Festival Online de la Data

PROCHAINES DATES

Jeudi 14 mai de 15h à 15h30
La Data Federation Open Source avec Dremio
Charly Clairmont, CTO #Synalteam

Jeudi 28 mai de 15h à 15h30
Sensibilisation à la DataViz
Franck Nguyen, consultant #Synalteam

Jeudi 4 juin de 15h à 15h30
Debezium, CDC Open Source en temps réel
Yabir Canario de la Mota, consultant #Synalteam

Jeudi 11 Juin de 15h à 15h30
Knowage, toute la BI en Open Source
Stefano Scamuzzo, Engineering
Charly Clairmont, Synaltic

Jeudi 18 Juin de 15h à 15h30
Préparation de Données rapide et agile dans les Clouds AWS, Azure et Google
Victor Coustonoble, Trifacta
Franck Nguyen, Synaltic

Jeudi 25 juin de 15h à 15h30
Comment mobiliser business et IT autour de leurs données les plus précieuses
Jean-Michel Franco, Talend
Romain Legrand, Synaltic

SYNALTIC EN QUELQUES MOTS

Acteur innovant, dénicheur de solutions et une équipe de collaborateurs engagés

Synaltic

30

30 collaborateurs formés et certifiés contribuant aux communautés d'utilisateurs.

15

Spécialiste en Data Management depuis plus de 15 ans et plus de **250 projets** réalisés.

160

Plus de **160 clients** dont certains s'appuient sur nous avec succès depuis plus de **8 ans**.

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

Dremio

Accéder à l’ensemble des données de votre Système d’Information

dremio

Synaltic group

Charly CLAIRMONT

CTO, Synaltic

@egwada

#OpenSource #BI #Hadoop #DataLake #Analytics #Innovation

1999 : développeur web, site d'informations sportives à large audience

2004 : création de Altic qui deviendra Synaltic

2011 : enseignant universitaire (Paris 6, Paris 13, UTT)

2012 : création du Hadoop User Group / Paris Spark Meetup

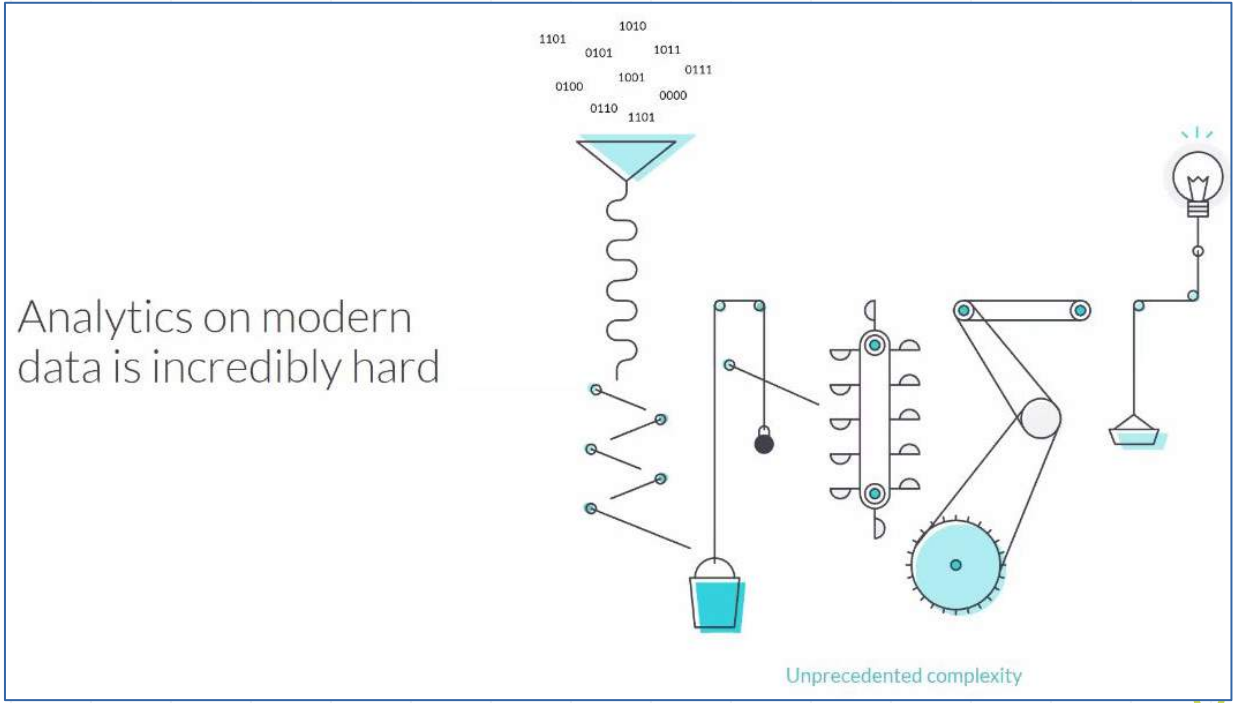
2019 : Synaltic une trentaine d'employés
Hug France : +1500 membres
Paris Spark Meetup : +3000 membres

L’analytique demeure trop complexe

Voilà des années que les plateformes telles que Facebook, Google existent…

Les utilisateurs s’y sont énormément habitués !

Les plateformes analytiques sont loin d’ être aussi simples.



dremio

Synaltic group

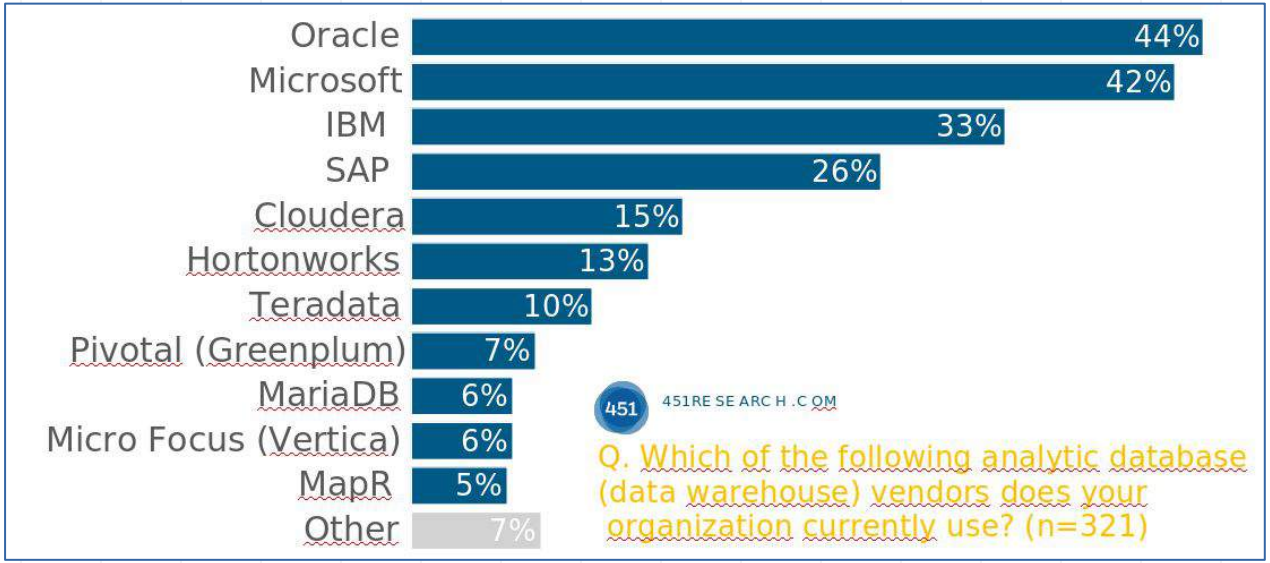
Les utilisateurs veulent un accès plus simple et direct à la donnée !

