

Festival Online de la Data 2021

DERNIER WEBINAIRE DE LA SAISON...



SYNALTIC EN QUELQUES MOTS

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



30

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

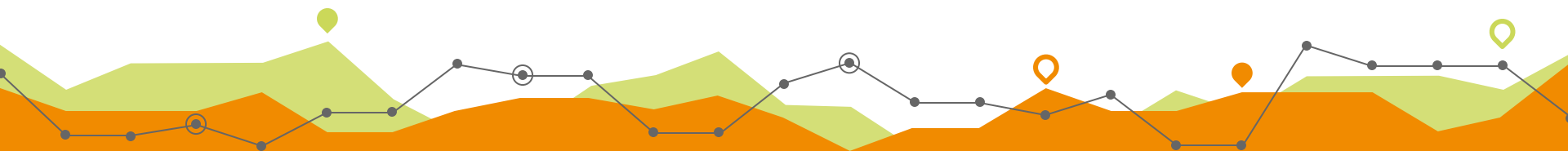
15

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 !



La Data Gouvernance par la pratique

UN WEBINAIRE ANIMÉ PAR CHARLY CLAIRMONT

Ecoutez un des co-fondateur de Synaltic



Charly CLAIRMONT

Co-Fondateur, Synaltic

@egwada

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

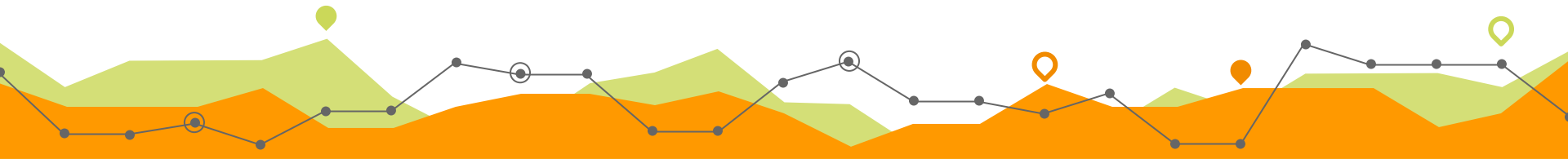
1999 : développeur web, site d'informations sportives à large audience

2004 : création de Altic qui deviendra Synaltic

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

2012 : création du Hadoop User Group / Paris Spark Meetup

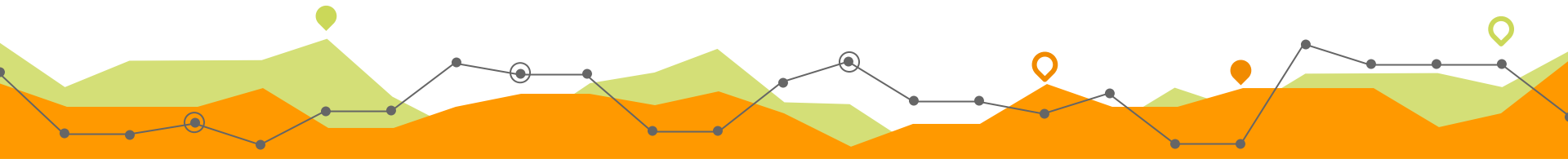
2019 : Synaltic une trentaine d'employés
Hug France : +1500 membres
Paris Spark Meetup : +3000 membres



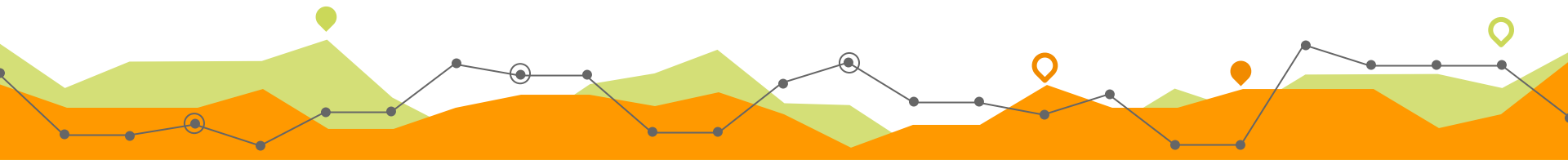
“La donnée” est partout !

On a déjà dit ça !

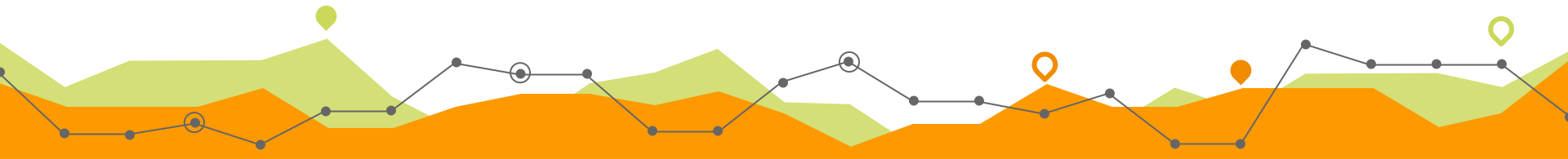
Webinaire précédent : Le Data Littératie, l'atout clé des entreprises du 21ème siècle



Toujours plus de données à gérer !



D'où l'idée d'aller au delà de sa
gestion mais de la Gouverner.



L'objectif de la gouvernance des données est d'assurer la **qualité**, la **disponibilité**, l'**intégrité**, la **sécurité**, la convivialité (facilité d'accès, facilité de l'appropriation) et la collaboration autour des données au sein d'une organisation.

A CHAQUE ORGANISATION SA CULTURE

adaptée la gouvernance des données à sa propre organisation

Chaque entreprise possède sa propre culture ce qui a pour conséquence des implémentations propres et uniques.

Toutes les expériences montrent qu'une gouvernance hyper-centralisée et bureaucratique n'engage pas les collaborateurs et finit par constituer un frein à la prise d'initiative.

