Contents

[Jour 1A 2](#_Toc136430351)

[Vocabulaire de la data 2](#_Toc136430352)

[1.1 Présentation du métier de data analyst 3](#_Toc136430353)

[1.2 Missions du data analyst 4](#_Toc136430354)

[2.1 Présentation du métier de data scientist 4](#_Toc136430355)

[2.2 Missions du data scientist 4](#_Toc136430356)

[3.1 Présentation du métier de data engineer 4](#_Toc136430357)

[3.2 Missions du data engineer 4](#_Toc136430358)

[4. Présentation du data hero (ambassador de data) 4](#_Toc136430359)

[5. Présentation du métier de data consultant 5](#_Toc136430360)

[6. Présentation du métier de DPO (déléguer à la protection des données) 5](#_Toc136430361)

[7. Présentation du métier du Head of Data (manager d’équipe data) 5](#_Toc136430362)

[8. Présentation du métier de CTO (chief technology officer) 5](#_Toc136430363)

[JOUR 2M-Maturité 5](#_Toc136430364)

[Comment évaluer la maturité data d'une entreprise ? 6](#_Toc136430365)

[L'évolution du niveau de maturité data d'une entreprise 6](#_Toc136430366)

[La gouvernance des données 6](#_Toc136430367)

[Quelles sont les roadmaps d'un projet data ? 7](#_Toc136430368)

[Comment devenir data driven ? 7](#_Toc136430369)

[Comment gérer les interactions des équipes tech et non tech ? 7](#_Toc136430370)

[Comment gérer une équipe de data analytics ? 7](#_Toc136430371)

[Commenter recruter des profils data ? 8](#_Toc136430372)

[Conseils pour se lancer dans la data 8](#_Toc136430373)

[The Modern Data Stack récap 8](#_Toc136430374)

[Grille de lecture - Data Maturity 8](#_Toc136430375)

[Data Maturity récap 8](#_Toc136430376)

[JOUR 2M- Infra 8](#_Toc136430377)

[👉 Objectifs 8](#_Toc136430378)

[Support de la présentation 8](#_Toc136430379)

# Jour 1A

## Vocabulaire de la data

Les mots de vocabulaire abordés :

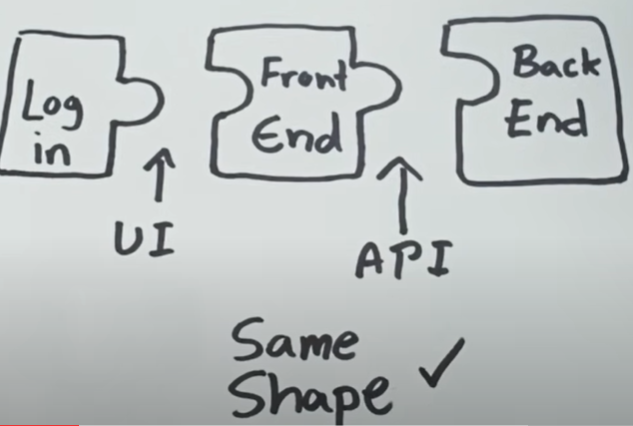
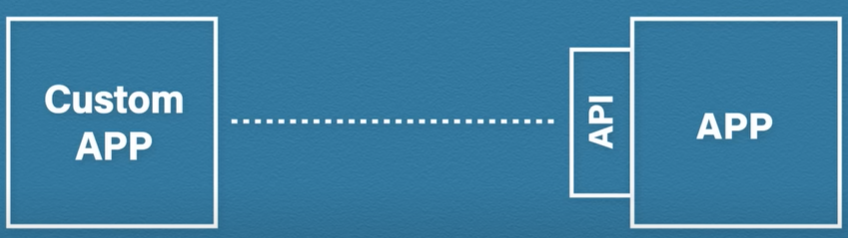
* API (application programing interface)

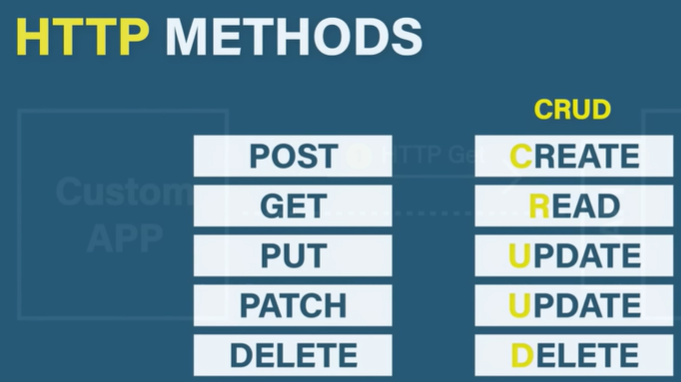
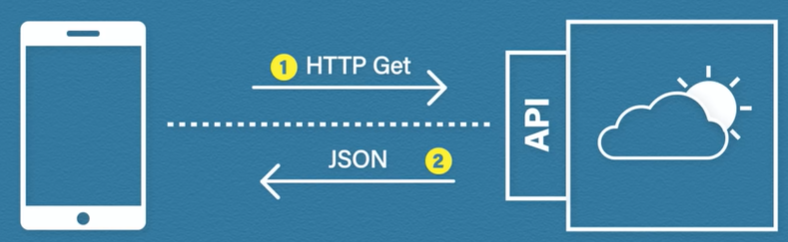
Opendatagouv

