
***BIG DATA : Traitez, analysez et conservez en toute
sécurité de gros volumes de données non
structurées !***

***Salon DOCUMENTATION
Mercredi 21 mars 2012***



Agenda

- 14h30 – 14h50 **IBM ECM : S'organiser face à l'explosion des volumétries de données !**
Marc Delaveau, Directeur Commercial, IBM Enterprise Content Management, IBM France

- 14h50 – 15h10 **Les premières applications métiers de Watson**
Patrice Poiraud, Directeur IBM Smarter Analytics, IBM France

- 15h10 – 15h25 **IBM Content Analytics : Trouver la bonne information, analyser le contenu, prendre des décisions plus pertinentes**
Patrick Hofleitner, Ingénieur Avant-Vente, IBM ECM, IBM France

- 15h25 – 15h40 **Illustrations et exemples d'utilisation chez nos clients**
Patrick Hofleitner, Ingénieur Avant-Vente, IBM ECM, IBM France

- 15h40 – 15h55 **BIG DATA & Gouvernance**
Renaud Finaz de Villaine, Directeur Marketing et Communication, Micropole

- 15h55 – 16h15 **Approche IBM sur la Gouvernance et cas client**
Solange Mathieu-Lartigue, Ingénieur Avant-Vente, IBM ECM, IBM France

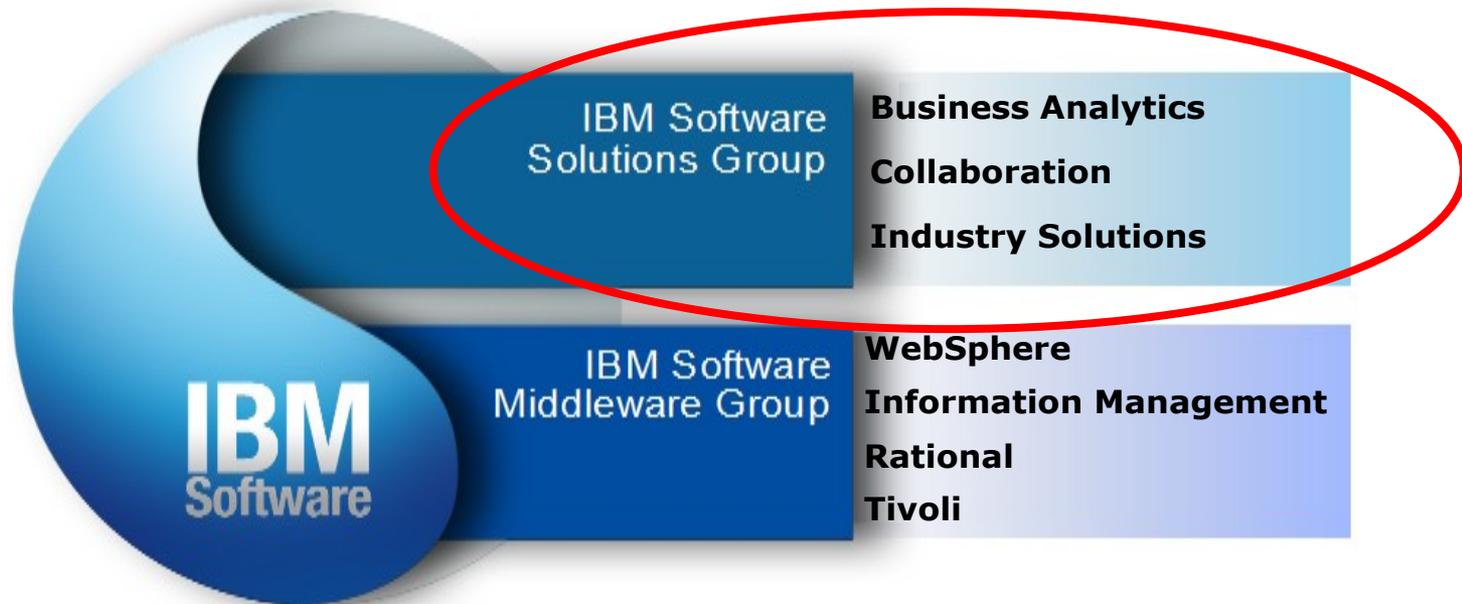
- 16h15 - 16h30 **Questions/réponses**

IBM ECM : S'organiser face à l'explosion des volumétries de données !

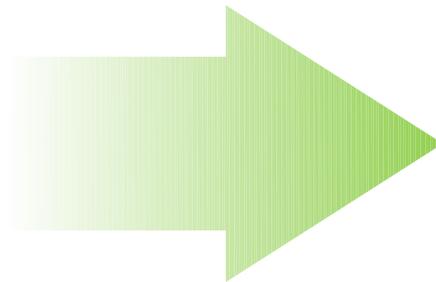
Marc Delaveau
Directeur Commercial
IBM Enterprise Content Management, IBM France



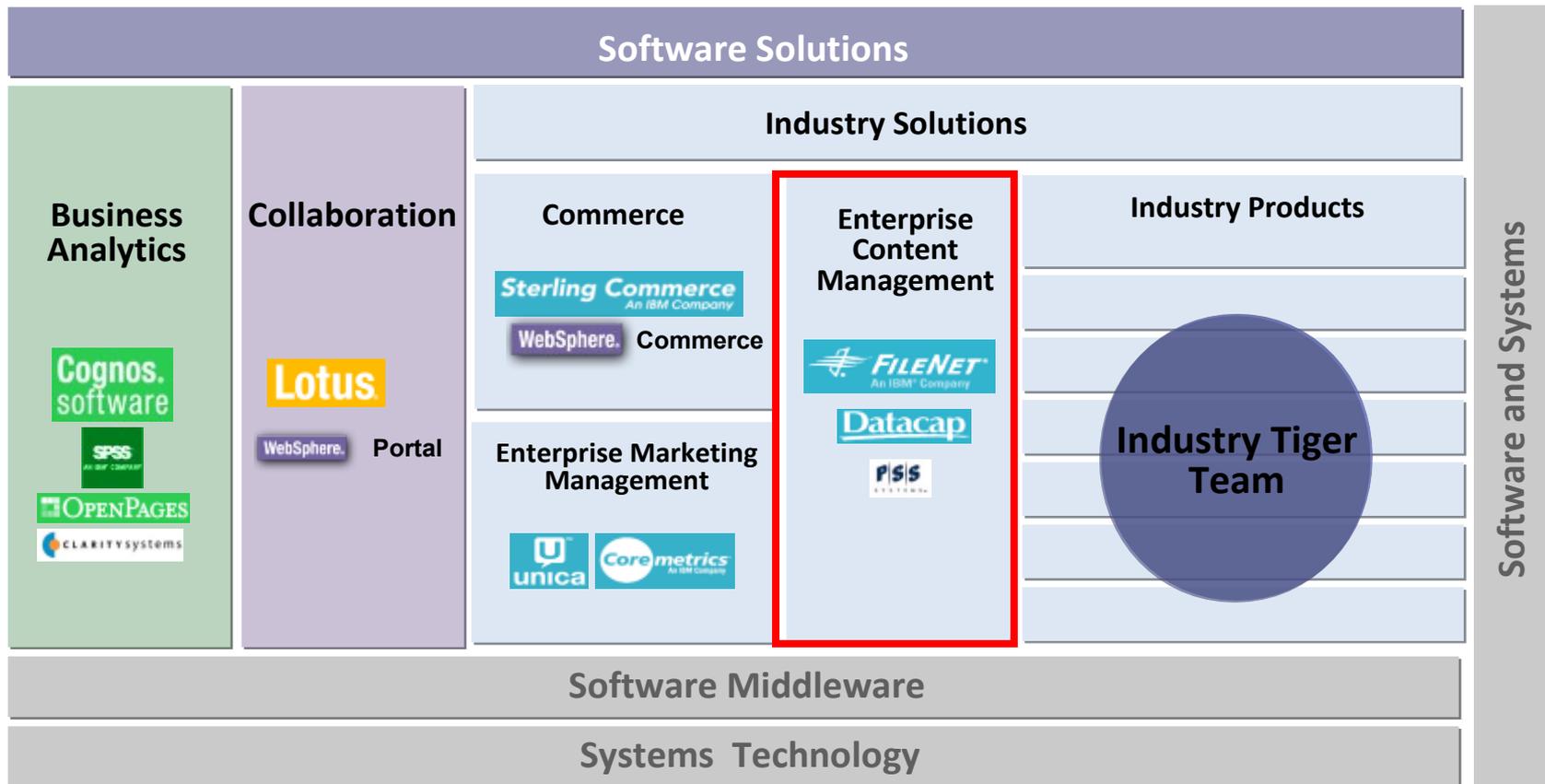
Evolution stratégique majeure - offrir des solutions métier



- **IBM n'est plus uniquement dans le domaine du middleware**
- **Evolution majeure de notre stratégie**
- **2 lignes de produit et de modèles d'investissement distincts**
- **Le « Solutions Group » est un élément stratégique majeur de notre plan 2015**



Software Solutions 2011



Transformer l'information en atout stratégique

Content Management



Master Data Management Santé



Business Analytics



Intégration de l'information Data Warehouse Master Data Management



Data Management



Gouvernance de l'information

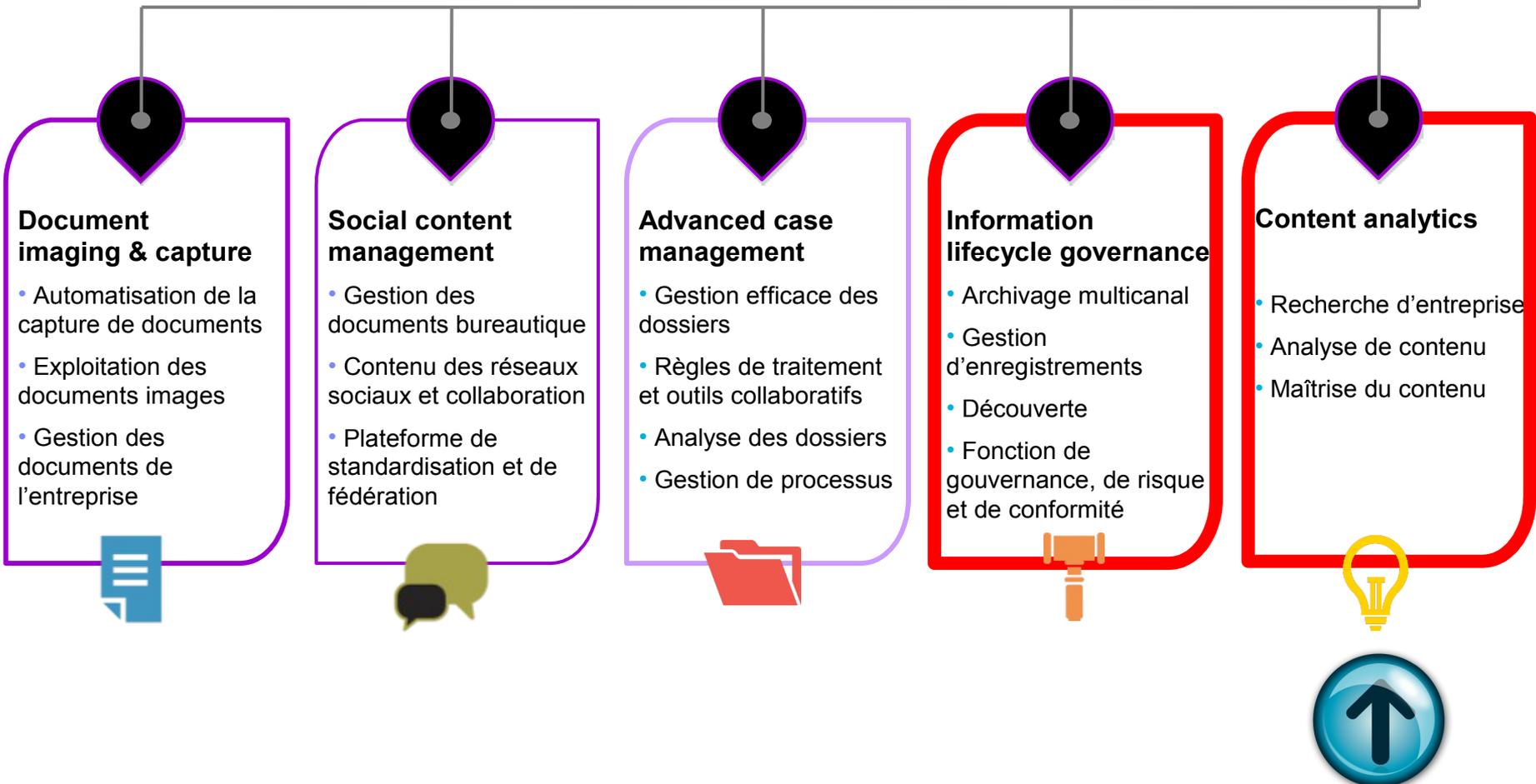


IBM Advances in Leader Quadrant in 2011 Gartner ECM Magic Quadrant!



