

Céline CARLIER, Marina LIN-SI, Paul THIBONNIER

Thème 4 -- Présentation



Sommaire

1. Introduction à la visualisation de données

- 1.1. Définition
- 1.2. Intérêts
- 1.3. Dashboard
- 1.4. Data Storytelling

2. Les types de représentation de données

- 2.1. Temporel
- 2.2. Hiérarchique
- 2.3. Réseau
- 2.4. Multidimensionnel
- 2.5. Géospatial

3. Quelques acteurs de la Data Visualisation

- 3.1. Tableau
- 3.2. PowerBI
- 3.3. Grafana
- 3.4. Kibana
- 3.5. Data Studio
- 3.6. D3.js
- 3.7. QuickSight

1.

Introduction à la visualisation de données



1. Introduction à la visualisation de données

1.1. Définition

« Une image vaut mille mots » - Confucius

- **Représentation graphique** d'informations et de données
- Transformation d'informations brutes (en apparence complexes) en **objets visuels** (points, barres, courbes, cartographies, ...)



1. Introduction à la visualisation de données

1.2. Intérêts

Q : Combien y a-t-il de 6 sur cette image ?

3	5	1	6	2	9	9	2	8	3	8	1	6	4	0	6
0	4	2	0	6	7	2	6	8	5	3	5	0	2	2	7
4	9	5	7	9	1	2	2	1	9	9	6	0	2	9	3
4	2	8	9	3	3	2	8	0	8	5	1	7	3	1	9
4	2	0	7	4	3	1	5	4	2	1	4	5	9	6	2
1	5	0	0	1	5	2	6	8	6	1	1	8	5	5	8
2	3	4	2	2	7	5	9	9	7	6	5	4	6	4	9
9	4	2	0	8	5	0	4	9	0	0	8	8	9	8	9
1	0	3	5	2	5	4	6	6	0	4	2	1	1	7	9
3	3	7	3	3	6	6	7	8	7	8	2	7	2	0	0
5	4	7	6	1	9	1	8	0	0	5	6	6	8	0	1
9	6	4	8	2	5	6	3	1	7	0	9	0	6	7	6
5	7	9	7	5	3	0	2	8	3	3	9	6	2	4	0
0	3	8	0	1	7	9	7	5	1	9	0	0	6	3	2
0	2	1	9	4	8	8	4	0	3	6	9	9	4	2	6
5	3	2	7	8	2	1	2	7	4	0	1	2	5	3	6
3	4	4	5	1	4	0	6	7	5	2	8	9	7	4	7
0	9	6	1	8	5	3	4	6	9	4	2	8	9	9	3
2	8	7	4	3	1	6	9	6	5	7	1	1	0	8	8
0	0	8	1	0	0	6	3	8	2	4	3	3	3	7	3

Q : Et maintenant ?

3	5	1	6	2	9	9	2	8	3	8	1	6	4	0	6
0	4	2	0	6	7	2	6	8	5	3	5	0	2	2	7
4	9	5	7	9	1	2	2	1	9	9	6	0	2	9	3
4	2	8	9	3	3	2	8	0	8	5	1	7	3	1	9
4	2	0	7	4	3	1	5	4	2	1	4	5	9	6	2
1	5	0	0	1	5	2	6	8	6	1	1	8	5	5	8
2	3	4	2	2	7	5	9	9	7	6	5	4	6	4	9
9	4	2	0	8	5	0	4	9	0	0	8	8	9	8	9
1	0	3	5	2	5	4	6	6	0	4	2	1	1	7	9
3	3	7	3	3	6	6	7	8	7	8	2	7	2	0	0
5	4	7	6	1	9	1	8	0	0	5	6	6	8	0	1
9	6	4	8	2	5	6	3	1	7	0	9	0	6	7	6
5	7	9	7	5	3	0	2	8	3	3	9	6	2	4	0
0	3	8	0	1	7	9	7	5	1	9	0	0	6	3	2
0	2	1	9	4	8	8	4	0	3	6	9	9	4	2	6
5	3	2	7	8	2	1	2	7	4	0	1	2	5	3	6
3	4	4	5	1	4	0	6	7	5	2	8	9	7	4	7
0	9	6	1	8	5	3	4	6	9	4	2	8	9	9	3
2	8	7	4	3	1	6	9	6	5	7	1	1	0	8	8
0	0	8	1	0	0	6	3	8	2	4	3	3	3	7	3

1. Introduction à la visualisation de données

1.2. Intérêts

Q : Quel est le pays le plus visité entre 2010 et 2014 ?

