

Solution “Data Lakehouse Management”

Dremio Cloud ou
Dremio On-premise



DREMIO-SYNALTIC • Décembre 2023



Agenda

1. Présentation Synaltic
2. Pourquoi Dremio
3. Notre offre Dremio-Synaltic
4. Tarification Publique
5. Cas Clients

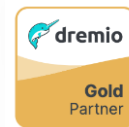


The background is a solid orange color with a pattern of thin, white, wavy lines that create a sense of movement and depth. The lines are more concentrated in some areas, creating a grid-like or mesh-like effect, while in other areas they are more sparse and flowing.

1

SYNALTIC

Au Sujet de Synaltic



Experts en Data Management, passionnés d'Open Source et d'Innovation !



19

20 ans d'expérience dans le **Data Management** et plus de **250 projets réalisés** en France et dans les pays limitrophes.

30

30 collaborateurs certifiés intervenant en France et dans les pays limitrophes, **contribuant aux communautés** d'utilisateurs.

14

Acteurs innovants par sa R&D, depuis plus de 14 ans, nous pratiquons le CIR : nous accueillons un doctorant et participons à des projets collaboratifs européens.



Synaltic, expert Dremio depuis 2018 en France



Sur ces quinze dernières années, les solutions répondant aux enjeux du big data n'ont cessé de bouleverser le marché. C'est l'open source et le cloud qui auront le plus apporté d'innovation dans cette ère nouvelle où la donnée est devenue le nouvel eldorado. Dans ce paysage, Dremio a fait un pari : apporter simplicité, collaboration, convivialité sans perdre de vue la performance attendue par les utilisateurs dans le cadre d'une interrogation ad hoc.

Ce sont pour toutes ses raisons que Synaltic fait très vite le choix de Dremio, se rapproche rapidement de l'éditeur et de la communauté, et forme ses collaborateurs dès 2018.

Consultant Dremio Expert

- 20 ans d'expérience.
- Java J2EE, Data Warehouse, SQL, Talend, MariaDB ColumnStore, Hadoop, Apache Kafka, Apache Airflow
- Dremio : Architecture, Cloud, Pont Hadoop, Déploiement Yarn, Déploiement Standalone, Accélération, Couches Sémantique, Apache Iceberg.

Consultant Dremio Confirmé

- 5 ans d'expérience
- Python, Machine Learning, SQL, Data Warehouse, Apache Kafka, Oracle, Snowflake
- Dremio : Couches Sémantiques, Accélération, Accompagnement client découverte et prise en main de Dremio

Formations Synaltic - Dremio

- Système Avancé de gestion de base de données, SQL, Talend, Tableau, Data Warehouse, Gestion agiles de projets de données, Data Engineering, Data Visualisation, Gouvernance, Sécurité
- Dremio essential (Dremio utilisateurs métier, Data Curation), Dremio Avancée (Gouvernance, Accélération)
- "Compagnonnage" autour des projets Dremio

Synaltic partage son expertise Dremio

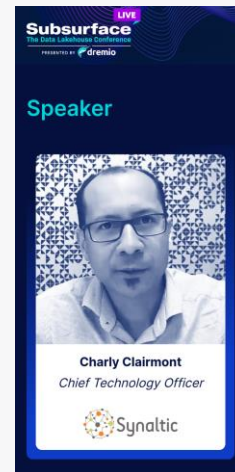
From Business Intelligence to Streaming Decisions

Subsurface - 2 Mars
2023

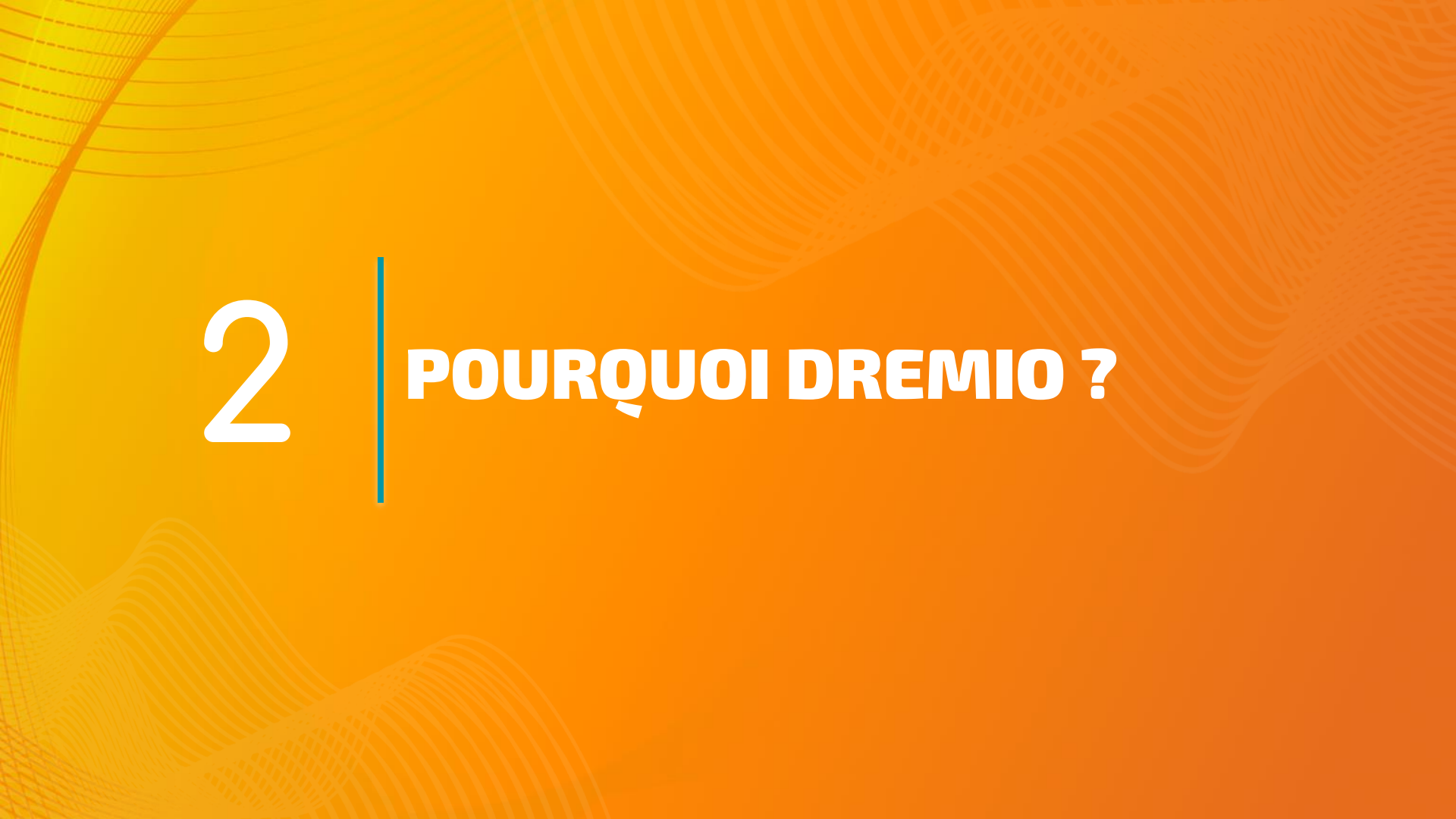
Retour sur 20 ans de
data warehousing.

