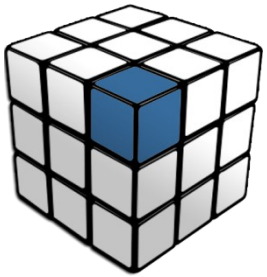


hurence

Big Data – get its magical power

Laurence.Hubert@hurence.com, CEO & CTO
http://www.hurence.com

Hurence : le 'pure player' Big Data

Une équipe de spécialistes du Big Data

Une équipe d'architectes, de data scientists, de développeurs séniors passionnés par le Big Data.
Une équipe innovante qui participe à la plupart des initiatives Big Data en France (projets d'investissements dans le cadre du Grand Emprunt).

Une approche indépendante

Une stratégie d'indépendance vis-à-vis des fournisseurs de solutions Big Data.
Une volonté de promouvoir l'intégralité des offres Big Data pertinentes pour nos clients et d'entretenir des relations privilégiées avec les plus grands acteurs (IBM, EMC, Oracle, Microsoft, HP, Dell, SAS, Teradata, Cloudera, Hortonworks..)

Une expérience importante

Plus de 30 clients « Big Data » en France.
La première société à se positionner sur une offre pure Big Data, autour de Hadoop et de son écosystème, en France.

Consulting & Services

ARCHITECTURE BIG DATA

Une offre de conseil en sélection d'outils et architectures Big Data, sur tout type de socle.

INSTALLATION ET GESTION

Une offre de dimensionnement, provisionnement, installation et gestion d'infrastructures matérielles et logicielles Big Data.

Technologies

LOGICIEL BIG DATA

Une offre logicielle basée sur l'écosystème Hadoop, dont plusieurs produits natifs Hadoop & HBase et des composants d'extraction Big Data.

SERVEUR D'ENRICHISSEMENT

Une offre logicielle dédiée analyse de logs dont un serveur d'enrichissement de données de logs.

Formation

FORMATIONS BIG DATA

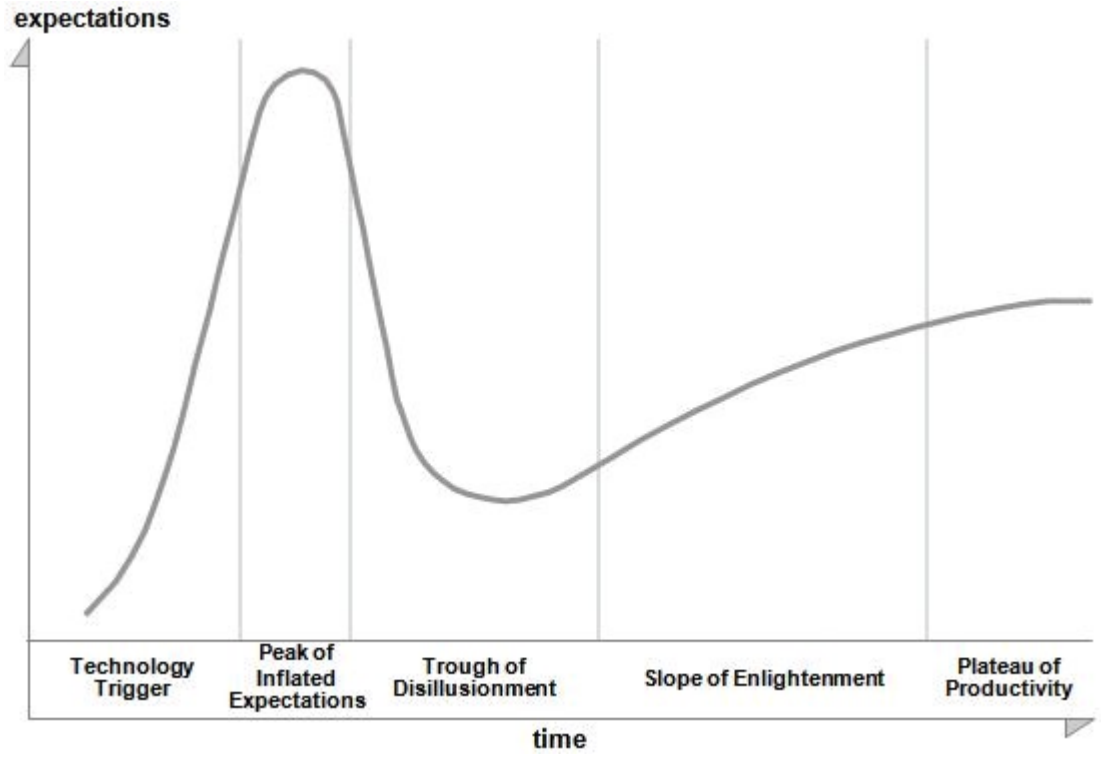
Une gamme de formation pour tous les profils et toutes les technologies Big Data open source.