Connect apps together

UI (user interface)

<https://www.youtube.com/watch?v=ZveW4_ZJtVY>



* Back-end vs Front-end
* Les méthodes agiles

Cette un séries d’utiles que permet les entreprise d’organiser l’équipe de projet

<https://www.youtube.com/watch?v=gYsU2VTESE8>

* Environnement de test vs Mise en production (MEP)

Test le model avant pour trouver des bug et deploiment

* Git
* ETL (Extract transform load)
  + Pour les données à faible volume
  + Il est utilisé pour les données structurées et relationnelles
  + La mise en œuvre du processus ETL est relativement simple.
  + Apporter des modifications aux données dans l'étape de transformation prend du temps, et plus il y a de données, plus la transformation prend de temps.
  + Étant donné que dans le processus ETL, les données doivent être sélectionnées, puis la transformation et le chargement sont effectués sur les données sélectionnées, cela nécessite donc une maintenance élevée.
* ELT (Extract load transform)
  + Il peut être utilisé pour les données à volume élevé
  + Il est utilisé pour les sources de données non structurées et les Clouds.
  + Cela demande des compétences particulières
  + La vitesse n'a rien à voir avec le volume de données.
  + Comme les données sont toujours disponibles, le besoin de maintenance est extrêmement faible.

<https://www.nemoudar.com/blog/elt-vs-etl/>

* Cloud (pour stoker les donnée)
* Composition d'une équipe produit

Il a tout le rôle nécessaire pour un développement de produit. I va faire la gestion de projet. Pour expliquer pus technique pour les développeur.

* Intelligence artificielle vs Data Science vs Machine Learning

ML( automatiser le machine), DS( statistique, optimisation), IA( les robot, permet de développer application de ML)

* Data Lake (centraliser la donnée)
* Data Warehouse (la base de données et le moteur qui permet en construire cette base de données pour obtenir une version plus exploitable des KPI (les indicateurs) pour tous les utilisateurs d’entreprise.

## 1.1 Présentation du métier de data analyst

* Le rôle du data analyst exploite les données de l'entreprise pour prendre des meilleures décisions
* Le data analyst travaille toujours pour le business

## 1.2 Missions du data analyst

Exemples de missions du data analyst :

* Réaliser un AB test
* Réaliser des dashboards
* Réaliser des analyses data ad hoc pour le business
* Lles langage : SQL, Excel, python, tableau, domo, metabase, powerBI, data hekou

## 2.1 Présentation du métier de data scientist

* Traditionnellement, le data scientist ne s'occupait que de créer des modèles
* Aujourd'hui, on attend aussi de lui de mettre en production les modèles qu'il a développés
* Les data scientists sont essentiellement des ingénieurs
* Les langages : python, java, scala, tensorfow

## 2.2 Missions du data scientist

* Le data scientist va principalement se concentrer sur la création et la performance de modèles
* Il peut faire de la recommandation, de la prédiction ou de l'analyse d'image / texte / son
* Exemple de projet concret sur lequel le data scientist peut travailler :
  + Repérer automatiquement les défauts sur des rails de chemin de fer

## 3.1 Présentation du métier de data engineer

* Le data engineer crée, alimente et maintient les bases de données
* Le data engineer s'occupe de l'ETL
* Le data engineer est un "développeur" dans la data
* Les langages : SQL, DBT, Cloud, AWS, Azur, python, Java, Scala, Go
* cloud sur les optimisation technique

## 3.2 Missions du data engineer

* Une mission du data engineer pourrait être de tagger la donnée avec des segments définis par les data analysts et les équipes marketing

## 4. Présentation du data hero (ambassador de data)

* La data hero n'est pas un métier en tant que tel
* Le data hero est un membre d'une équiper métier qui s'est formé sur des aspects techniques
* Le data hero est un réel ambassadeur et facilitateur data

## 5. Présentation du métier de data consultant

* Un consultant data peut avoir n'importe quel rôle data dans l'entreprise
* Il permet de développer une expertise en interne
* Un exemple de mission sur lequel il pourrait intervenir : embaucher des consultants data pour traiter la mise en conformité de la RGPD

## 6. Présentation du métier de DPO (déléguer à la protection des données)

* Le DPO conseille les entreprises dans la mise en conformité RGPD (règlement européen pour e protection des données)
* Il accompagne dans le respect des règles européennes mais aussi dans la mise en place des documents en cas de contrôle
* Un exemple concret dans lequel le DPO agirait serait le partage d'informations sensibles dans le cadre d'une entreprise opérant dans le recrutement

## 7. Présentation du métier du Head of Data (manager d’équipe data)

* Son rôle est de recruter et de former les équipes data
* Il va être aussi l'ambassadeur data dans l'entreprise
* Il est vraiment le pilier central de la stratégie "data-driven" de l'entreprise

## 8. Présentation du métier de CTO (chief technology officer)

* Son rôle est de donner la direction technique/technologique de l'entreprise. Il peut être :
  + Le manager des équipes techniques
  + L'expert technique de l'entreprise
  + Les deux

# JOUR 2M-Maturité

👉 Objectifs

* Evaluer la maturité data d'une entreprise
* Développer une approche data driven
* Quels sont les profils à recruter

## Comment évaluer la maturité data d'une entreprise ?

* Pour analyser la maturité data d'une entreprise on peut regarder :
  + Les typologies et les rôles des profils présents dans l'entreprise
  + La taille des équipes data
  + Les technologies utilisées (cloud, tooling, ...)
  + Les process mis en place