Le mot d'ordre : s'approprier les données et les analyser soi-même.

C'est l'âge d'or du

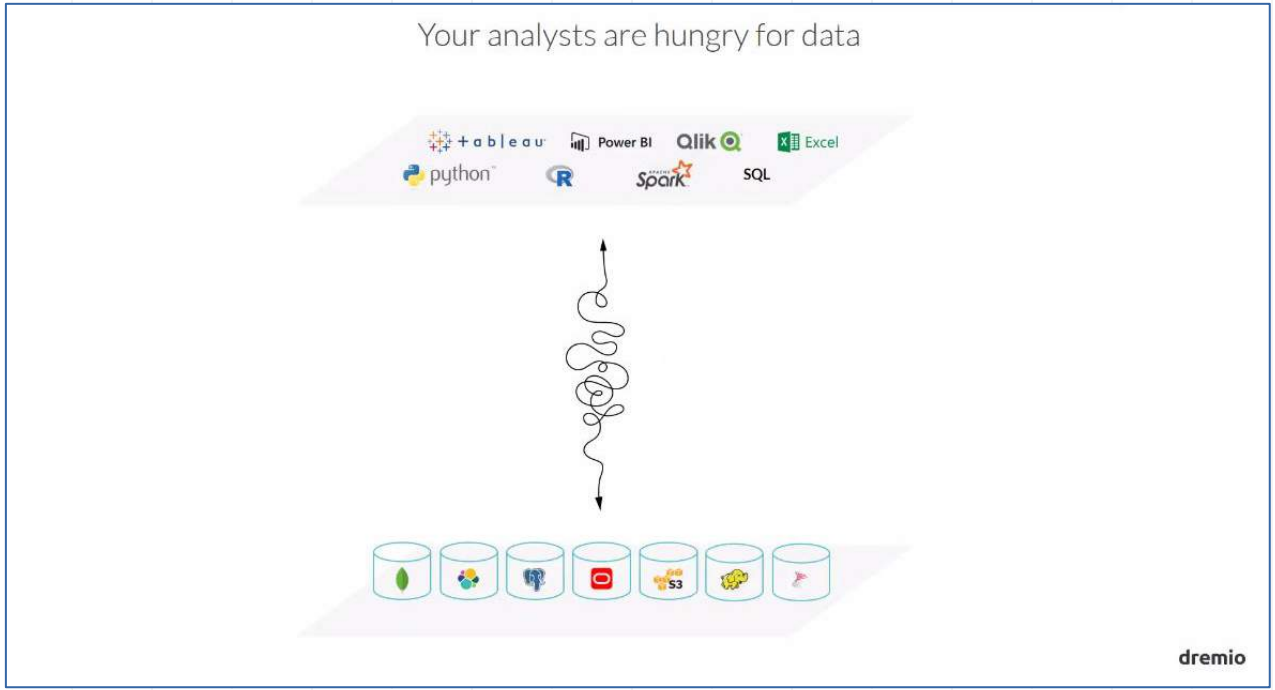
Self Service

Et les organisations n'ont définitivement pas qu'une source de données !



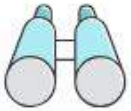
The Next Wave: Data, AI and Analytics in 2019 and Beyond
Matt Aslett, Research Vice President

Peu de solutions offrent un tel accès direct et en self-service à la donnée

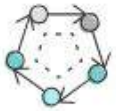


Une approche unique des projets de données, trop éloignée des besoins et de la temporalité des métiers

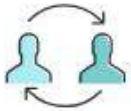
COMMENT FÉDÉRER LA DONNÉE ? COMMENT LA RENDRE AUX MÉTIERS PLUS FACILEMENT ET PLUS RAPIDEMENT ?



Search and discover data from a central data catalog.



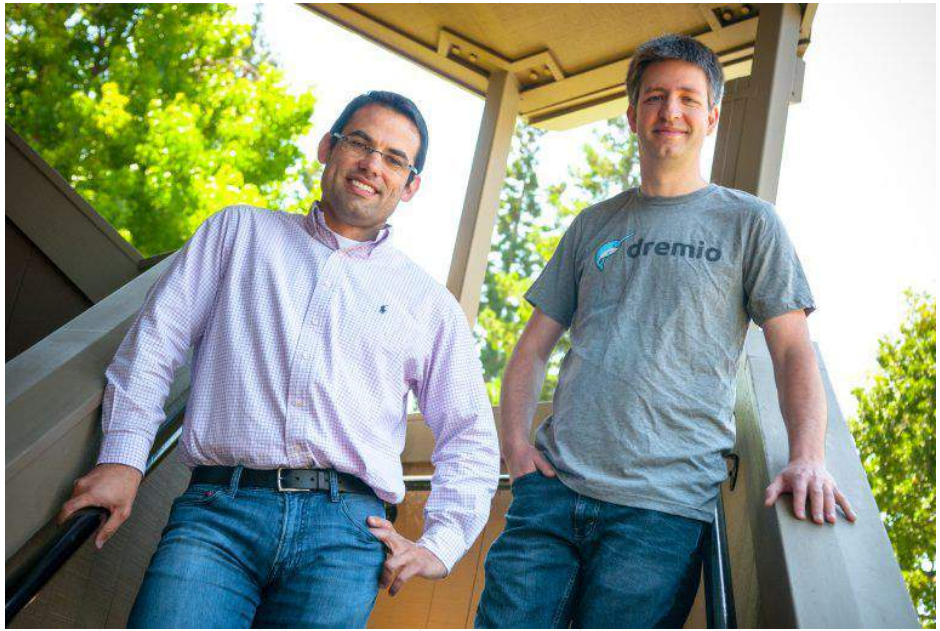
Filter, transform, and join any source to curate data for your needs.