Service

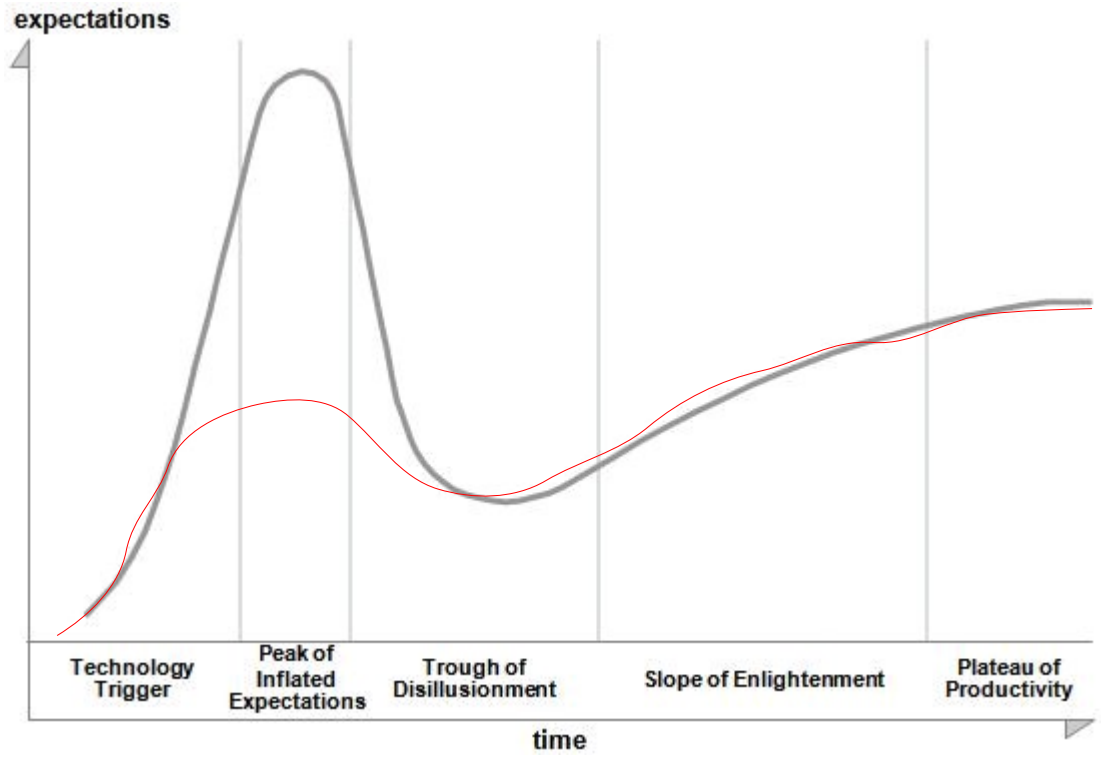
EXPRESS DATA SERVICE

Une offre de traitement de données sur **nos** clusters Hadoop.