Mise en avant d'une
architecture de livraison
de données en continue
vers un data lake S3
alimenté via Apache
Kafka au format Apache
Iceberg.

Dremio actualise en
temps réel les jeux de
données pour les livrer
aux métiers.



OW2Con 2019, Paris Open Source Summit
2019, Big Data Paris 2020, Modern Data
Stack France

The background is a solid orange color with a pattern of thin, white, wavy lines that create a sense of movement and depth. The lines are more concentrated on the left side and fade out towards the right.

2

POURQUOI DREMIO ?

Au Sujet du Partenariat

Les fondateurs de Dremio ont l'open source chevillé au corps ! La motivation de créer une solution telle que Dremio remonte au début de leur carrière quand ils observaient à quel point les utilisateurs métier avaient un mal fou à accéder et à consulter les données. Rapidement, ils participent aux projets Hadoop. En 2015 ils commencent les développements pour créer "Dremio, qui a pour objectif de faciliter l'accès aux données en libre-service pour les analystes et les data scientists, de la même manière que AWS rend l'infrastructure en libre-service pour les développeurs".

Aujourd'hui, Dremio est un Data Lakehouse dont le moteur ultra-rapide fournit les performances nécessaires permettant la mise place de tableaux de bord interactifs même pour de larges volumes de données. Dremio, apporte une couche sémantique qui permet aux utilisateurs non-techniques de créer, partager et réutiliser des jeux de données grâce à une interface client intuitive.

Dremio réduit la complexité des architectures pour permettre aux entreprises de toutes tailles de s'approprier toutes leurs données autant celles qui se trouvent dans les data lake que celles qui sont dans les bases de données.

L'aventure de Dremio a débuté en France avec des acteurs comme OVHcloud, RTE, Le Point. Depuis 2017, Synaltic promeut cette technologie. Aujourd'hui, d'autres acteurs font partie de la communauté Dremio : Bic, Michelin, BNP PARIBAS. Tous les clients de Dremio sont unanimes concernant cette technologie : on note une augmentation de la productivité et les données disponibles sont mises à disposition des utilisateurs tellement plus rapidement.

Au-delà de la performance, Dremio est bâti avec une architecture ouverte qui offre une vraie liberté à ses utilisateurs : il n'est point question de les enfermer dans un quelconque format propriétaire ou porté par une poignée d'organisations. Cette ouverture est propice à l'innovation car Dremio continue et continuera de créer de la valeur pour ses utilisateurs. Dremio est l'une des rares plateformes de données à être directement prête pour le cloud hybride. Dremio est aussi l'une des rares plateformes à rendre la donnée "fluide", apportant une expérience de type Git pour les données et optimisant automatiquement les données pour garantir des analyses de haute performance.

Au Sujet de Dremio

Le Data Lakehouse Facile et Ouvert

- **Fonctionnalités et performances type Data Warehouse** directement sur le Data Lake
- **Accélération des requêtes** pour éliminer les copies et les extraits de BI
- **Couche sémantique** pour permettre du libre-service gouverné
- **Multi-connecteurs** de base de données pour permettre des requêtes fédérées

Enterprise Adoption

- **+ de 1000 références** d'entreprises de tous les secteurs
- **5** des 10 entreprises du Fortune 10