<https://www.nemoudar.com/blog/data-maturity-model/>

چهار مرحله مدل بلوغ داده

Data Aware (آگاهی به داده)

چیرگی بر داده ها (Data Proficient)

درک داده‌ها (Data Savvy)

داده محور (Data Driven)

## L'évolution du niveau de maturité data d'une entreprise

* Stade 0 : Utilisation de simples fichiers Excel sur un Drive
* Stade 1 : Recruter une première personne en data et mise en place d'une base de données et d'un outil de visualisation open source
* Stade 2 : Formation des équipes au SQL et au dashboarding pour utiliser et valoriser la donnée
* Stade 3 : Apparition de problématiques plus complexes. Nécessité d'acculturation et de formation en Python par exemple
* Stade 4 : Intégration d'outils nouveaux et plus performants

## La gouvernance des données

La gouvernance de données peut se définir en 5 points :

* Les data contracts : définir la donnée et son emplacement
* Définir les KPIs principaux
* L'accès aux données : les "data catalog"
* La sécurité : qui a accès à quoi ?
* La compliance (RGPD)

## Quelles sont les roadmaps d'un projet data ?

* Roadmap d'un projet data sciences : prédiction de volume pour la supply chain
  + Cadrer votre projet
  + Développer vos algorithmes
  + Déployer vos modèles
  + Piloter vos résultats
* Roadmap d'un projet data analyse : manque de visibilité sur les habitudes des consommateurs
  + Cadrer votre projet (5 Why's)
  + Analyser, récupérer et agréger vos données
  + Créer votre dashboard et  restituer vos analyses
* Roadmap d'un projet data engineer : l'équipe finance n'a pas de visibilité sur certains indicateurs
  + Identifier où est la donnée
  + Mettre en place un data contract
  + Récupérer et transformer la donnée
  + Monitorer la performance du projet

## Comment devenir data driven ?

* Identifier des cas d'usage
* Former des ambassadeurs
* Former et acculturer vos collaborateurs à la data

## Comment gérer les interactions des équipes tech et non tech ?

* Enjeu de vocabulaire : il faut parler la même langue
* Adapter la complexité des analyses et des recommandations à son public
* Instaurer des data hero et des process

## Comment gérer une équipe de data analytics ?

* 2 flows de travail
  + Des gros projets
  + Des analyses ad hoc avec un système de ticketing
* Les outils comme Trello ou Jira peuvent être utilisés

## Commenter recruter des profils data ?

* Matérialiser et valoriser l'impact d'un profil data sur le business
* Etre clair sur le rôle en question et la montée en compétences
* Embaucher des profils curieux et débrouillards

## Conseils pour se lancer dans la data

## The Modern Data Stack récap

PDF

## Grille de lecture - Data Maturity

PDF

## Data Maturity récap

PDF

# JOUR 2M- Infra

👉 Objectifs

Pour ce module c'est [Alexandre](https://www.linkedin.com/in/alexandre-miny-de-tornaco/), cofondateur et directeur pédagogique de Databird 🦜, qui va t'expliquer les infrastructures data modernes.

Via ce module, tu vas donc :

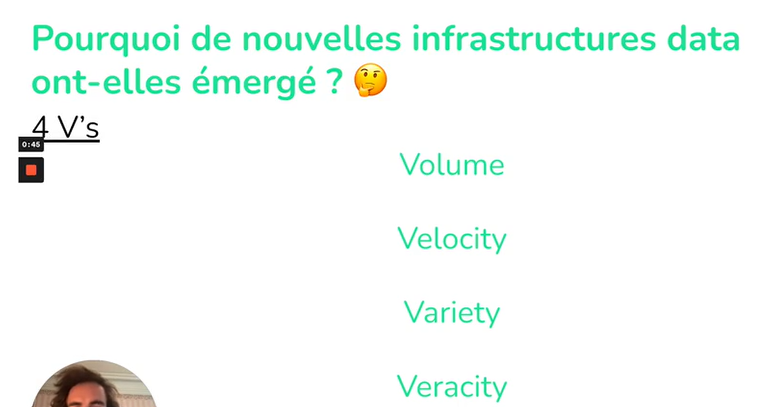
* apprendre de quoi est constitué une infrastructure data moderne
* apprendre quelle est le rôle de chaque outils utilisé
* quelles sont les grandes tendances de marchés sur "l'infra" et le "tooling" data