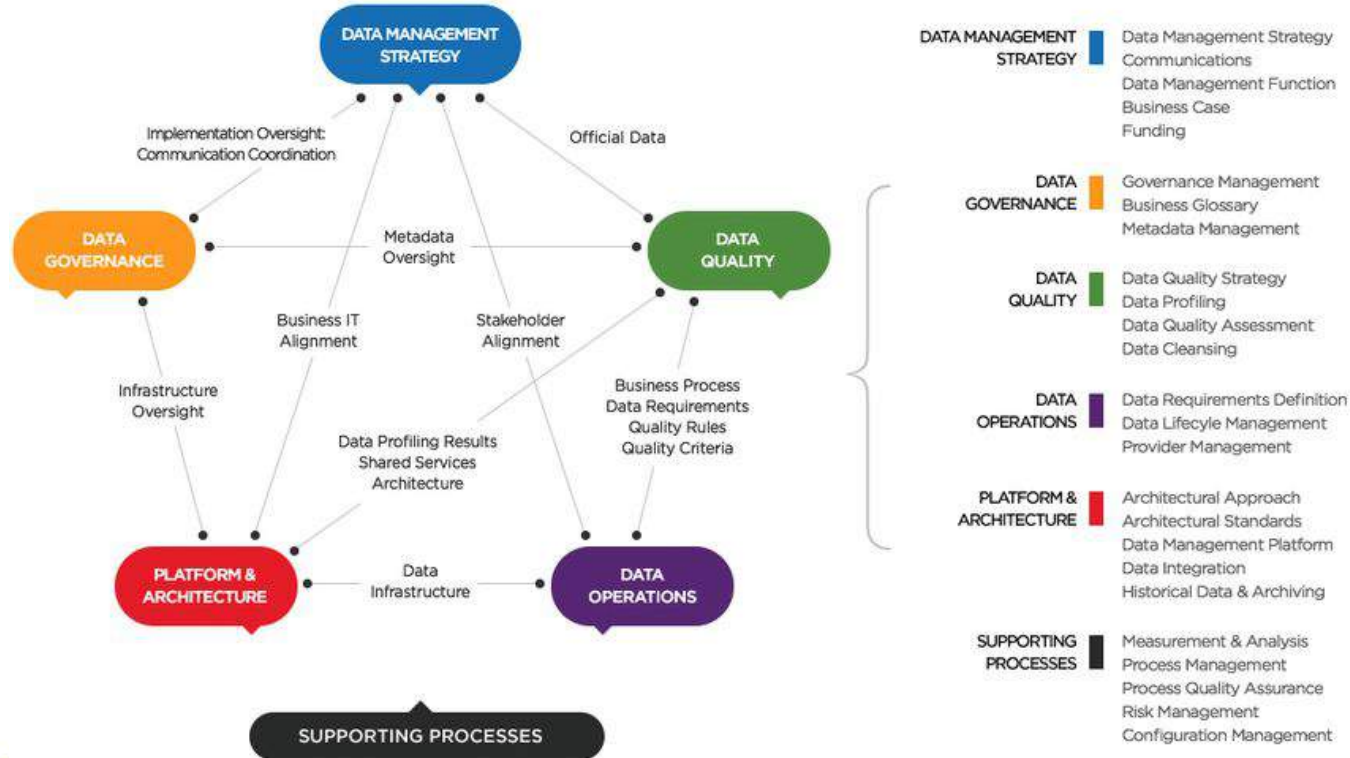
Il faut se sentir à l'aise dans cette gouvernance !

Chez Synaltic nous privilégions une approche de la gouvernance des données **pragmatique, simple, agile**, et **décentralisée** !

La "**gouvernance agile**", est avant tout au service des utilisateurs ; elle se concentre sur leur motivation à s'appliquer à bien faire, à suivre ce qu'ils font et d'employer leurs efforts là où c'est d'abord nécessaire.

GOUVERNANCE DES DONNÉES, DE NOMBREUSES DIMENSIONS

Un cercle vertueux de construction de la connaissance



La gouvernance des données est aux données ce que les ressources humaines sont au personnel.

CADRE GOUVERNANCE DE DONNÉES

La gouvernance des données est au cœur de toutes les activités de gestion des données.

La Gouvernance de donnée c'est tout le cycle de vie de la donnée et ce sur le long terme.



SUCCÈS D'UNE GOUVERNANCE AGILE DES DONNÉES

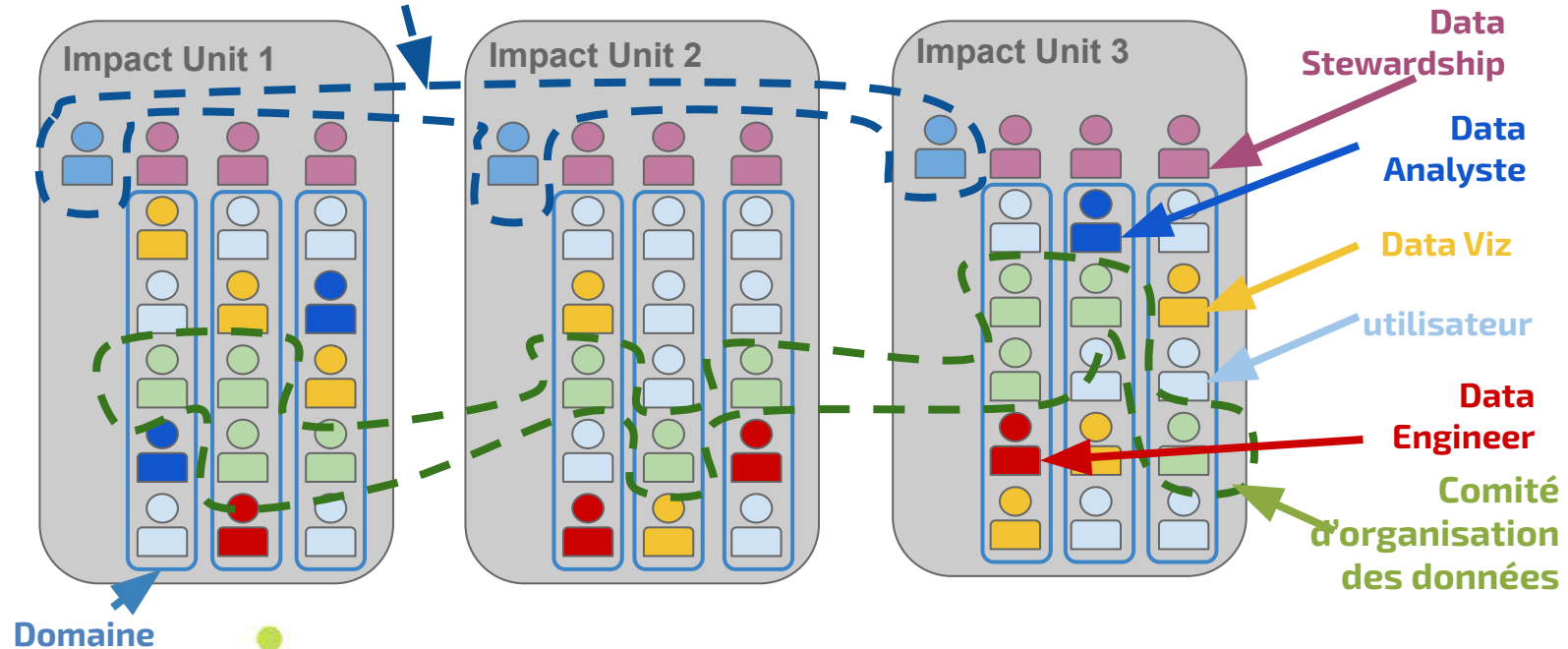
Les clés du succès pour une gouvernance agile des données

- ❑ La gouvernance des données s'inscrit dans la stratégie globale
- ❑ Effort de gouvernance maîtrisée et continue
- ❑ Avoir des objectifs clairs et quantifiables
- ❑ Mesurer et observer honnêtement les résultats
- ❑ Laisser la place à l'initiative
- ❑ **Eduquer ! Eduquer ! Eduquer !**

GOUVERNANCE AGILE, ORGANISATION AGILE

simplicité organisationnelle

Conseil de Gouvernance des Données



GOUVERNANCE AGILE, CADRE ORGANISATIONNEL (1/2)

simplicité organisationnelle, décentralisation des initiatives

Conseil de Gouvernance des Données (CGD)



Établir et former des membres de la CGD
Convenir d'une feuille de route pour la gouvernance des données
Déterminer les membres et habiller la COD
Définir la politique de gouvernance des données
Rencontrer sur une base récurrente

Comité d'organisation des données (COD)



Recevoir une formation et une éducation
Exécuter la feuille de route de la CGD
Définir le plan de communication de la CGD
Sélectionner et responsabiliser les équipes d'intendance
Définir les rôles et responsabilités
Définir la charte
Rencontrer / Communiquer fréquemment