Open Source et Communauté

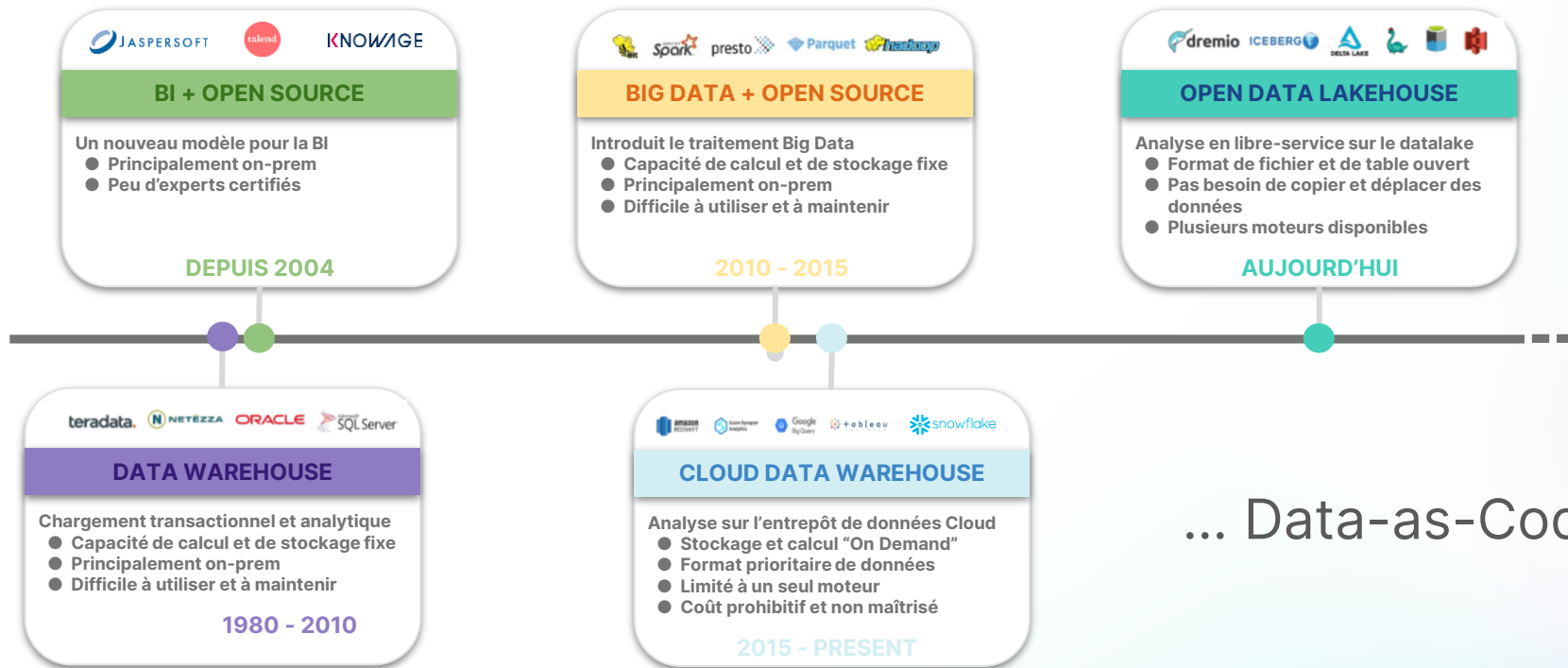
- Apache Arrow (**60M+** downloads/m), Apache Iceberg, Nessie
- Créateur et organisateur de la conférence **Subsurface LIVE**