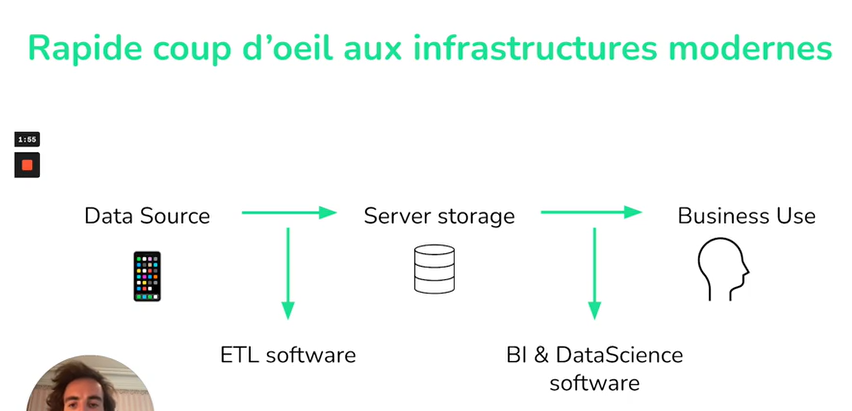
## Support de la présentation

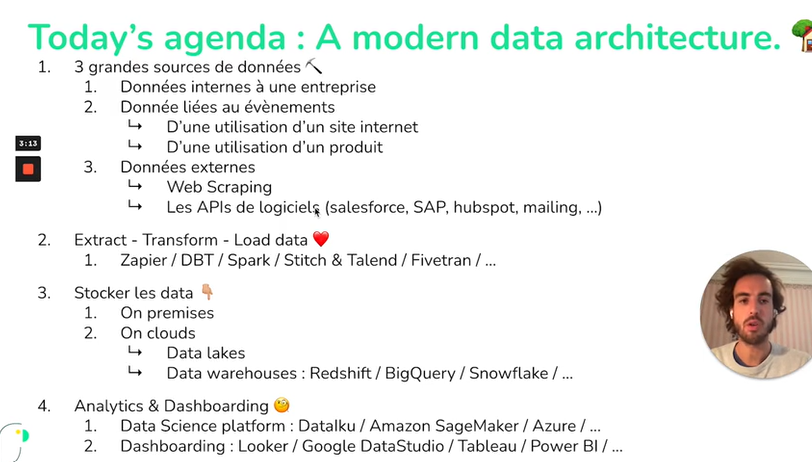
PDF Culture tech et data

## Intro - Modern data Stack

* Une infrastructure data moderne permet de centraliser des données éparses (internes / d'évènements / externes)  via des ETL (ou ELT)
* Afin de fédérer les données on peut utiliser des APIs qui vont agir comme des connecteurs de données
* Une fois les données centralisées dans un data warehouse, des analyses peuvent être faites, soit en SQL, soit via des plateformes de data analytics ou de dashboarding







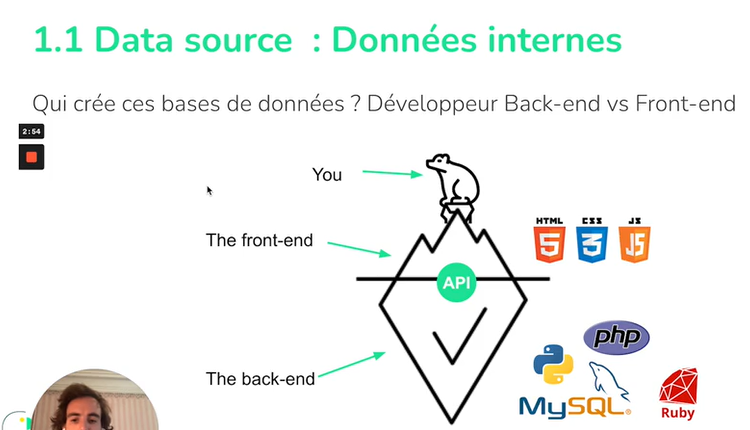
## Les sources de données

Nous avons segmenté les sources de données en 3 :

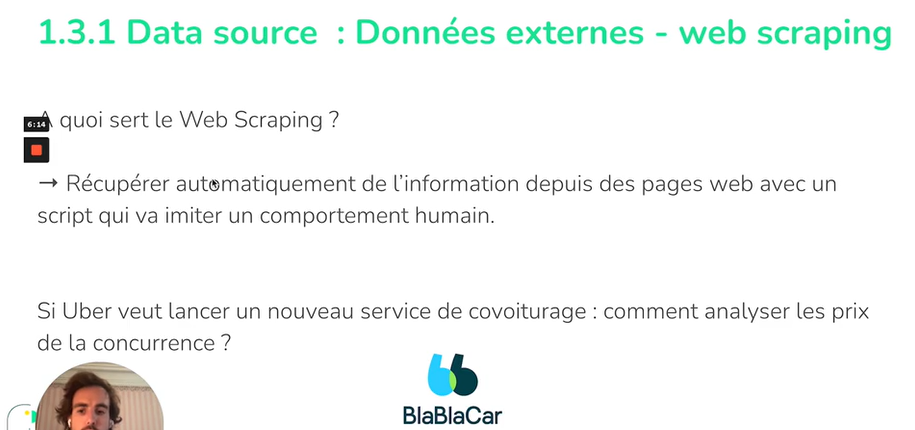
* Les données internes à l'entreprise (compte client / RH / Finance ...)
* Les données d'évènements (comportement utilisateur sur un site ou un produit)
* Les données externes (WebScraping ou contenues dans des SaaS)

Les APIs :