Visits per countries

	2010	2011	2012	2013	2014
U.K.	30	35	40	25	30
Belgium	10	15	20	20	15
France	35	40	20	20	25
Italy	10	10	15	10	20
Norway	20	25	25	35	45
Spain	5	15	10	15	20
Sweden	20	30	30	45	40
Germany	40	50	40	35	40
Finland	35	40	40	35	45
Danemark	5	5	15	20	20

Q : Et maintenant ?



1. Introduction à la visualisation de données

1.2. Intérêts

Problématique : L'ère du Big Data est en train de passer à la vitesse supérieure.
Comment donner sens aux nombreuses données générées chaque jour ?

La visualisation de données :

- Constitue un **gain de temps** conséquent dans la recherche, l'analyse et l'explication des données
- Permet un **accès aux données rapide, précis et compréhensible**
- Constitue un **outil de communication** puissant
- Facilite la **prise de décision**

*Des exemples de Data Visualisation :
<https://public.tableau.com/fr-fr/gallery/>*

1. Introduction à la visualisation de données

1.3. Dashboard

2 types d'indicateurs :

- **Indicateur opérationnel** : permet un **suivi de l'activité**. Est dit « préventif » car permet de détecter les écarts de performance, réagir et corriger les résultats à venir.
- **Indicateur de résultat** : permet une **mesure périodique de l'activité** pour établir des constats à des échéances fixes. Est dit « réactif » car permet de constater un résultat et d'en tirer des conclusions.

Un **dashboard** regroupe des indicateurs opérationnels.

Un **reporting** regroupe des indicateurs de résultats.

1. Introduction à la visualisation de données

1.3. Dashboard

Un **dashboard** :

- Traduit une **nécessité opérationnelle** : celle de transformer les analyses de données en langage courant dans le but de comprendre et décider.
- Est un ensemble d'indicateurs de performance (KPIs) qui permettent de **suivre une activité**, si possible en temps réel, afin de **réagir en conséquence** le plus rapidement possible.