Fondée en 2014, Version 1.0 en 2017

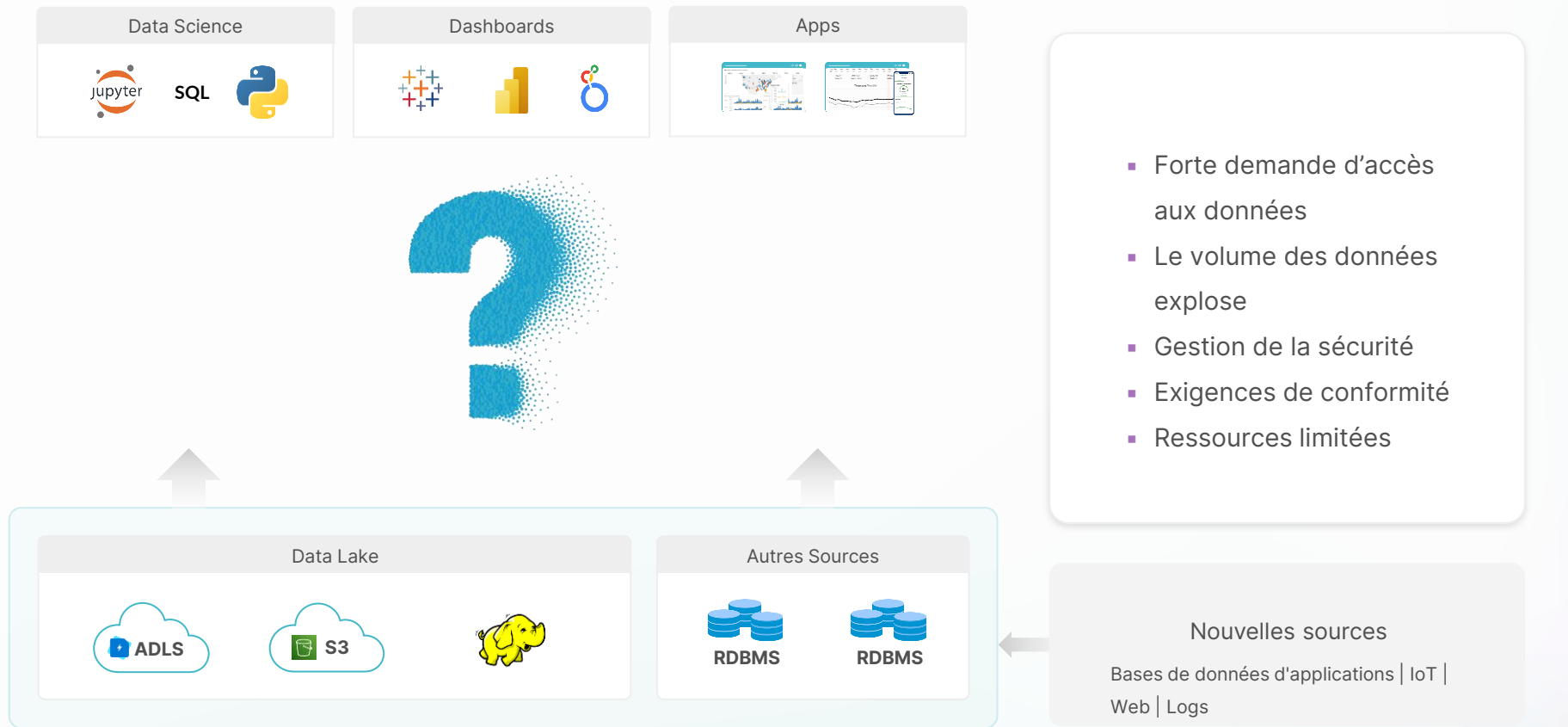


Historique de la Data Analytics

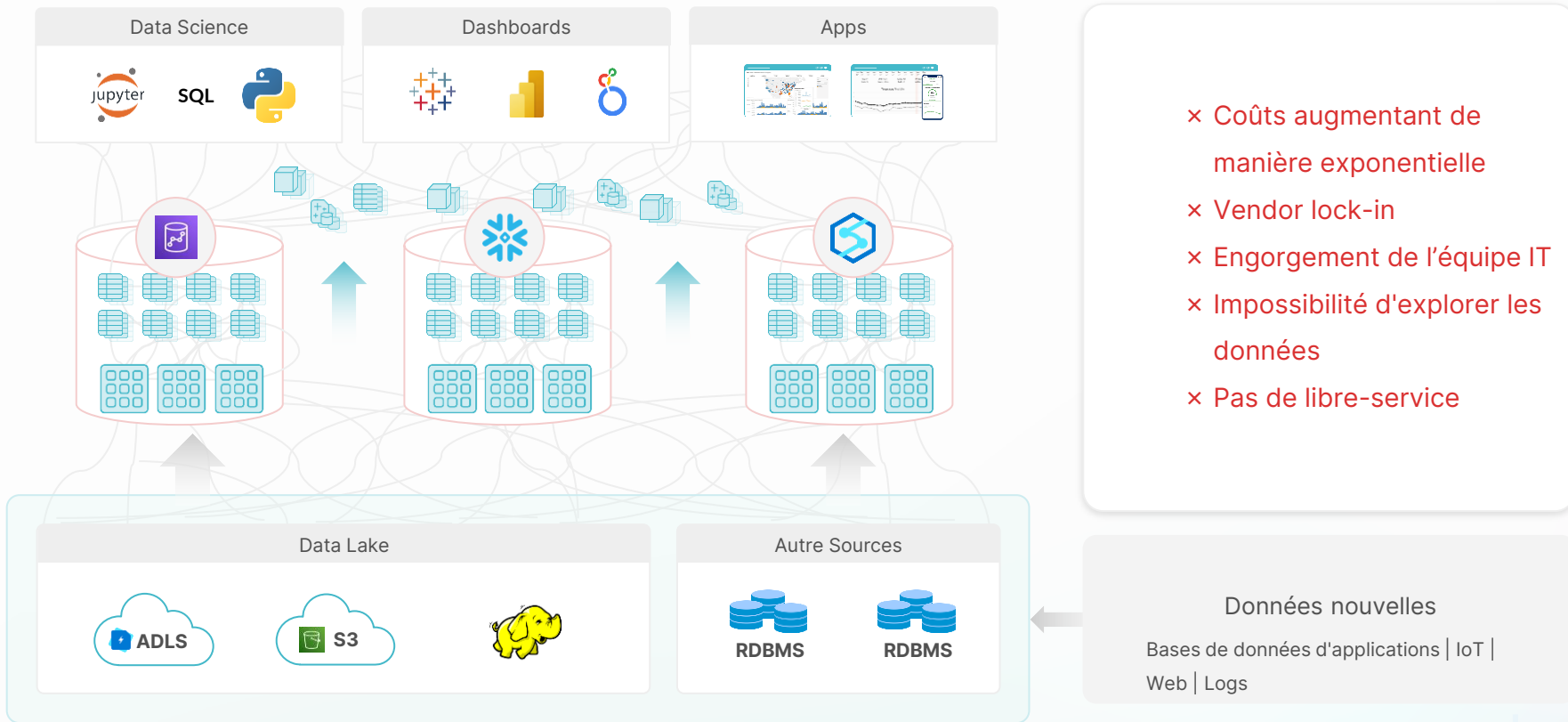
Du Data Warehouse au ...



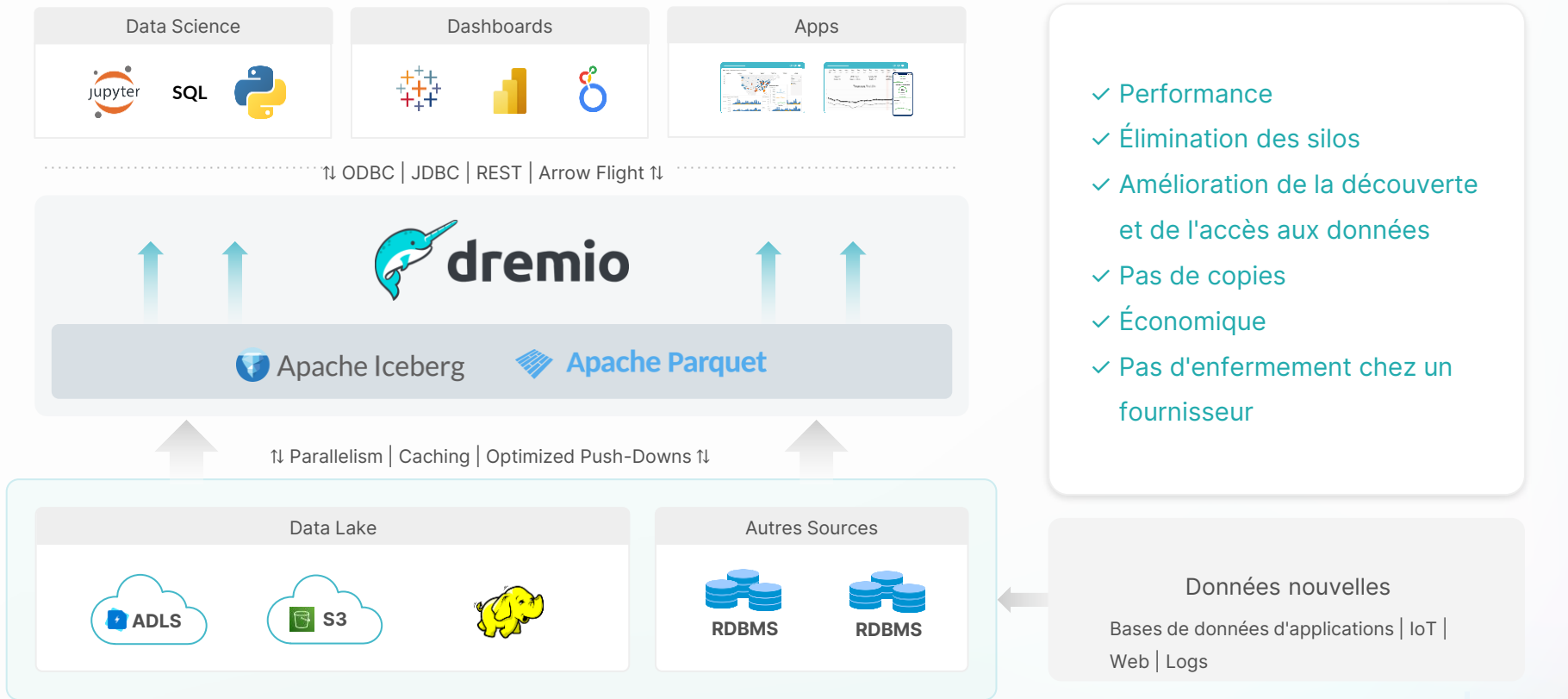
Les entreprises veulent démocratiser les données... Mais comment ?