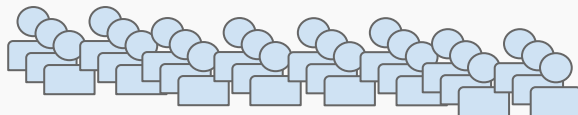
Data Stewardship

Data Viz, Data Analyste, Data Engineer



Membres métiers, techniques,
Recevoir une formation et une éducation
Exécuter la feuille de route de la CGD

utilisateur (Consom'Acteur)



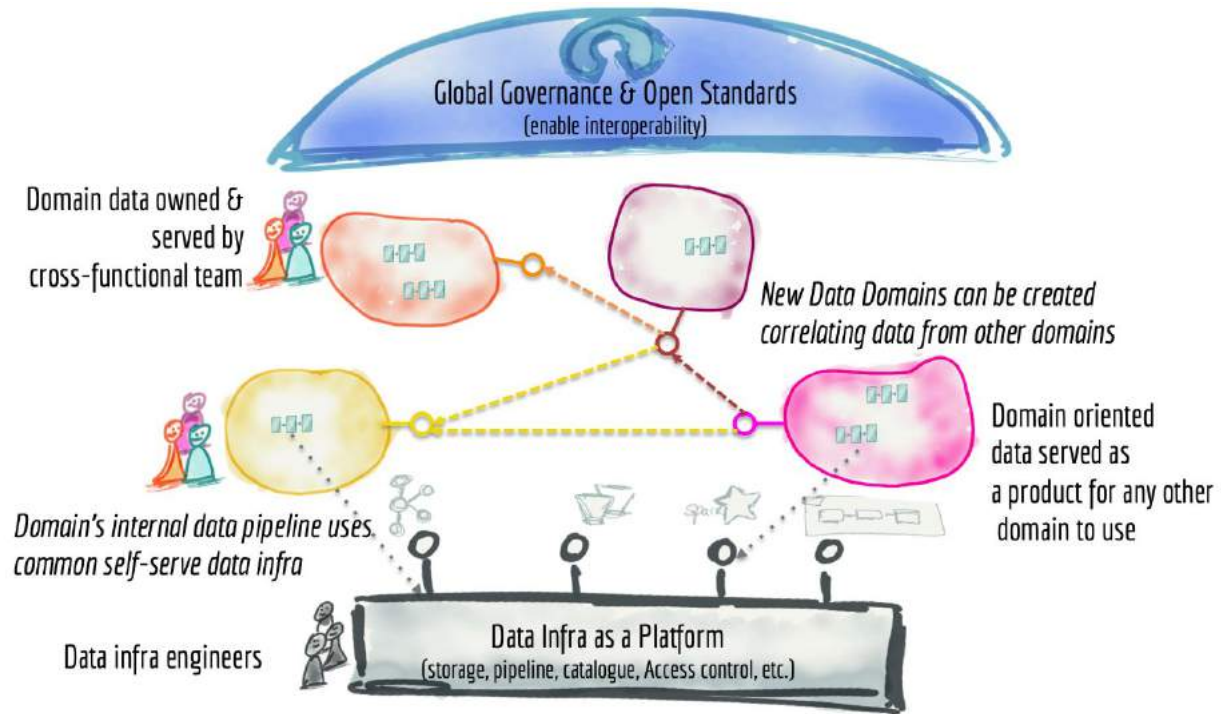
Membres professionnels
Recevoir une formation et éducation

GOUVERNANCE AGILE, CADRE ORGANISATIONNEL (2/2)

simplicité organisationnelle, décentralisation des initiatives

Favoriser une
approche libre,
agile,
indépendante des
projets

Data Mesh



GOUVERNANCE DES DONNÉES, ACCÉDER AUX DONNÉES

La gouvernance des données doit rester pratique !

Rechercher 	Tracer les changements 	Construire un réseau agile 
Où est la table, le tableau de bord ? Quel en est le contenu ?	J'ai changé un schéma d'une table qui est le responsable et quels sont les utilisateurs ?	Je voudrais suivre un utilisateur clé de mon équipe
Est-ce que cette analyse existe déjà ?	La livraison du jeu de données est en retard. Je voudrais notifier tout le monde	Je voudrais mettre en favoris une table et être notifié des modifications de schémas, Des nouvelles données, des incidents.

Les analystes et data scientists passent $\frac{1}{3}$ de leur temps à découvrir les données.

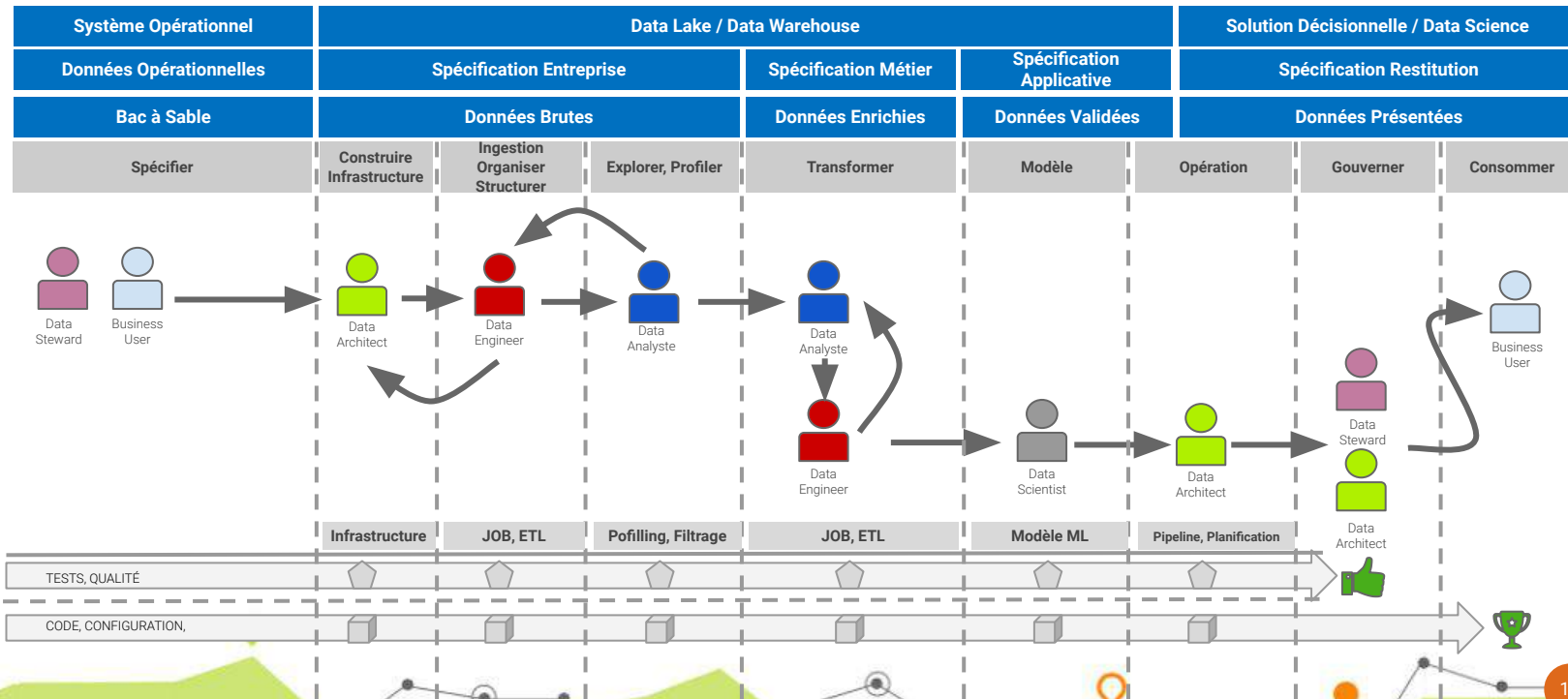
Eviter le "select *" : trop de ressources utilisées, trop de colonnes à comprendre