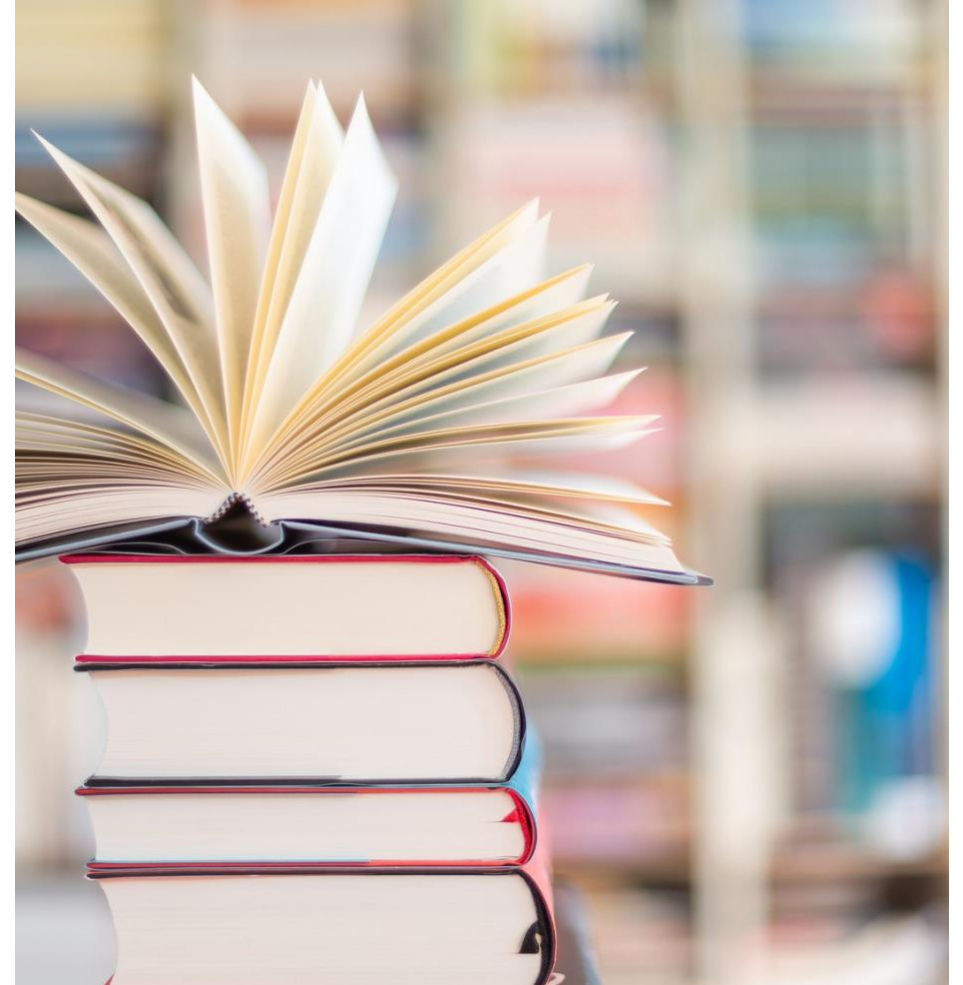


1. Introduction à la visualisation de données

1.4. Data Storytelling

- Une histoire est 22 fois mieux retenue qu'une statistique (d'après Jennifer Aaker, psychologue et professeur de marketing à Stanford)
- Exemples :
 - Au moyen âge, les théâtres de rue se chargeaient de compléter l'enseignement des églises à la population illettrée à travers des récits. Cette narration permet d'imager des concepts complexes en facilitant la transmission d'informations clés.
 - Les contes pour enfants suivent cette logique et permettent la transmission des valeurs morales à travers des histoires.

*L'histoire est un puissant **véhicule de l'information***



1. Introduction à la visualisation de données

1.4. Data Storytelling

La Data Storytelling consiste à :

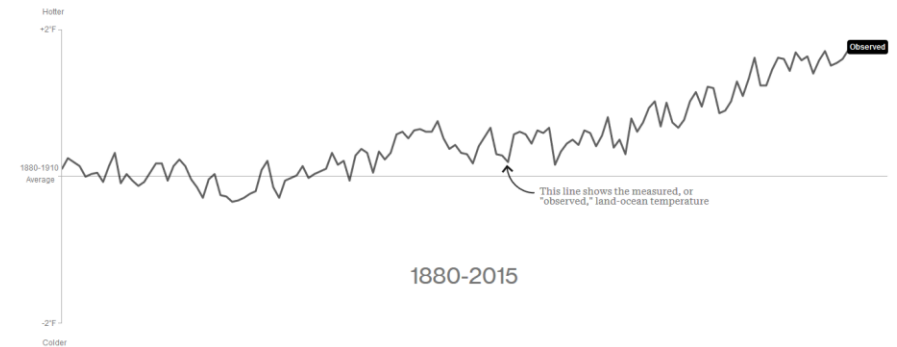
- Transformer des **résultats abstraits en histoires**
- Adapter l'information à faire passer en fonction de son **audience**
- **Hiérarchiser les informations** à présenter
- S'adresser aux **émotions** pour faire passer un message et faciliter la prise de décision

La Data Storytelling c'est raconter à partir de données choisies plutôt que de cumuler les graphiques et indicateurs

What's Really Warming the World?