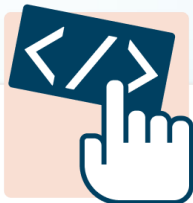
Data Warehouses : Coûteux, Propriétaires, Complexes



Dremio Data Lakehouse: Facile, Ouvert, 1/10ème du coût



L'avantage Dremio



Analytique en libre-service

Interface utilisateur moderne et intuitive

Vue unifiée des données
(en local, hybride et Cloud)

Requêtes fédérées



Données ouvertes, pas de lock-in

Basé sur des normes impulsées par la communauté, notamment :

- Apache Parquet
- Apache Iceberg
- Apache Arrow



Performance à 1/10ème du coût

Requêtes performantes

Haute concurrence

Pas de copies de données coûteuses à gérer

The background is a solid orange color with a pattern of thin, white, wavy lines that create a sense of movement and depth. The lines are more concentrated on the left side and fade towards the right.

3

ACCOMPAGNEMENT

Valeur des Services Professionnels Dremio



Expertise

Tirer partie des plans et des méthodologies

Exploiter les compétences et l'expertise

Découvrir la meilleure voie à suivre

->

Déployer la production plus rapidement

->

Risque de livraison minimisé

->

Confiance dans la solution



Education

Être formé sur les fonctionnalités de Dremio

Apprendre à intégrer avec les sources de données et les applications clientes

Se familiariser avec les modèles de conception Dremio



Économies

Économiser du temps, des coûts et des ressources en tirant partie des meilleures pratiques

Réduire le temps de développement et d'intégration

Lorsque les délais doivent être respectés et ne peuvent être retardés, exploitez les Services Professionnels Dremio (Dremio PS)

Silver - 24x7 Support compris dans la souscription

Ressources en libre-service

- Documentation produit & Université Dremio

Excellence du service

- Accès 24x7 aux nouvelles versions Entreprise
- Accès 24x7 à la base de connaissances et aux forums communautaires
- Notification proactive des corrections importantes et des améliorations
- Intégration/initiation au support avec la direction du support
- Accès **24x7** au support Dremio pour les problèmes **urgents**

Caractéristiques

- Engagement de **support 24x7** dans les centres de support EMEA, USA, APAC
- Support basé sur le Web pour signaler et hiérarchiser les problèmes
 - Temps de réponse initiale d'**1 heure** pour les problèmes **urgents**
 - **Cinq (5) contacts de support** pouvant déposer / escalader des tickets auprès du support Dremio
- Analyse et remédiation complètes par des ingénieurs hautement qualifiés

Abonnement de support Silver
(24x7x365)



Support Features [Silver]	Service Level
Forums Communautaires	24x7x365
Base de Connaissances	24x7x365
Niveau support	24x7 Enterprise
Contacts Support	5
SLA de réponse initiale	Urgent: 1 heure High: 4 heures ouvrables Normal: 1 jour ouvrable Low: 2 jours ouvrables

Gold - 24x7 Support

Toutes les fonctionnalités de Silver plus :

Gestionnaire de compte technique

- On-boarding approfondie avec une documentation personnalisé du déploiement

Excellence de service

- Kick-Off et formation avec un gestionnaire de compte technique assigné
- **Accès 24h/24 et 7j/7** au support Dremio pour les problèmes urgents / **prioritaires** élevés
- Revue **mensuel** des tickets et rapport **trimestriel** sur les performances du support

Services techniques personnalisés

- Gestion d'incidents critiques et traitement stratégique des tickets
- Programme de planification et d'exécution de mise à niveau annuelle unique
- Un "healthcheck" annuel

Caractéristiques

- Engagement de support **24h/24** dans les centres de support EMEA, USA, APAC
- Temps de réponse initiale **d'une heure** pour les problèmes prioritaires urgents
- Temps de réponse initiale de **quatre heures** pour les problèmes prioritaires élevés
- Ligne téléphonique directe pour signaler les problèmes urgents / prioritaires élevés
- **Dix (10) contacts** désignés qui peuvent soumettre / escalader des tickets auprès du support Dremio.

