ACTIAN
a division of HCLSoftware

WHITE PAPER

Comment atteindre une observabilité complète des données, sans exploser votre budget

Découvrez une nouvelle approche découplée pour surveiller les pipelines de données et garantir leur qualité à grande échelle, sans surcharger vos systèmes ni grever vos dépenses.

Par Ole Olesen-Bagneux
Chief Evangelist, PhD

Table des matières

3 Pourquoi les organisations ont-elles besoin de l'observabilité des données ?

4 Qu'est-ce que l'observabilité des données ?

5 Comment l'observabilité des données est-elle assurée avec Actian ?

7 Résumé exécutif : le point de vue d'Actian sur l'observabilité des données

8 Actian Data Observability en action

8 À propos d'Actian

La vitesse et l'ampleur des flux de données actuels dépassent les capacités des contrôles qualité traditionnels. Actian propose une approche moderne et intelligente de l'observabilité des données, pour instaurer la confiance, renforcer la fiabilité et optimiser les coûts des opérations de données.

À propos de l'auteur

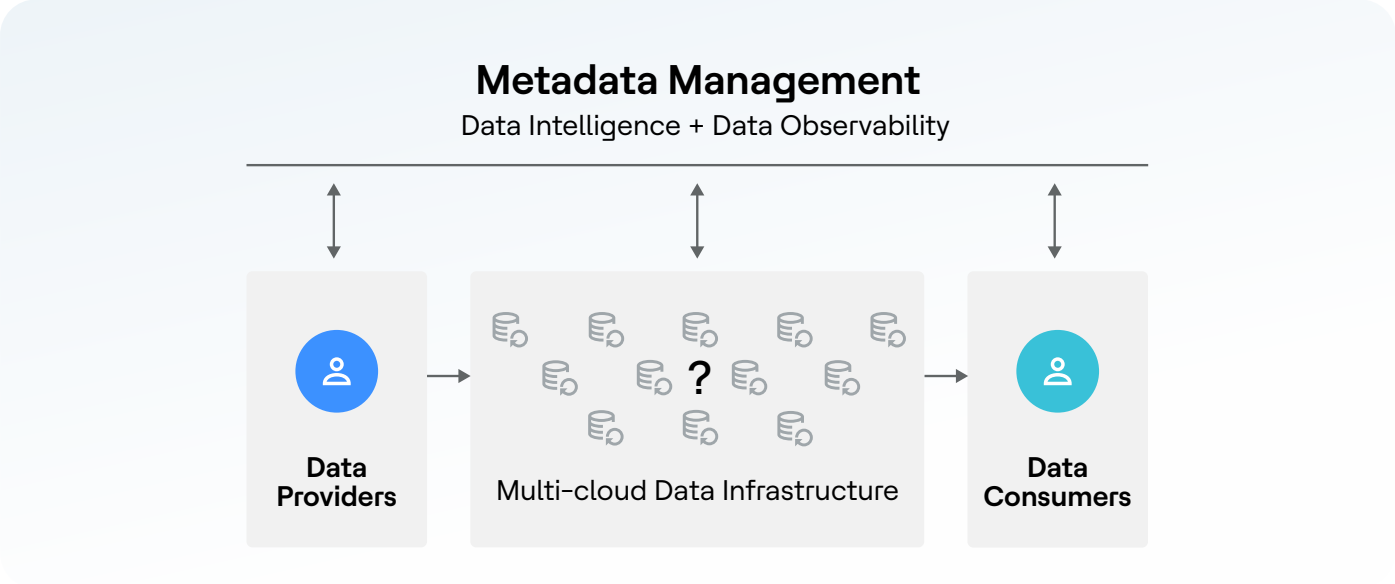
Ole Olesen-Bagneux est un expert mondialement reconnu dans la gestion des métadonnées et l'architecture des données d'entreprise. En tant que vice-président et évangéliste en chef chez Actian, il sensibilise le secteur à l'adoption d'approches modernes en matière d'intelligence des données, en s'appuyant sur sa vaste expertise dans la gestion des données, les métadonnées, les catalogues de données et les architectures décentralisées. Auteur accompli, Ole a publié The Enterprise Data Catalog (O'Reilly, 2023) et travaille actuellement sur Fundamentals of Metadata Management (O'Reilly, 2025). Titulaire d'un doctorat en bibliothéconomie et sciences de l'information de l'Université de Copenhague, il propose une vision unique à la croisée de la science de l'information traditionnelle et de la gestion moderne des données.

