

DataBFC

Ouvrir et gérer les données de la recherche
en Bourgogne Franche-Comté

BIG DATA

Problématique Big Data : définition et enjeux

Christophe Cruz | Besançon | 13 novembre 2017

Big Data - 2008



WELCOME TO THE PETACENTRE

What does it take to store bytes by the tens of thousands of trillions? **Cory Doctorow** meets the people and machines for which it's all in a day's work.

Ten seconds after I stepped into the roar of the data centre at the UK Wellcome Trust Sanger Institute, in rural Cambridgeshire, my video camera croaked: **CARD FULL**. Impossible. That morning, I'd tossed a handful of thumbnail-sized 32-GB memory cards into my pocket, each one good for a couple of hours' worth of high-definition video. Yet this one had filled up in seconds.

I fumbled with my camera while Phil Butcher, the Sanger Institute's head of information technology (IT), politely waited, grinning in the shower of cold air washing down from the air conditioning. It took only a couple of embar-



watt that you put into retrieving data and calculating with them comes out in heat, whether it be on a desktop or in a data centre; in the United States, the energy used by computers has more than doubled since 2000. Once you're conducting petacalculations on petabytes, you're into petaheat territory. Two floors of the Sanger data centre are devoted to cooling. The top one houses the current cooling system. The one below sits waiting for the day that the centre needs to double its cooling capacity. Both are sheathed in dramatic blue glass; the scientists call the building the Ice Cube.

pétaoctet \pe.ta.ɔk.tɛ\ masculin

(Métrologie) (Informatique) Unité de mesure de quantité d'information numérique, valant 10^{15} octets, et dont le symbole est Po.

Définition (1)

Les « mégadonnées » ou « données massives »

Vocabulaire de l'informatique

[Masquer le panneau de navigation](#)

[Imprimer](#)

Navigation

AVIS DIVERS du

▶ [Version initiale](#)

JORF n°0193 du 22 août 2014 page 13972
texte n° 89

Vocabulaire de l'informatique

NOR: CTNX1419323X
ELI: Non disponible

mégadonnées, n.f.pl.

Domaine : Informatique.

Définition : Données structurées ou non dont le très grand volume requiert des outils d'analyse adaptés.

Note :

On trouve aussi l'expression « données massives ».

Équivalent étranger : big data.

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000029388087&dateTexte=&categorieLien=id>



Définition (2)

C'est la capacité de traiter de gros volumes d'information avec des moyens informatiques de plus en plus standards.

La révolution Big Data, Les données au cœur de la transformation de l'entreprise,
J.-C. Cointot, Y. Eychenne, Dunod Ed.

Changement de paradigme (1)

Facteurs de son émergence



➤ Les **réseaux sociaux** (consommateur passif/actif)

➤ La **mobilités** (développement de la téléphonie)

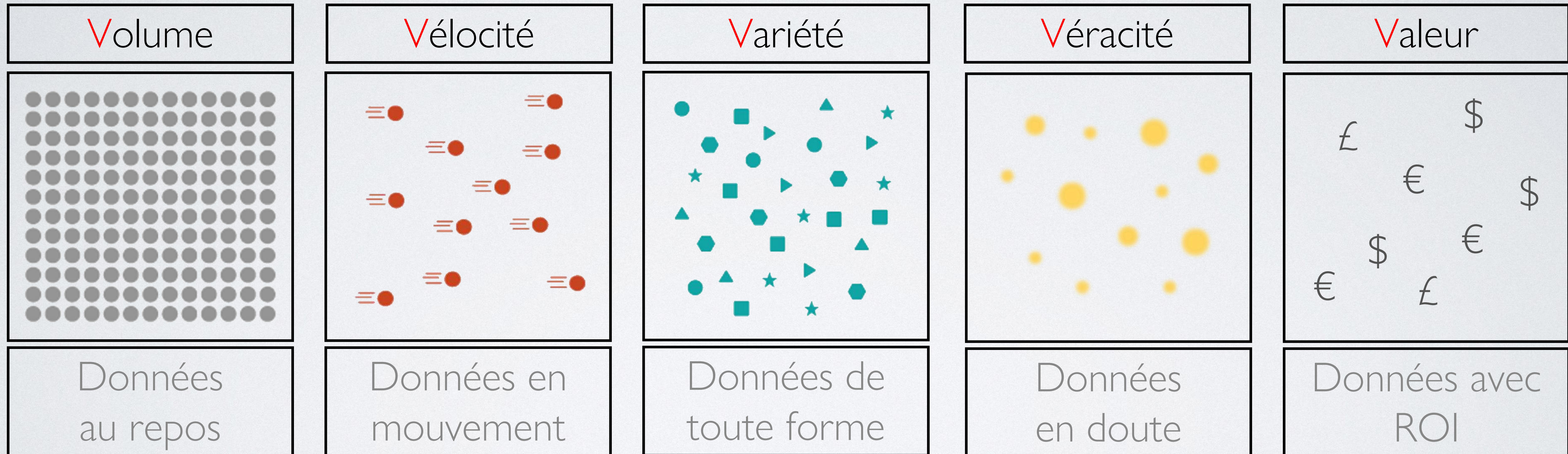
➤ L'**ubiquité numérique** (le cloud, échange et partage, public/privé)