Abonnement Gold (24x7x365)

Prix fixé à € 100,000 par an



Support Features [Gold]	Service Level
Forums Communautaires	24x7x365
Base de Connaissances	24x7x365
Niveau support	Gold
Contacts Support	10
SLA de réponse initiale	Urgent: 1 heure High: 4 heures Normal: 1 jour ouvrable Low: 2 jours ouvrables

Support - Matrice Comparative

Service	Silver	Gold
Heures de support	24x7x365	24x7x365
Nombre de contacts de support	5	10
Base de connaissances et forums communautaires	✓	✓
Formation en support	Local Support Mngr	TAM
Temps de réponse initiale - Urgent	1 hr	1 hr
Temps de réponse initial - Élevé	4 hr ouvrables	4 hrs
Temps de réponse initial - Normal	1 jour ouvrable	1 jour ouvrable
Temps de réponse initial - Faible	2 jours ouvrables	2 jours ouvrables
Ligne de téléphone d'escalade de support	✓	✓
Gestionnaire de compte technique	N/A	✓

S'APPROPRIER DREMIO



Engagement
Kickoff &
Technical
Deep Dive

Gestion de projet / Livraison

Mois 1

Mois 2

Mois 3

Formation /
Conseil

Lancement

Utilisation des Clients

Couche Sémantique

Conception des accélérations

Administration Dremio

Sources de données

Exécution des requêtes

Optimisation des requêtes

**Synaltic /
Dremio**

Création d'un cas d'utilisation initial + conseils sur des cas d'utilisation supplémentaires

Conception &
Dev.

Intégrer les sources de données
Conception de la couche sémantique
Conception de réflexion
Configuration de la sécurité des données

Intégration avec les outils de consommation
Conception et optimisation des requêtes
Optimisation de la gestion de la charge de travail
Accélération tableau de bord

Assistance à la production
et à la mise en service

Dremio

Installation et déploiement hors production

Installation et déploiement de production

Activation des opérations et de la maintenance

Infrastructure.

Configurer l'environnement de
développement
Configurer l'authentification Dremio

Configuration de l'environnement de production : sur site /
cloud
Configurer l'authentification Dremio
Configurer la haute disponibilité (le cas échéant)
Configurer les règles de gestion de la charge de travail
initiales

Aider à intégrer les pipelines CI/CD des clients avec
Dremio
Configuration / intégration de la surveillance Dremio
Configurer et tester la sauvegarde et la restauration
Dremio
Configurer et tester la haute disponibilité de Dremio
(HA)
Définition de la stratégie de reprise après sinistre (DR)

Dremio

Série d'Ateliers Dremio

Accompagner votre appropriation de Dremio

1

Initiation Dremio

Architecture Dremio
Objets Basiques
Accélération des requêtes

2

Cas Clients

Dremio UI
SQL and REST API
BI dashboards

3

Data Sources

Data lakes & C3
Iceberg et Nessie
External sources

4

Couche Sémantique

Utilisateurs et rôles
Privilèges
Conception de couches

5

Exécution de Requêtes

Workload management
Moteurs Elastiques
Maximiser la performance

6

Création de Reflections

Stratégies de conception
Processus d'évaluation
Options d'actualisation

7

Optimisation Requête

Query architecture
Query profiles
Scenarios

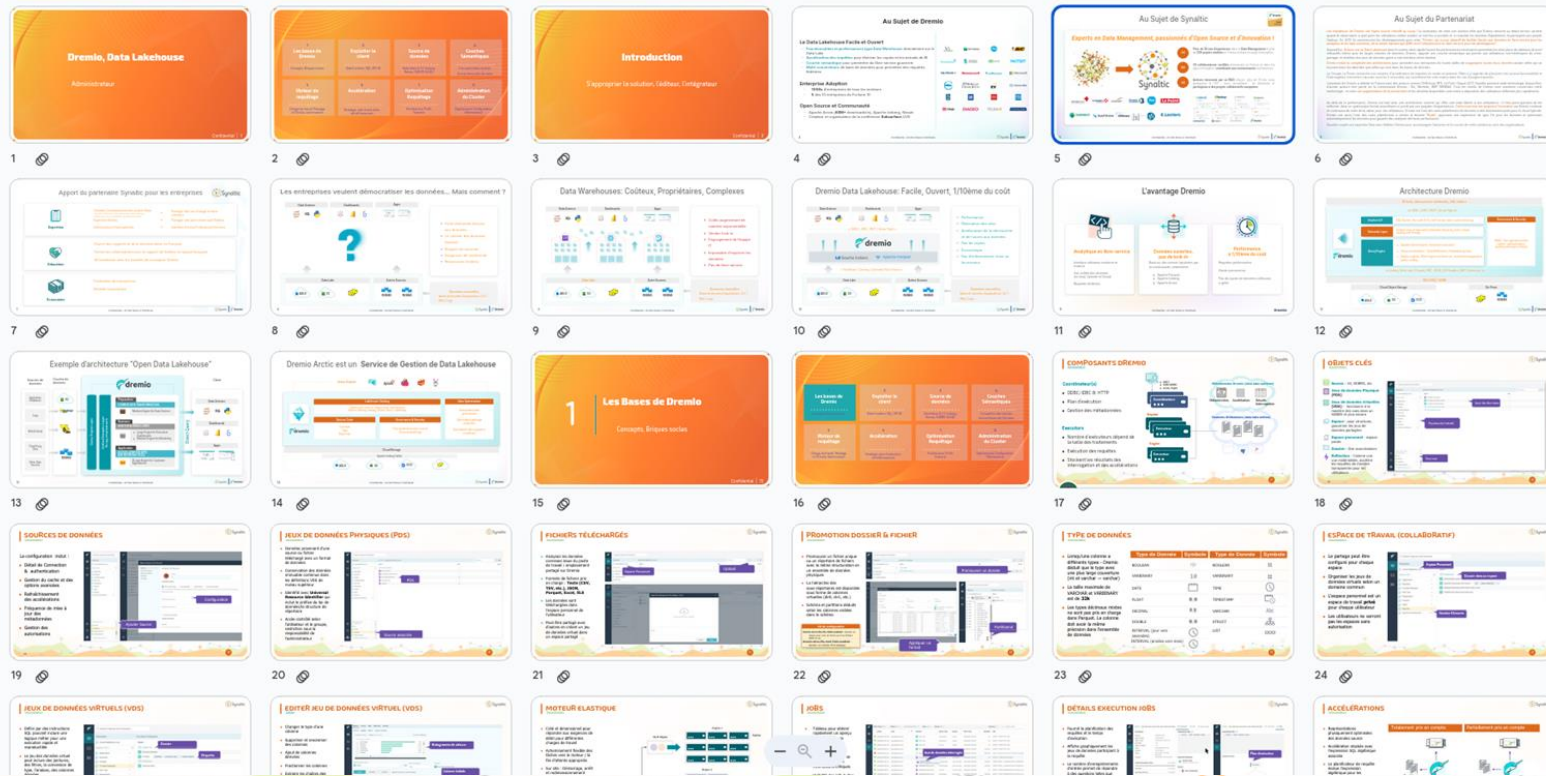
8

Cluster Admin

Options de déploiement
Configuration
Maintenance

FORMER LES DIFFÉRENTS ACTEURS

Supports de formation disponibles en français



The background is a solid orange color with a pattern of thin, white, wavy lines that create a sense of movement and depth. The lines are more concentrated on the left side and fade out towards the right.