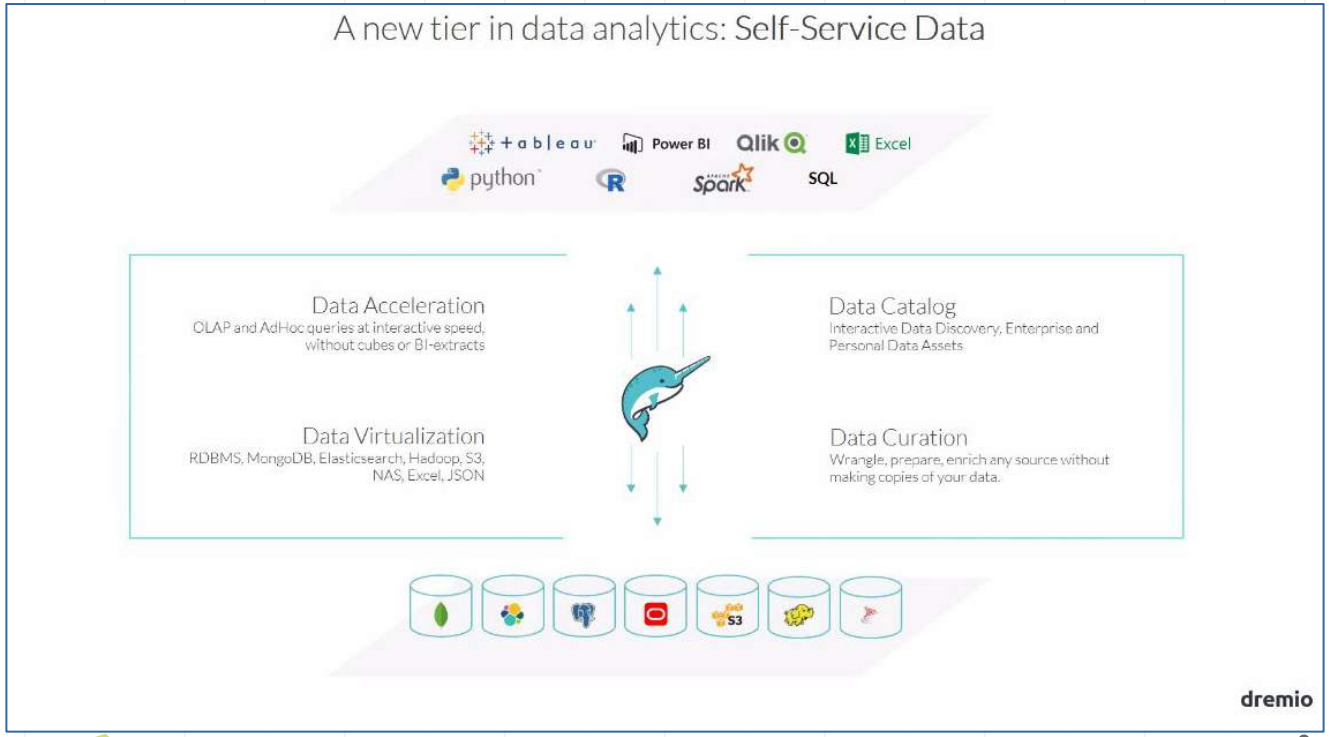
Share your work with your team to build together.

DREMIO, SON ÉQUIPE

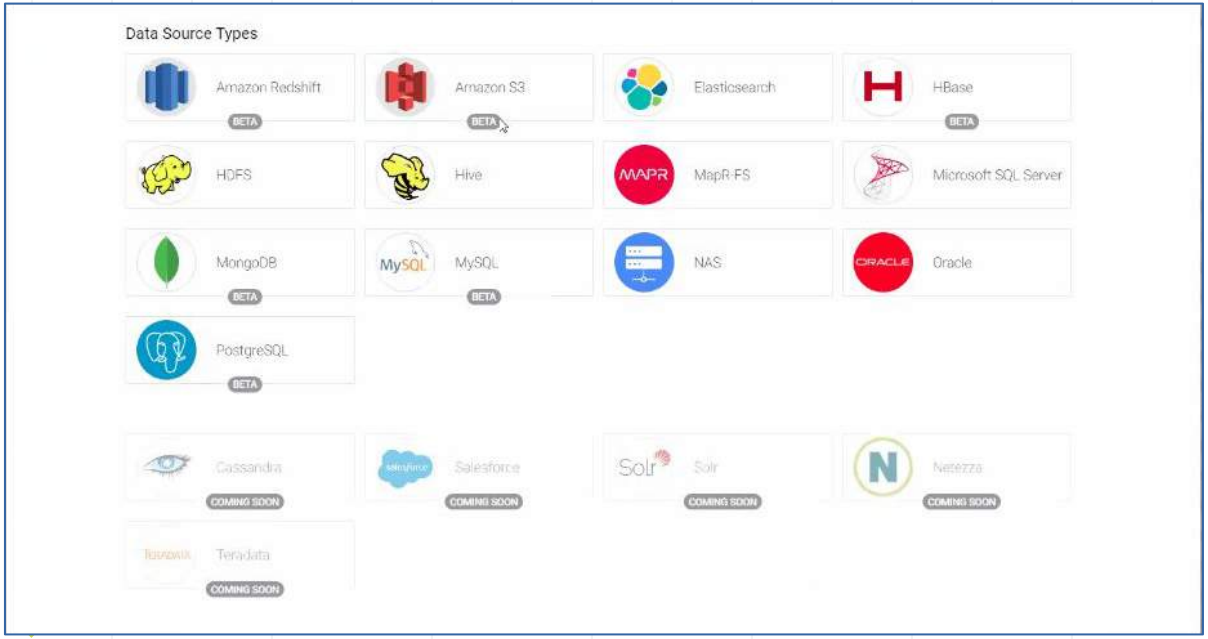
- Fondée en 2015
- Apache Parquet, Apache Calcite et Apache Arrow
- Version 1.0 en 2017
- Actuellement en version 4.2