ACTIAN™
a division of HCLSoftware

Pourquoi les organisations ont-elles besoin de l'observabilité des données ?

La réponse est simple : les entreprises gèrent d'immenses volumes de données stockées et transformées dans une multitude de technologies. Pour que ces données soient réellement utiles, il est essentiel de les observer avec rigueur. L'observabilité des données est justement née pour répondre à ce besoin, à la fois comme discipline et comme technologie. [1]

Certains collaborateurs sont donc chargés de fournir ces données, qu'ils doivent transporter et transformer afin de les rendre exploitables par d'autres utilisateurs : les consommateurs de données. Pour relier les producteurs et les consommateurs, les entreprises finissent par mettre en place une infrastructure de données complexe, souvent multi-cloud. Et c'est là que les problèmes commencent : cette infrastructure est difficile à contrôler car elle combine de nombreuses technologies issues de divers fournisseurs, manipulant d'énormes volumes de données transférées de multiples façons. Sa gouvernance devient très complexe. Pour y remédier, les entreprises s'appuient sur la gestion des métadonnées pour gouverner leur infrastructure de données. [2]



Pensez aux responsables de systèmes dans votre organisation et aux données qu'ils gèrent : il peut s'agir de données RH, de production, de R&D, de finances, etc. Ces données ont une valeur immense : elles alimentent des cas d'usage innovants et peuvent déboucher sur de nouvelles offres, soutenant ainsi la croissance et la pérennité de votre entreprise.

Cette couche de métadonnées, placée au-dessus de l'infrastructure multi-cloud, leur apporte l'intelligence nécessaire pour piloter leurs données à l'aide de plusieurs fonctionnalités, dont l'observabilité des données.

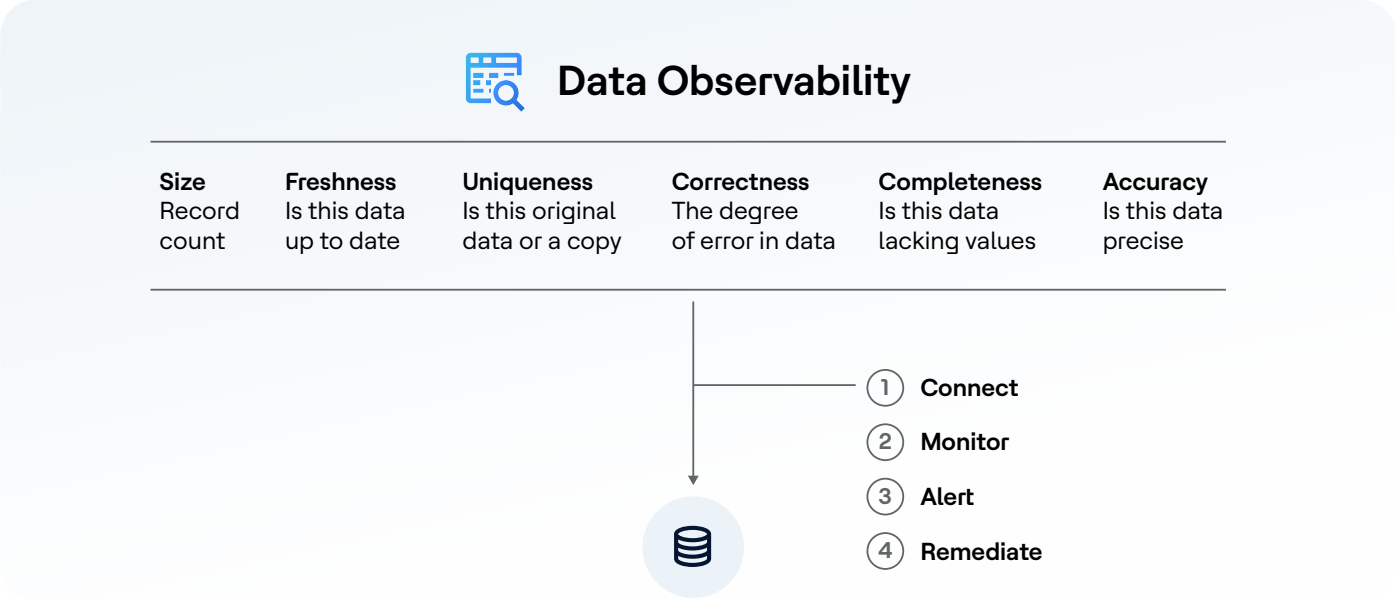
En résumé : les organisations ont besoin de l'observabilité des données parce qu'il est tout simplement impossible de contrôler manuellement la masse de données en circulation. L'observabilité garantit aux consommateurs de données des informations toujours fraîches, correctes et de haute qualité.