BIG DATA

Le Big Data – une définition ?

du 3V au 5V ou plus ... Volume + variété + vitesse + véracité + valeur ...



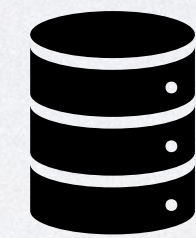
BIG DATA

Des métiers ...

Infrastructures, Stockage et Analyse de données Massives



Data Scientist



Data Analyst



Data Architect



Data Curator

BIG DATA

Des technologies ...

<p>Vertical Apps</p>	<p>Ad/Media Apps</p>	<p>Business Intelligence</p>	<p>Analytics and Visualization</p>
<p>Log Data Apps</p>			
<p>Data As A Service</p>			
<p>Analytics Infrastructure</p>	<p>Operational Infrastructure</p>	<p>Infrastructure As A Service</p>	<p>Structured Databases</p>
<p>Technologies</p>			



Big Data Landscape (Version 2.0)



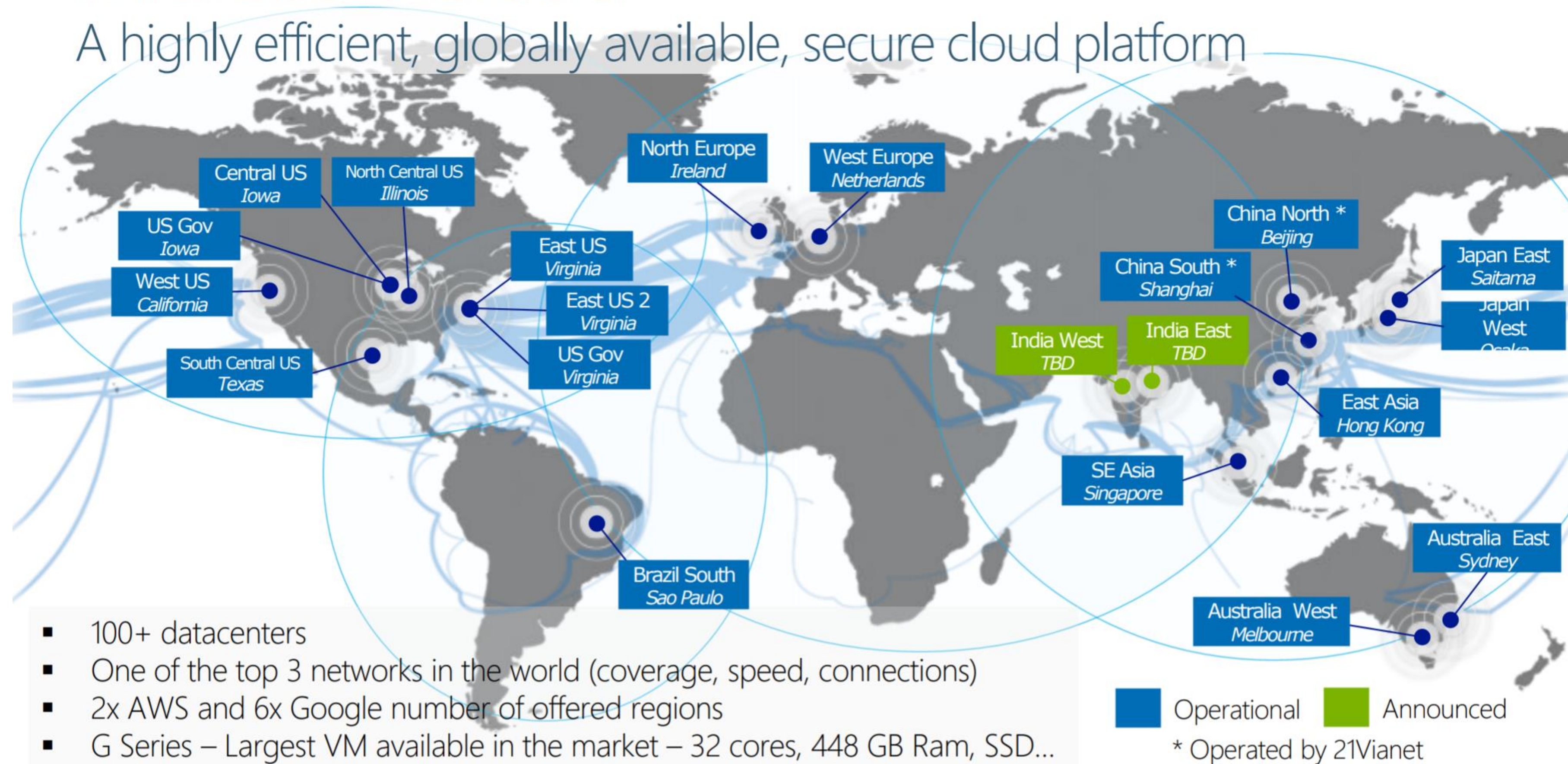
Big Data Landscape 2016 (Version 2.0)