Gartner Hype Cycle



French curve



Big Data challenges and opportunities

- The Big Data « me too » phenomenon
- The Big Data «NIH»
- Big Data is data !
- Big Data is hard !
- Big is Big !
- Big Data is sometimes poor data !
- Big Data is Hadoop but not only Hadoop
- Big Data is not just data... it can be beautiful
- Big Data is … ambition

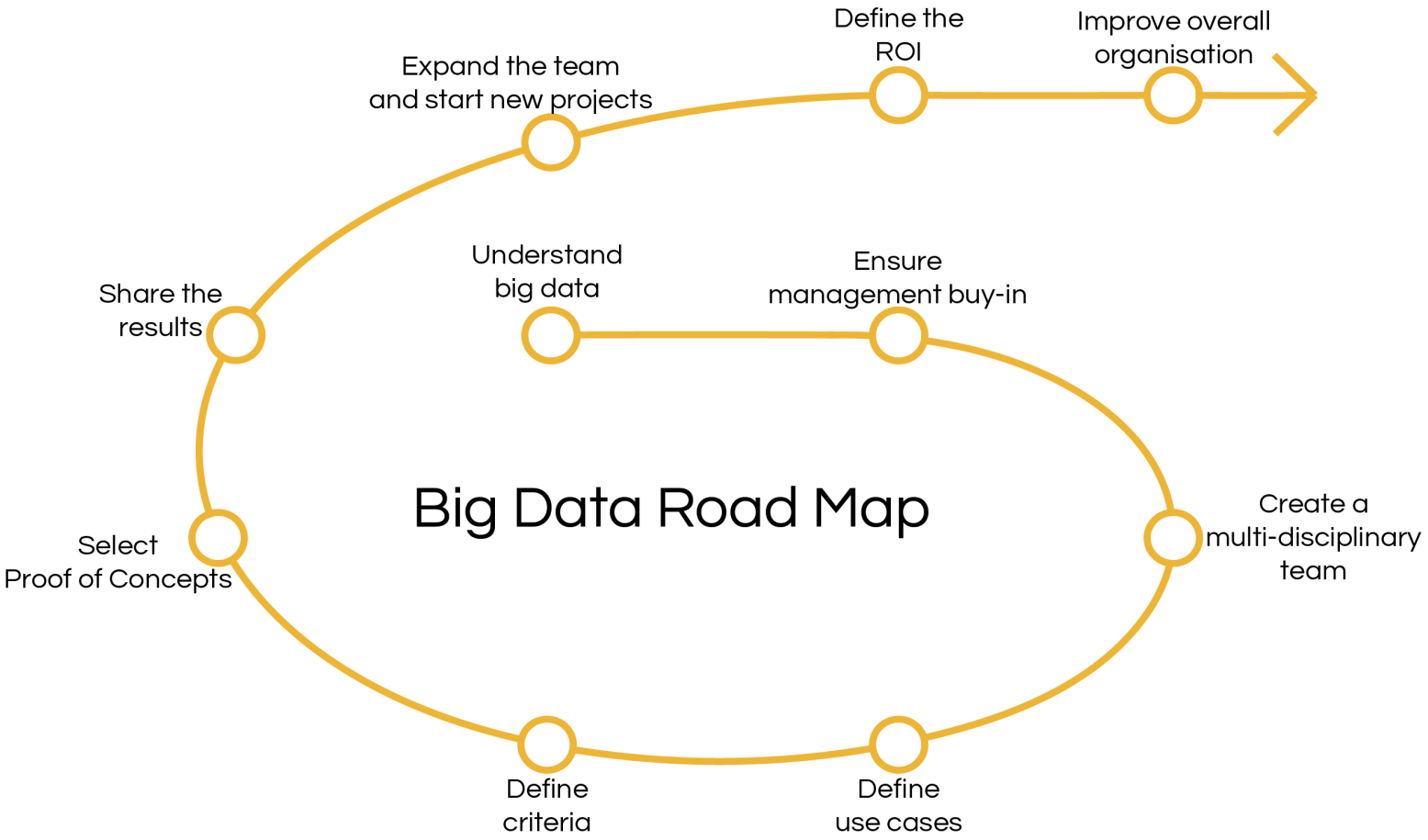
The Big Data « me too » phenomenon

- Vendors and consultants tend to use the same « Big Data » technologies and use cases... if not this is research or free POCs !
  - Customer Relationship Management through analysis of social networks, Advertising and campaign management, predictive churn etc.
  - Security and Fraud Management

Big Data cloning :)

- Managers engage in Big Data because they think they have to...
  - First consulting action is to train them
  - Second consulting action is to help them build an ambitious and innovative Big Data plan with respect to « their » business!