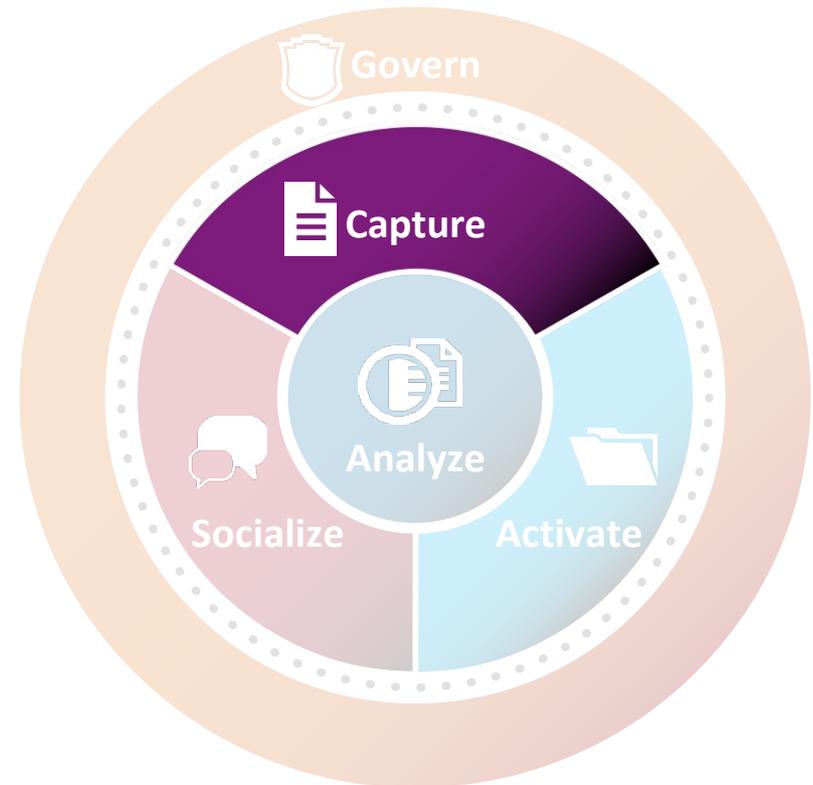
IBM Enterprise Content Management

Selon les besoins, accès à la plus-value de la solution par plusieurs **points d'entrée**



Dématérialisation et gestion de documents de production

- Capture, gère et partage du contenu partout où il existe.
- Plateforme intégrée et unifiée, de la dématérialisation à la gestion de contenu et de processus.
- Elimine le document papier et délivre en ligne un accès rapide aux états et rapports pour des millions d'utilisateurs.

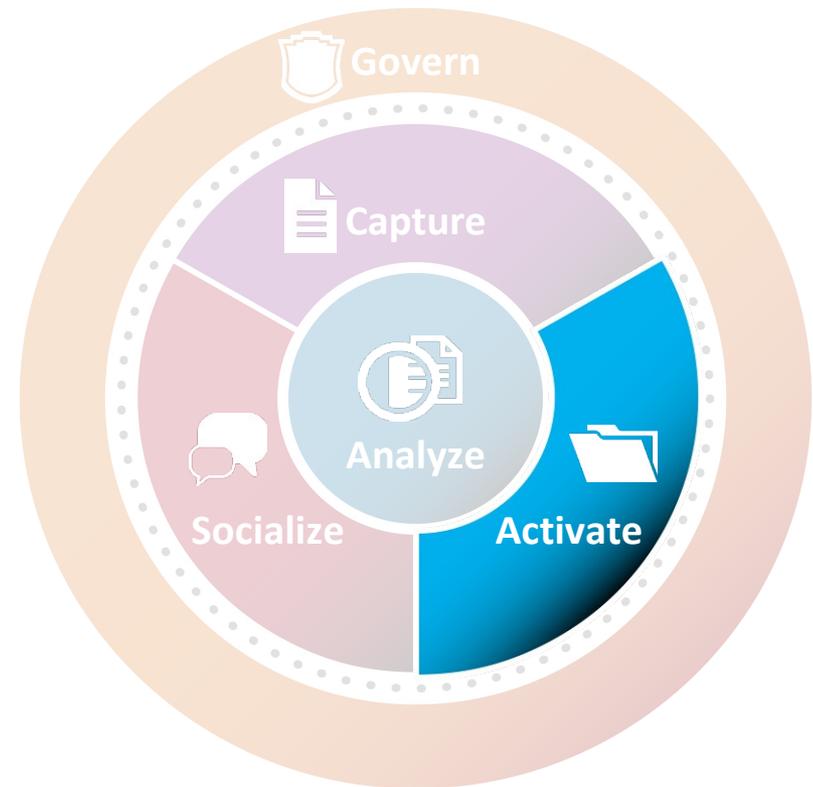


En 2011:

- **IBM Datacap: Document Capture Automation**
- **IBM Production Imaging Edition**

Advanced Case Management

- Les clients dans le monde entier réalisent des gains de productivité et des ROI substantiels
- Les gestionnaires peuvent rapidement tirer partie des meilleures pratiques de leur industrie
- Les partenaires IBM délivrent maintenant un ensemble de solutions spécifiques aux différents domaines industriels.

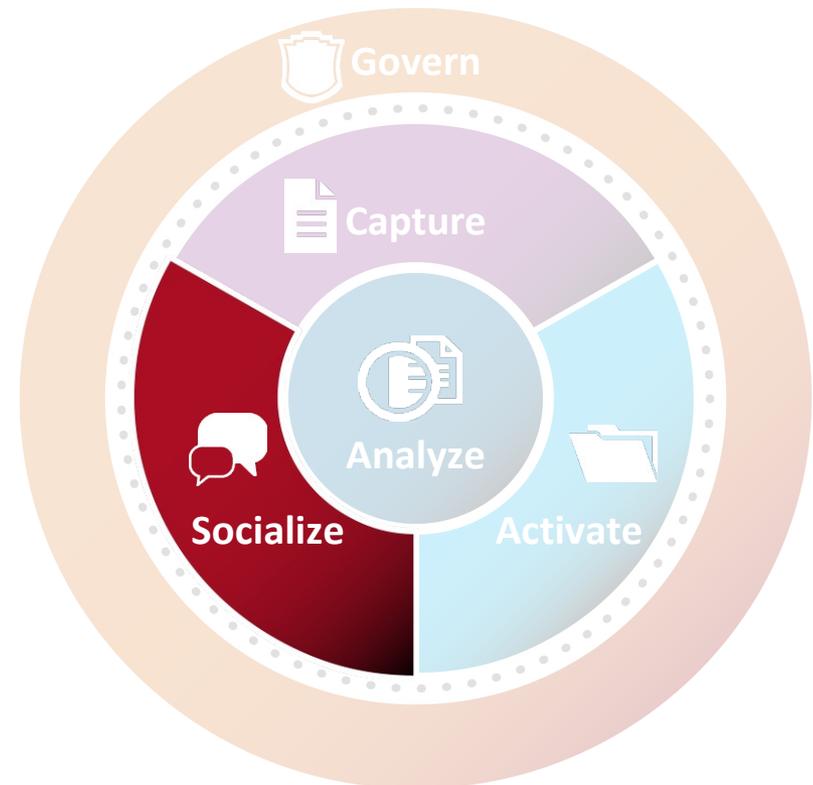


En 2011:

- **IBM Case Manager v5.1**

Gestion de contenu et réseaux sociaux

- Connexion des experts aux producteurs / consommateurs de contenu
- Connexion des contenus appropriés aux bonnes personnes et processus.
- Collaboration autour du document



En 2011 :

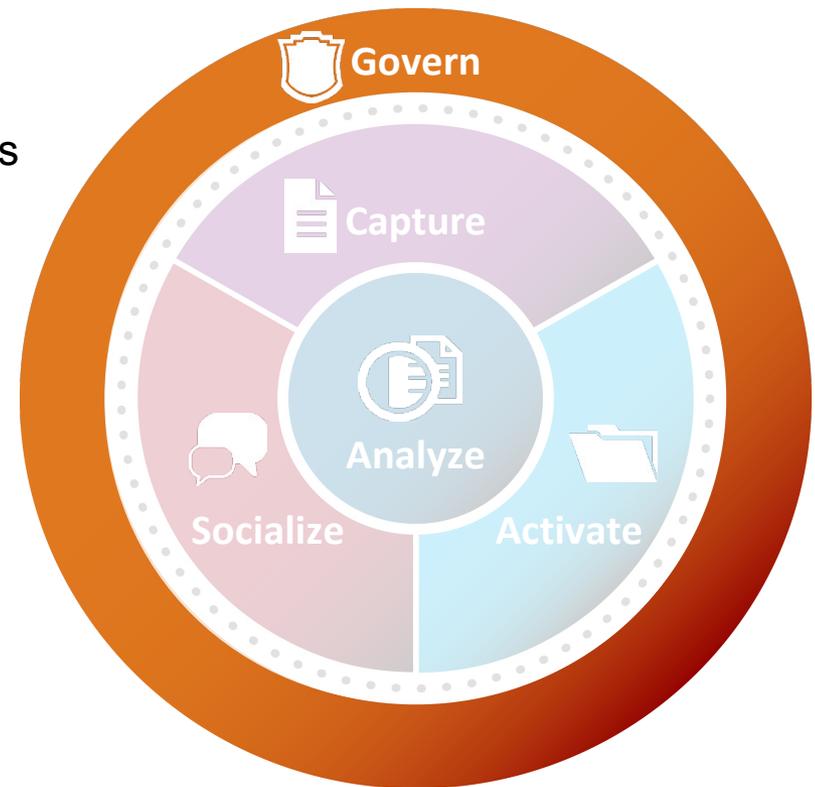
- **IBM Connections Enterprise Content Edition**
 - **Sécurité et gouvernance pour le contenu social**

Gouvernance du cycle de vie de l'information

- Une proposition de valeur unifiée pour l'archivage, la conservation, la découverte et la suppression des données structurées et non structurées
- Se conformer aux exigences réglementaires et répondre aux litiges avec efficacité
- Un différenciateur clé pour les entreprises qui souhaitent réduire le coût, la complexité et les risques