By Eric Roston and Blacki Migliorzi | June 24, 2015

Skeptics of manmade climate change offer various natural causes to explain why the Earth has warmed 1.4 degrees Fahrenheit since 1880. But can these account for the planet's rising temperature? Scroll down to see how much different factors, both natural and industrial, contribute to global warming, based on findings from NASA's Goddard Institute for Space Studies.



Un exemple de Data Storytelling :
<https://www.bloomberg.com/graphics/2015-whats-warming-the-world/>



2.

Les types de représentation de données

2. Les types de représentation de données

Il existe **5 types de représentation** de données :

- Temporel
- Hiérarchique
- Réseau
- Multidimensionnel
- Géospatial

Il est important de **choisir le bon type** de représentation de données pour faciliter :

- L'interprétation de ce qui est observé
- La prise de décisions

Catalogue des graphiques de données :
<https://datavizcatalogue.com/index.html>



2. Les types de représentation de données

2.1. Temporel

Type de données :

Données généralement linéaires et unidimensionnelles qui évoluent au cours du temps

Information mise en avant :

Les tendances ou changements au fil du temps

Exemples :



Nuage de points



Graphique linéaire



Diagramme de zones polaires



Graphique chronologique

2. Les types de représentation de données

2.2. Hiérarchique

Type de données :

Données provenant souvent d'un même point d'origine

Information mise en avant :

La hiérarchie / l'organisation de groupes de données au sein de groupes plus grands

Exemples :

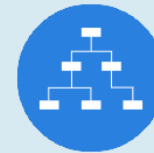


Diagramme
en arbre



Treemap



Diagramme
sunburst

2. Les types de représentation de données

2.3. Réseau

Type de données :

Données fortement connectées entre elles

Information mise en avant :

Les relations entre les données au sein d'un réseau ou la corrélation entre deux variables ou plus

Exemples :



Diagramme
en radar



Carte
thermique



Diagramme
réseau

2. Les types de représentation de données

2.4. Multidimensionnel

Type de données :

Données à plusieurs dimensions (couches et jeux de données simultanées)

Information mise en avant :

Visualisation de données 3D

Plus visuel et accrocheur, permet de décomposer les nombreuses données en points clés

Exemples :



Nuage de points



Diagramme circulaire



Histogramme



Histogramme empilé



Diagramme Venn

2. Les types de représentation de données

2.5. Géospatial

Type de données :

Données se rapportant à des emplacements physiques réels

Information mise en avant :

Visualisation de données sur des zones géographiques

Souvent utilisé pour les ventes et acquisitions, les campagnes politiques ou la pénétration du marché dans les sociétés multinationales

Exemples :



Carte de points



Carte des flux



Carte
choroplèthe

3.

Quelques acteurs de la Data Visualisation

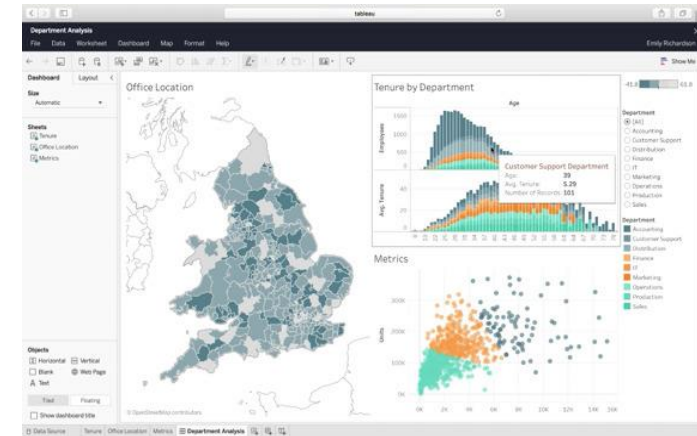


3. Quelques acteurs de la Data Visualisation

3.1. Tableau

Présentation :

- Plateforme d'analyse as a service
- Tableau server et online
- Différents profils (créateur, explorateur ou spectateur)
- Création de tableaux, graph, infographies à partir d'un large panel de sources de données (SQL, hadoop, excel, etc)



Lien officiel : <https://www.tableau.com/fr-fr>

Avantages :