ACTIAN™ a division of HCLSoftware

Qu'est-ce que l'observabilité des données ?

Creusons un peu plus. **L'observabilité des données consiste à surveiller des critères tels que la taille, la fraîcheur, l'unicité, la correction, la complétude et la précision des données. Ces métriques [3] peuvent être définies, pilotées et appliquées grâce à des connexions, des systèmes de surveillance, des alertes et des mécanismes de remédiation**

Transport des données dans les pipelines. Le transport des données via des outils ETL (extract-transform-load) ou ELT (extract-load-transform) est au cœur du travail des ingénieurs data. [4] L'observabilité leur permet d'évaluer la portée et la stabilité de leurs pipelines, de réduire leur coût grâce à des étapes plus simples, et même d'anticiper d'éventuels échecs.



L'observabilité des données permet de :

Transporter les données via des pipelines logiques :

- Suivre les performances des pipelines de données
- Analyser leurs coûts
- Anticiper leur fiabilité

Faire évoluer les architectures de données :

- Garantir la fiabilité des produits de données
- Faciliter la mise en œuvre de contrats de données
- Maximiser la réussite des projets d'IA

Renforcer la gouvernance et la conformité :

- Améliorer la qualité des données
- Repérer les données personnelles ou sensibles
- Enrichir le catalogue de données (le « moteur de recherche » de la couche de métadonnées)

Voyons cela plus en détail.

Évolutivité des architectures data. Les entreprises adoptent de plus en plus les data products, car leur gestion est plus évolutive que celle des entrepôts ou des Data Lakes traditionnels. L'observabilité est une composante clé pour faire évoluer les architectures data grâce à des pipelines solides, capables de produire des produits de données de qualité, faciles à consommer via des contrats de données automatisés – un levier essentiel pour réussir les projets d'IA de votre entreprise. [5]