En 2011 :

- **IBM SmartCloud Archive**
- **IBM Atlas policy syndication to SharePoint & Optim**

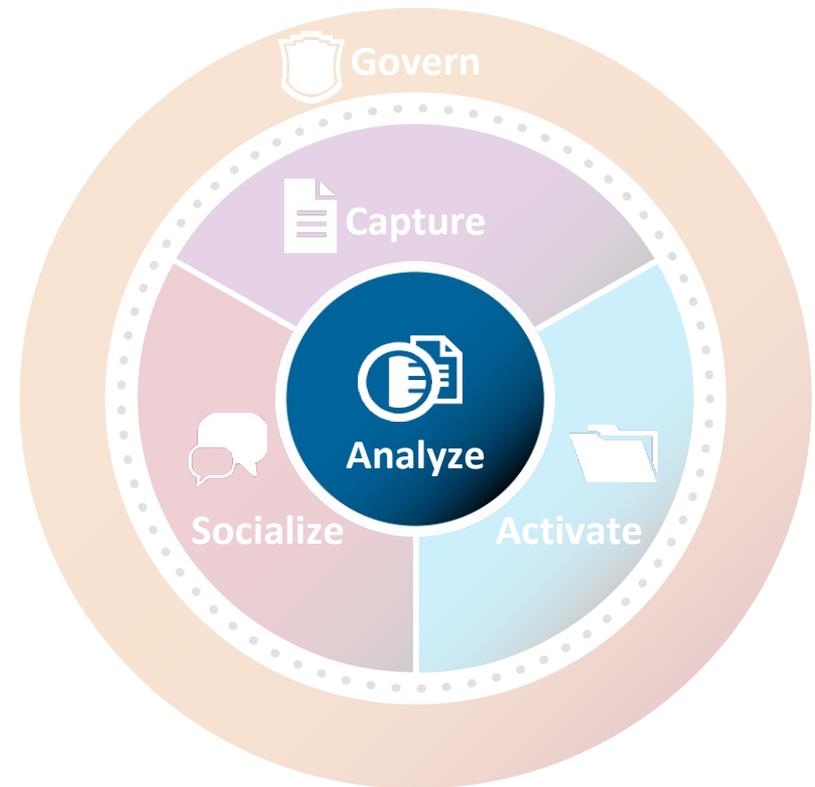


Analyser le Contenu

- Dérivée une vue à 360° en ajoutant l'analyse et la connaissance des contenus non structurés
- Rechercher, évaluer et analyser de grands volumes de texte pour comprendre et déterminer rapidement les informations pertinentes
- Dégager rapidement de nouvelles connaissances propres aux besoins d'une industrie et/ou de clients spécifiques

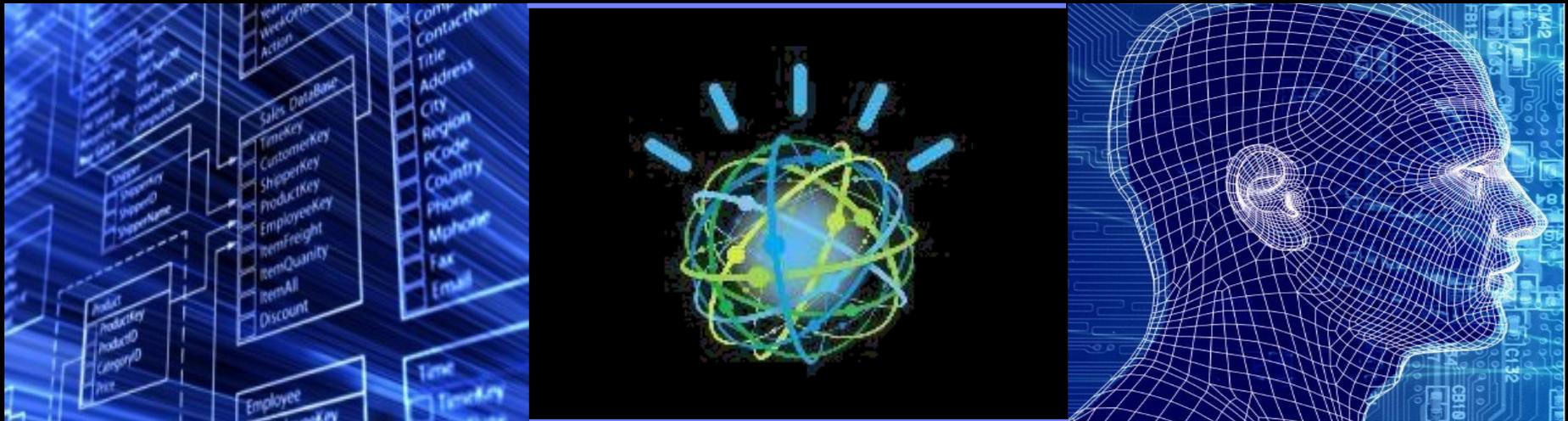
En 2011 :

- **IBM Content & Predictive Analytics for Healthcare**
- **Intégration OmniFind, Netezza & BigInsights**



Les premières applications métiers de Watson

Patrice Poiraud
Directeur, IBM Smarter Analytics



Agenda

Qu'est-ce que Watson pour IBM et pourquoi est-ce important ?

Comment IBM a pu concrétiser Watson ?

Que peut-on en attendre ?

Le 14 février 2011, IBM a changé l'histoire en présentant Watson capable de rivaliser avec l'être humain en répondant à des questions posées en langage naturel, avec la vitesse, la précision et le degré de confiance nécessaire.

- Watson a remporté la finale !
- Plus importante partie de Jeopardy en 5 ans
 - 34.5 Millions de spectateurs
 - 1.3 Milliards de pages vues
- Plus de 10 000 retombées presse
- 11,000 retransmissions des matchs
- + de 2.5 Millions de Videos vues
- 12,582 Twitter
- 25,763 Facebook Fans



Derrières les coulisses ...

System Specifications

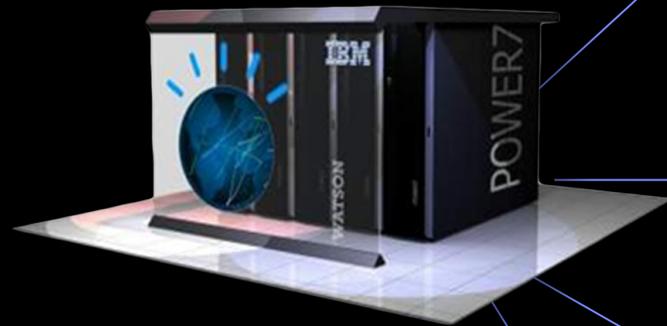
 **2880 Processing Cores**

 **90 IBM P750 Servers** 

 **16 Terabytes Memory (RAM) – 20TB Disk**

 **80 Teraflops (80 trillion operations per second)**

 **Workload Optimized Systems**



IBM Technology Depth



Content Analytics



Business Analytics



Big Data



Databases / Data Warehouses

Dans les 5 dernières années, IBM a investi plus de 14 Milliards de US\$ en acquisitions et consacre un budget annuel de 6 Milliards de US\$ à la Recherche et au Développement

Single thread performance is slowing dramatically

Dans 12 ans, Watson sera de la taille d'un portable



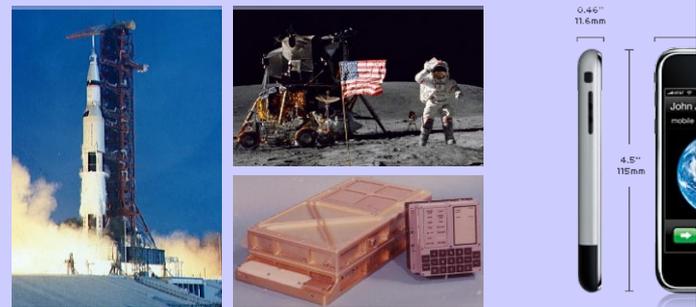
80 TFlops

- 0.5 2010
- 1 2013
- 2 2013
- 4 2016
- 8 2016
- 16 2019
- 32 2019
- 64 2019
- 128 2023

1/2 TFlops



Compute Power 1969 and 2007



Apollo Guidance Computer

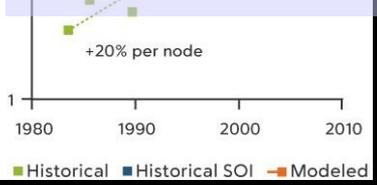
72K Read only Memory
 8K of RAM
 2MHz
 32 Kg
 61cm x 32cm x 15cm
 80 Watt

iPhone

16GB RAM
 1.2GHz,
 135 g
 115mmx66mm
 600mW

→X 2 million
 →X 600
 →/200
 →/350
 →/130

Sources:
 AGC: <http://www.ddj.com/184404139>
 iPhone: <http://www.apple.com/iphone/specs.html>
 ARM11: <http://www.physorg.com/newsh5139.html>



5'000 iPods



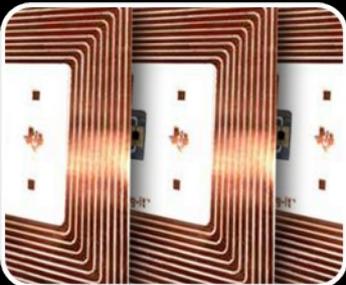
1 Game Console



Les données non structurées prolifèrent . . .

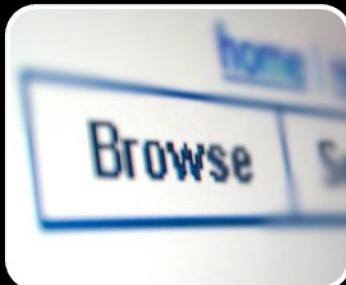
... 249 Milliards d'e-mails (2.8M/sec) and 200 Millions de Tweets chaque jour

... 35 heures de vidéo chargées sur You Tube toutes les minutes



En 2005 il y avait 1.3 Milliards de Tags RFID en circulation ...