- Interface intuitive
- Principe de "glisser-déposer"
- Graphiques mis à jour en temps réel et partageables (de manière sécurisée)
- Disponible en français
- Utilisation immédiate et sans contrainte technique

Inconvénients :

- À partir de 70\$/mois par utilisateur
- Manque d'un outil pour transformer les données en amont

3. Quelques acteurs de la Data Visualisation

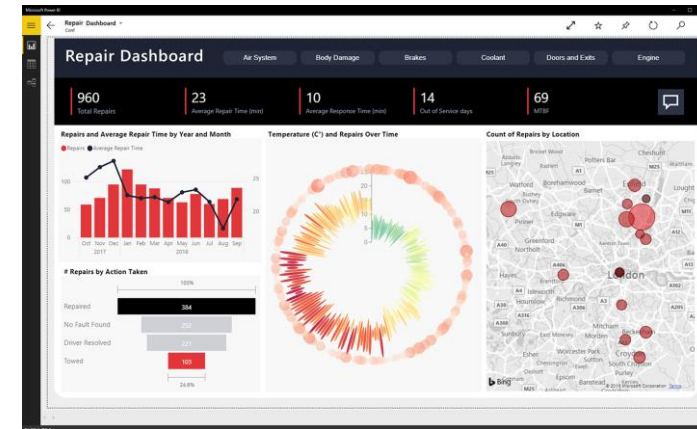
3.2. Power BI

Présentation :

- Solution BI cloud de Microsoft
- Agréger, analyser et visualiser les données provenant de multiples sources
- Tableaux de bords préfabriqués personnalisables, pas besoin de compétence technique
- Cloud, logiciel, mobile app

Avantages :

- Outil intuitif permettant de dégager des infos exploitables
- Meilleur rapport qualité/prix du marché
- Simple d'utilisation
- Nombreuses fonctionnalités ajoutées régulièrement
- Nombreuses sources de données



Lien officiel : <https://powerbi.microsoft.com/fr-fr/>

Inconvénients :

- Accès limité aux utilisateurs d'un même nom de domaine
- Capacité de **stockage limitée**
- Souvent utilisé en complément d'autres logiciels
- **Coût plus élevé** pour le logiciel bureau

3. Quelques acteurs de la Data Visualisation

3.3. Grafana

Présentation :

- Outil Open Source de visualisation utilisé en complément de BDD
- Le + souvent avec Graphite et Influxdb
- Interface en remplacement de Graphite Web (+ de fonctionnalités, meilleure qualité)

Avantages :

- Pour l'analyse de métriques
- Graphes versatiles
- Personnalisable



Lien officiel : <https://grafana.com/grafana/>

3. Quelques acteurs de la Data Visualisation

3.4. Kibana

Présentation :

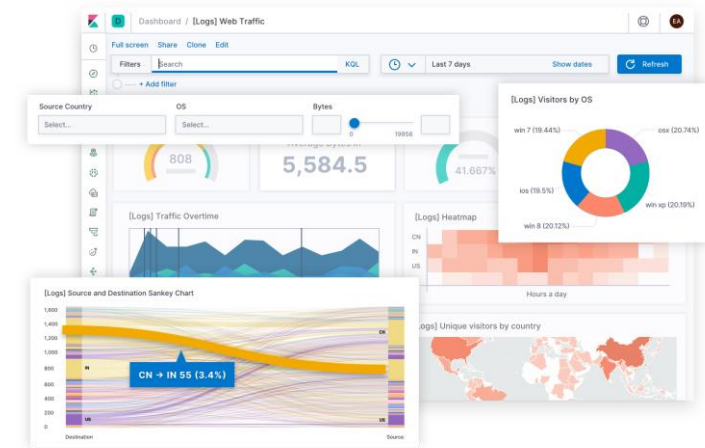
- 'K' dans ELK Stack plateforme Open Source la plus populaire d'analyse de log
- Outil pour explorer, visualiser et créer des tableaux de bord pour les données logs stockées dans les clusters ElasticSearch
- Feature principale : data querying and analysis
- Recherche dans les données indexées dans ElasticSearch d'événements spécifiques
- Visualisation des résultats à travers des tableaux, graphiques, maps géo etc