Efficacité à trouver les jeux de données utiles aux analyses.
⇒ Augmenter la productivité

GOUVERNER LES DONNÉES PAR LA PRATIQUE : DATA OPS (2/2)

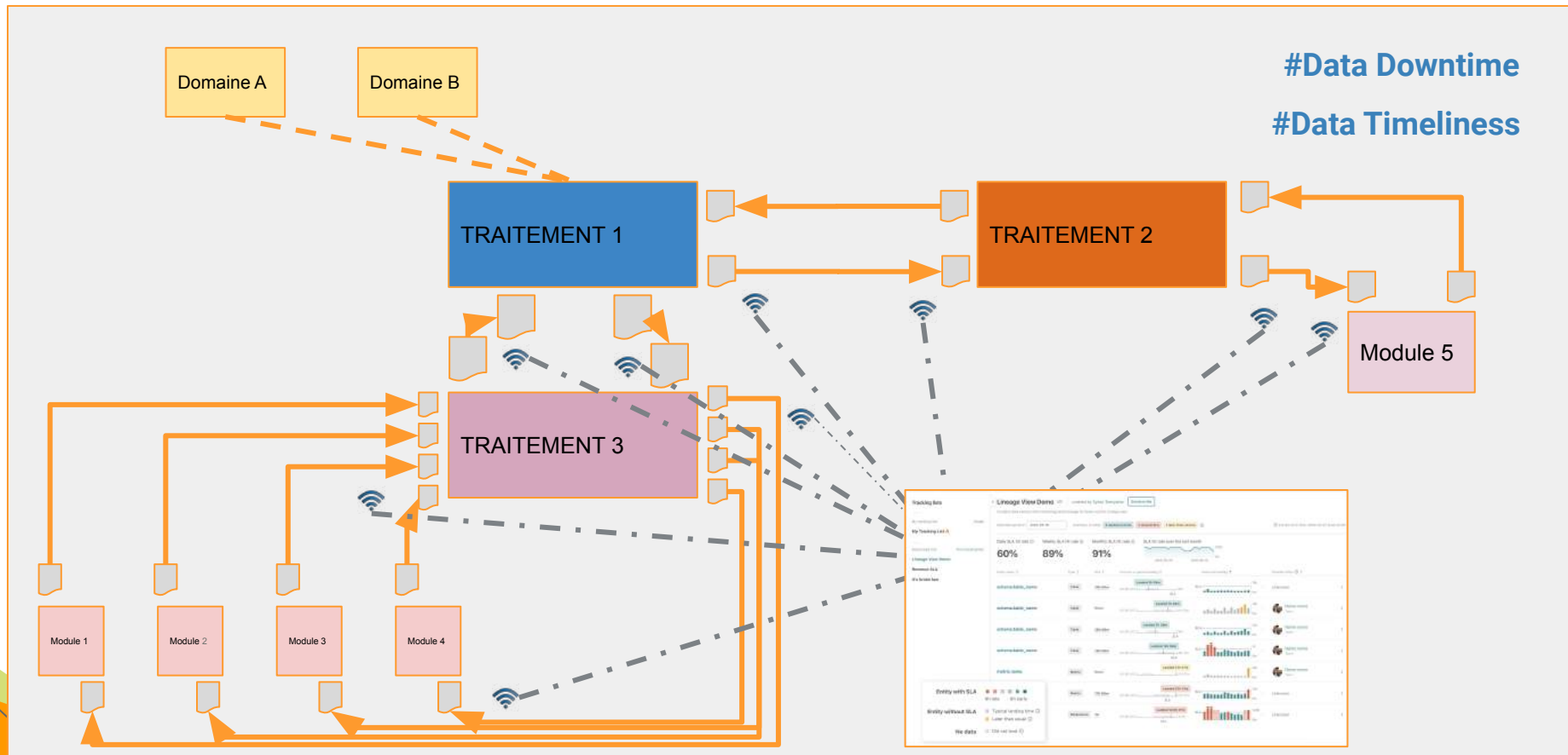
La gouvernance des données reste pratique et pragmatique !

Data Ops, par la pratique, pour simplifier et la gouvernance et la qualité de données



OBSERVABILITÉ DES TRAITEMENTS ET DES DONNÉES

Suivre la qualité de données prise en compte au sein des traitements



En résumé

La donnée accessible à tous

Quelle donnée ? Qui a accès ? Où est la donnée ? Quelle qualité ?



Expert DWH



Analyste



Data Scientist



Utilisateur
Métier



Direction



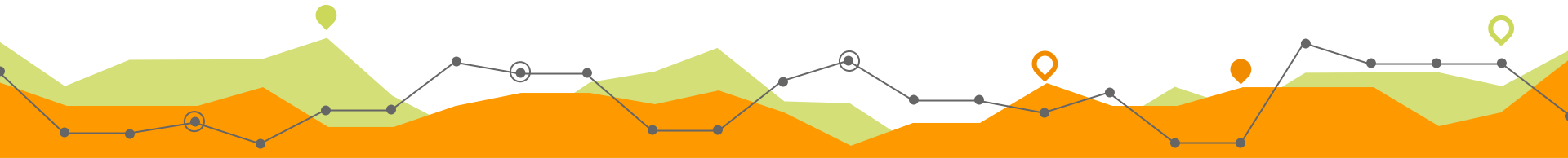
Expert ETL,
Développeur



Etude,
Expérimentation

Plateforme de données

En pratique ça donne quoi ?