A Big Data Roadmap...



http://www.bigdata-startups.com/the-big-data-roadmap/

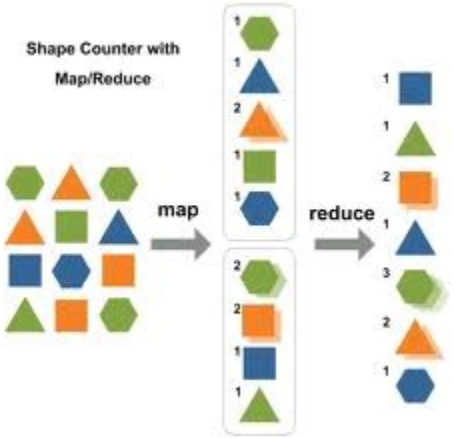
The Big Data «NIH»

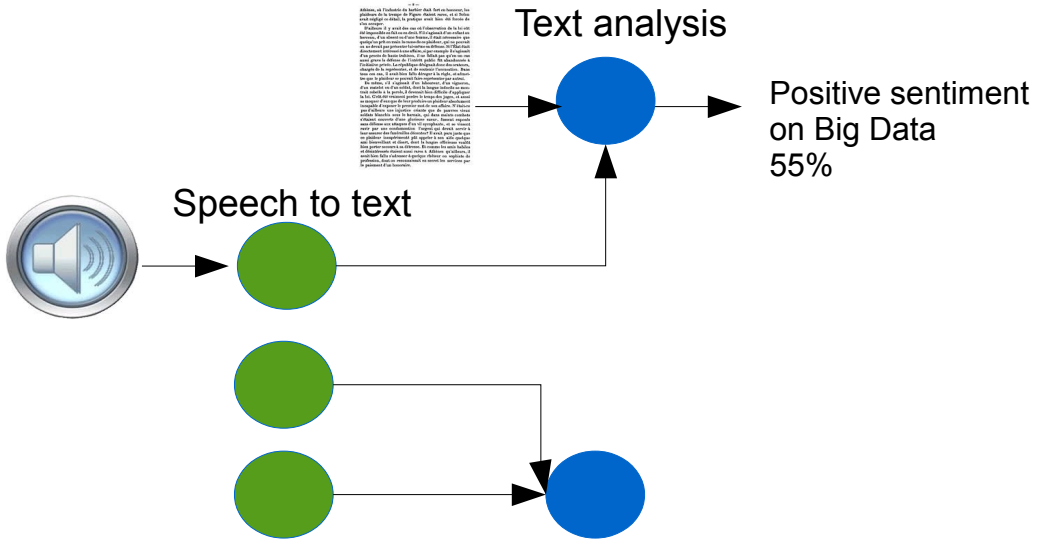
- Big Data was marketed as a « rare resource » which it is
  - trend : a lot of French managers are trying to turn their teams into « Big Data specialists »
  - but... the maturity of Big Data tools does not yet allow this
  - and... not sure it will ever allow this...
- Big Data is « private » data sometimes => analysis is done in-house
- Not every engineer can turn into a Big Data scientist : on 100 engineers we train, less than 20 have the « Big Data spirit » !

- Siloed data, confidential data
  - Sometimes HUGE political and managerial problems to access data
- Data is power !
- Data is money !

Big Data is hard

- Thinking parallel is inherently hard
  - **Map Reduce** is « fairly » easy but current implementation in Hadoop has limitations (latency, re-entrance, long-life jobs)
  - **MPI** (Message Passing Interface) in other words **stream processing** is much more difficult to setup and program (IBM streams, Storms)





Big Data is hard

- Advanced sentiment analysis or text mining or web crawling
  - Not just counting positive or negative words !
  - Involves linguistic knowledge and resources (domain specific)
  - Text Analytics is a domain for specialists

Big Data Big Brother ! Super!