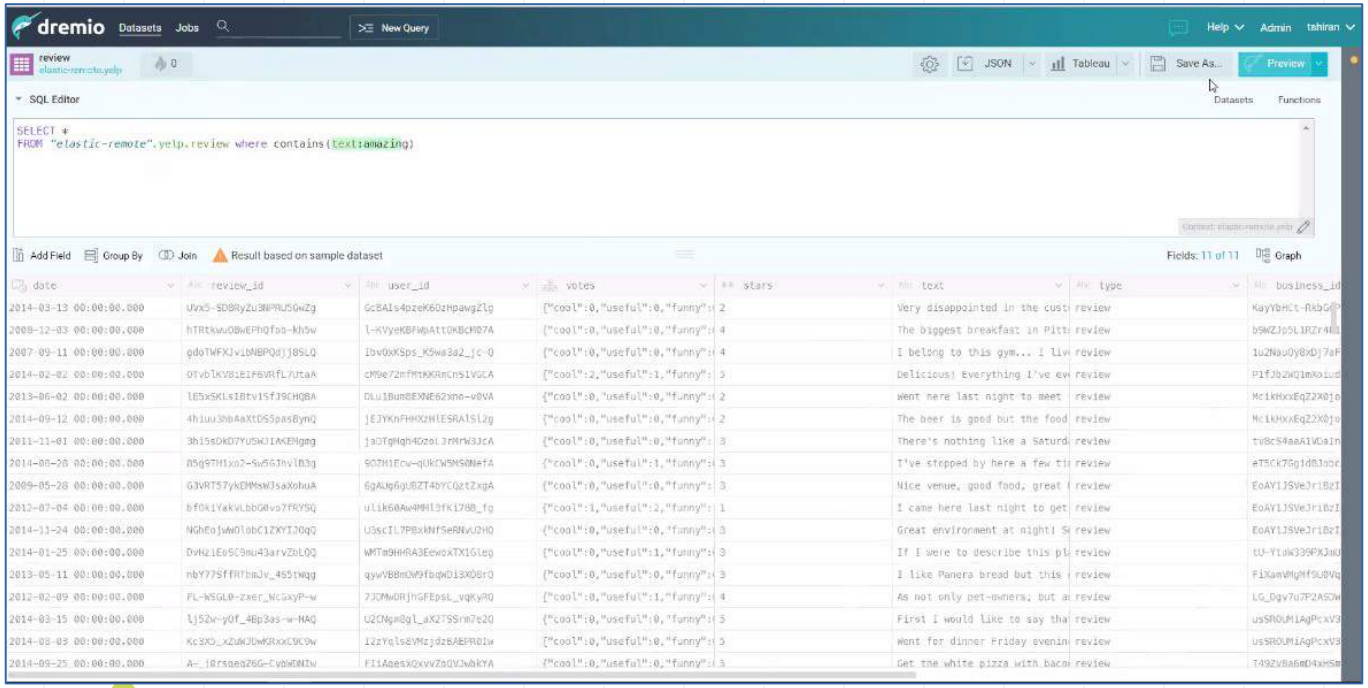
Dremio change la donne : une approche plus directe !



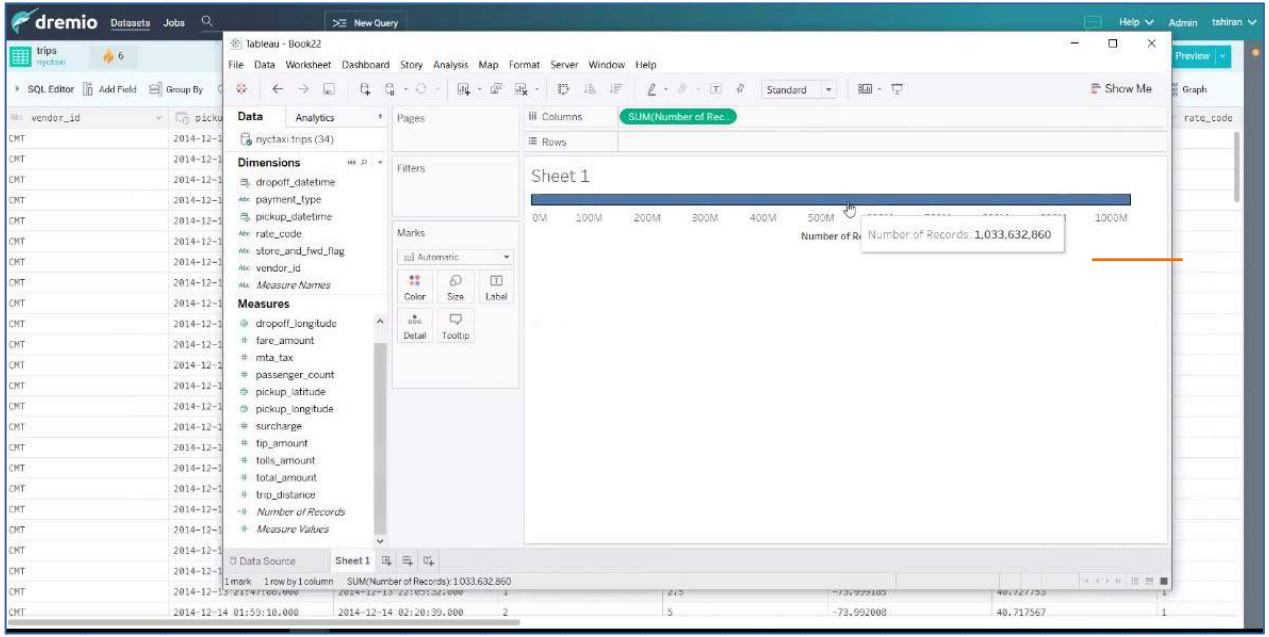
Dremio, parce que vous accédez à toutes vos sources…



Interroger Dremio (via SQL ou API REST)

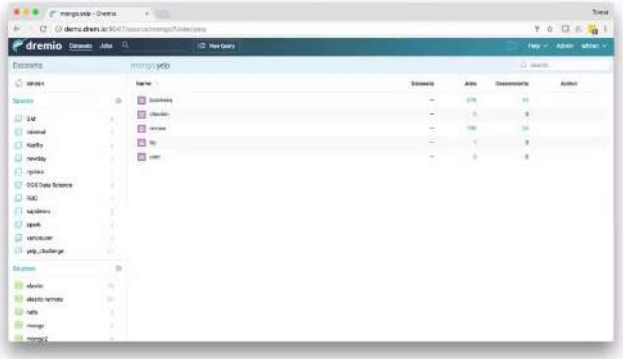


Dremio + Tableau (ou n’importe quelle solution #BI)