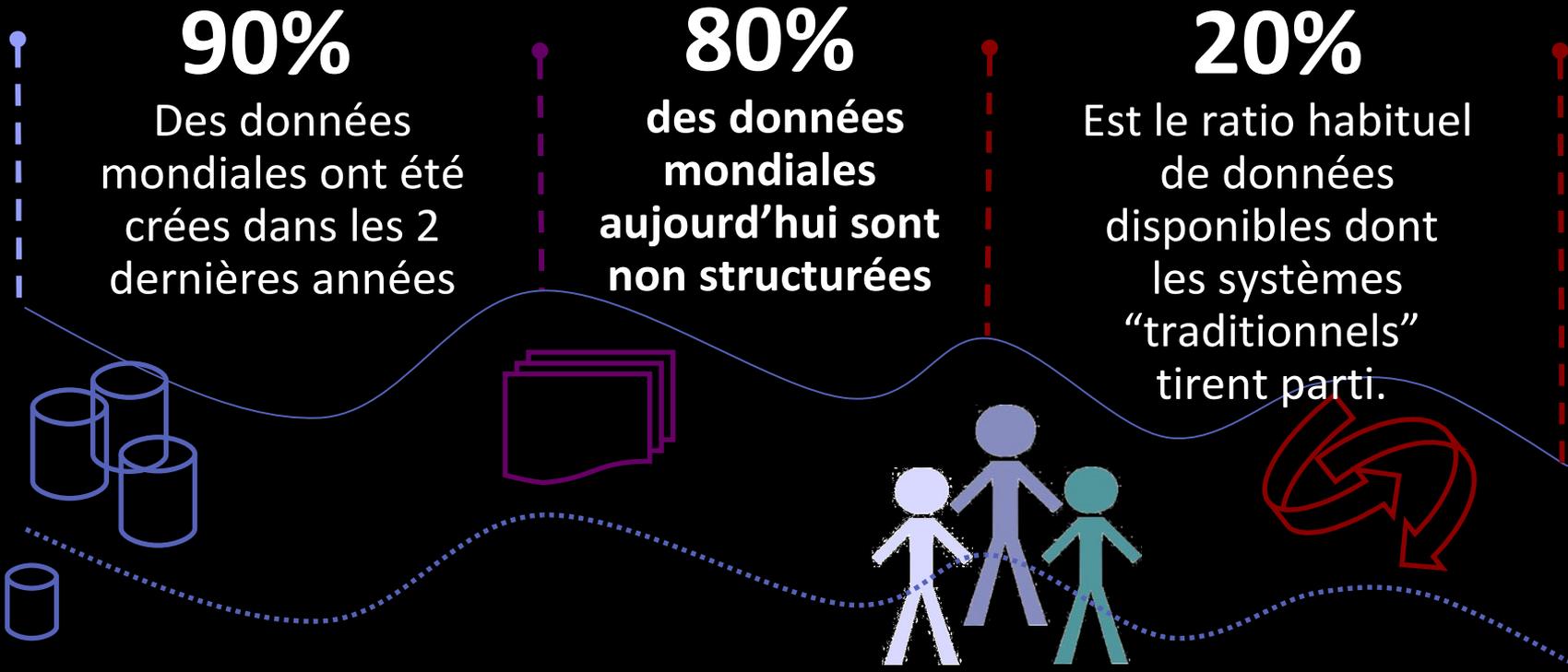
... En 2010, nous étions à 33 Milliards.



Environ 2 Milliards de personnes accèdent au Web en 2011 ...

... et 1000 Milliards d'objets connectés (voitures, appareils domestiques, téléphones, routes, pipelines, ...)

Les organisations actuelles souffrent d'une pénurie d'information tout en étant submergées par un océan de données



90%

Des données mondiales ont été créées dans les 2 dernières années

80%

des données mondiales aujourd'hui sont non structurées

20%

Est le ratio habituel de données disponibles dont les systèmes "traditionnels" tirent parti.

1 sur 2

Décideurs n'ayant pas accès à l'information dont ils ont besoin

83%

Des DSI citent la Business Intelligence et l'Analytique comme l'une de leur préoccupation principale

5.4X

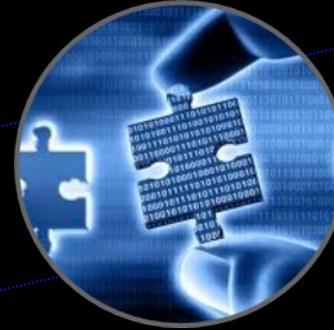
Les entreprises performantes utilisant l'Analytique sont 5,4 fois plus nombreuses que les autres

Les défis actuels obligent les organisations à se remettre en cause pour continuer à exister demain



Informatique traditionnelle

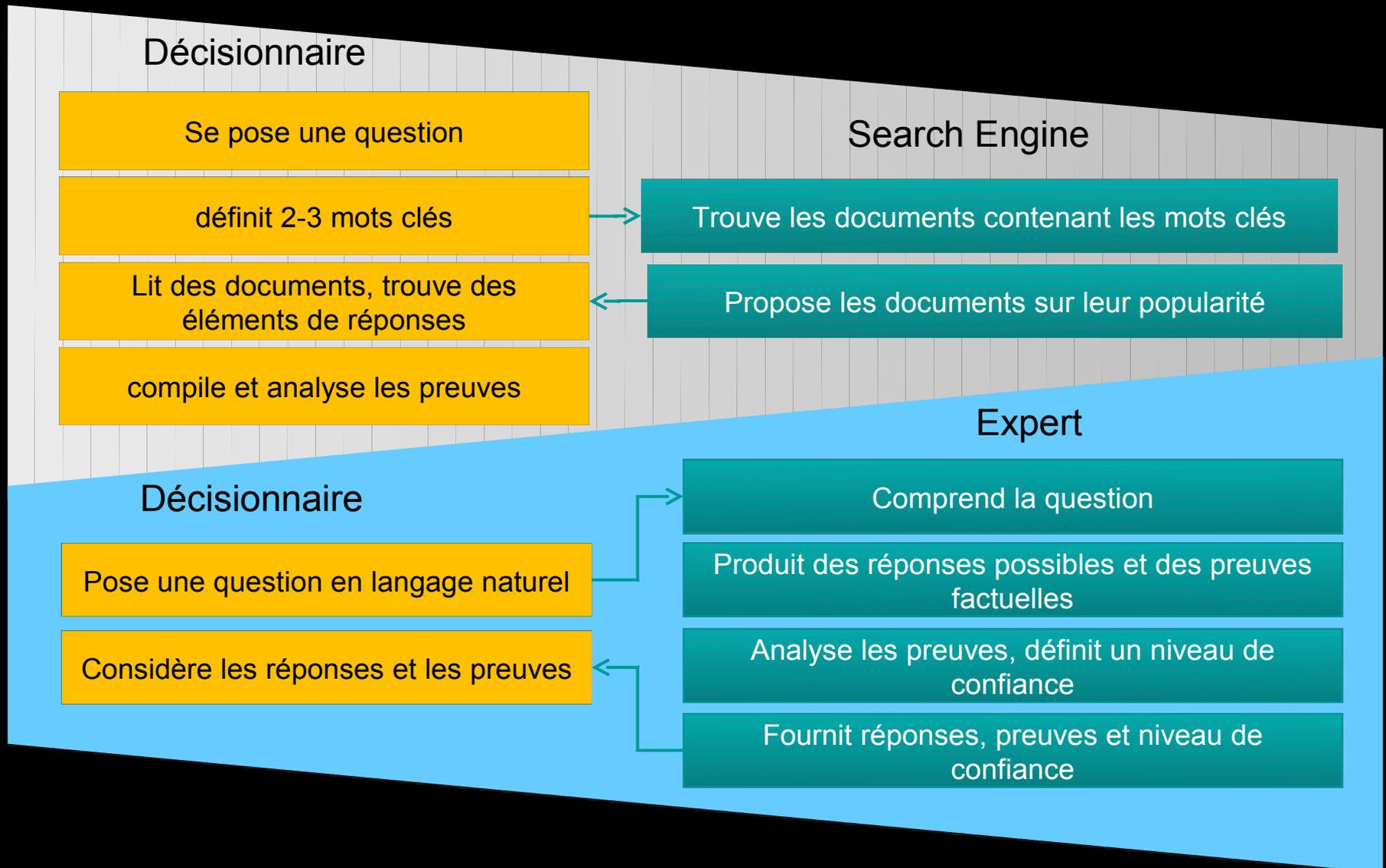
- Données structurées (localement)
- Applications déterministes
- Orienté sur le "Search"
- Peu de données
- Langage machine



Informatique "Nouvelle Génération"

- Données structurées et non structurées (globales)
- Applications probabilistiques
- Orienté sur le mode "Discovery"
- Volumes de données divers (dont Big Data)
- Langage naturel

Watson fonctionne comme un expert



Jeopardy! Et après ...

Il est maintenant question d'influencer la vision, les actes et les opérations des organisations.

Santé

Assistance au diagnostic / traitement, basé sur des preuves , approche de la médecine collaborative

Services financiers

Planification des investissements et retraits, des axes de négociation et aide à la décision

Services clients

Centres d'appels techniques et commerciaux, gestion des connaissances, recommandations selon le profil du client

Secteur Public

Sécurité publique, partage de l'information entre les services, sécurité

IBM Watson a la capacité de pouvoir répondre à de grands enjeux économiques et sociaux

Le secteur de la santé doit faire face l'un des challenges en terme de valorisation de l'information les plus conséquents



Le volume d'informations médicales double tous les 5 ans, la plupart étant en format non structuré



81% of des médecins déclarent consacrer ,mensuellement , 5 heures ou moins à la lecture de revues médicales



1 in 5

diagnosis that are estimated to be inaccurate or incomplete



1.5 million

errors in the way medications are prescribed, delivered and taken in the U.S. every year



44,000 -98,000

of Americans who die each year from preventable medical errors in hospitals alone

“Medicine has become too complex (and only) about 20 percent of the knowledge clinicians use today is evidence-based.”

Steven Shapiro, Chief Medical and Scientific Officer, UPMC

En quoi la technologie Watson est-elle idéale dans le domaine de la santé ?

Compréhension du langage naturel



Dans quelles conditions souffre-t-on d'yeux rouges, de douleurs, d'inflammation, d'une vision floue, de tâches au niveau de la vision, et de sensibilité à la lumière ?

Analyse de grands volumes de données non structurées



Avis médicaux, revues médicales, essais cliniques, études pathologiques, blogs, wikipédia

Génère et évalue différentes hypothèses



<u>Diagnostic possible</u>	<u>% confiance</u>
Uvéite	91%
Iritis	48%
Keratite	29%

Propose des réponses validées selon leur certitude



Intègre les échanges itératifs pour affiner les résultats



Histoire de famille, entretien avec le patient, examen physique, médicaments actuellement prescrits

Auto-apprend des résultats obtenus au fur et à mesure du temps

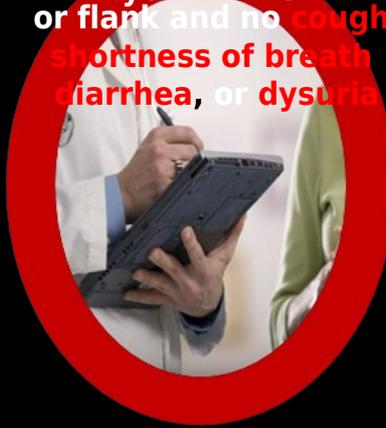


Quelles mesures ont été prises ?
Quels traitements ont été prescrits ?
Pour quel résultat ?

Disposer de toutes les pièces en même temps au point d'impact peut être un élément décisif

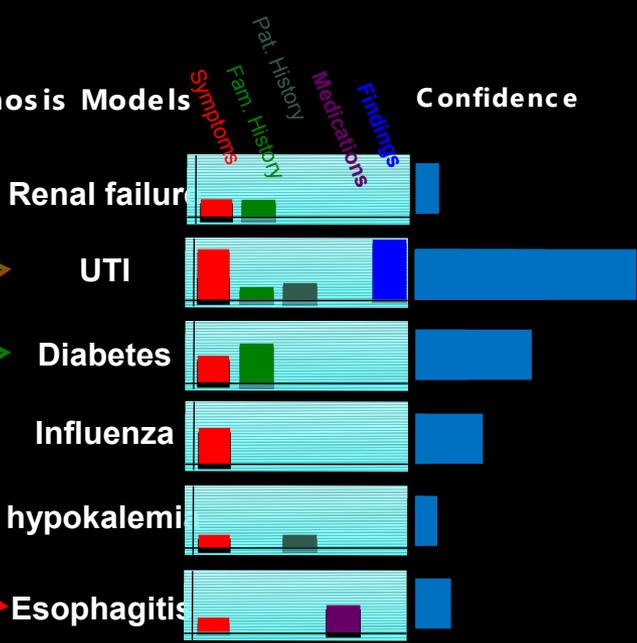
Findings

A 68-year-old female with a history of hypertension, hyperlipidemia, and hypothyroidism. She presented with a 3-day history of difficulty swallowing, fever, dry mouth, thirst, anorexia, frequent urination, and dizziness. She also reported no abdominal pain, no back pain, no cough, and no diarrhea. Her medical history includes oral cancer, bladder cancer, and purpura. Her family history is notable for Graves' Disease (Thyroid Autoimmune). Her current medications include Alendronate, pravastatin, levothyroxine, and hydroxychloroquine. Her physical examination revealed a heart rate of 88 bpm and a urine culture positive for E. Coli.



