexes provenant de plusieurs sources de données pour détecter les anomalies (valeurs incorrectes, valeurs manquantes, doublons et données désuètes), la qualité et la fiabilité des données continuent d'être un problème pour les lacs de données, qui peuvent facilement devenir un dépotoir de données. Certaines données sont plus précises que d'autres. Cela peut poser un réel problème pour quiconque utilise plusieurs ensembles de données et prend des décisions fondées sur des analyses effectuées avec des données de différents degrés de qualité.



### **Défis liés aux marécages de données et au rendement et à la flexibilité des données**

Les données stockées dans les lacs de données peuvent parfois devenir *boueuses* lorsque de bonnes données sont mélangées à de mauvaises données. L'infrastructure de lacs de données est conçue pour stocker et traiter de grandes quantités de données, généralement dans des fichiers de données massifs. Un lac de données n'est pas optimisé pour un grand nombre d'utilisateurs ou pour des charges de travail diverses et simultanées en raison de tâches de requêtes intensives. Cela peut entraîner une dégradation du rendement et les pannes sont fréquentes lors de l'exécution simultanée de tâches d'extraction, de transformation et de chargement. Les lacs de données sur site font face à d'autres défis de rendement dans la mesure où leur configuration est statique.

### **Capacité d'accumulation et de stockage des données**

Les données stockées dans les lacs de données peuvent ne jamais être utilisées en production et rester inutilisées indéfiniment dans le lac de données. En stockant des quantités massives de données historiques, le lac de données infini peut fausser l'analyse avec des données qui ne sont plus pertinentes par rapport aux priorités de l'entreprise. En conservant les données historiques, les métadonnées qui les décrivent doivent également être comprises. Cela diminue le rendement du lac de données en augmentant la charge de travail globale des employés pour nettoyer les ensembles de données qui ne servent plus à l'analyse.

Le stockage de quantités de plus en plus massives de données pour une durée illimitée entraînera également des défis d'évolutivité et de coûts. Les défis d'évolutivité sont moins risqués dans les environnements infonuagiques publics, mais le coût reste un facteur. Les lacs de données sur site sont plus sensibles aux défis de coûts. En effet, leurs nœuds de grappe nécessitent les trois dimensions du calcul (stockage, mémoire et traitement). Les organisations de toutes sortes génèrent des quantités massives de données (y compris des métadonnées) et leur nombre augmente de façon exponentielle.

La capacité de stockage de toutes ces données (et des données futures) sera un défi permanent et nécessitera une gestion constante. Bien que les lacs de données puissent être stockés dans le nuage, et ils le seront, SPC, en tant que courtier en infonuagique pour le gouvernement du Canada, devra fournir l'infrastructure et l'évolutivité appropriées aux clients.

### **Utilisateurs avancés nécessaires**

Les lacs de données ne sont pas une plateforme qui doit être explorée par tout le monde. Ils présentent une vue non raffinée des données que seuls les analystes les plus compétents sont habituellement en mesure d'explorer et d'affiner, indépendamment de tout autre système d'enregistrement officiel, tel qu'un entrepôt de données.

Dans une organisation, tout le monde n'est pas assez compétent en matière de données pour tirer profit de grandes quantités de données brutes ou non conservées.

En réalité, seule une poignée de membres du personnel sont assez compétents pour naviguer dans un lac de données. Puisque les lacs de données stockent des données brutes, leur valeur opérationnelle est entièrement déterminée par les compétences de leurs utilisateurs. Ces compétences font souvent défaut dans une organisation.

### **Sécurité des données**

Les données contenues dans un lac de données manquent d'une protection standard en matière de sécurité, avec un système relationnel de gestion des bases de données ou une base de données d'entreprise. Dans la pratique, cela signifie que les données ne sont pas cryptées et qu'elles ne font pas l'objet d'un contrôle d'accès. La sécurité n'est pas seulement une solution binaire. Nous avons divers degrés de sécurité (non classifié, secret, très secret, etc.) et tous nécessitent des approches différentes. Cela présentera inévitablement des défis en matière d'utilisation réussie des données provenant des lacs de données. Pour lutter contre cela, les organisations devront adopter un nouveau cadre de sécurité compatible avec les lacs de données et les scientifiques des données.

## **Considérations**

Services partagés Canada (SPC) a une excellente occasion de tirer parti de son mandat de fournir des services de stockage des données aux autres ministères du gouvernement du Canada. En tant que fournisseur de services du gouvernement du Canada, SPC pourrait éventuellement mettre en œuvre un lac de données du gouvernement du Canada centralisé et permettre aux scientifiques des données du gouvernement du Canada d'accéder à ces données centrales par l'intermédiaire d'une seule interface de lacs de données unifiée. Toutefois, il s'agit d'un projet qui devrait être mis en œuvre après l'adoption de l'infonuagique et la migration des centres de données d'entreprise afin de fournir une infrastructure et une mise à l'échelle adéquates.

Les lacs de données ne devraient pas être confondus avec les bases de données conventionnelles, bien que tous deux stockent des informations. Un lac de données aura toujours un rendement insatisfaisant si on lui confie les tâches d'une base de données conventionnelle. Pour lutter contre cela, SPC doit créer des architectures de données qui définissent l'application correcte des lacs de données. Trop souvent, les lacs de données souffrent d'un manque de prévoyance quant à ce qu'ils sont censés accomplir. La création d'un lac de données devient l'objectif, plutôt que l'atteinte d'un objectif stratégique. Services partagés Canada (SPC) devrait envisager de concevoir l'infrastructure des lacs de données en fonction des accords sur les niveaux de service (ANS) afin de maintenir les efforts sur la bonne voie en matière de lacs de données. Il faut notamment s'assurer que SPC a établi des objectifs clairs pour les lacs de données avant le déploiement.