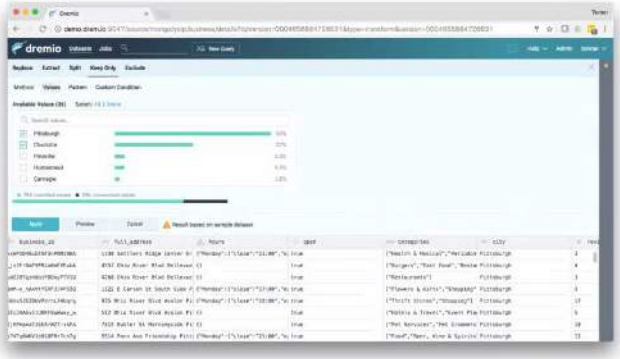
Une approche collaborative de la gestion des données (1/2)

« catalogue de données »



"Google Docs for Data"

Find, manage and share data regardless of size & location



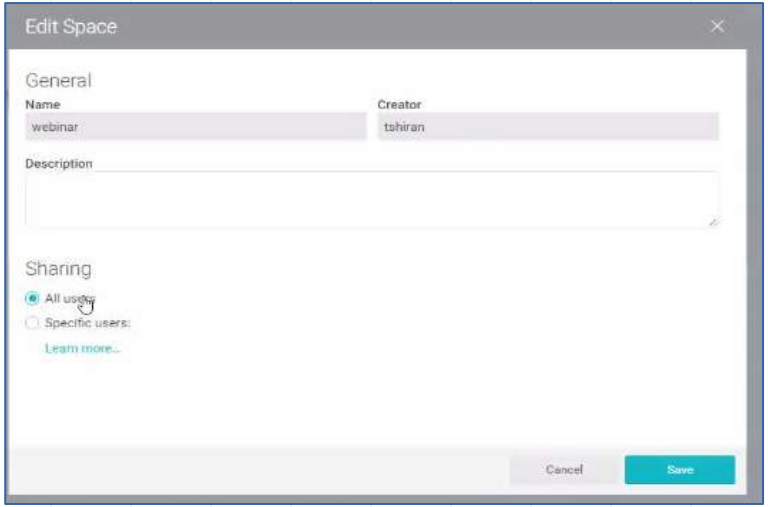
Live Data Curation

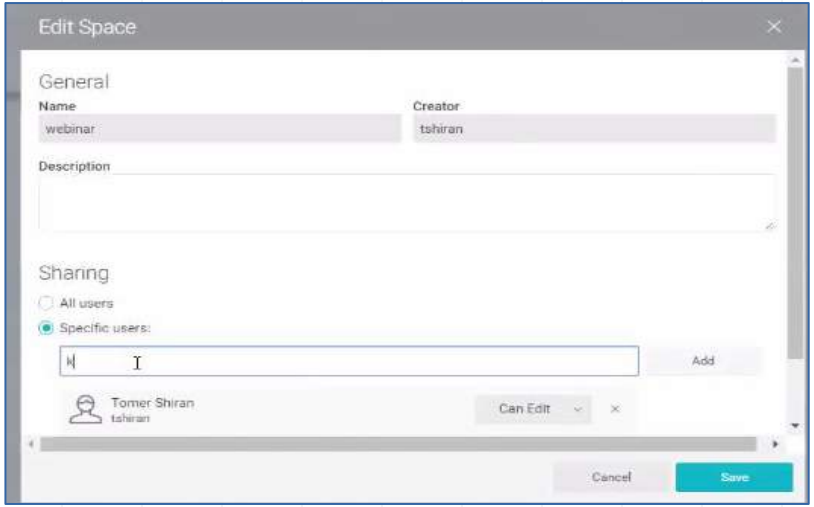
AI-powered curation of data without creating a single copy

dremio

Une approche collaborative de la gestion des données (2/2)

« partage des espaces de travail »





Dremio travaille à toutes les échelles :

d'une seule machine à un cluster multi-nœuds

On Premise ou Cloud

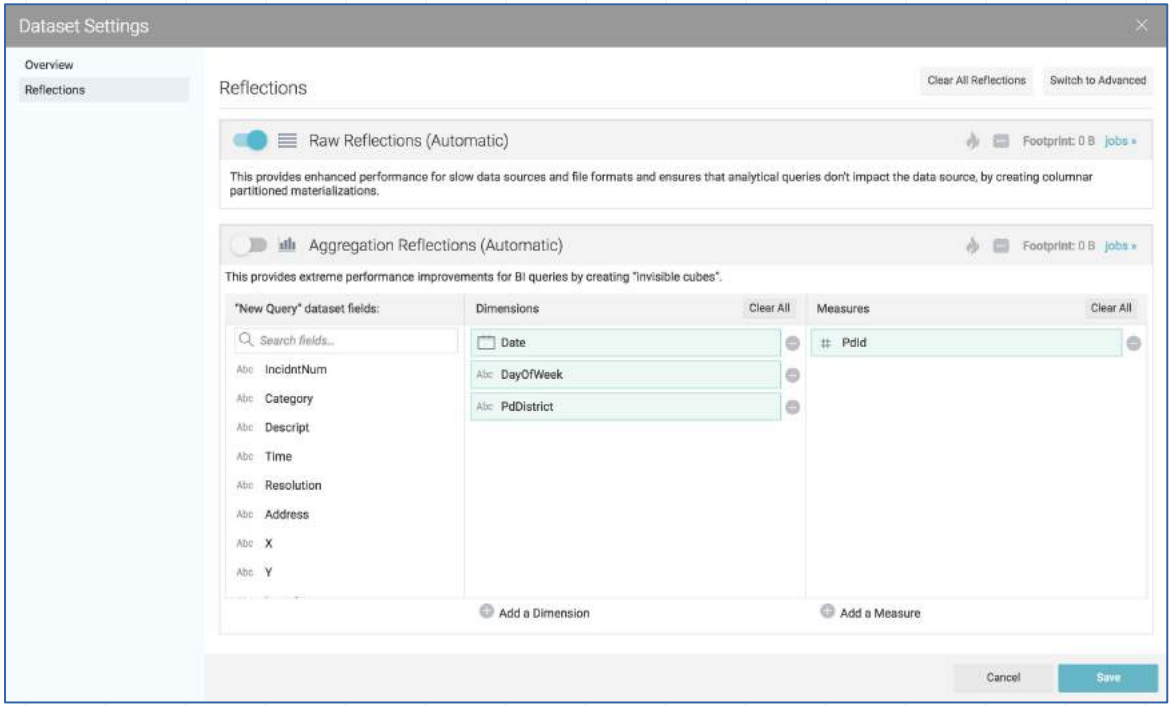
Deploy On Yarn

Deploy via Kubernets

Deploy via Docker

Dremio, accélère l'accès aux données

mise en cache des données



Dremio, trace l'accès aux données et détaille les calculs

dremio Datasets Jobs New Query Help Admin tshiran

amazingreviews webinar 0 JSON Tableau Save Preview

SQL Editor Add Field Group By Join Result based on sample dataset Fields: 27 of 27 Graph