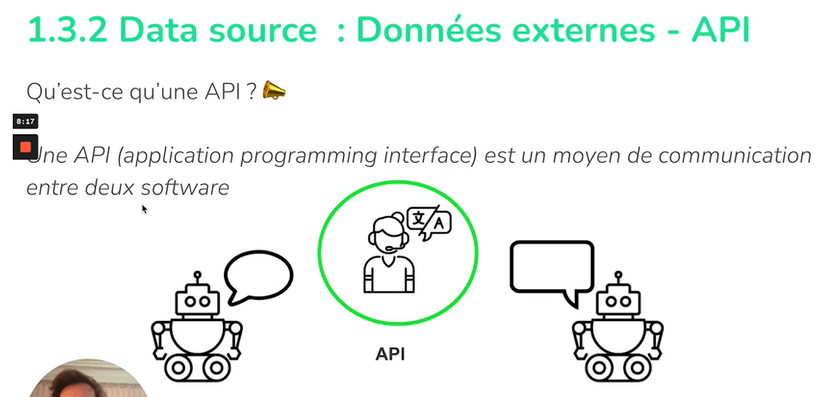
* Sont un moyen de communiquer entre deux programmes informatiques
* Agissent comme des agents de sécurité et des traducteurs
* Suivent les principes REST à 80% d'entre elles

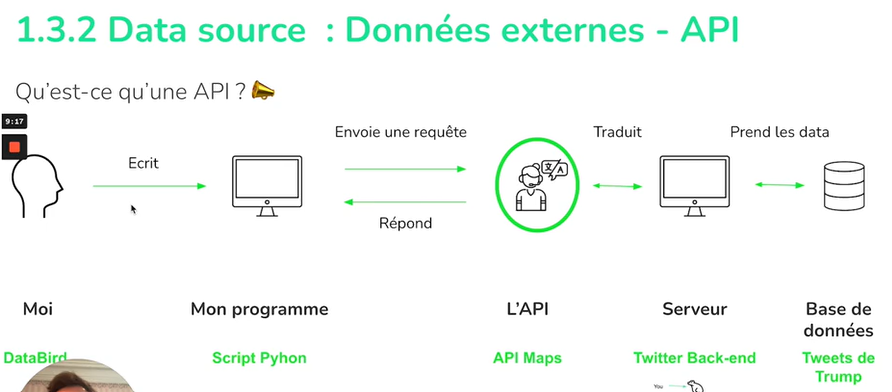


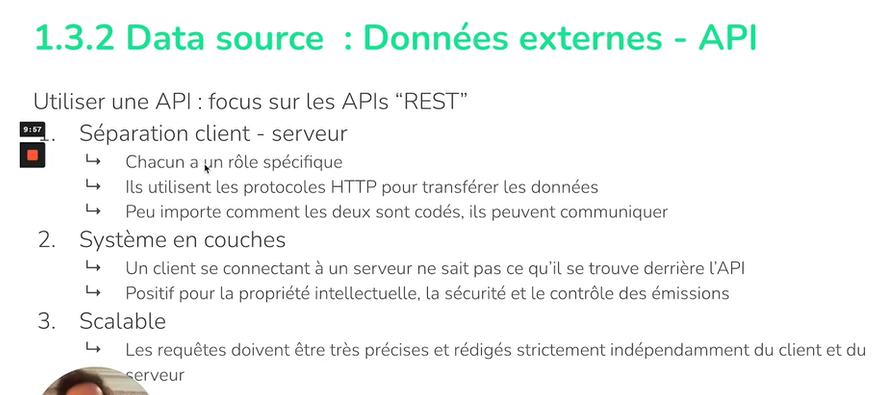






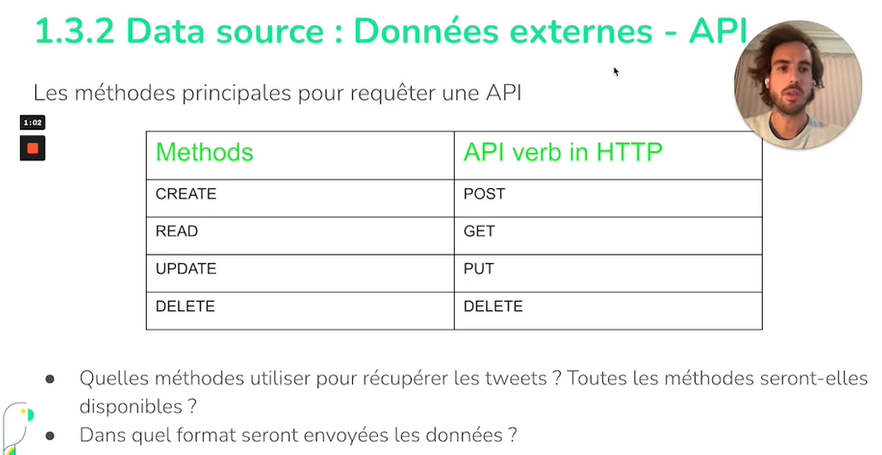






## Les APIs

* GET permet de récupérer de l'info dans la base de données qu'on interroge
* POST permet de créer une nouvelle ligne dans la base de données qu'on interroge
* PUT permet de mettre à jour une ligne dans la base de données qu'on interroge
* DELETE permet de supprimer une ligne dans la base de données qu'on interroge