Athènes, où l'industrie du barbier était fort en honneur, les plaideurs de la trempe de Figure étaient rares, et si Solon avait négligé ce détail, la pratique avait bien été forcée de s'en occuper.

D'ailleurs il y avait des cas où l'observation de la loi eût été impossible en fait ou en droit. S'il s'agissait d'un enfant ou berceau, d'un absent ou d'une femme, il était nécessaire que quelqu'un prît en main la cause de ce plaideur, qui ne pouvait ou ne devait pas présenter lui-même sa défense. Si l'Etat était directement intéressé à une affaire, si par exemple il s'agissait d'un procès de haute trahison, il ne fallait pas qu'en un cas aussi grave la défense de l'intérêt public fût abandonnée à l'initiative privée. La république désignait donc des orateurs, chargés de la représenter, et de soutenir l'accusation. Dans tous ces cas, il avait bien fallu déroger à la règle, et admettre que le plaideur ne pourrait faire représenter par autrui.

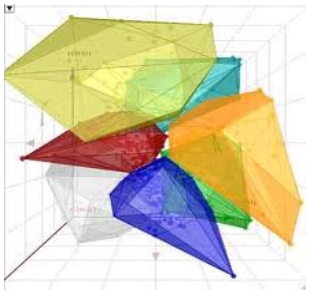
De même, s'il s'agissait d'un laboureur, d'un vigneron, d'un matelot ou d'un soldat, dont la langue inhabile se montrait rebelle à la parole, il devenait bien difficile d'appliquer la loi. C'eût été vraiment perdre le temps des juges, et aussi se moquer d'eux que de leur produire un plaideur absolument incapable d'exposer le premier mot de son affaire. N'était-ce pas d'ailleurs une injustice criante que de pauvres vieux soldats blanchis sous le harnais, qui dans maints combats s'étaient couverts d'une glorieuse sueur, fussent exposés sans défense aux attaques d'un vil sycophante, et se vissent ravir par une condamnation l'argent qui devait servir à leur assurer des funérailles décentes? Il avait paru juste que ce plaideur inexpérimenté pût appeler à son aide quelque ami bienveillant et disert, dont la langue officieuse voulût bien porter secours à sa détresse. Et comme les amis habiles et désintéressés étaient aussi rares à Athènes qu'ailleurs, il avait bien fallu s'adresser à quelque rhéteur ou sophiste de profession, dont on reconnaissait en secret les services par le paiement d'un honoraire.

Mix of ML and traditional tools

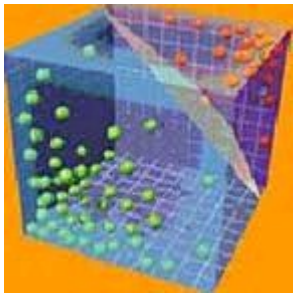
- Parsing / Semantic analysis
- Part Of Speech Tagging
- Dictionary Lookup
- Sentence / Paragraph Detection
- Tokenizer (split the text to words)

Big Data is hard

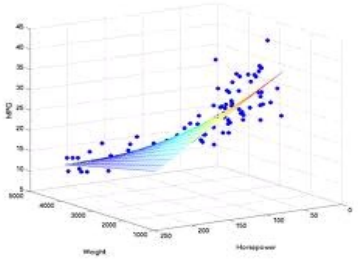
- Advanced Data Mining (beyond traditional BI)
  - Not just drag and droping data with nice tools !
  - Need to define the right variables
  - Your variables model somehow your hypothesis about the world – so you need to have an hypothesis !
  - Big Data is just here to compute !
  - This is also a domain !



K-Means



Support Vector Machines



Linear Regressions

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Bayesian models

Etc.