Le cloud (informatique en nuage) ...

Microsoft Cloud

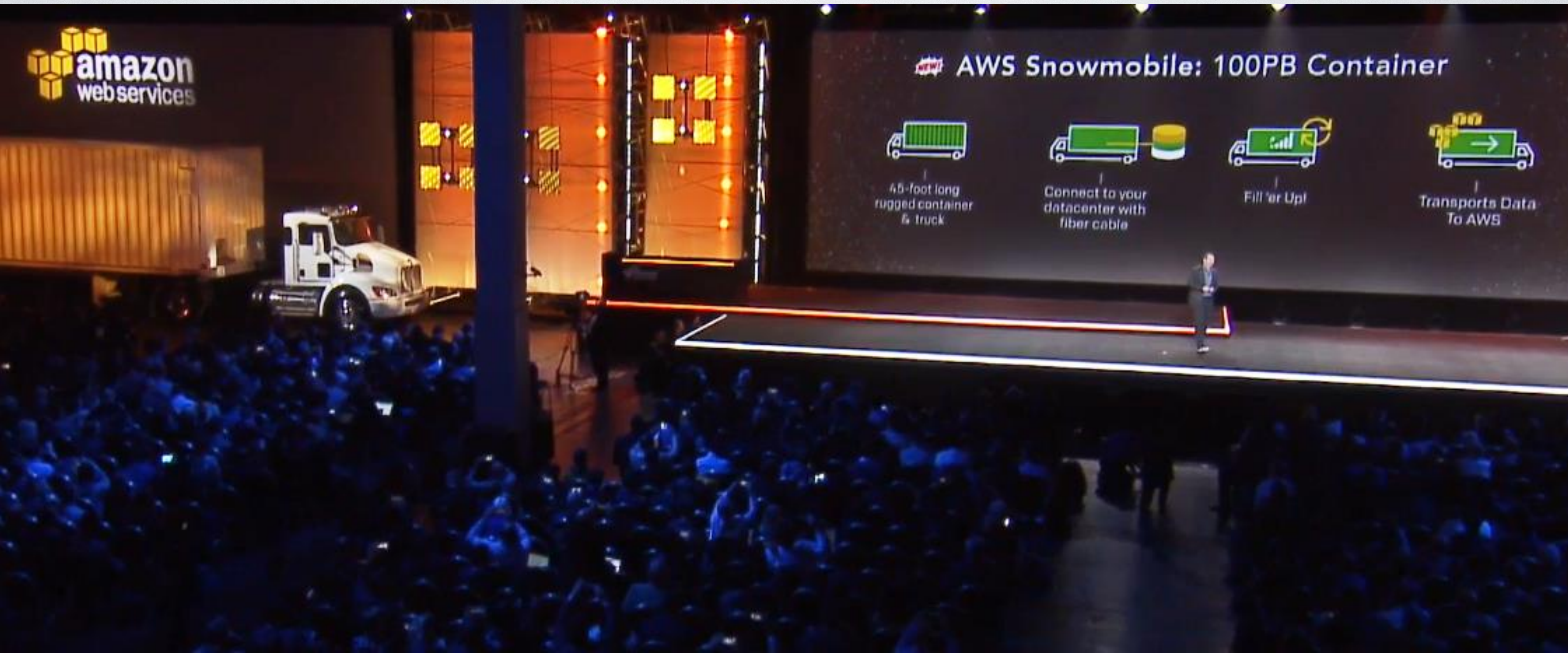
A highly efficient, globally available, secure cloud platform