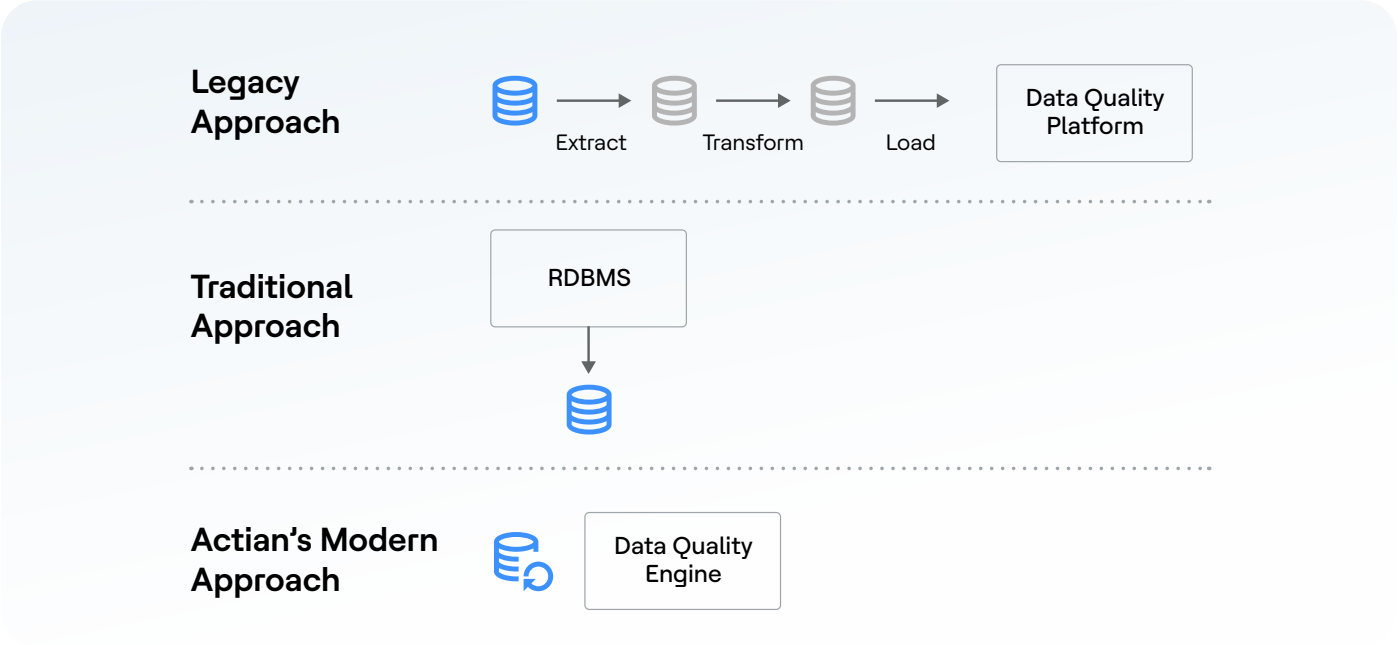
Gouvernance et conformité. L'observabilité joue un rôle central dans la gouvernance des données : elle permet de faire respecter des politiques qualité via une surveillance active, notamment pour l'usage des données sensibles et des informations personnelles identifiables (PII). Elle peut aussi être intégrée au « moteur de recherche » de la couche de métadonnées (le catalogue de données), afin que les utilisateurs ne découvrent que des jeux de données fiables. [6]

ACTIAN™ a division of HCLSoftware

Comment l'observabilité des données est-elle assurée avec Actian ?

Chez Actian, nous défendons une approche entièrement nouvelle et moderne de l'observabilité des données. Aujourd'hui, cette observabilité repose le plus souvent sur une méthode traditionnelle, et certaines organisations utilisent encore une approche héritée. Vous pouvez voir ces différentes approches dans le diagramme ci-dessous.

Pour comprendre en quoi la solution d'Actian se démarque, examinons les trois grands modèles d'observabilité des données : l'approche héritée, l'approche traditionnelle et l'approche moderne.



L'approche héritée. L'approche héritée de l'observabilité des données consiste à mettre en place un pipeline de données classique, qui extrait, transforme et charge les données dans une plateforme de gestion de la qualité, où une combinaison de contrôles automatiques et manuels est appliquée. Ce dispositif est très chronophage, n'est pas synchronisé avec les données qu'il observe (des changements peuvent survenir et rendre obsolètes les constats effectués), et il ne peut pas passer à l'échelle car il nécessite trop d'interventions humaines.

L'approche traditionnelle. Les approches traditionnelles de l'observabilité des données s'appuient sur un système de gestion de base de données relationnelle (SGBDR) placé au-dessus de la base pour interroger et observer les données, de la même manière que pour tout autre type de requête effectué via le SGBDR.

Cette méthode présente certains avantages par rapport à l'approche héritée, mais elle reste confrontée à de sérieuses limites. Il existe une contrainte de puissance de calcul qui limite le nombre de requêtes pouvant être exécutées, et pour rester rentable, ces requêtes ne peuvent pas être lancées sur l'ensemble des données.