CATALOGUER, GOUVERNER, DOCUMENTER, REQUÊTER, PARTAGER

ACCÉLÉRER VOS PROJETS DE DONNÉES

Une donnée fraîche et accessible dès sa création

DÉLIVRER VOTRE DATA LAKE PLUS RAPIDEMENT



SQL

Data Scientists



Utilisateur BI

3

Prolifération Cubes,
BI Extracts, &
Aggregation Tables

2

Propriétaire & Cher
DW/Data Marts

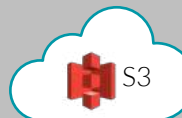
1

Trop Complexe
ETL/ELT

Data Lake
Stockage



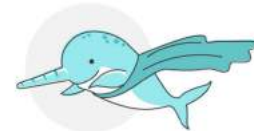
or



or

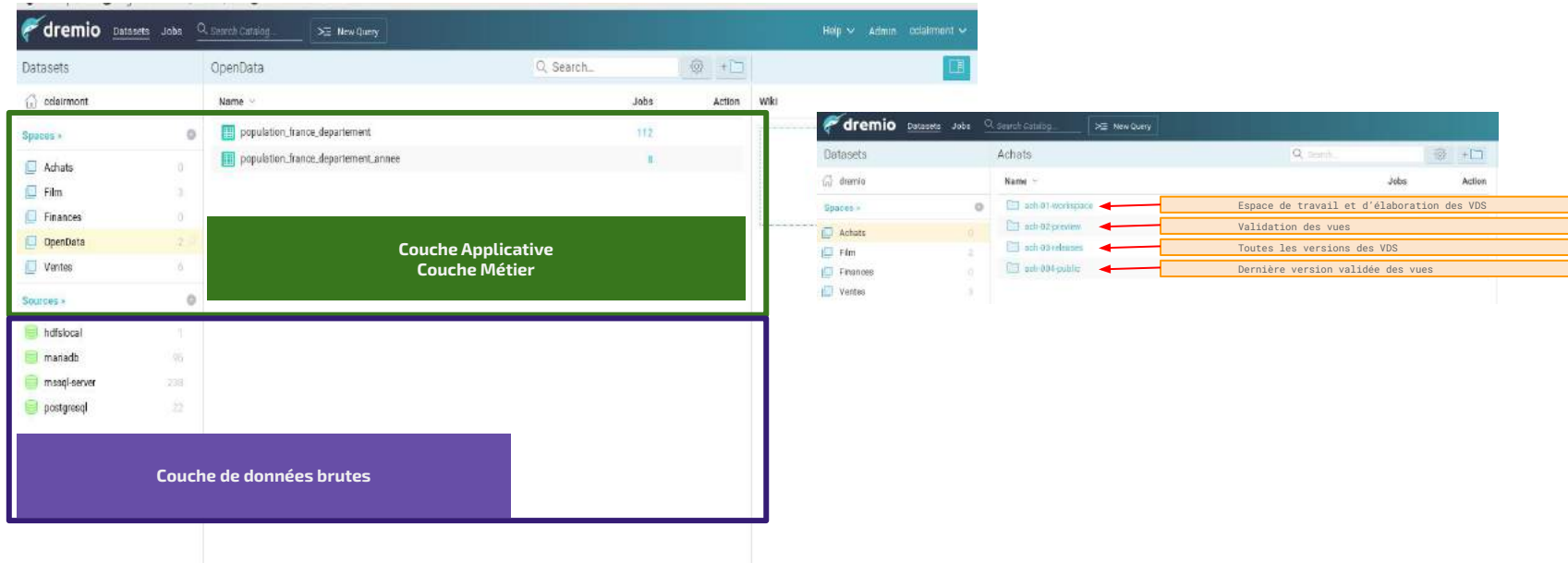


Accélérer la donnée et la productivité



DREMIO : CATALOGUER, GOUVERNER

Un point central pour toutes les données



The image shows two screenshots of the Dremio web interface, illustrating the data layer architecture.

Left Screenshot (OpenData Catalog):

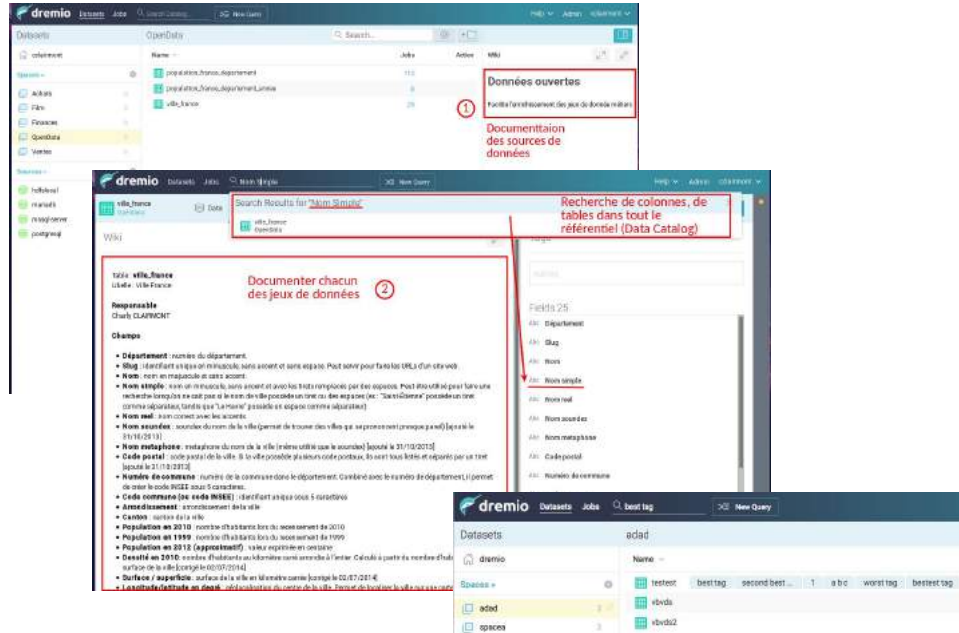
- Spaces:** A list of data spaces including Achats, Film, Finances, OpenData (highlighted), and Ventes.
- Sources:** A list of data sources including hdfslocal, mariadb, msasql-server, and postgresql.
- Central Panel:** Displays datasets under the 'OpenData' space, including 'population_france_departement' and 'population_france_departement_annee'.
- Annotations:**
 - A green box labeled "Couche Applicative Couche Métier" (Application Layer / Business Layer) is overlaid on the dataset list.
 - A purple box labeled "Couche de données brutes" (Raw Data Layer) is overlaid on the sources list.

Right Screenshot (Achats Catalog):

- Spaces:** A list of data spaces including ach-01-workspace, ach-02-preview, ach-03-releases, and ach-004-public.
- Annotations:**
 - Four orange boxes with red arrows pointing to the spaces list:
 - ach-01-workspace: "Espace de travail et d'élaboration des VDS" (Workspace for VDS development)
 - ach-02-preview: "Validation des vues" (View validation)
 - ach-03-releases: "Toutes les versions des VDS" (All versions of VDS)
 - ach-004-public: "Dernière version validée des vues" (Last validated version of views)

DREMIO : DOCUMENTER

Un point central pour toutes les données



The image displays three screenshots of the Dremio interface, illustrating its capabilities for documenting data sources.

Top Screenshot: Shows the 'Datasets' view with a table named 'ville_france'. A red box highlights the 'Données ouvertes' (Open Data) section, indicating the source of the data. A red circle with the number 1 is placed next to the 'ville_france' dataset.

Middle Screenshot: Shows the 'Search Results for "ville_france"' view. A red box highlights the 'Recherche de colonnes, de tables dans tout le référentiel (Data Catalog)' (Search for columns, tables in the entire Data Catalog) section. A red circle with the number 2 is placed next to the 'Documenter chacun des jeux de données' (Document each dataset) section.