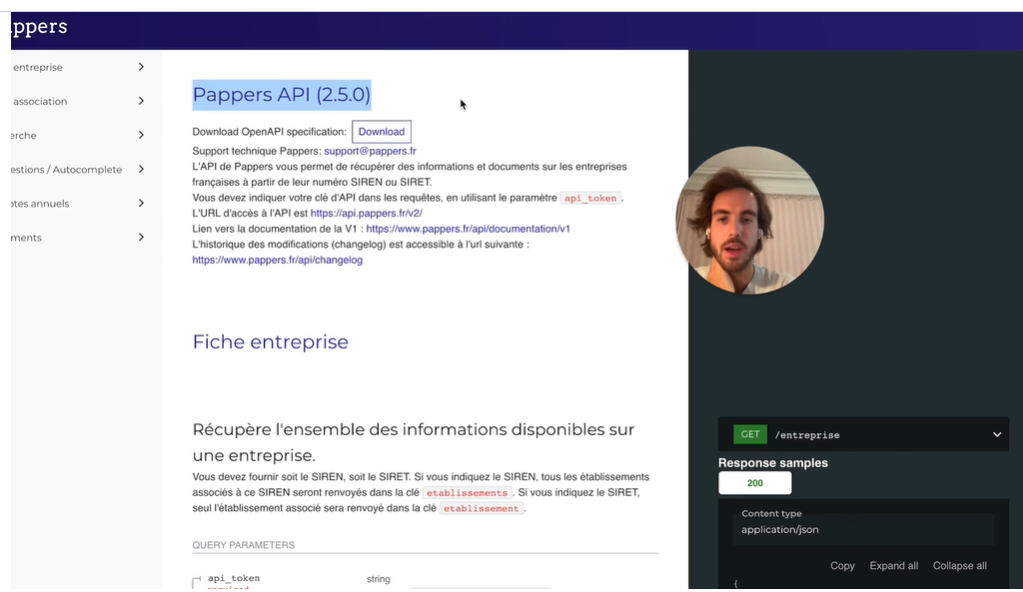


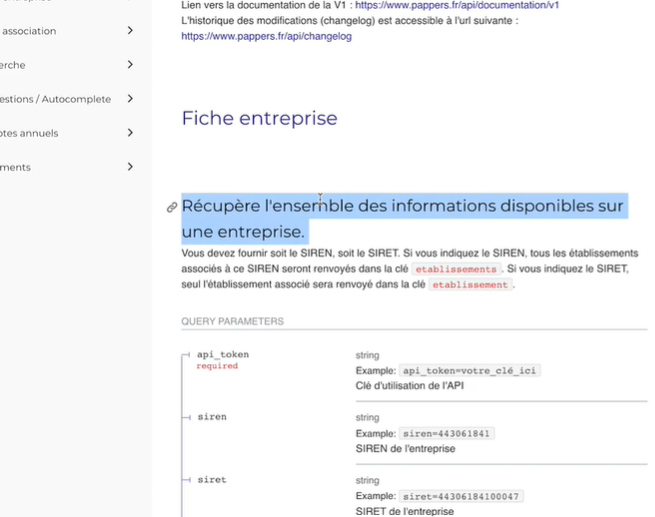
## APIs Partie 2

Les sites internets et outils utilisés :  
<https://www.pappers.fr/>  
https://www.pappers.fr/api/documentation#tag/Recherche  
<https://www.postman.com/downloads/>  
  
La requête API utilisée sur PostMan :  
<https://api.pappers.fr/v2/recherche?api_token=d6ef957beb40690c8884e6cae24967d71068f7dbbb36bfcd&siren=883538100>  
  
Key takeaways :

* Un outil comme Postman permet d'interroger des APIs
* Le Token fourni vous identifie
* Des paramètres supplémentaires peuvent être fournis
* Les APIs utilisent le protocole HTTP pour envoyer de l'info et souvent des données au format JSON

Il faut aller dans la documentation : API 🡪 documentation : <https://www.pappers.fr/api/documentation>

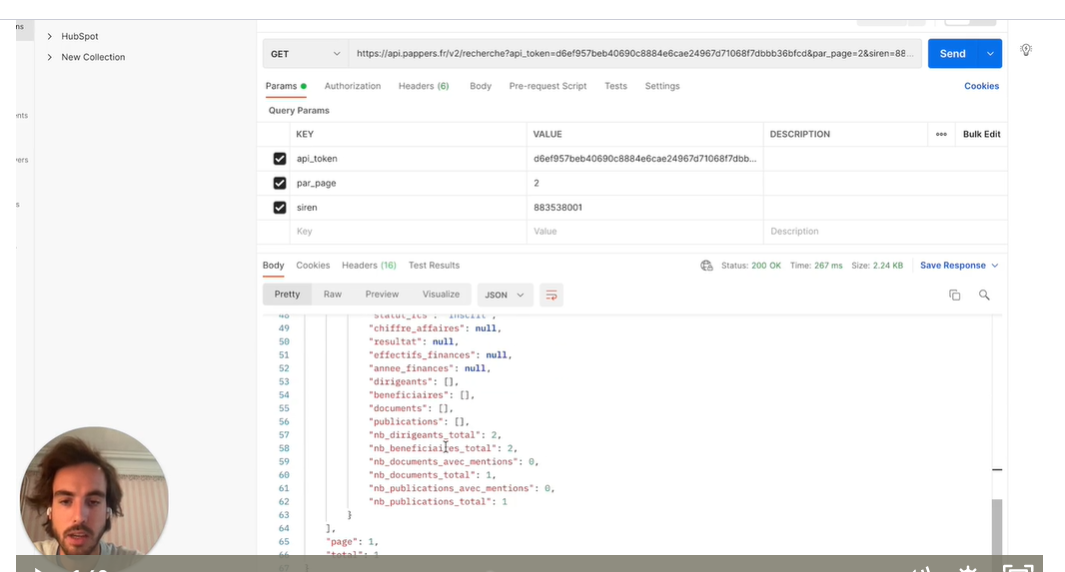






Et sur le PostMan ca donne ca ;