SPC devrait également envisager la création d'un groupe spécial d'experts qui se concentrerait sur l'analyse avancée et la découverte des tendances en matière de données expérimentales dans les lacs de données. Bien que selon l'hypothèse fondamentale qui sous-tend le concept de lacs de données, toutes les personnes qui

accèdent à un lac de données sont modérément ou extrêmement qualifiées pour manipuler et analyser les données, la plupart ne le sont pas en réalité. SPC devrait envisager un investissement important dans la formation du personnel aux compétences nécessaires, comme la science des données, l'intelligence artificielle, l'apprentissage automatique ou l'ingénierie des données.

SPC devrait être conscient que les attentes en matière de lacs de données sont démesurées. Des attentes démesurées entraînent des cas d'utilisation vagues et ambigus, et augmentent les risques de pannes catastrophiques. En tant que fournisseur de services, SPC doit être strict dans l'établissement d'objectifs clairs pour les efforts d'approvisionnement en lacs de données avant le déploiement. SPC devrait se méfier des tentatives visant à remplacer l'élaboration de la stratégie par l'infrastructure. Un lac de données peut être une composante technologique qui appuie une stratégie de données et d'analyse, mais il ne peut remplacer cette stratégie.

SPC devrait se préoccuper de la fourniture et de l'exploitation de l'infrastructure; les ministères eux-mêmes sont responsables des données qu'ils mettent dans le lac de données. Cependant, en tant que fournisseur de services, SPC devrait surveiller le lac de données en ce qui concerne la gouvernance des données, le cycle de vie des données pour l'hygiène des données, et ce qui se passe dans le lac de données en général. Selon la technologie, SPC devra être très clair sur la façon de surveiller les activités dans les lacs de données qu'il fournit au gouvernement du Canada.

SPC devrait considérer un projet de mise en œuvre d'un lac de données comme un moyen d'introduire ou de revigorer un programme de gestion des données en faisant des capacités de gestion des données une condition préalable à la mise en œuvre d'un lac de données réussi.

Les données devront être qualifiées avant d'atteindre le lac de données, ce qui peut et doit être fait avant tout dans un système d'enregistrement. De cette façon, les données peuvent être organisées pour s'intégrer à la mise en œuvre du lac de données.

SPC devrait élaborer des politiques sur la façon dont les données sont gérées et nettoyées dans le lac de données. Des technologies automatisées de gouvernance des données devraient être ajoutées pour prendre en charge les analyses avancées. La normalisation d'un type particulier d'outil de gouvernance est un problème qui doit être résolu. De plus, la planification d'une gestion efficace des métadonnées, tenant compte de la découverte des métadonnées, du catalogage et de la gestion des métadonnées d'entreprise, appliquée à la mise en œuvre des lacs de données est essentielle. Une application rigoureuse de la discipline et de l'hygiène des données est nécessaire. Pour lutter contre cela, SPC devrait utiliser des outils de gestion des données et élaborer des politiques sur la façon dont les données sont gérées et nettoyées dans le lac de données. La majorité des analystes des lacs de données préféreront travailler avec des données propres, enrichies et fiables. Cependant, la qualité des données est liée à la tâche à accomplir. Des données de faible qualité peuvent être acceptables pour une analyse à faible incidence ou des prévisions à distance, mais inacceptables pour une analyse tactique ou à forte incidence. Les évaluations de SPC devraient tenir compte de cela.

Il faut concevoir des lacs de données avec les éléments nécessaires pour fournir des résultats d'analyse fiables à divers consommateurs de données. L'objectif est d'accroître l'utilisation intersectorielle afin d'offrir des perspectives d'analyse avancées, de créer des lacs de données pour des unités opérationnelles ou des applications d'analyse particulières, plutôt que d'essayer de mettre en œuvre une vague notion de lac de données d'entreprise unique. Cependant, les autres architectures, comme les centres de données, sont souvent mieux adaptées à la mise en commun des données au sein d'une organisation.

## Références

<https://www.dataversity.net/data-lakes-101-overview/#>

<https://www.youtube.com/watch?v=v3yv88h68GY>

<https://www.pcmag.com/article/347020/data-lakes-explained>

<https://hortonworks.com/blog/enterprise-hadoop-journey-data-lake/>

[https://en.wikipedia.org/wiki/Google\\_File\\_System](https://en.wikipedia.org/wiki/Google_File_System)

<https://www.sqlchick.com/entries/2016/7/31/data-lake-use-cases-and-planning>

<https://www.sqlchick.com/entries/2016/7/31/data-lake-use-cases-and-planning>

<https://dataconomy.com/2018/07/six-reasons-to-think-twice-about-your-data-lake-strategy/>

<https://www.arcadiadata.com/blog/the-top-six-reasons-data-lakes-have-failed-to-live-up-to-expectations/>

<https://aws.amazon.com/solutions/data-lake-solution/>

---

<sup>i</sup> Treasury Board of Canada Secretariat. (March 29<sup>th</sup>, 2019). *Digital Operations Strategic Plan: 2018-2022*. Government of Canada. Treasury Board of Canada Secretariat. Retrieved 26-May-2019 from: <https://www.canada.ca/en/government/system/digital-government/digital-operations-strategic-plan-2018-2022.html>

<sup>ii</sup> Ibid.

<sup>iii</sup> Brisson, Yannick, and Craig, Sheila. (November, 2018). *ESDC Data Lake – Implementation Strategy and Roadmap Update*. Government of Canada. Employment and Social Development Canada – Data and Analytics Services. Presentation. Last Modified on 2019-04-26 15:45. Retrieved 07-May-2019 from GCDocs: <https://gcdocs.gc.ca/ssc-spc/llisapi.dll?func=ll&objaction=overview&objid=36624914>