Focus "amazingreviews"

Sources (2)

Parents (2)

review
elastic-remote.yelp
Jobs: 251
Descendants: 23
Fields: 11

business
mongo.yelp
Jobs: 969
Descendants: 56
Fields: 16

amazingreviews
webinar
Jobs: 0
Descendants: 0
Fields: 27

date
review_id
user_id
votes
stars
text
type
_index
_type
_uid
business_id
business_id0
_id
full_address
hours

10:42 AM
10/13/2017

Auditer tout ce qu'il s'est passé sur Dremio

« traçabilité des traitements »

dremio Datasets Jobs New Query Help Admin tshiran

Jobs

Start Time : All Time | Status: All | UI, External Tools | User: All | Contains text... | Order by Start Time

Dataset	User	Start Time	Duration	End Time
trips nyctaxi	tshiran	10/13/2017 10:49:16	<1s	10/13/2017 10:49:17
trips nyctaxi	tshiran	10/13/2017 10:49:16	<1s	10/13/2017 10:49:16
trips nyctaxi	tshiran	10/13/2017 10:48:44	<1s	10/13/2017 10:48:44
trips nyctaxi	tshiran	10/13/2017 10:48:44	<1s	10/13/2017 10:48:44
trips nyctaxi	tshiran	10/13/2017 10:48:14	<1s	10/13/2017 10:48:14
trips nyctaxi	tshiran	10/13/2017 10:48:14	<1s	10/13/2017 10:48:14
trips nyctaxi	tshiran	10/13/2017 10:48:08	<1s	10/13/2017 10:48:08
trips nyctaxi	tshiran	10/13/2017 10:48:08	<1s	10/13/2017 10:48:08

Completed

Overview | Details | Profile

Summary

Query Type: ODBC Client (execute prepared statement)
Duration: <1s
Start Time: 10/13/2017 10:49:16
End Time: 10/13/2017 10:49:17
User: tshiran
Job ID: 261f0362-cfde-4bce-a2dc-46dd5289fa00

Query

Parents: trips nyctaxi 6

Input		Output	
Input Bytes:	1.50 MB	Output Bytes:	0 B
Input Records:	48,316	Output Records:	0

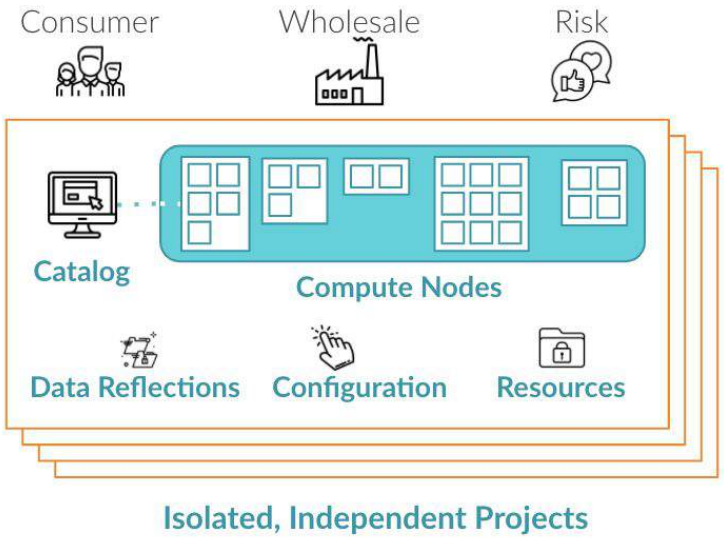
SQL:

```
SELECT AVG( "trips"."tip_amount") AS "avg_tip_amount_ok",
  SUM( 1) AS "sum_Number_of_Records_ok",
  {fn TIMESTAMPADD( SQL_TSI_MONTH,CAST( {fn TRUNCATE( ( 3 * ( CAST( {fn TRUNCATE( {fn QUARTER(
"trips"."dropoff_datetime"},0)} AS INTEGER) - 1)),0)} AS INTEGER),{fn TIMESTAMPADD( SQL_TSI_DAY,CAST( {fn
TRUNCATE( ( -1 * ( {fn DAYOFYEAR( "trips"."dropoff_datetime"}) - 1)),0)} AS INTEGER),CAST(
"trips"."dropoff_datetime" AS DATE)})})} AS "tqr_dropoff_datetime_ok"
FROM "nyctaxi"."trips" "trips"
GROUP BY {fn TIMESTAMPADD( SQL_TSI_MONTH,CAST( {fn TRUNCATE( ( 3 * ( CAST( {fn TRUNCATE( {fn QUARTER(
"trips"."dropoff_datetime"},0)} AS INTEGER) - 1)),0)} AS INTEGER),{fn TIMESTAMPADD( SQL_TSI_DAY,CAST( {fn
TRUNCATE( ( -1 * ( {fn DAYOFYEAR( "trips"."dropoff_datetime")] - 1)),0)} AS INTEGER),CAST(
"trips"."dropoff_datetime" AS DATE)})})}
```