Par conséquent, les données sont observées de manière partielle, car seules des observations sur échantillons sont possibles, ce qui rend l'observabilité fondamentalement peu fiable. Elle ne peut pas s'étendre à un niveau permettant de surveiller toutes les données à la source. De plus, cette approche présente d'importantes restrictions en matière de types de sources observables, excluant notamment les données semi-structurées, comme les fichiers JSON imbriqués ou YAML. Enfin, comme elle repose intensivement sur des requêtes, cette approche est également coûteuse. Et souhaitez-vous vraiment une observabilité des données qui soit incomplète, peu fiable et onéreuse ?

ACTIAN™ a division of HCLSoftware

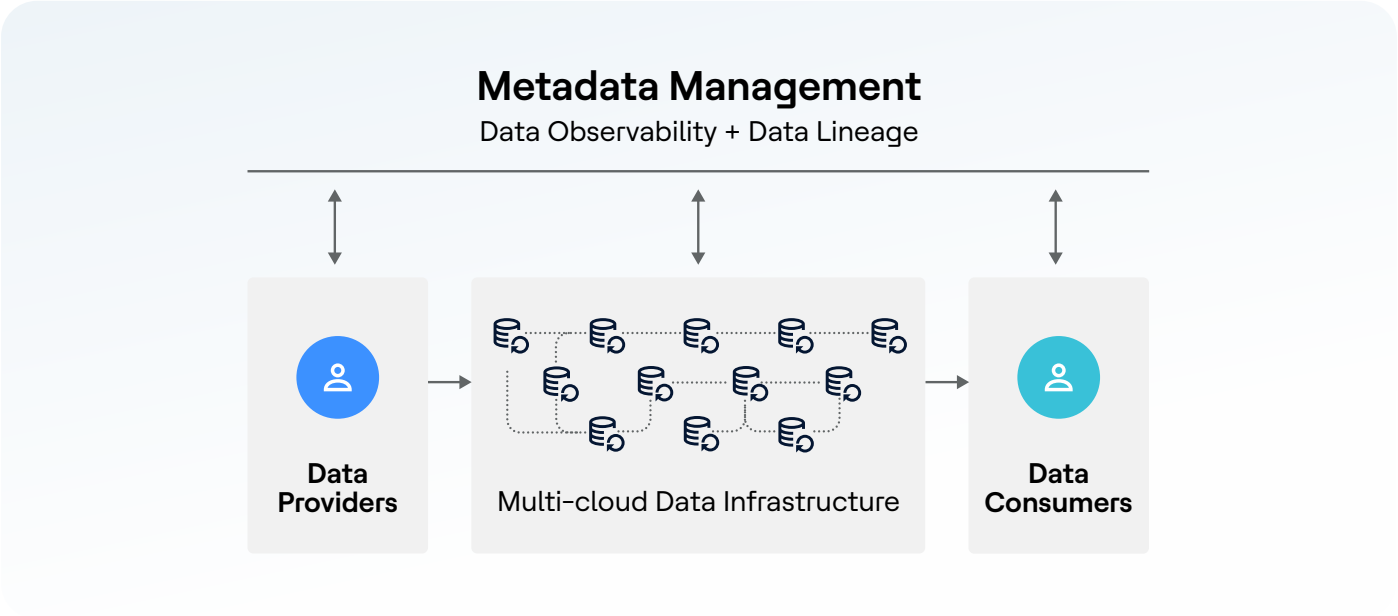
L'approche moderne d'Actian

Pour observer les données, Actian Data Observability déploie un moteur de qualité des données, léger et découplé, placé à proximité des données, en mémoire, et non au-dessus d'elles via un SGBDR. [7] Ainsi, les données ne sont pas transportées pour être observées, comme dans l'approche héritée. Elles ne sont pas non plus interrogées par des requêtes SQL, comme dans l'approche traditionnelle. Grâce à cette approche moderne, reposant sur un moteur découplé, Actian Data Observability peut se connecter à toutes vos sources de données, qu'elles soient structurées ou semi-structurées – des fichiers JSON, Parquet, YAML, jusqu'à la plupart des formats de tables ouverts et des flux – souvent inaccessibles via l'approche traditionnelle. L'un de nos plus grands atouts réside dans le fait qu'Actian Data Observability rend la qualité des données totalement indépendante des systèmes sous-jacents. Cette approche agnostique simplifie considérablement l'adoption pour les clients disposant de piles technologiques hétérogènes. De plus, cette architecture découplée pérennise les processus qualité, en évitant les casse-têtes liés aux migrations lorsque les systèmes évoluent. Puisque le moteur de qualité fonctionne aux côtés des données, il ne sollicite pas les systèmes opérationnels en exécutant des requêtes sur leurs bases en continu.