- Symptoms:** difficulty swallowing, fever, dry mouth, thirst, anorexia, frequent urination, dizziness, no abdominal pain, no back pain, no cough, no diarrhea.
- Family History:** Oral cancer, bladder cancer, melanochromatosis, purpura, Graves' Disease (Thyroid Autoimmune).
- Patient History:** cutaneous lupus, osteoporosis, hyperlipidemia, frequent UTI, hypothyroidism.
- Medications:** Alendronate, pravastatin, levothyroxine, hydroxychloroquine.
- Findings:** urine dipstick: leukocyte esterase positive, spine 120/80 mm HG, heart rate: 88 bpm, urine culture: E. Coli.

Diagnosis Models



Most Confident Diagnosis: Esophagitis

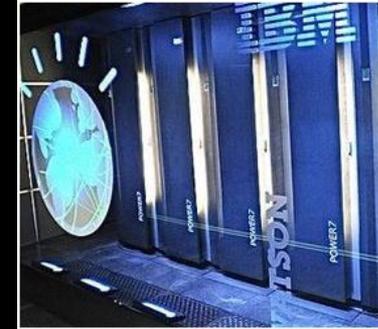
- Extract **Medications**
- Use database of drug side-effects
- Together, multiple diagnoses may best explain symptoms
- Extract **Findings**: Confirms that UTI was present

Le Cedars Sinai Hospital à Los Angeles intègre l'IBM Watson à son équipe de médecins



LeMondelInformatique.fr

Toute l'info et les tendances du monde IT



Le supercalculateur Watson d'IBM doit commencer son travail d'évaluation en matière de traitement du cancer, à l'hôpital Cedars-Sinai de Los Angeles. Il pourra suggérer aux médecins les traitements les plus adaptés en quelques secondes.

IBM et WellPoint, le plus grand programme de santé du Blue Cross Blue Shield, ont mis au point des applications qui vont transformer le supercalculateur Watson en conseiller spécialisé, au service des oncologues du département Samuel Oschin Comprehensive Cancer Institute de l'hôpital Cedars-Sinai de Los Angeles.

Après la santé ...



Santé

Assistance au diagnostic / traitement, basé sur des preuves , approche de la médecine collaborative



Services financiers

Planification des investissements et retraits, des axes de négociation et aide à la décision

Les organisations financières font face à certains des challenges les plus importants en termes de valorisation de l'information

- Reuters publie l'équivalent de 9000 pages d'informations financières chaque jour¹
- 5 nouveaux documents d'analyse et de recherche sont publiés par Wall Street toutes les minutes¹
- Les gestionnaires d'actifs reçoivent jusqu'à 1 000 e-mails chaque jour¹



Only 1/3 of the ~10,000 US public companies are covered by meaningful Wall Street research¹



Productivity: Top traders can do \$4M/hour in trades²



2011 NYSE average: 4.3M shares traded per minute³

“Le volume du flux d'information rend la prise de décision avec certitude difficile -Adam Margolis, Citi trader

Services financiers – Simulations basées sur l'exemple



Conseiller en investissements financiers

Utilise plus d'informations, avec plus d'efficacité et d'efficience pour confirmer ses décisions en termes d'investissement et de crédit



Conseiller de banque de détail

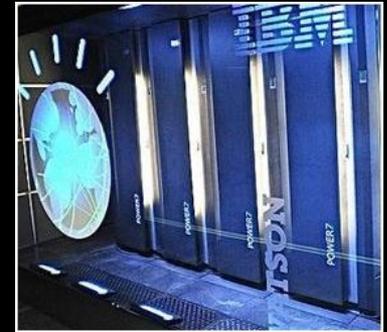
Fournit en temps réel les conseils en placement aux conseillers financiers en fonction du profil et de l'historique des clients



Gestionnaire de risques

Utilise les fonctionnalités pour détecter les fraudes potentielles et se conformer aux exigences de conformité réglementaire

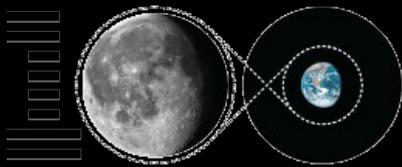
Citigroup embauche le superordinateur Watson



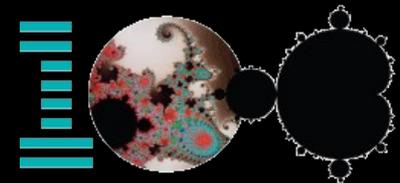
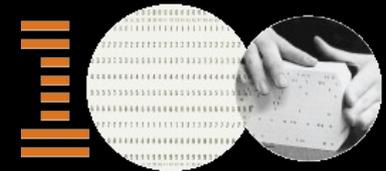
La troisième banque américaine va utiliser la puissance d'analyse du superordinateur d'IBM. Watson s'était fait connaître en février 2011 en devenant le premier ordinateur à battre un adversaire humain au jeu télévisé Jeopardy. (...)

Le troisième groupe bancaire américain a précisé, lundi 5 mars, qu'il comptait utiliser Watson pour "analyser les besoins des clients et les opportunités financières qu'il est possible de leur proposer". Concrètement, Watson pourra compiler et résumer en quelques minutes, pour les gestionnaires de portefeuilles de Citigroup, les résultats financiers de centaines d'entreprises ainsi que les informations qui circulent sur d'éventuelles opportunités d'investissement à la vitesse du Web sur les réseaux sociaux.

(...)



MERCI



IBM Content Analytics : *Trouver la bonne information, analyser le contenu, prendre des décisions plus pertinentes*

Patrick Hofleitner
Ingénieur Avant-Vente
IBM ECM, IBM France



Les approches traditionnelles convergent

La recherche par mots-clés n'est pas suffisante

“Pouvoir rechercher dans les informations non-structurées est une nécessité pour toutes les applications, de l'intranet aux référentiels de documents.”

– Whit Andrews, Rita Knox Gartner

Analyser le contenu non-structuré est indispensable

“L'accès aux données structurées, même valorisées par de l'Informatique Décisionnelle ou de l'Analyse Prédictive, manque d'indices contextuels pour la bonne prise de décision. Services clients, finance, ... impliquent une grande diversité de canaux”

– Craig Le Clair Forrester

Impact croissant dans l'activité

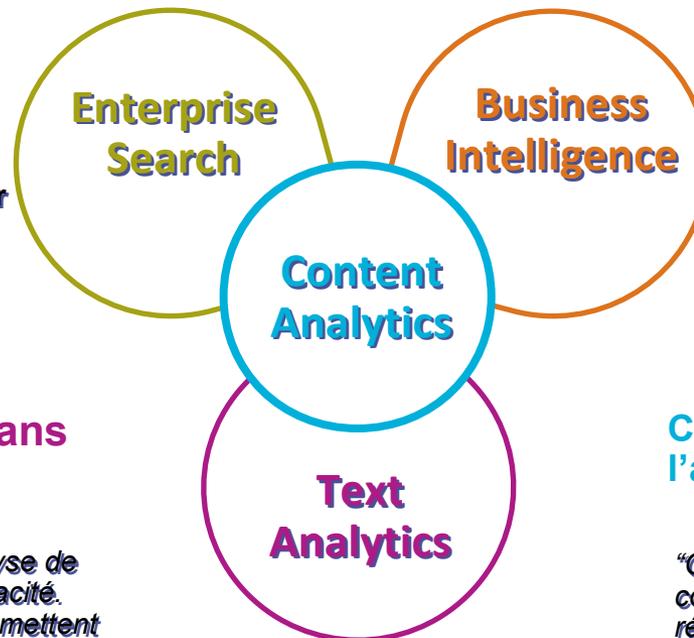
“Les pionniers utilisant l'analyse de texte ont déjà gagné en efficacité. Les autres organisations se mettent en retrait.”

– Sue Feldman IDC

Convergence vers l'analyse de contenu

“Chaque organisation doit comprendre comment l'analyse de contenu peut répondre à ses préoccupations cruciales; dès maintenant pour bénéficier de la disponibilité de ces nouveaux outils.”

– Rita Knox Gartner



De l'information brute à une appréciation efficace

Découvrir la valeur du contenu par une approche visuelle unique

Agréger et extraire de sources différentes

... pour générer des collections depuis des sources multiples, internes et externes, mais aussi de types différents : référentiels de documents, données structurées, réseaux sociaux, internet et autres.

Organiser, analyser et visualiser

... documents et données de l'organisation en identifiant des tendances, modèles, corrélations, anomalies, contextes.

Chercher et explorer pour puiser

... pour trouver confirmation ou découvrir de nouveaux éléments sans avoir à développer des modèles et déployer des systèmes complexes.



IBM Content Analytics une solution pour une appréciation efficace du contenu

- Transformer une information brute en **contenu exploitable** sans avoir à développer des modèles et déployer des systèmes complexes.
- Déduire en quelques **heures** ou **jours** ... au lieu de semaines ou mois. Outil de **recherche** et **exploration** du contenu **d'utilisation aisée** par tout professionnel.
- **Flexible** et **extensible** pour toute évolution.



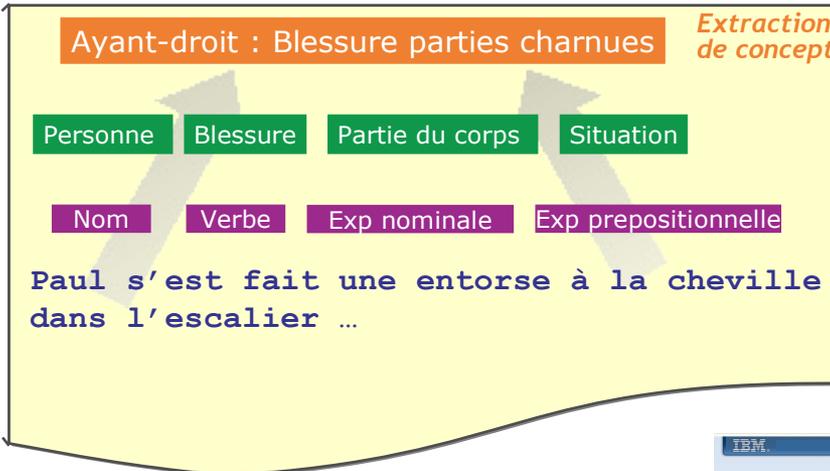
**Documents et Données, Internes et Externes
incluant Réseaux Sociaux et autres sources**

IBM Content Analytics permet des analyses inconcevables



Les sources

(ECM, Fichiers, Web, BdD, ...)