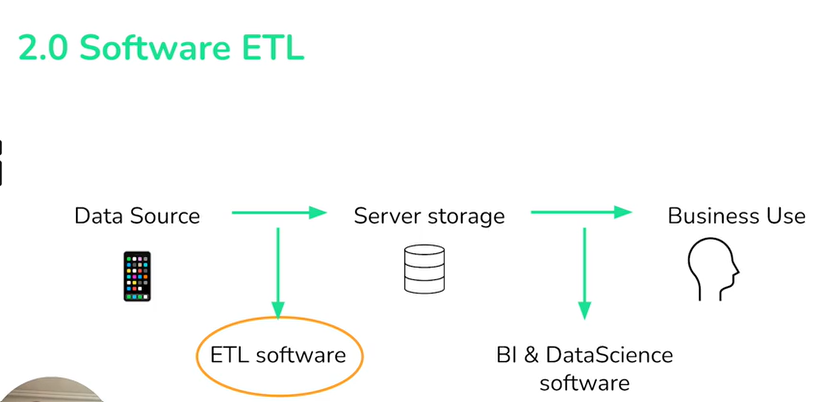
## Les APIs de SaaS

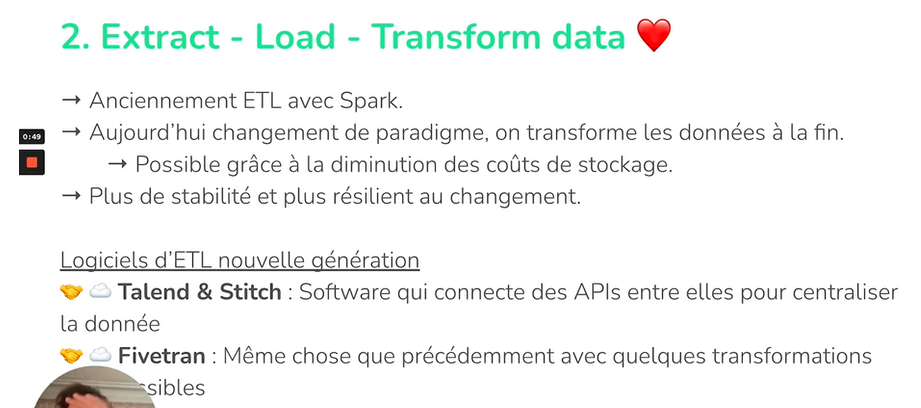
* Tous les SaaS qu'une entreprise utilise possède une API
* Interroger l'API de votre SaaS vous permet de récupérer de manière automatique toute la donnée qu'il contient
* Beaucoup de logiciels d'ETL se résument maintenant à des connecteurs d'APIs



## Les ETL

* On passe d'ETL à ELT dû aux coûts de stockage
* Dans un mon idéal, c'est le consommateur de la donnée qui va la modéliser car personne ne sait mieux que lui ce dont il a besoin
* Des logiciels comme DBT permettent de faire des transformations de données en SQL (à la place de Python, Java ou Scala)



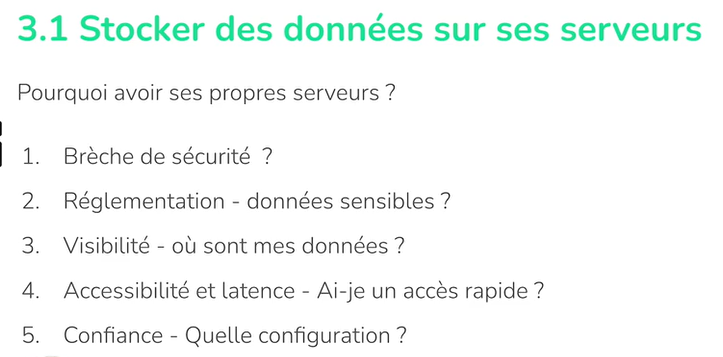


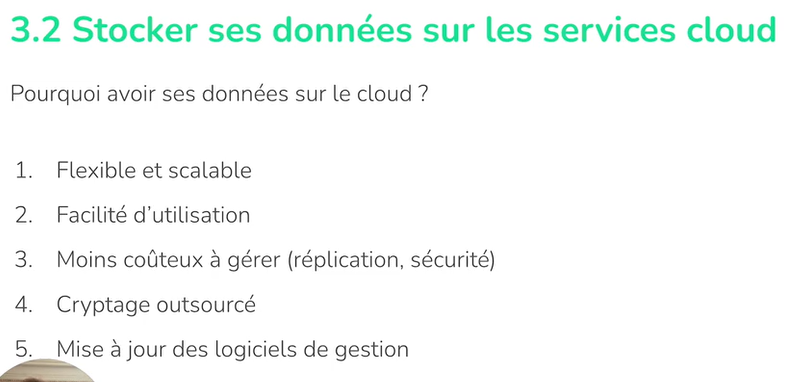


## Le stockage

* Un data lake permet de stocker toute sorte de données
* Un data warehouse est à la fois un endroit de stockage et un super calculateur
* Airflow permet d'orchestrer toutes les opérations