Big Data is Big

- The 4Vs are for marketing : Big is Big
- Big Data technologies like Hadoop are overkilling for small and medium data
  - Hadoop YARN (2.0) should improve this
  - Hadoop focused on making Hadoop « consumable » through SQL (Impala/Hive/Stinger) but is slow on improving on the real-time an small/medium data for Map Reduce jobs and to propose other parallel paradigms (MPI).
  - Still big fans of Hadoop but a bit deceived on how Hadoop has evolved over the last 2 years to prevent « advanced users » desillusions.
- Avoid the 5 VMs nodes Hadoop cluster and the 20 mega tests !

Big Data is sometimes poor data

GetMore

Enrichissement classique

- IP → **latitude**, **longitude**, **ville**, **densité de population**, **altitude moyenne**,…
- Ville française → densité, altitude, …
- Villes françaises → **distance** qui les sépare
- Date (+ heure), latitude, longitude → **météo**

Où?
Geololisation
Localisation

Qui?
Profiling visiteur

Qui?
Profiling produit

Enrichissement custom

- Association d'un groupe de mots à plusieurs **caractéristiques**
- Url : http://www.myhealth/node/6/ → produits, fitness, femme,…
- Produits : chips allégée → chips, régime, snack,…

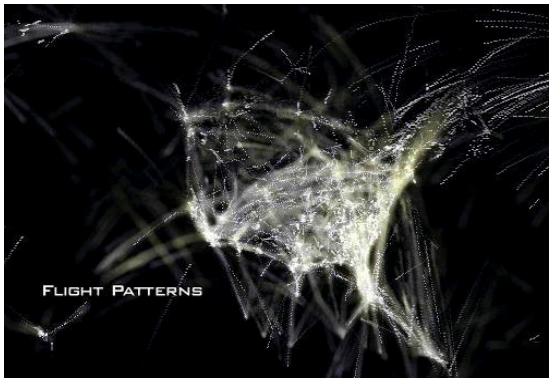
Contexte?
Météo

Big Data is not only Hadoop (NoHadoop ? After NoSQL !)

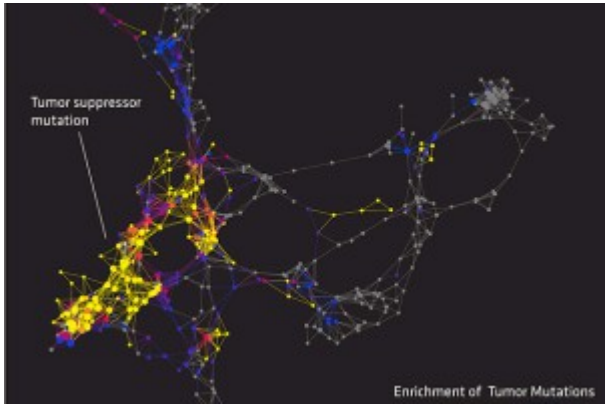
- Diamonds in the open source domain : Elastic Search is one of them
- You can index the world with Elastic Search and even geolocalize it !

Big Data can be beautiful

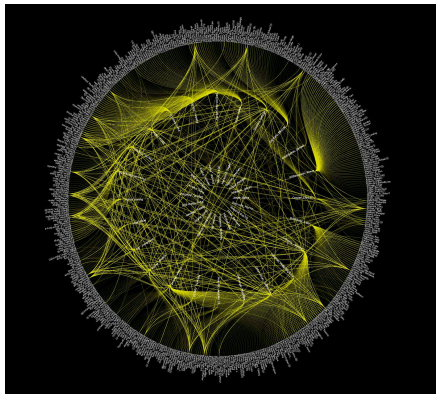
- http://www.bigdata-startups.com/the-beauty-is-in-the-big-data/



US Flight patterns by FAA



Tumor mutations



Connections between Oscar winners
http://www.pitchinteractive.com

Big Data won't make you an innovator but...

- If you are an innovator, this is your technology !
  - Big Data players dreamed « Google Search » , « Google Earth » and « Google Map » and figured out the tools they needed for that...
  - They made the work for us … now our challenge is « just » to have ambition !!!
- The magic powers you get ?
  - **The possibility to store, describe and analyse the world !**

Thank you!

- Laurence Hubert
- Email : Laurence.Hubert@hurence.com
- Web : http://www.hurence.com
- Twitter : @hurence