Avantages :

- Pour l'analyse de logs
- Graphes versatiles
- Personnalisable

Inconvénients :

- Limité à ElasticSearch



Lien officiel : <https://www.elastic.co/fr/products/kibana>

3. Quelques acteurs de la Data Visualisation

3.5. Data Studio

Présentation :

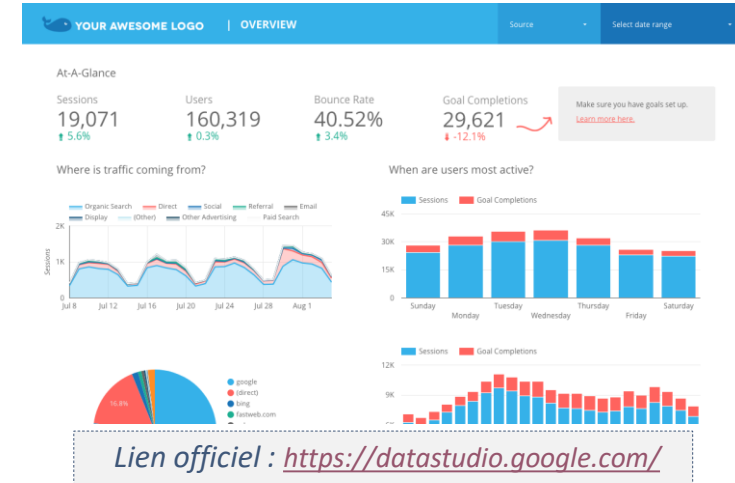
- Outil proposé par Google intégré à la suite Google analytics 360
- Entièrement gratuit
- Créer des reporting à partir des données issues de la suite google
- Connection possible avec d'autres sources externes
- Création de rapport from scratch ou utilisation de templates

Avantages :

- Ergonomie soignée
- Interface en ligne intuitive
- Interactif et collaboratif
- Personnalisable
- MAJ rapide

Inconvénients :

- Propriété de données vague
- Manque de connecteurs
- Fonctionnalités élémentaires
- Connexion mySQL non sécurisée
- Exportation difficile (pas prévu pour)

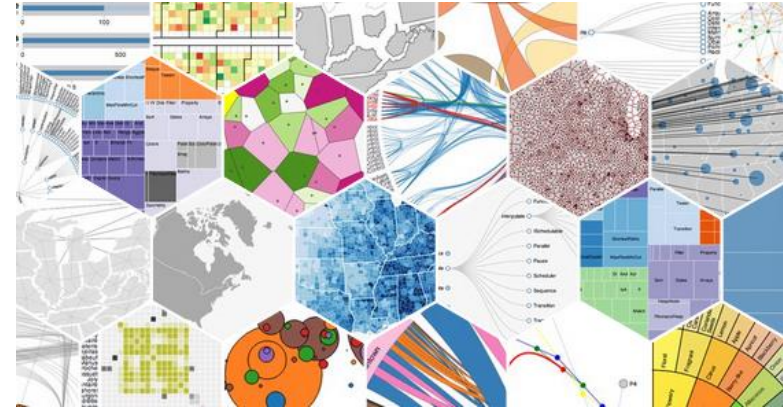


3. Quelques acteurs de la Data Visualisation

3.6. D3.js

Présentation :

- Développé par un data scientist du NY Times
- Bibliothèque open source de graphiques JavaScript
- D3 = data-driven-documents
- Beaucoup de tutos
- Gestion de gros volumes de données
- Création de visualisations de données interactives et dynamiques directement en SVG à l'intérieur du DOM



Lien officiel : <https://d3js.org/>

Avantages :

- Indépendant
- Compatibilité des navigateurs
- Compatibles nombreux frameworks
- Facilite l'automatisation de la création de représentation de graphiques
- Interactive, rapide, flexible

Inconvénients :