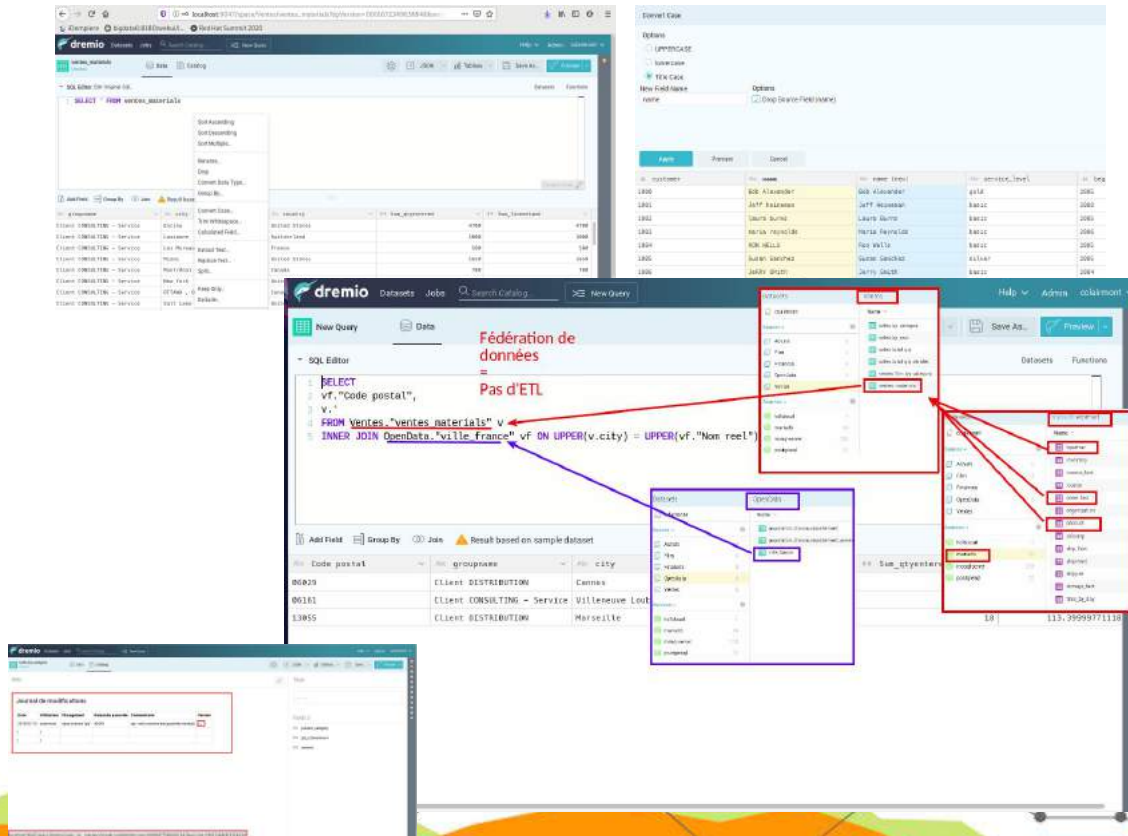
Bottom Screenshot: Shows the 'Datasets' view with a table named 'adad'. A red box highlights the 'Documenter chacun des jeux de données' (Document each dataset) section.

Annotations:

- Données ouvertes
- Documenter chacun des jeux de données
- Recherche de colonnes, de tables dans tout le référentiel (Data Catalog)
- Documenter chacun des jeux de données

DREMIO : REQUÊTER

Un point central pour toutes les données



The screenshot displays the Dremio web interface. The main window is the SQL Editor, showing a query that joins data from 'ventes_materials' and 'OpenData_ville_france'. The query is as follows:

```
SELECT
  vf."code postal",
  V."v",
  FROM Ventes."ventes_materials" v
  INNER JOIN OpenData."ville_france" vf ON UPPER(v.city) = UPPER(vf."Nom reel")
```

Annotations on the screenshot include:

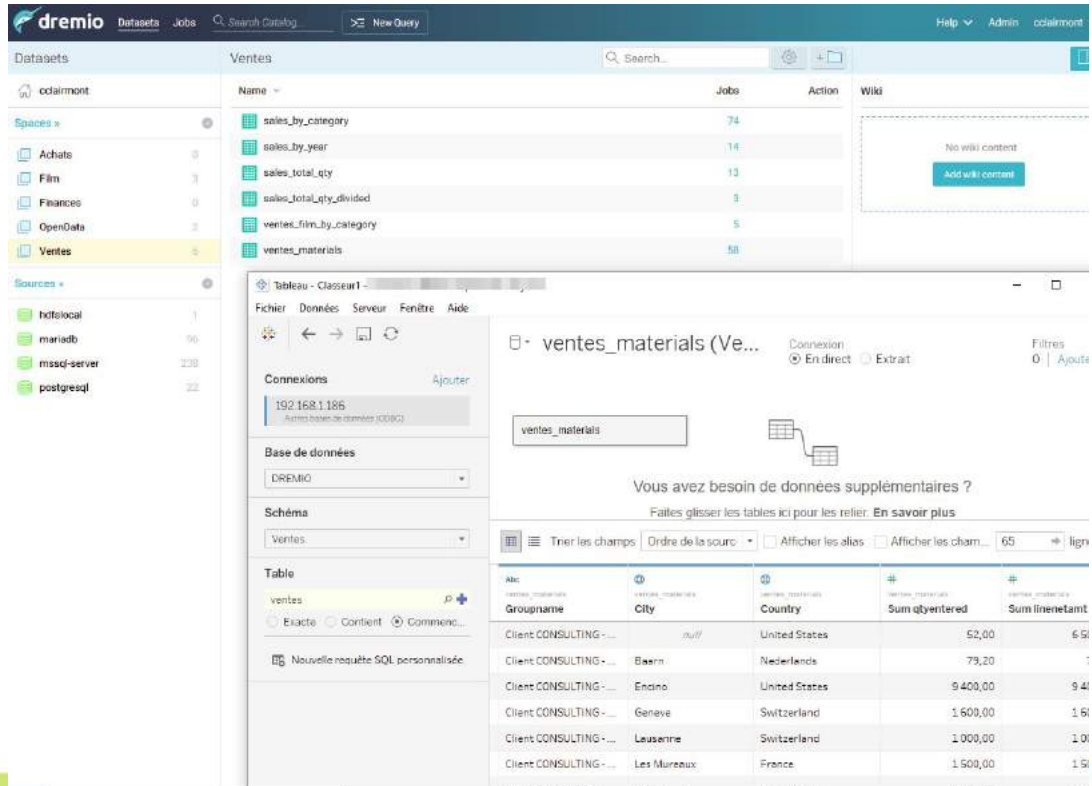
- A red box highlights the text "Fédération de données = Pas d'ETL" (Data Federation = No ETL).
- A red box highlights the "Datasets" panel on the right, showing a list of datasets including "ventes_materials" and "ville_france".
- A red box highlights the "Query Results" panel at the bottom, showing a table with columns "Code postal", "groupe", and "city".
- A red box highlights the "SQL Editor" panel, showing the query text.
- A red box highlights the "Query Results" panel, showing the query results.
- A red box highlights the "Query Results" panel, showing the query results.

The "Query Results" panel shows the following data:

Code postal	groupe	city
06009	Client DISTRIBUTION	Comman
06161	Client CONSULTING - Service	Villeuve Lou
13055	Client DISTRIBUTION	Marseille

DREMIO : PARTAGER

Un point central pour toutes les données



The screenshot displays the Dremio web interface. On the left, the 'Datasets' sidebar shows a tree structure with 'Ventures' selected. The main panel shows a list of datasets under 'Ventures' with columns for Name, Jobs, and Action. A 'Wiki' panel on the right indicates 'No wiki content'. Below the dataset list, a 'Tableau - Classeur1' window is open, showing a query for 'ventes_materials'. The query is visualized as a diagram with a table icon and the text 'ventes_materials'. Below the diagram, a message asks 'Vous avez besoin de données supplémentaires ?' (Do you need more data?). The bottom section shows a table of results with columns: Groupname, City, Country, Sum qtyentered, and Sum linetamt.