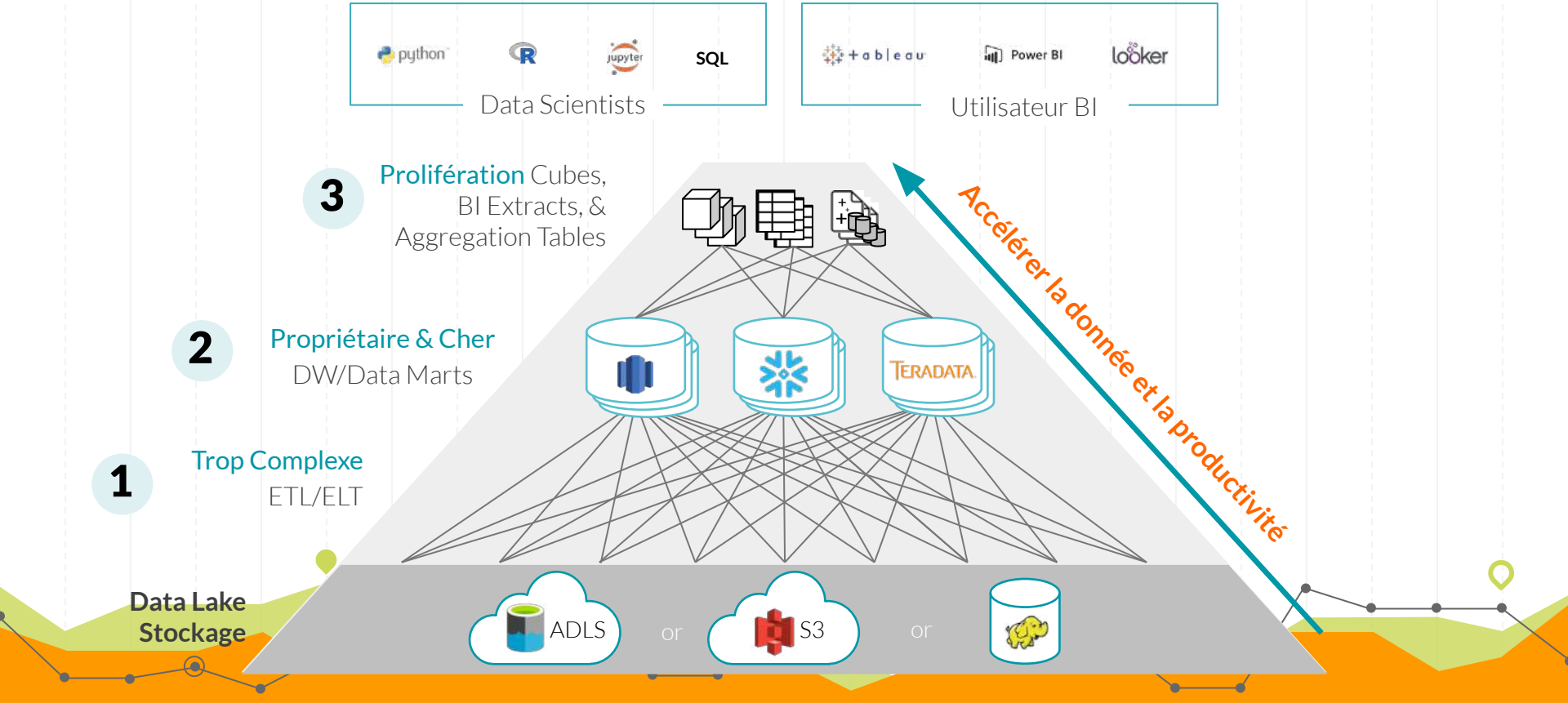
demo.dremio:9047/jobs

Les cas d’usage

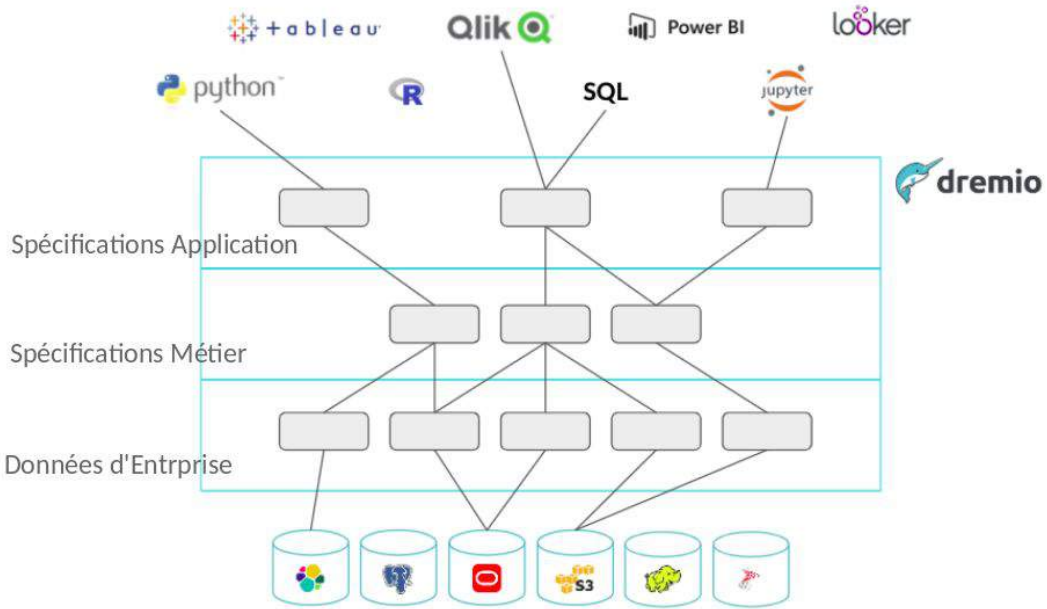
Infrastructure “Dynamique” et moderne de données



Accélérer votre Business Intelligence

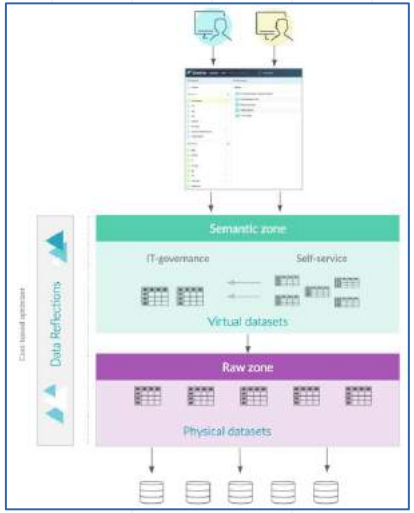


Catalogage & Gouvernance de données (1)



Catalogage & Gouvernance de données (2)