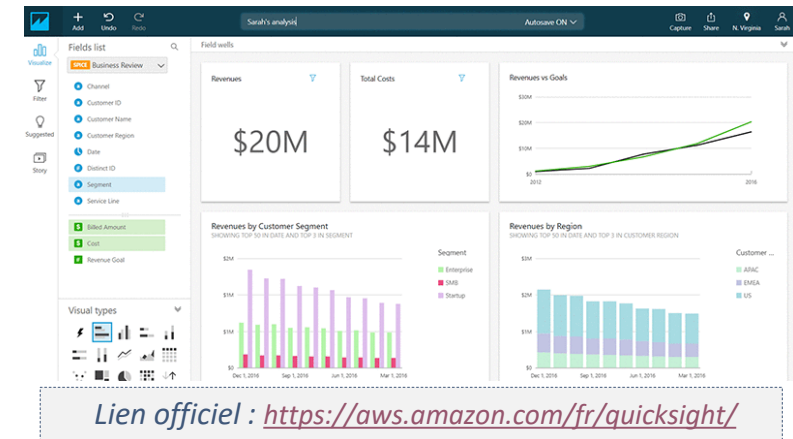
- Ne convient pas pour des visualisations plus poussées

3. Quelques acteurs de la Data Visualisation

3.7. QuickSight

Présentation :

- Service de BI sur le cloud d'Amazon Web Services
- Fondé sur sa propre BDD dite "SPICE" (Super-Fast, Parallel, In-Memory Calculation Engine)
- In memory, avec son propre système de gestion de fichiers (possibilité de paralléliser les requêtes)



Avantages :

- Moins cher que ses concurrents traditionnels (9\$/mois/utilisateur)
- Super rapide pour l'exécution

Inconvénients :

- Long pour rapatrier les données
- Pas pour des besoins "en temps réel"

3. Quelques acteurs de la Data Visualisation

	Tableau Online	Power BI	Data Studio	Kibana	Grafana	D3JS	QuickSight
Fournisseur	Tableau	Microsoft	Google	Elastic	Grafana Labs	D3JS	Amazon Web Services
Capacité de données	Pas de limitation	10 Go / utilisateur	Pas de limitation	Pas de limitation	Pas de limitation	Pas de limitation	10Go / utilisateur
Sources de données	Cloudera Hadoop Oracle AWS Redshift Google BigQuery Microsoft SQL Server Salesforce Google Analytics	60 connecteurs : Access Excel Dynamics SQL Server Azure SQL Database Hadoop Salesforce Google Analytics	Solutions Google: Google Sheets BigQuery AdWords Analytics Fichiers .csv	ElasticSearch	30 connecteurs : Graphite InfluxDB ElasticSearch Prometheus AWS Cloud Watch	Fichiers JSON Fichiers CSV Fichiers GeoJSON	Produits Amazon Apache Spark MariaDB Microsoft SQL Server MySQL PostgreSQL Presto Snowflake Teradata
Mobilité	Cloud On premise Version mobile	Version mobile	Rapports graphiques en responsive design	Exports pdf et csv Système d'alerte par mail (fréquence et contenu personnalisables)	Export JSON Lien partageable Snapshot interactif	Différents navigateurs web, exports pages HTML	Exports CSV Envoi par mail Publication de rapports
Sécurité	Authentification unique SAML 2.0 Synchro Active Directory	Azure Active Directory	Standards Google : échanges sécurisés SSL conformité ISO 27001	Chiffrement TLS Contrôle d'accès utilisateurs	NC	NC	Microsoft Active Directory Contrôle d'accès basé sur les rôles Authentification unique Prise en charge possible de différentes normes
Version gratuite	Version d'essai de 14j	Limitation à 1Go/utilisateur Actualisation 1x/jour	Gratuit	Open source	Open source	Open source	1 an gratuit avec un compte AWS
Tarifs	500\$/an/utilisateur	9\$/mois/utilisateur	Gratuit	Open source	Open source	Open source	A partir de 9\$/mois/utilisateur

4.

Démonstration





Merci pour votre attention