Google (Council Bluffs, Iowa)

An overhead view of the interior of a Google data center in Council Bluffs, Iowa, where the search giant houses one of the heaviest concentrations of its Internet infrastructure.

La clés USB pour Data center



amazon
web services

NEW! AWS Snowmobile: 100PB Container

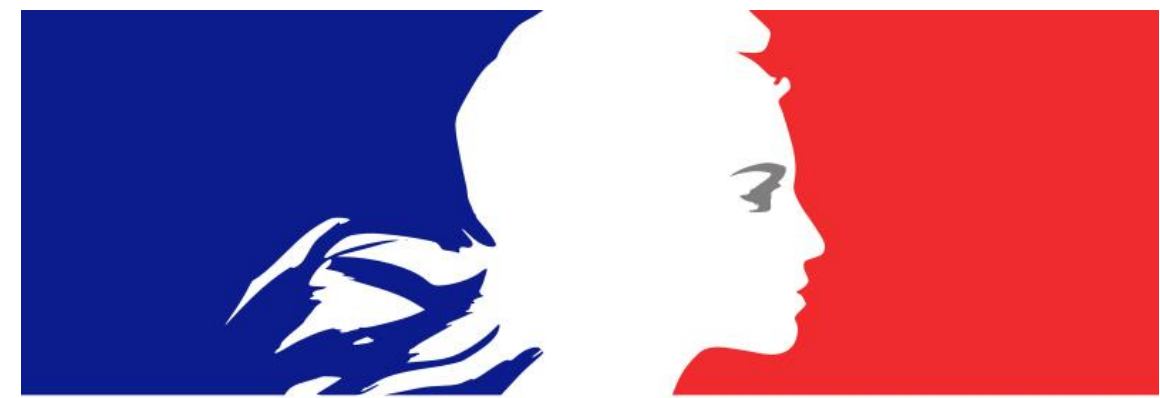

45-foot long
rugged container
& truck


Connect to your
datacenter with
fiber cable


Fill'er Up!


Transports Data
To AWS

Enjeux pour l'ESR



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère de l'Enseignement supérieur,
de la Recherche et de l'Innovation

Enjeux pour l'enseignement



Des évolutions ...



d'après M. D. Pistilli, EDUCAUSE 2015

Enjeux pour l'enseignement



Des évolutionsmais certaines constantes !

d'après M. D. Pistilli, EDUCAUSE 2015

Opportunité

Learning Analytics

Définition (Long and Siemens 2011)

« Mesure, collecte, analyse et traitement des données relatives aux apprenants et à leurs environnements dans le but de comprendre et optimiser l'apprentissage et les environnements dans lesquels il se produit. »



Mieux comprendre les apprenants et optimiser l'apprentissage.

BIG DATA <http://projet-cela.fr>

Les données et le numérique au cœur des sciences participatives

Le rapport de février 2016



- ① **Etat des lieux** et méthodes
- ② **Bonnes pratiques** pour les porteurs de projet
- ③ **Recommandations** aux institutions
- ④ **Annexes**

Rapport de François Houllier, PDG de l'Inra et président de l'Alliance nationale de recherche pour l'environnement, remis le 4 février 2016 à Najat Vallaud-Belkacem, ministre de l'Éducation nationale, de l'enseignement supérieur et de la recherche, et à Thierry Mandon, secrétaire d'État chargé de l'enseignement supérieur et de la recherche.