Groupname	City	Country	Sum qtyentered	Sum linetamt
Client CONSULTING - ...	Null	United States	52,00	6 5
Client CONSULTING - ...	Basen	Netherlands	79,20	1
Client CONSULTING - ...	Endino	United States	9 420,00	9 4
Client CONSULTING - ...	Geneve	Switzerland	1 600,00	1 6
Client CONSULTING - ...	Leusenne	Switzerland	1 000,00	1 0
Client CONSULTING - ...	Les Mureaux	France	1 500,00	1 5



RÉPERTORIER, DOCUMENTER, DÉCOUVRIR

AMUNDSEN : RÉPERTOIR

Répondre à la fois à la mise en place du Data Lake et à la gouvernance

Connecteurs

Table Connectors

- [Amazon Athena](#)
- [Amazon Glue](#) and anything built over it (like Databricks Delta - which is a work in progress).
- [Amazon Redshift](#)
- [Apache Cassandra](#)
- [Apache Druid](#)
- [Apache Hive](#)
- [CSV](#)
- [Delta Lake](#)
- [Google BigQuery](#)
- [IBM DB2](#)
- [Microsoft SQL Server](#)
- [MySQL](#)
- [Oracle](#) (through `dbapi` or `sql_alchemy`)
- [PostgreSQL](#)
- [Trino](#) (formerly [Presto SQL](#))
- [Vertica](#)
- [Snowflake](#)

Amundsen can also connect to any database that provides `dbapi` or `sql_alchemy` interface (which most DBs provide).

ETL Orchestration

- [Apache Airflow](#)

Dashboard Connectors

- [Mode Analytics](#)
- [Redash](#)
- [Tableau](#)

BI Viz Tool

- [Apache Superset](#)

AMUNDSEN : DOCUMENTER

Répondre à la fois à la mise en place du Data Lake et à la gouvernance

AMUNDSEN

Search for data resources...

Browse

test_schema.test_table1

Datasets • Hive • gold

Airflow Github Preview

Documenter la table

Description

1st test table

Update Cancel

Propriétaires

Owners

chris@example.org

roald.amundsen@example.org

Utilisateurs de la table

Frequent Users

Last Updated

Oct 05, 2019 1am CEST

Date Range

From: Apr 22, 2017

To: Sep 30, 2019

Tags

tag1 tag2 pii high_quality

expensive low_quality

Quality Service

Qualité des données

Quality Report:

Ipsus enim. Ipsus enim ipsus lorem.

build preview

Columns (5)

Documenter les colonnes

col1

Description

col1 description

Update Cancel

Column Statistics Stats reflect data collected between May 22, 2015 and Jul 05, 2019.

NAME	TYPE
col1	string
col2	string
col3	string
col4	string
col5	float

Amundsen was last indexed on December 1st 2020 at 5:36:01 pm

AMUNDSEN : DÉCOUVRIR

Répondre à la fois à la mise en place du Data Lake et à la gouvernance



Search within a category using the pattern with wildcard support 'category:*searchTerm*', e.g. 'schema:*core*'.
Current categories are 'column', 'database', 'schema', 'table', and 'tag'.

Browse Tags

tag1 1 tag2 1

My Bookmarks

 **test_schema.test_table2**  dynamo
2nd test table

Popular Tables

 **test_schema.test_table1**  Hive
1st test table

Amundsen was last indexed on December 18th 2019 at 2:07:17 pm

Datasets

 test_schema.test_table2 2nd test table	dynamo
 test_schema.test_table1 1st test table	Hive

[See all 2 Datasets results](#)

AMUNDSEN

Ernest

Resource	RESOURCES	SOURCE	SAMPLES
 Databases	15		
Source	public.city Clear Apply	Postgres	
Column	public.address Clear Apply	Postgres	
Schema	public.country Clear Apply	Postgres	
Table	public.city Clear Apply	Postgres	
Tag	public.inventory Clear Apply	Postgres	
	public.payment Clear Apply	Postgres	
	public.survey Clear Apply	Postgres	
	public.customer Clear Apply	Postgres	

Amundsen was last indexed on December 18th 2019 at 2:49:33 pm



CATALOGUER, GOUVERNER, QUALITÉ

TALEND : CATALOGUER

De l'intégration des données à la gouvernance

ADD A NEW DATASET

Connection*

s3

Description

List of restaurants, bars, clubs, cafés and their reviews

S3 DATA INPUT

Bucket*

big-business

Object*

/big_business.avro

Encrypt Data At Rest

FORMAT CONFIG

AUTO DETECT

Format*

Avro

CANCEL VIEW SAMPLE VALIDATE

DATA INVENTORY

ADD DATASET
Drop your files or browse

Rating < 50 x Add filter

Name	Quality	Connection	Author	Updated
Nova	54%	Nova	Ruprecht Wadham	2 min ago
Oakridge	32%	Oakridge	Erv Charte	36 min ago
Mitsner-Kunonsson	90%			
Koch-Batz	76%			
Pfeffer, Hamill and Cremin	82%			
Gulgowski Group	61%			
Sauer, Deckow and Gislason	12%			
Kulas-Spinka	95%			
	72%			