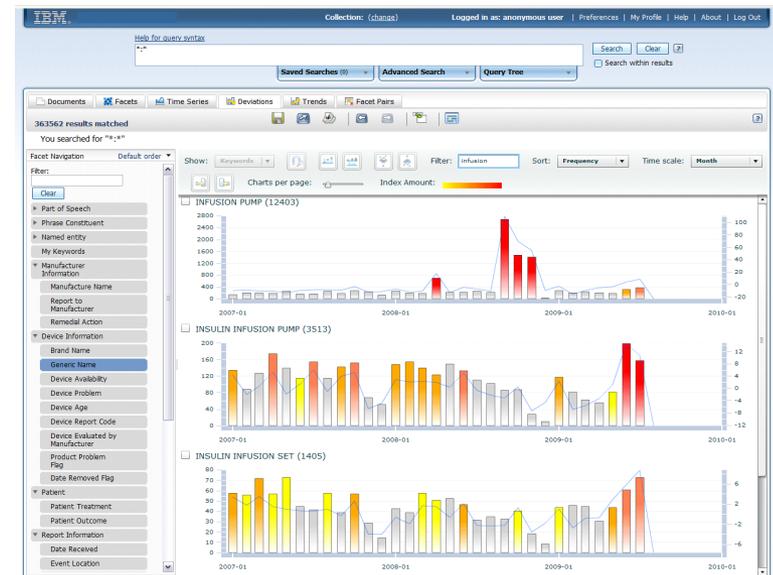


Documents analysés
Avec identification de concepts



Visualisation automatique
Concepts et sources sont présentés dans l'interface utilisateur

- **Analyse document par document**
- **Consolidation des conclusions**
- **Evaluation sur de gros volumes d'informations**
- **Inconcevable par traitement manuel ou hors de coût et délai**

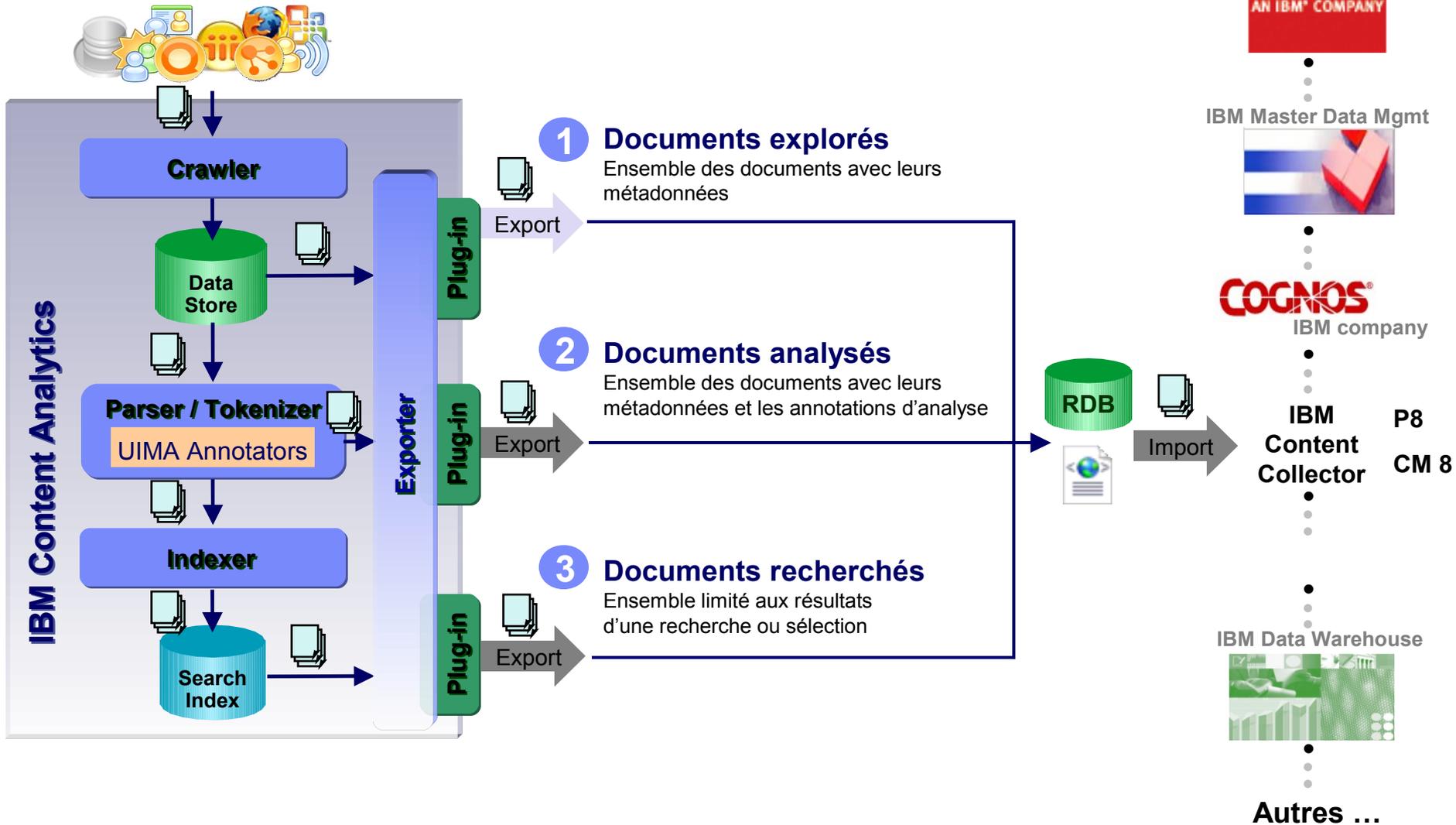


Qu'est-ce que NLP et UIMA ?

- **Natural Language Processing (NLP)**, la pierre angulaire pour rendre possible l'interaction directe entre les ordinateurs et les langages humains
 - Watson utilise **IBM Content Analytics** pour les fonctions NLP
- **Unstructured Information Management Architecture (UIMA)**, une architecture ouverte pour le traitement de texte et la définition de solutions d'analyse de texte
 - Mise en oeuvre dans de nombreuses solutions IBM
 - Standard Open Source OASIS
 - Projet supporté par Apache Software Foundation



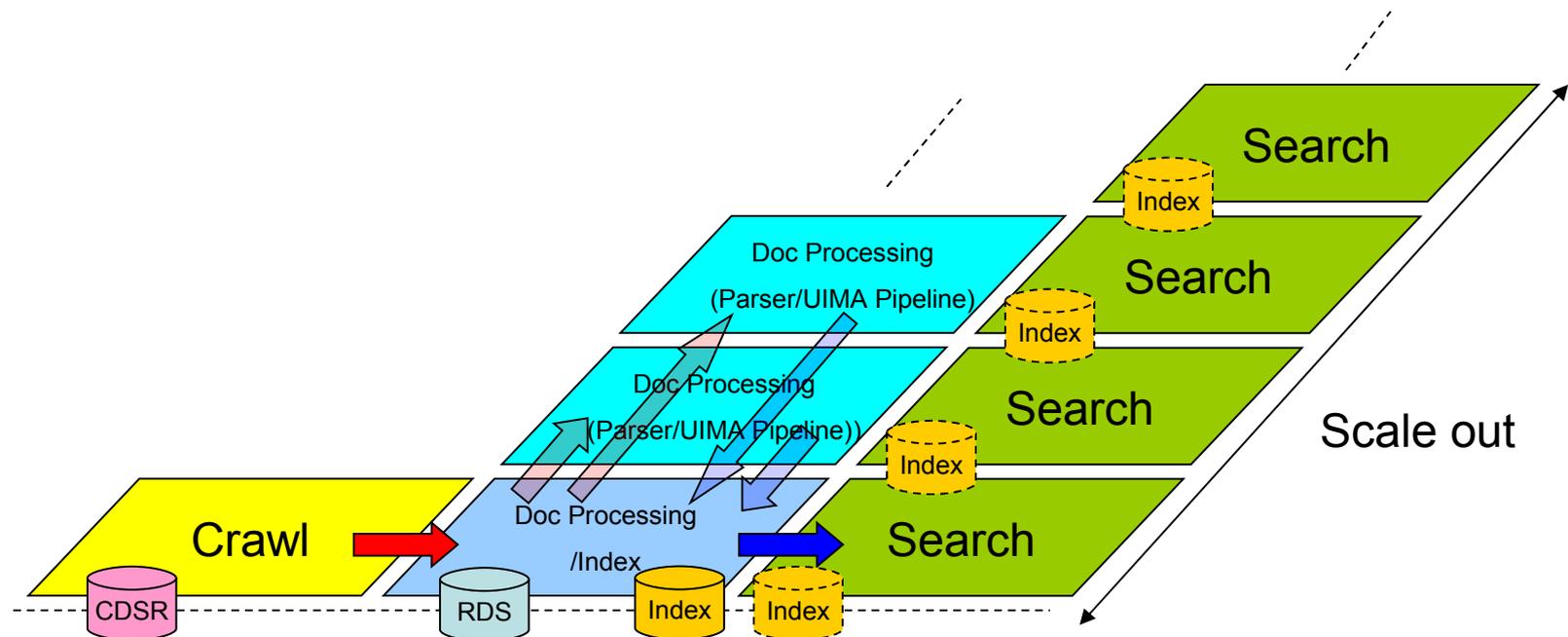
Fonctions d'intégration



Système évolutif

Environnement adaptable aux besoins (1 noeud à n noeuds)

- Traitement de documents n noeuds
- Recherche de documents n noeuds



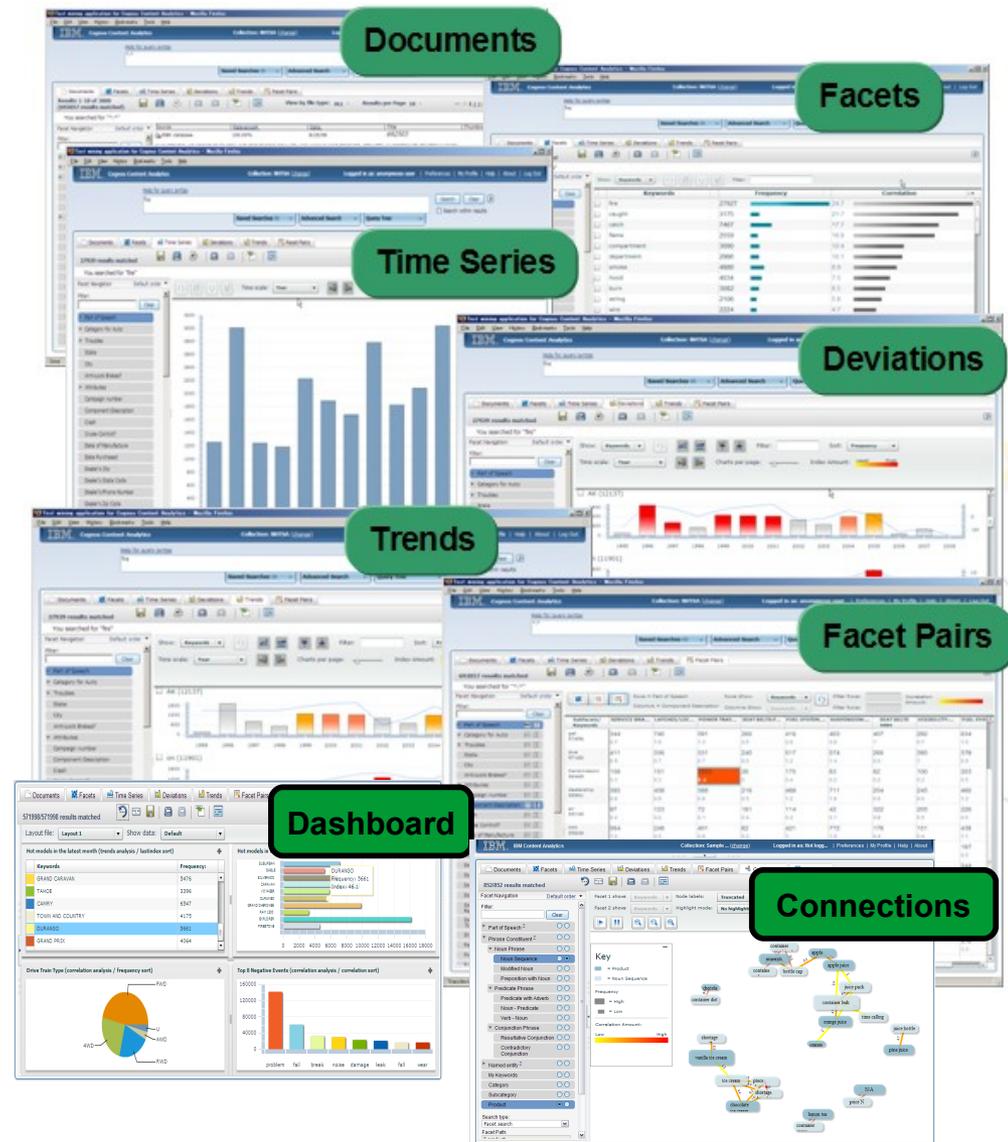
Explorer l'information

• Analyse interactive

- Faciliter la recherche de documents qui ont des valeurs d'index spécifiques par rapport à certaines facettes / dimensions
- Vues multiples en fonction de la nature des analyses
- Graphiques interactifs pour restituer l'information : zoom, tri, filtre