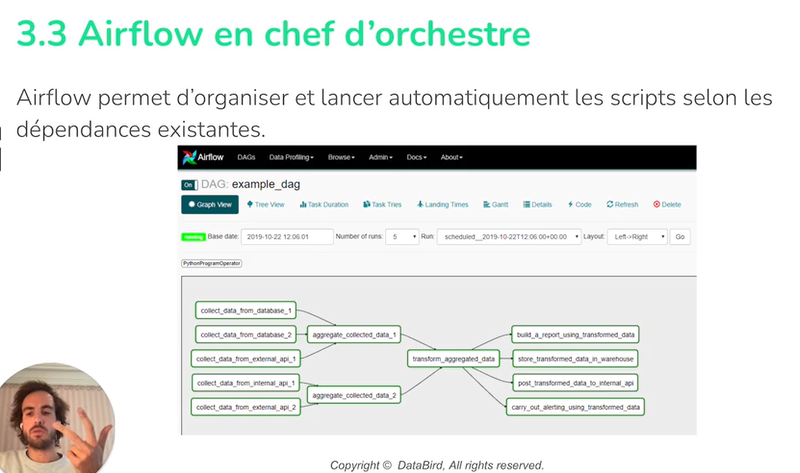






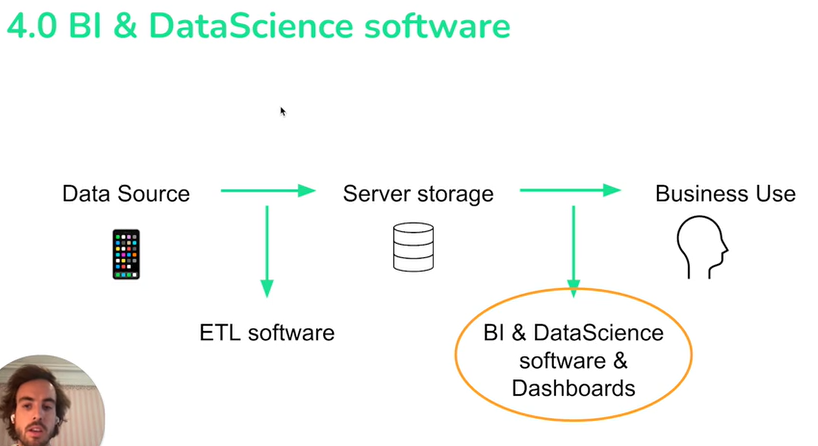


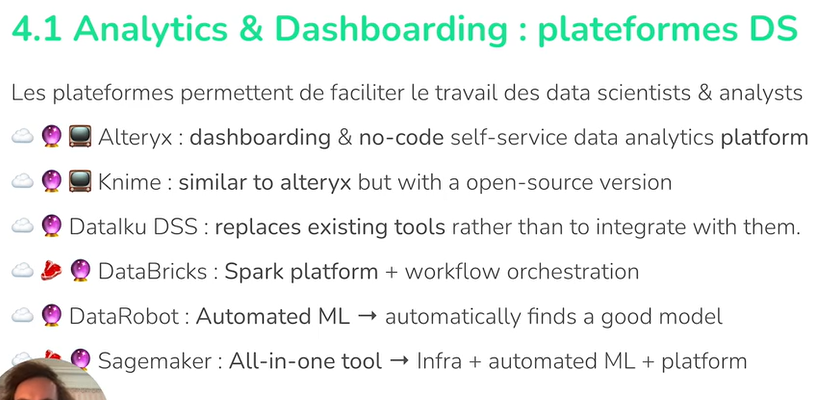




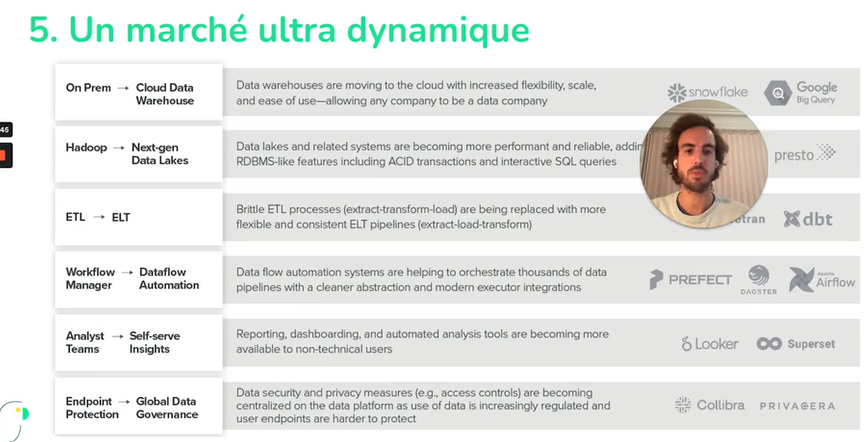
## Analytics Platform

* Les plateformes de data sciences permettent de simplifier et fluidifier le travail des data scientists
* Les logiciels de dashboarding se connectent directement au cloud et se mettent à jour automatiquement, ce qui permet un reporting plus facile
* C'est un marché dynamique et qui a vocation à encore grandir









## Recap

Pour résumer ce module, lis ces 3 articles (classés par ordre d'importance) afin de bien comprendre les outils évoqués !

* <https://technically.dev/posts/what-your-data-team-is-using>
* <https://towardsdatascience.com/modern-unified-data-architecture-38182304afcc>
* <https://medium.com/castor-app/what-if-you-had-to-build-a-data-stack-from-scratch-9700c8ec558c>

# JOUR 2A

PDF : Cas\_Macro\_ParisCode-220207-121217