catalogue de données



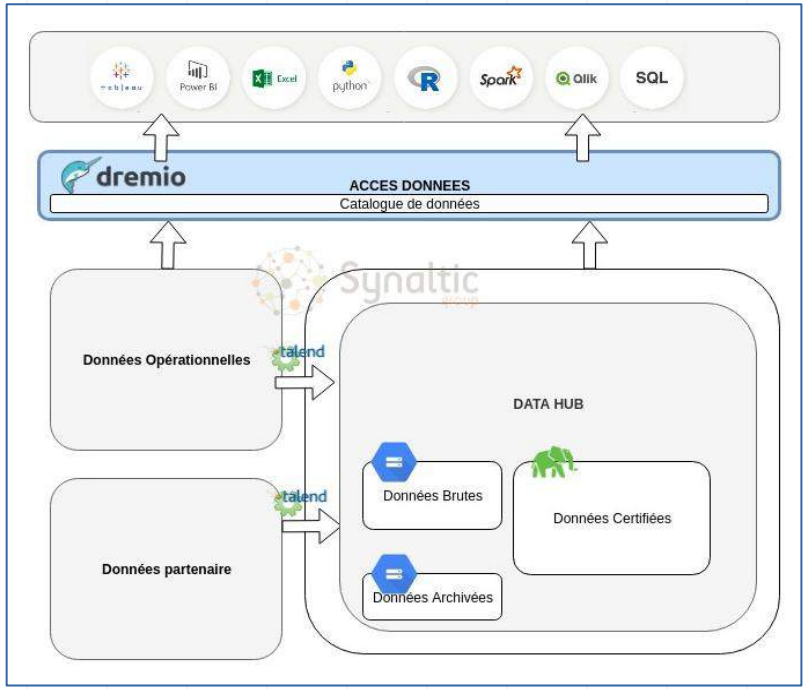
catalogue de données

Retours d’Expérience Clients

Un accès unique à la donnée

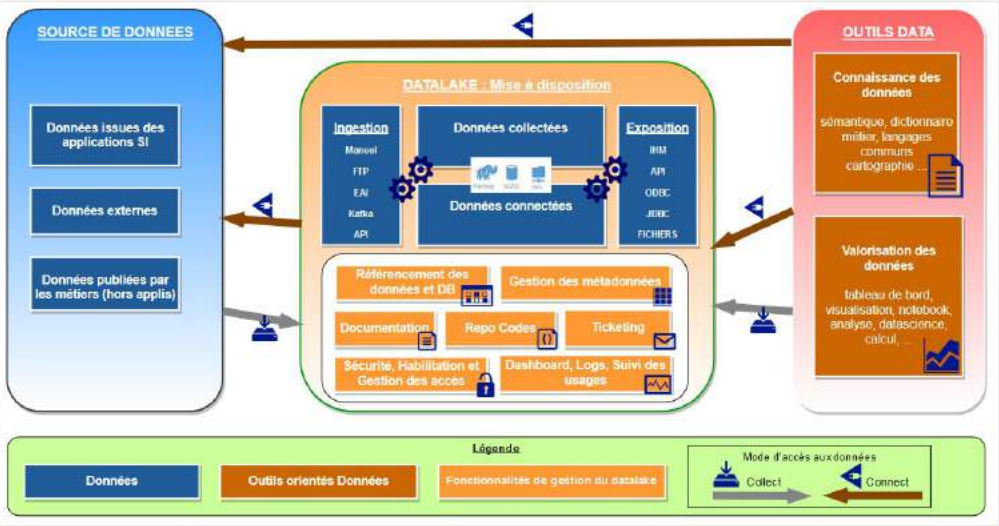
Le Point

Le service IT du journal Le Point a besoin de créer une **plate-forme de centralisation de ses données** qui permet d'avoir une vision 360 des abonnés. Afin de donner un accès *rapide* aux data analystes et data scientistes, mais aussi à *n'importe quelle personne qui a le droit et le besoin* de monitorer la performance de son activité.



Audit et Suivi de tout traitement

Une des promesses de Dremio est la performance, notre vrai intérêt c'est la **capacité à sécuriser, à l'utilisateur, l'accès à toutes nos sources de données**, et suivre chaque traitement de chaque utilisateur.



DEMO

Merci !

Questions?

@Synaltic / @egwada

https://www.synaltic.fr

Festival Online de la Data

PROCHAINES DATES

Jeudi 14 mai de 15h à 15h30
La Data Federation Open Source avec Dremio
Charly Clairmont, consultant #Synalteam

Jeudi 28 mai de 15h à 15h30
Sensibilisation à la DataViz
Franck Nguyen, consultant #Synalteam

Jeudi 4 juin de 15h à 15h30
Debezium, CDC Open Source en temps réel
Yabir Canario de la Mota, consultant #Synalteam

Jeudi 11 Juin de 15h à 15h30
Knowage, toute la BI en Open Source
Stefano Scamuzzo, Engineering
Charly Clairmont, Synaltic

Jeudi 18 Juin de 15h à 15h30
Préparation de Données rapide et agile dans les Clouds AWS, Azure et Google
Victor Coustonoble, Trifacta
Franck Nguyen, Synaltic

Jeudi 25 juin de 15h à 15h30
Comment mobiliser business et IT autour de leurs données les plus précieuses
Jean-Michel Franco, Talend
Romain Legrand, Synaltic

Produits

	Community Edition	Enterprise Edition
Powerful & Intuitive UX	✓	✓
Scales From 1 to 1000+ Nodes	✓	✓
Data Connectors for RDBMS, Elasticsearch, MongoDB, HBase, HDFS, Amazon S3, MapR-FS, ADLS	✓	✓
Dremio Reflections™	✓	✓
High Availability & Disaster Recovery	✓	✓
Enterprise Data Connectors for DB2, Teradata		✓
Role-based Authorization; LDAP/AD Authentication; End-to-end encryption over TLS; Secure Impersonation; Apache Ranger integration		✓

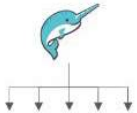
	Community Edition	Enterprise Edition
Multi-tenant Workload Controls		✓
Crowdsourced Reflection Prioritization		✓
Advanced Reflection Substitution Logic		✓
Reflection Management Console		✓
Support	Community	Enterprise
License	Apache 2.0	Commercial

Dremio, accélère l'accès aux données (1/2)

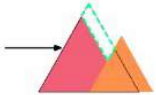
Dremio optimizes your data and your queries automatically for 10x-1000x acceleration

Apache Arrow Execution
From 1 to 1000+ nodes, run on dedicated infrastructure or in your Hadoop cluster, via YARN.

Dremio Reflections™
Optimized physical data structures for row and aggregation operations.



Native Push-Downs
Optimized query semantics for each data source: relational, NoSQL HDFS, Amazon S3, and more.



Universal Relational Algebra
Query Planner automatically substitutes plans to make optimal use of cache fragments.

dremio

Dremio, accélère l'accès aux données (3/3)

mise en cache des agrégations selon axes et mesures

Dataset Settings

Overview
Reflections

Reflections

Clear All Reflections

Raw Reflections | Aggregation Reflections

Raw Reflections

New Reflection

Unnamed Reflection

Footprint: 4.33 MB (8.66 MB) jobs »

Search fields...

Fields	Display	Sort	Partition	Distribution
Abc IncidntNum	✓	–	–	–
Abc Category	✓	–	–	–
Abc Descript	✓	–	–	–
Abc DayOfWeek	✓	–	–	–
Date	✓	–	–	–
Abc Time	✓	–	–	–
Abc PdDistrict	✓	–	–	–
Abc Resolution	✓	–	–	–
Abc Address	✓	–	–	–
Abc X	✓	–	–	–
Abc Y	✓	–	–	–

Cancel | Save