Filter by rating:
50%
Include datasets without rating
APPLY

ADD DATASET
Drop a file or browse

customer
Add filter

Search in
Dataset name
Connection name
Description

TALEND : GOUVERNER

De l'intégration des données à la gouvernance

taxi data
Owner Talend User

ALL USERS AND GROUPS

Jane Doe

Marc Myers

Alan Maroon

Peter Baillou

Remy Wood

Talend Admin

CURRENT COLLABORATORS

Talend_group Editor

Justin Doe Owner

John Doe Viewer

Jimmy Doe Editor

Owner

Editor

Viewer

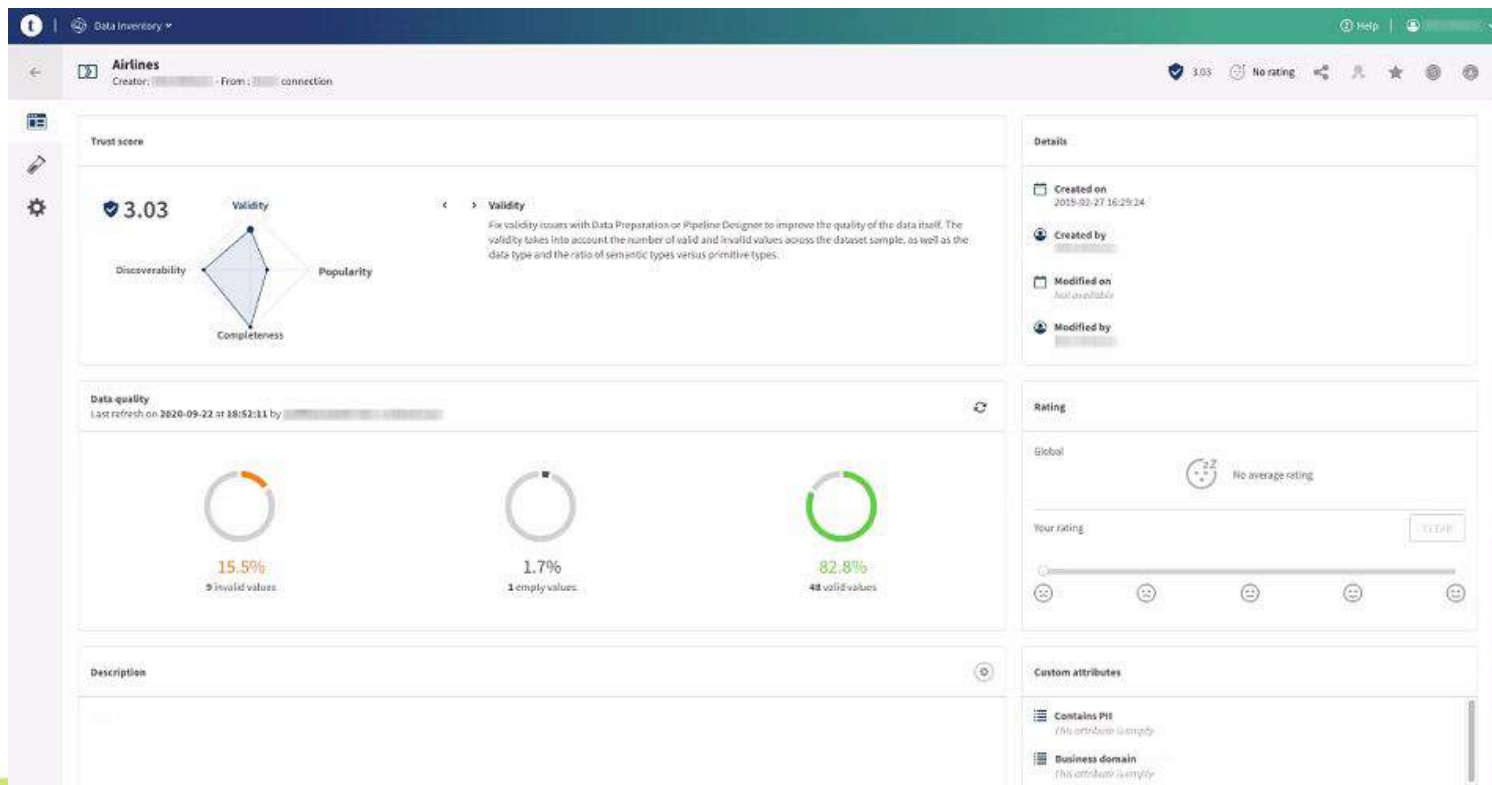
CANCEL

ADD AS ...

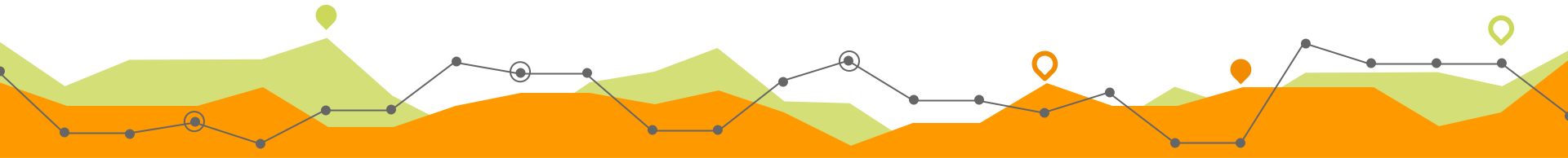
SHARE

TALEND : QUALITÉ

De l'intégration des données à la gouvernance



Mais encore...



UN PETIT LANDSCAPE

...

Data Catalog / Data Gouvernance



Cloud



Data Observabilité



Français



Solutions



Open Source



Adopter un “Data Facilitateur”

Faciliter l'évolution de vos pratiques avec la donnée

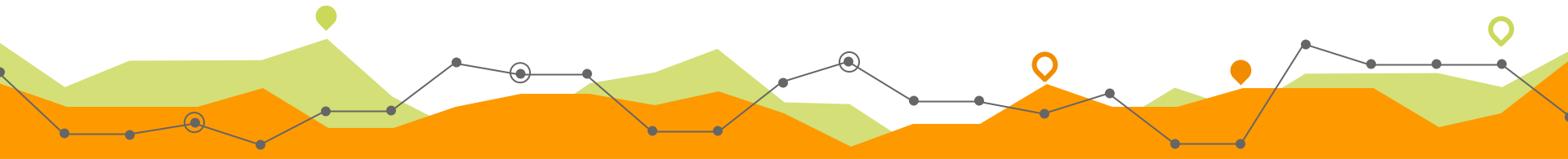
Connaissances

Culture de la donnée
Système d'information
Processus métier
Agit selon les méthodes agiles
Créativité et innovation

Animateur / Négociateur

Favoriser l'engagement de vos équipes,
Promouvoir la culture de la donnée
Assurer la transparence des choix,
Organiser les rendus,
Transformer les idées en projets, les projets en produit

Contactez nous, nous avons spécialement conçu ce rôle pour vous accompagner !

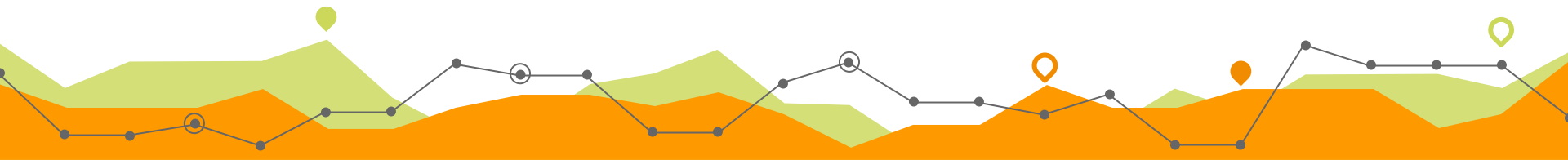


Merci !

Questions?

[@Synaltic / @egwada](#)

<https://www.synaltic.fr>



Festival Online de la Data 2021



Synaltic

 **Synaltic**

OFFRES | PARTENAIRES | CLIENTS | **WEBINAIRE** | BLOG | A PROPOS | CONTACT | 

Elastic Cloud on Kubernetes

Animé par Galla TOPALIAN et Alexandre NASRY, SYNALTIC

Vous allez découvrir comment le déploiement d'un cluster **Elasticsearch** et du visualiseur **Kibana** peut être réalisé en moins de 10 minutes avec Kubernetes. De l'installation à votre premier Dashboard en 20 minutes, c'est l'objectif de cette présentation en forme de tutoriel.

[VOIR LE REPLAY](#)

Data Asso

Animé par Marc SALLIERES, Président, SYNALTIC

Né d'une initiative citoyenne, **Data-Asso** est soutenu par Synaltic, en partenariat avec le Ministère de l'éducation nationale, de la jeunesse et des sports, chargé notamment de la vie associative. En effet, le projet a bénéficié en partie des fonds du Plan d'Investissement d'Avenir (PIA).

[VOIR LE REPLAY](#)

De Tableau Prep à la visualisation sur Tableau

Animé par Aya MHADHBI, Data Analyst, SYNALTIC

Préparer vos données en quelques clics pour se lancer dans l'analyse rapidement est un vrai atout. Avec **Tableau Prep**, vous allez gagner du temps en réduisant considérablement le délai entre la collecte et les données propres.

[VOIR LE REPLAY](#)

LE REPLAY VOUS ATTEND SUR NOTRE SITE !