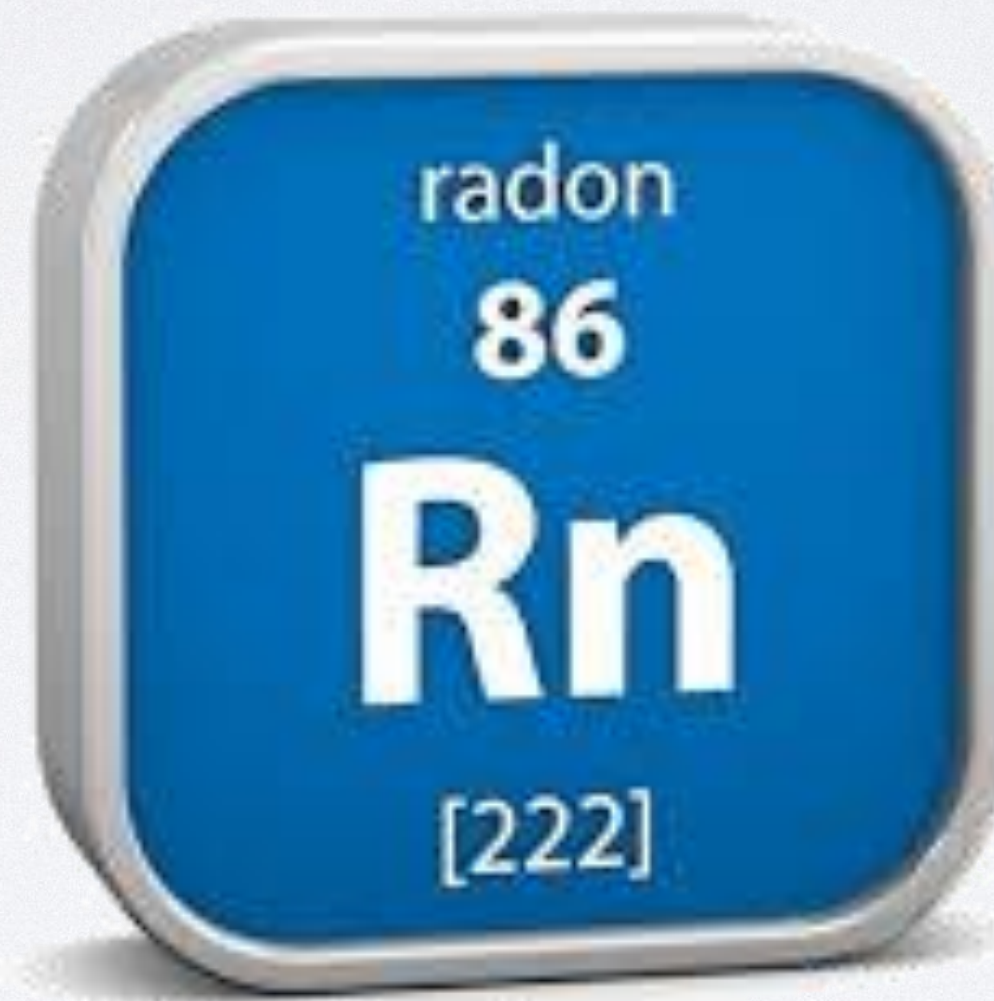
Enjeux pour les données

Souveraineté des données, souveraineté numérique

La souveraineté des données ou souveraineté numérique désigne « l'application des principes de **souveraineté** au domaine des technologies de l'information et de la communication (TIC), c'est-à-dire à l'informatique et aux télécommunications ».

Big data, opportunité ou danger ?

Quand le Radon était swag !



Big data, opportunité ou danger ?

DIGESTIVE RECONSTITUANTE FORTIFIANTE

EAU MINÉRALE
NATURELLE GAZEUSE
RADIO-ACTIVE
DE
BUSSANG

ÉTABLISSEMENT FONDÉ EN 1677
SOURCE SALMADE (VOSGES)
Indiquée dans toutes les convalescences et anémies
DÉCLARÉE D'INTÉRÊT PUBLIC
Autorisée par décret du 7 avril 1866

FERRUGINEUSE - ARSENIÉE - ALCALINE - MANGANESÉE - PHOSPHATÉE - LITHINÉE
CONTRE : ANÉMIE, MALADIES D'ESTOMAC, DES REINS ET DE LA VESSIE, GOUTTE, DIABÈTE
NEURASTHÉNIE, DÉBILITÉ, LYMPHATISME

DERUFFE PUBLICITE

CRÈME
COLD-CREAM
POUR PEAU SÈCHE

CRÈME
MAIRESQUE
FOND DE TEINT

POUDRE
SAVON

LAIT DE TOILETTE
DEMAQUILLANT

ROUGE À LEVRES
DENTIFRICE

THO-RADIA

MÉTHODE
SCIENTIFIQUE
DE
BEAUTÉ

EN PHARMACIE EXCLUSIVEMENT



Big data, opportunité ou danger ?