Cela permet aussi de réduire considérablement les coûts de l'observabilité, grâce à une surveillance précise et condensée en mémoire. Enfin, cette approche permet d'assurer une observabilité complète de l'ensemble des données à la source, sans avoir à se soucier du coût des requêtes. Vous bénéficiez ainsi d'une observabilité des données qui est complète, simple à mettre en place, économique – et surtout, intégrale.

La solution complète : la traçabilité au service de l'observabilité

À elle seule, une solution d'observabilité ne permet pas de suivre les mouvements des données entre les différentes sources. Elle se contente d'examiner chaque source indépendamment, sans les comparer entre elles. Cela signifie qu'il est impossible de suivre l'évolution de la qualité des données au fil de leur parcours dans l'infrastructure, à moins de cartographier manuellement les sources. C'est pourquoi Actian Data Observability intègre désormais une fonction de traçabilité des données (data lineage), qui suit automatiquement les données entre les différentes sources. Cette combinaison rend possible une observabilité parfaite, de la source à la consommation, du producteur de données au consommateur.



ACTIAN™ a division of HCLSoftware

Résumé exécutif : le point de vue d'Actian sur l'observabilité des données

Chez Actian, nous sommes convaincus que l'observabilité des données peut et doit s'appuyer sur les valeurs technologiques qui nous sont chères : elle doit être fiable, flexible et simple à utiliser. C'est précisément ce que nous proposons avec Actian Data Observability.

En tant que responsable data ou décideur technologique, vous devez retenir ceci : votre organisation mérite une solution d'observabilité moderne, économique et réellement complète. Actian Data Observability incarne exactement cela, dans un marché dominé par des technologies qui sont, malheureusement, coûteuses, complexes et, tout simplement, incomplètes.

Vous ne pouvez pas compter sur les solutions traditionnelles d'observabilité des données pour mener à bien vos projets stratégiques – ceux dont dépend l'avenir de votre entreprise. Ces technologies ne sont pas conçues pour évoluer à grande échelle et, dans la plupart des cas, elles ne parviennent même pas à s'imposer efficacement dans les organisations. En revanche, elles épuiseront vos systèmes sources et feront grimper les coûts.

Avec Actian Data Observability, combinée aux capacités de traçabilité des données d'Actian, votre entreprise peut instaurer la confiance dans ses données à l'échelle de l'ensemble du système d'information – tout en assurant la continuité des systèmes de production, avec un minimum de frais.

Références

[1] Nous abordons ici l'observabilité des données en nous appuyant sur la définition, que nous partageons, proposée par Gartner®, dans « Quick Answer: 'What is Data Observability' », Ankush Jain, 8 juin 2022 : « L'observabilité des données est la capacité d'une organisation à disposer d'une visibilité étendue sur son patrimoine de données et sur les dépendances multi-niveaux (comme les pipelines, l'infrastructure, les applications de données) à tout moment, afin d'identifier, contrôler, prévenir, escalader et résoudre rapidement les interruptions de données dans les limites fixées par les SLA. »

[2] Le diagramme est inspiré de la Figure 1-1. de Piethein Strengholt: Building Medallion Architectures (O'Reilly, 2025) – l'observabilité des données y est également identifiée comme une capacité essentielle.