4

TARIFICATION

Les versions de Dremio, résumé (*)

Platform Edition	DREMIO CLOUD	DREMIO SOFTWARE	
	Enterprise	Community (CE)	Enterprise (EE)
	Cloud managed by Dremio	Communautaire Edition	On Premise
Cloud Service	✓		
Governed, Self-Service Experience	✓	✓	✓
Security	✓		✓
Identity Management & Authentication	✓		✓
Regulatory Compliance	✓	✓	✓
Data Sources	✓	✓	✓
SQL Support, Performance, and Scalability	✓	✓	✓
Interfaces and Client Connectivity	✓	✓	✓

(*) Se reporter au site éditeur pour les détails : <https://www.dremio.com/pricing/>

(1) €0.39/DCU

(2) 15 000 € par noeud

Modèle Souscription - Prix Public Dremio "On Premise"

Cluster Production Plateforme Dremio

150.000 €

- 2 nœuds coordonnateurs, 10 nœuds exécuteurs
- Utilisateurs illimités
- Sources de données illimitées
- Données et réflexions illimitées
- Cœurs et CPU illimités



Nœuds de NON Production

75.000 €

- 2 nœuds coordonnateurs, 5 nœuds exécuteurs
- Utilisateurs illimités
- Sources de données illimitées
- Données et réflexions illimitées
- Cœurs et CPU illimités

Exemples:

- ☐ Un cluster de 25 nœuds : $150\ 000 + 15 \times 12\ 000 = 330\ 000\text{€}$
- ☐ Deux clusters de 30 nœuds : $2 \times 150\ 000 + 15 \times 12\ 000 + 25 \times 11\ 000 = 755\ 000\text{ €}$
- ☐ Trois clusters de 5 nœuds : $3 \times 150\ 000 = 450\ 000\text{ €}$

Conditions:

- ☐ Un « Nœud » désigne une machine physique ou virtuelle disposant d'un maximum de 128 Go de RAM alloués au Logiciel. Si plus de 128 Go de RAM sont alloués au Logiciel, le nombre de Nœuds augmentera proportionnellement (par exemple, le Client achètera deux (2) Nœuds s'il alloue 256 Go de RAM au Logiciel).

Dremio Cloud, estimation CPU

DCU - Dremio Consumption Units :

Il s'agit de l'outil de mesure (U.O) de Dremio Cloud basé sur le nombre de nœuds et la taille des moteurs (size) qui permettent de déterminer le niveau de consommation en termes de traitement (puissance et heure de consommation).

Dans notre exemple, nous optons pour 2 nœuds avec un moteur de taille "small" soit 16 DCU par heure de traitement (voir schémas ci-contre).

Exemple de calcul de DCU :

- pour le traitement quotidien,
2h chaque jour, soit **224 DCUs** - ($2*7*16$) à la semaine.
- pour le traitement hebdomadaire :
2h chaque weekend soit **32 DCUs** ($16*2$) par semaine.

Au total, nous sommes sur une estimation de **14.976 DCU** pour les traitements réguliers à l'année.

A cela, nous ajoutons une utilisation moyenne de 5h/semaine ce qui revient à 1h par jour ouvré, soit 4 160 DCU soit au total, nous estimons à **20.000 DCU** pour l'année 2023.

Ces chiffres sont bien sur des estimations.

Type d'instance	DCU/h	Prix sur demande* (\$/heure)
XX Petit	4	1,56 \$
X petit	8	3,12 \$
Petit	16	\$6.24
Moyen	32	12,48 \$
Grand	64	24,96 \$
X Grand	128	49,92 \$
2X Grand	256	99,84 \$
3X Grand	512	199,68 \$

Taille	Nombre de nœuds	Type d'instance AWS	DCU par heure de disponibilité du moteur
XX petit (taille par défaut)	1	m5d.4xlarge	4
XSmall	1	m5d.8xlarge	8
Petit	2	m5d.8xlarge	16
Moyen	4	m5d.8xlarge	32
Grand	8	m5d.8xlarge	64
X Grand	16	m5d.8xlarge	128
2X Grand	32	m5d.8xlarge	256
3X Grand	64	m5d.8xlarge	512

The background is a solid orange color with a pattern of thin, white, wavy lines that create a sense of movement and depth. The lines are more concentrated on the left side and fade out towards the right.

6

CAS CLIENTS

Dremio permet aux analystes de TransUnion de réaliser des analyses en libre-service tout en accélérant les requêtes



La plateforme PRAMA de TransUnion permet aux clients d'explorer eux-mêmes une longue histoire de grands volumes de données, sans avoir besoin d'une intervention IT. L'objectif est de rendre les pétaoctets de données faciles à accéder et à utiliser avec une expérience rapide.

Problème

- Besoin de gérer 30 pétaoctets de données provenant de 90 000 sources de données et de plus de trois milliards de mises à jour par mois de leurs fournisseurs de données.
- Les clients s'attendent à une réponse rapide aux requêtes, et les performances médiocres de SQL sur Hadoop ont entraîné une expérience médiocre pour les analystes et les clients finaux.
- Les employés ont eu l'impression que les services informatiques ralentissaient les choses. Ils ont été contraints de consacrer des équipes agiles entières à la maintenance.