DIGESTIVE RECONSTITUANTE FORTIFIANTE

EAU MINÉRALE
NATURELLE GAZEUSE
RADIO-ACTIVE
DE
BUSSANG

ÉTABLISSEMENT FONDÉ EN 1677
SOURCE SALMADE (VOSGES)
Indiquée dans toutes les convalescences et anémies
DÉCLARÉE D'INTÉRÊT PUBLIC
Autorisée par décret du 7 avril 1866

FERRUGINEUSE - ARSENIÉE - ALCALINE - MANGANESÉE - PHOSPHATÉE - LITHINÉE
CONTRE : ANÉMIE, MALADIES D'ESTOMAC, DES REINS ET DE LA VESSIE, GOUTTE, DIABÈTE
NEURASTHÉNIE, DÉBILITÉ, LYMPHATISME

DERUFFE PUBLICITE

CRÈME
COLD-CREAM
POUR PEAU SÈCHE

CRÈME
MAIRESQUE
POUR DE TEINT

POUDRE
SAVON

LAIT DE TOILETTE
DEMAQUILLANT
ROUGE À LEVRES
DENTIFRICE

THO-RADIA

MÉTHODE
SCIENTIFIQUE
DE
BEAUTÉ

EN PHARMACIE EXCLUSIVEMENT

IRADIA

**SOUS-VÊTEMENTS
RADIOACTIFS**
DU DOCTEUR
BAURAY

IMP. DE LA MÉDITERRANÉE

Big data, opportunité ou danger ?

DIGESTIVE RECONSTITUANTE FORTIFIANTE

EAU MINÉRALE NATURELLE GAZEUSE RADIO-ACTIVE DE **BUSSANG**

ÉTABLISSEMENT FONDÉ EN 1677
SOURCE SALMADE (VOSGES)
Indiquée dans toutes les convalescences et anémies
DÉCLARÉE D'INTÉRÊT PUBLIC
Autorisée par décret du 7 avril 1866

FERRUGINEUSE - ARSENIÉE - ALCALINE - MANGANESÉE - PHOSPHATÉE - LITHINÉE
CONTRE : ANÉMIE, MALADIES D'ESTOMAC, DES REINS ET DE LA VESSIE, GOUTTE, DIABÈTE,
NEURASTHÉNIE, DÉBILITÉ, LYMPHATISME

DERUFFE PUBLICITE



CRÈME
COLD-CREAM
POUR PEAU SÈCHE
CRÈME
MAINTIENSQUE
FOND DE TEINT
POUDRE
SAVON
LAIT DE TOILETTE
DEMAQUILLANT
ROUGE À LEVRES
DENTIFRICE

THO-RADIA

MÉTHODE
SCIENTIFIQUE
DE
BEAUTÉ

EN PHARMACIE EXCLUSIVEMENT



IRADIA

SOUS-VÊTEMENTS
RADIOACTIFS
DU DOCTEUR
BAURAY

IMP. DE LA MÉDITERRANÉE



GILBERT nuclear physics
No. U-238

ATOMIC ENERGY LAB

PERFORMS OVER 150 EXCITING EXPERIMENTS!
FOR THE JUNIOR SCIENTIST

- MOST MODERN SCIENTIFIC SET EVER CREATED!
- SEE PATHS OF ALPHA PARTICLES SPEEDING AT 13,300 MILES PER SECOND!
- WATCH ACTUAL ATOMIC DISINTEGRATION—RIGHT BEFORE YOUR EYES!
- PROSPECT FOR URANIUM WITH GEIGER-MULLER COUNTER!

Exciting! Safe! ANOTHER GILBERT HALL OF SCIENCE PRODUCT

GILBERT ATOMIC ENERGY LAB DEIONIZER

GILBERT CLOUD CHAMBER

ELECTROSCOPE

DataBFC

Ouvrir et gérer les données de la recherche
en Bourgogne Franche-Comté

BIG DATA

Questions ?