[3] Nous n'entrerons pas ici dans les détails mathématiques des métriques. Pour une source utile sur le sujet, voir Andy Petrella : Fundamentals of Data Observability, (O'Reilly, 2023), chapitre 4

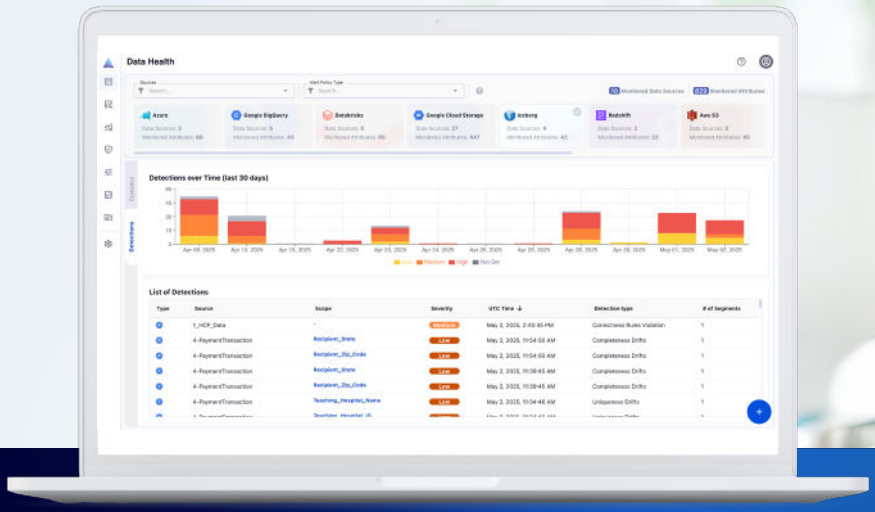
[4] L'ouvrage de référence sur l'ingénierie des données est celui de Joe Reis et Matthew Housley : Fundamentals of Data Engineering (O'Reilly, 2022)

[5] Selon une étude, plus de 60 % des projets IA échoueront d'ici 2026 faute de données prêtes pour l'IA, voir : Gartner® : « A Journey Guide to Delivering AI Success through "AI-Ready" Data » par Ehtisham Zaidi, Roxane Edilali, 18 octobre 2024

[6] Ole Olesen-Bagneux : The Enterprise Data Catalog, (O'Reilly, 2023), chapitre 7

[7] Les détails techniques de cette configuration sont importants, mais trop denses pour ce livre blanc. en savoir plus, consultez : **actian.com/data-observability**

ACTIAN™ a division of HCLSoftware



Vous souhaitez voir Actian Data Observability en action ?

Commencez dès maintenant à réduire vos coûts, à renforcer la confiance et à simplifier la gouvernance de l'ensemble de votre architecture de données, plus rapidement que vous ne l'auriez imaginé

Grâce à l'approche moderne et découplée d'Actian, vous bénéficiez d'une visibilité complète sur la qualité de vos données, sans alourdir vos systèmes ni votre budget. Que vous cherchiez à faire évoluer vos produits de données, à vous préparer à l'IA ou à appliquer une gouvernance à grande vitesse, Actian vous aide à garder le contrôle en toute confiance.

> Rendez-vous sur **actian.com** pour **demander une démonstration** et découvrir comment Actian peut transformer votre manière d'observer, de fiabiliser et de gérer vos données.

À propos d'Actian

Actian permet aux entreprises de gérer et de maîtriser leurs données en toute confiance et à grande échelle. Les organisations font confiance aux solutions de gestion et d'intelligence des données d'Actian pour rationaliser les environnements de données complexes et accélérer la fourniture de données prêtes pour l'IA. Conçues pour être flexibles, les solutions Actian s'intègrent de manière transparente et fonctionnent de manière fiable dans les environnements sur site, cloud et hybrides. Pour en savoir plus sur Actian, la division Données de **HCLSoftware**, rendez-vous sur le site **actian.com**.

Actian France 27 Avenue de l'Opéra, 75001 Paris, France Tel: +33 1 73 19 12 87 **Actian.com**

© 2025 Actian Corporation. Actian is a trademark of Actian Corporation and its subsidiaries. All other trademarks, trade names, service marks, and logos referenced herein belong to theirrespective companies.

All Rights Reserved. V1-2025-6

ACTIAN™
a division of HCLSoftware