Solution

- Dremio est implémenté entre les outils BI et de data science et le data lake.
- Les réflexions de données et les ensembles de données virtuels offrent une accélération et une "passation de relais entre les services informatiques et les utilisateurs finaux".
- Il est facile de joindre les ensembles de données des clients à la solution et elle s'échelonne facilement avec leur cluster Hadoop.

Impact Business

- Dremio donne aux analystes la capacité **d'explorer les données en libre-service** sans avoir à attendre les ingénieurs de données.
- Dremio offre aux analystes et aux clients des **tableaux de bord interactifs personnalisés**.
- Dremio offre un gain de **performance immédiat de 5 à 10 fois**, avant la mise en place des réflexions
- Cela réduit les coûts généraux et augmente l'agilité du développement de produits, **permettant de redéployer 14 ingénieurs de données de la maintenance à la création de nouveaux produits**.

Démocratisation des données : Permettre du libre-service



Construire une plateforme d'analyse stratégique dans le cloud pour permettre le libre-service pour les utilisateurs techniques et non techniques tout en réduisant leur complexité architecturale et en permettant un temps d'analyse plus rapide.

Problème

- Il était **difficile pour les utilisateurs non techniques de consommer les données** d'ADLS, ce qui limite la valeur commerciale dérivée de ces données.
- Repose sur des cubes stockés en mémoire qui sont **statiques, coûteux et difficiles à gérer**.
- **Mauvaise performance sur des requêtes** plus importantes et plus complexes
- Difficile de rejoindre les données d'Oracle et d'ADLS.

Solution

- Être capable **d'interroger directement le lac de données**, le tout avec une performance élevée.
- Utiliser la **couche sémantique** conviviale de Dremio **pour démocratiser l'accès** aux utilisateurs techniques et non techniques.
- **Éviter de s'appuyer sur la copie** de données dans des cubes
- Être capable de **joindre et d'interroger** des données stockées dans Oracle et ADLS pour obtenir des informations intelligentes.

Impact Business

- Favoriser une culture de démocratisation des données chez Michelin - "**modèle de données pour tous**".
- Rendre l'ingénierie des données plus productive et efficace - réduire la complexité.
- Éliminer le backlog des demandes et libérer l'équipe IT



LE POINT

MISE EN PLACE D'UN DATAHUB COMME POINT D'ACCÈS UNIQUE



➤ LE CLIENT

- Média d'actualité depuis 1972
- +150 employés
- +60 millions € en 2017

➤ SES ENJEUX

- Fournir des analyses quotidiennes
- Avoir toutes les données à disposition pour fournir de futures analyses

LA PROBLÉMATIQUE

OLIVIER BOST souhaitait disposer d'une plateforme hyper unifiée afin de pouvoir l'alimenter, étape par étape, avec toutes les données non encore présentes dans le SI ou devant l'être par la suite.

➤ NOTRE INTERVENTION (3 ans)

- Mise en place du Cluster Hadoop
- PoV de 2 mois débouchant sur un démonstrateur, puis le GO du projet !
- Définition de la stratégie en termes d'architecture
- Alimentation des données sources :
 - Abonnement papier et numérique,
 - Logs de comportement de lecture,
 - Data Marketing
- Ingestion de ces données dans un datalake avec Airflow et Talend
- Approche hybride du stockage avec MariaDB ColumnStore
- Exploration et diffusion avec Tableau Software

➤ STACK TECHNOLOGIQUE

- AirFlow
- Talend
- Dremio
- Tableau Software
- Hadoop HDP
- MariaDB AX (ColumnStore)
- Google Cloud Platform



➤ LE CLIENT

- 105 000 Km de Lignes Électriques
- 9 586 collaborateurs
- 4,9 milliards CA (2022)

➤ SES ENJEUX

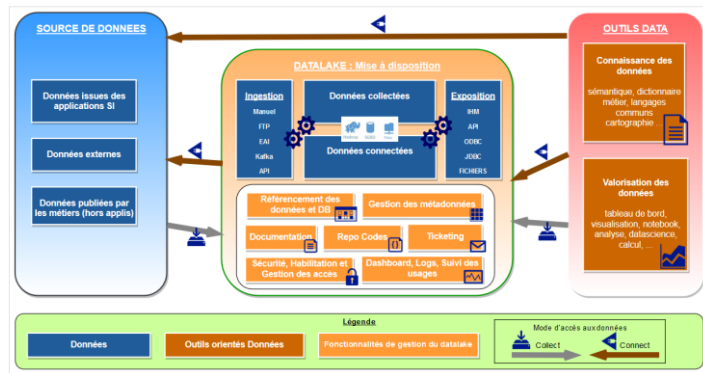
- Gouvernance des données
- Catalogage les données
 - sources de données
 - jeux de données
- Suivi
 - de toutes les données
 - de tous les traitements

LA PROBLÉMATIQUE

Plus que la performance, la problématique est la capacité à sécuriser, à l'utilisateur, l'accès à toutes sources de données, et la capacité de suivre chaque traitement de chaque utilisateur qui emploie divers jeux de données issus de diverses sources.

➤ NOTRE INTERVENTION (depuis 4 ans)

- Expertise autour de la plateforme Dremio
- Mise en oeuvre méthodologie de gouvernance technique basée sur les méthodologies agiles
- Autonomie et responsabilité des différentes équipes (métier) quand l'infrastructure des données est gérée par la DSI



➤ STACK TECHNOLOGIQUE

- Cloudera CDH
- **Dremio**
- RS Studio



Notre téléphone

01 55 28 89 40



Nos bureaux

24 Rue de l'Église, 94300 Vincennes



Notre site web

www.synaltic.fr



Marc SALLIERES

Synaltic
Président
msallieres@synaltic.fr

Charly CLAIRMONT

Synlatic
Directeur Technique
cclairmont@synaltic.fr

Guillermo KIRKPATRICK

Dremio
guillermo.kirkpatrick@dremio.com