• 8 Vues

- **Documents** : liste les documents en fonction de la recherche
- **Facettes**: liste les mots-clés d'une facette
- **Série chronologique** : montre la fréquence dans le temps
- **Déviations** : montre les déviations entre les mots-clés
- **Tendances** : détecte les variations importantes dans le temps
- **Paires** : montrent la corrélation entre 2 dimensions / facettes
- **Connections** : montrent les connections entre les valeurs de facettes
- **Tableaux de bord** : vue de synthèse



21 mars 2012



IBM Content Analytics :

Illustrations et exemples d'utilisation chez nos clients

Patrick Hofleitner
Ingénieur Avant-Vente
IBM ECM, IBM France



Trouver l'information, décider en conseil municipal Ville de KARLSRUHE

OmniFind fournit une interface Intranet de recherche et d'accès aux informations et services de la ville de Karlsruhe

Les besoins :

La ville de Karlsruhe, Allemagne, désire accélérer la prise de décision et favoriser le travail en équipe. Adoption d'un nouveau portail, d'outils de collaboration et de recherche pour le conseil municipal et le management. Ces solutions permettent un accès unique à l'ensemble des informations pour permettre une exploitation améliorée de l'information et une vue globale sur l'ensemble de l'organisation.

La solution :

IBM WebSphere® Portal Server, accès unique à l'ensemble des applications

IBM Lotus Notes® ,Domino et Quickr, messagerie et outil collaboratif

IBM OmniFind® Enterprise Edition

Les bénéfices :

- Amélioration de la communication et de la collaboration
- Réduction des coûts et des délais

La solution :

IBM® OmniFind® Enterprise Edition

- IBM Lotus® Domino®
- IBM Lotus Quickr
- IBM WebSphere® Portal Server

Partenaire :

Profi Engineering Systems AG



Améliorer la qualité de service clients chez NTT DOCOMO

“Nous pouvons facilement identifier les tendances et les modèles de comportement de nos clients et apporter le meilleur service client.”

*Makoto Ichise, Manager of Information System
Department Group, NTT DOCOMO*

industrie : Télécommunications

besoin : Améliorer le service client, fidélisation

Le Challenge : IBM Content Analytics

Une stratégie orientée clients pour offrir les meilleurs produits et services en étant à l'écoute des attentes réelles exprimées par les clients.

La fidélisation

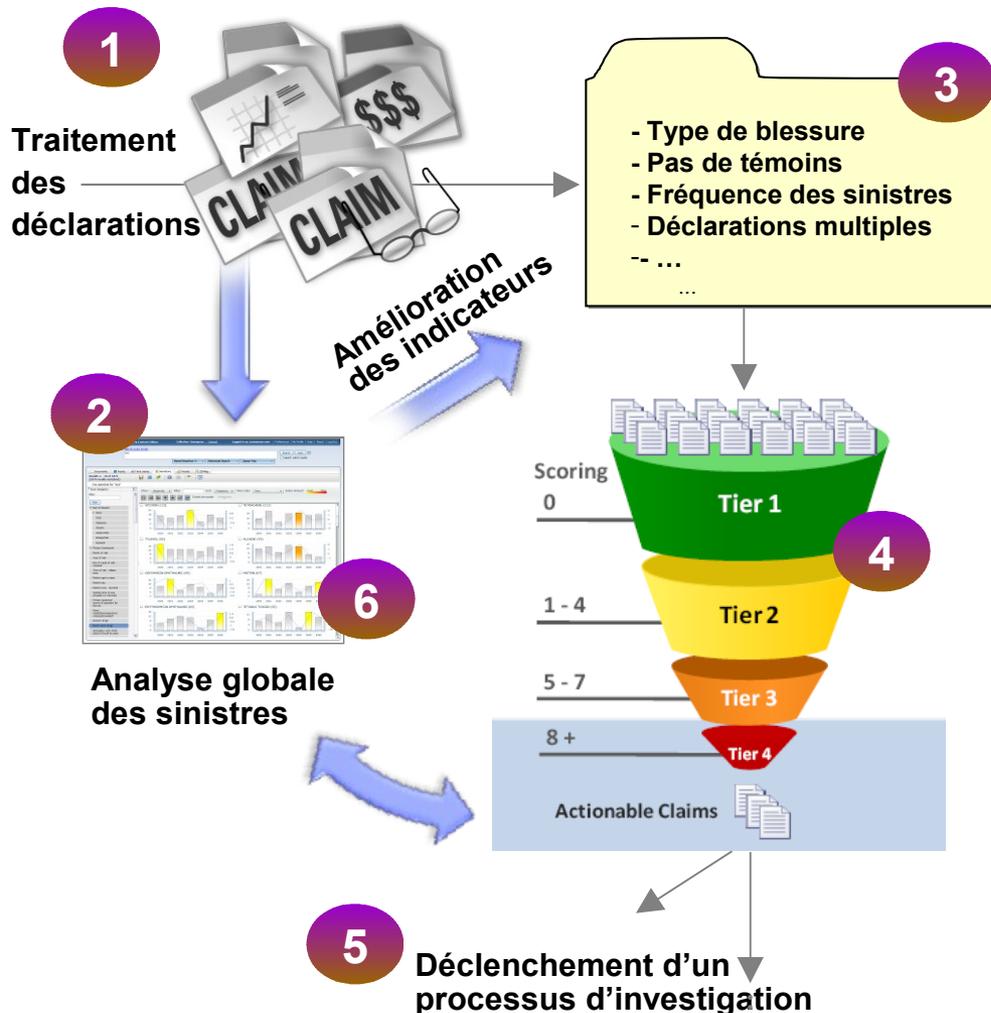
Analyse des mails clients et des rapports du centre de support . Des règles d'analyse textuelles d'IBM Content Analytics permettent de déterminer les clients susceptibles de changer d'opérateur. Des alertes sont envoyées à un département chargé de traiter ces cas.

Les améliorations

Meilleures offres et services aux clients fidèles.
Programme Premium Club.
Paramétrage par défaut des téléphones.

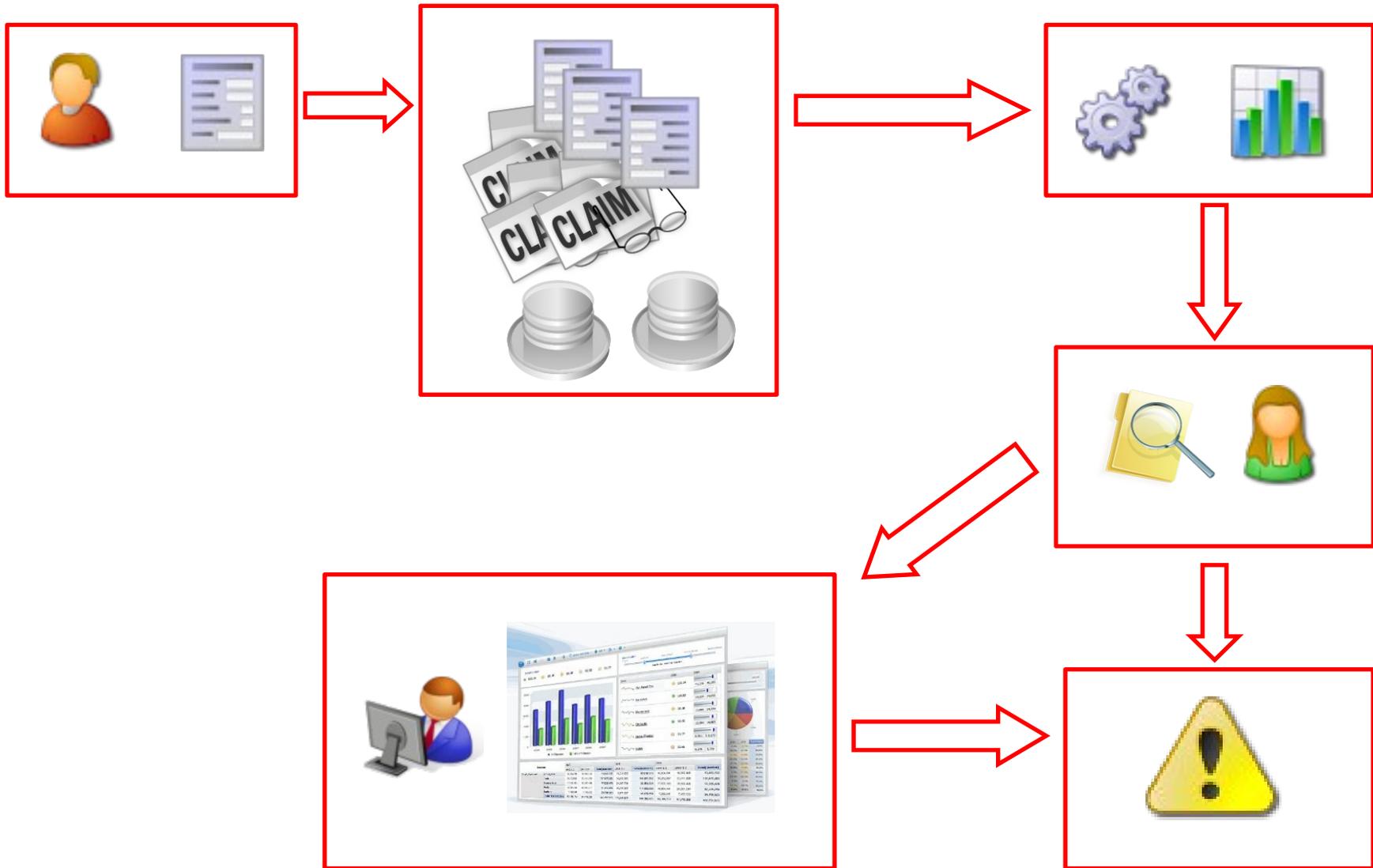


Dossiers d'assurance, détection de fraude et analyse des risques



1. Agrégation automatique de l'ensemble des informations, structurées et non structurées concernant le traitement des sinistres.
2. Détermination de corrélations, tendances, anomalies, comportements... visualisation et exploration pour de nouvelles perspectives du traitement des dossiers.
3. Indicateurs dans le traitement en ligne des dossiers. Identification de tentative de fraude et évaluation des risques.
4. Qualification des risques.
5. Traitement par exception des dossiers à risques.
6. Amélioration continue des indicateurs par réitération.

Illustration de la solution analyse d'incident santé



Analyse d'incident santé sources d'informations

U.S. Department of Health and Human Services
MedWatch
 The FDA Safety Information and Adverse Event Reporting Program
 General Instructions

Form Approved: OMB No. 0910-0291
 See OMB statement on reverse.

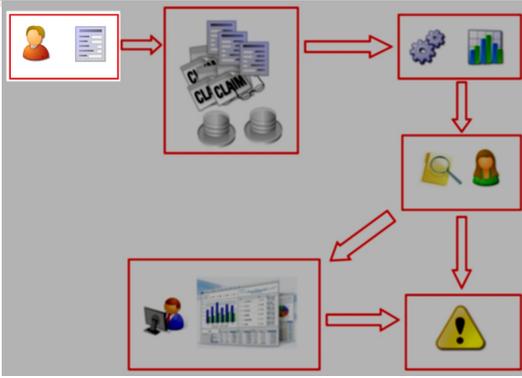
For VOLUNTARY reporting of adverse events, product problems and use errors

Page 1 of 2

FDA USE ONLY

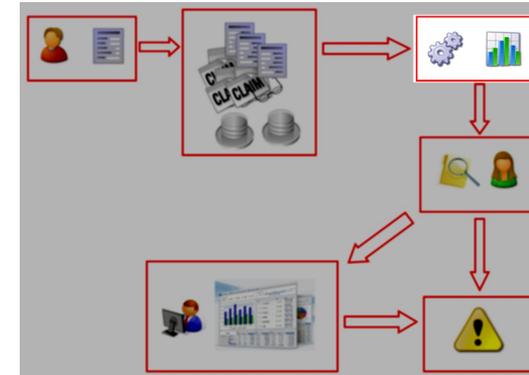
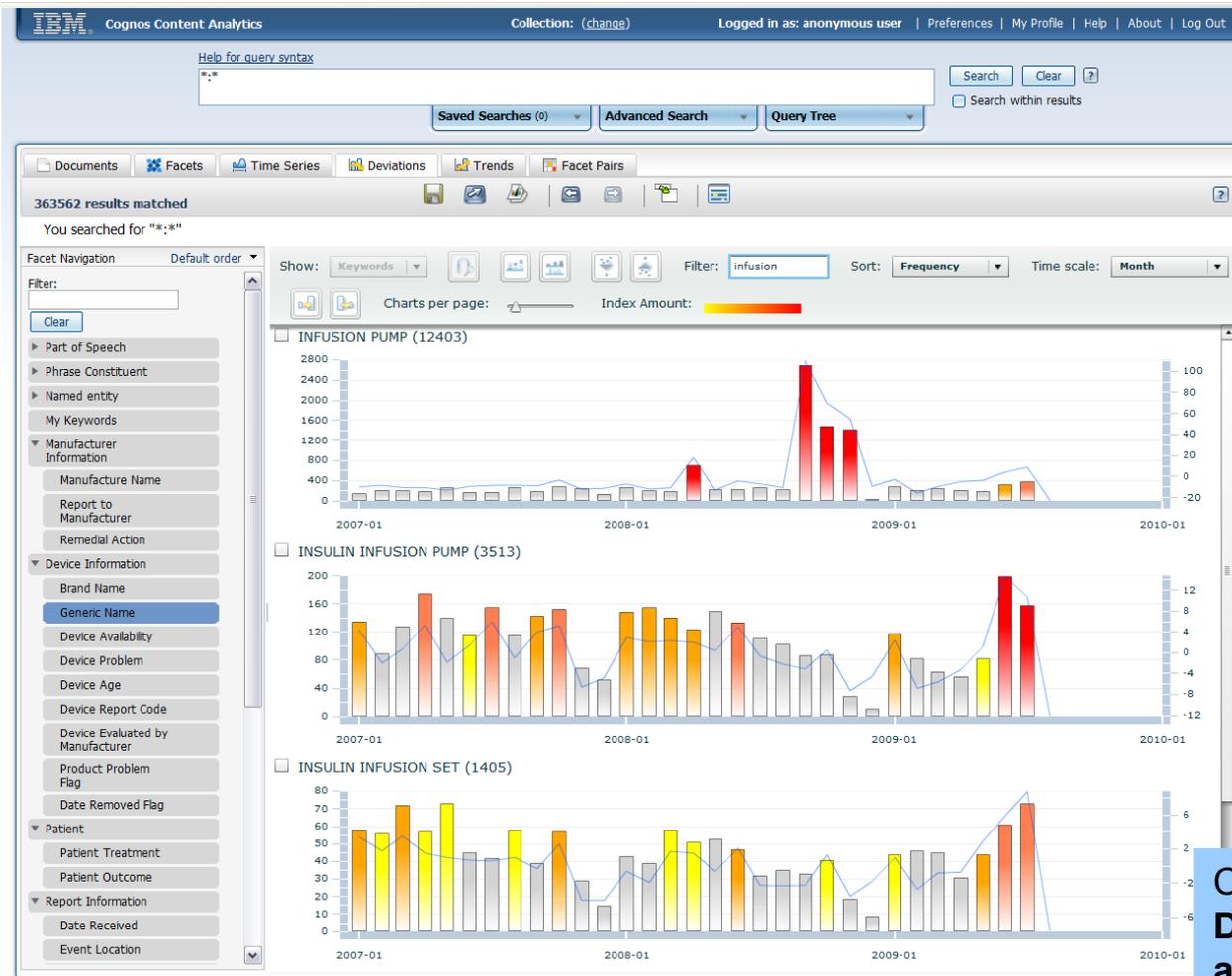
Triage unit # _____
 Sequence # _____

A. PATIENT INFORMATION				E. SUSPECT MEDICAL DEVICE	
1. Patient Identifier _____ <small>(In confidence)</small>	2. Age _____ <small>(In confidence)</small>	3. Sex <input type="checkbox"/> Female <input type="checkbox"/> Male	4. Weight _____ Lb or _____ Kg	1. Brand Name COLLEAGUE 3CX VOLUMETRIC INFUSION PUMP	
B. ADVERSE EVENT, PRODUCT PROBLEM OR ERROR				2. Common Device Name INFUSION PUMP	
1. Date of Event 2006-03-15		2. Date of this Report 2006-03-15		3. Manufacturer Name BAXTER HEALTHCARE PTE. LTD.	
3. Describe Event, Problem, or Product Use Error THE FACILITY REPRESENTATIVE REPORTED A BAD BATTERIES DURING BIOMED TESTING ON SITE. THE HOSPITAL REPRESENTATIVE DID NOT HAVE INFORMATON REGARDING WHETHER THERE HAVE BEEN ANY REPORTS OF ANY PATIENT INJURY OR MEDICAL INTERVENTION. NO ADDITIONAL INFORMATION IS AVAILABLE.				4. Device Problem _____ _____	5. Manufacturer Evaluation EVALUATION SUMMARY: THE REPORTED CONDITION OF DEPLETED BATTERIES WAS CONFIRMED. THE BATTERIES WERE REPLACED DUE TO POTENTIAL DAMAGE. REVIEW OF THE COMPLAINT HISTORY REVEALS SIMILAR REPORTS HAVE BEEN RECEIVED FOR THIS PRODUCT FOR THE REPORTED ISSUE. THIS ISSUE IS BEING INVESTIGATED UNDER CAPA. EVALUATION WAS CONDUCTED ON SITE



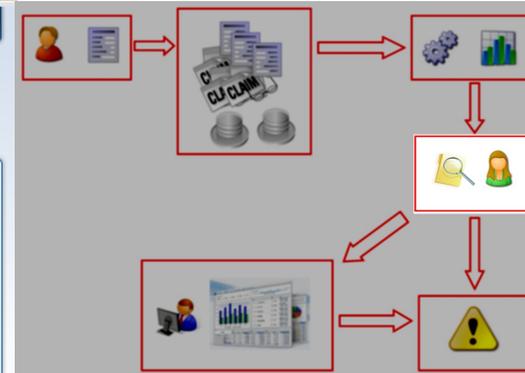
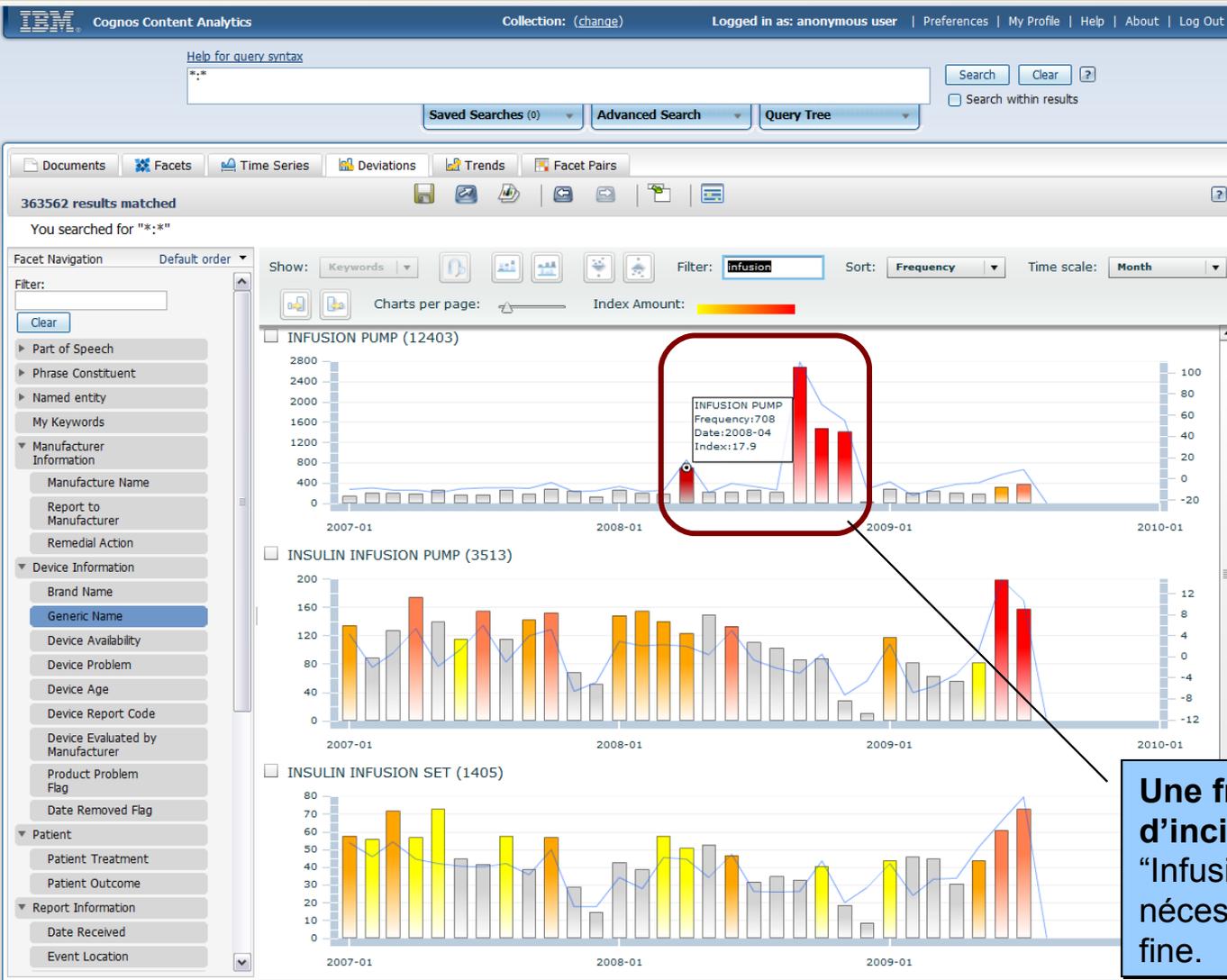
Les rapports d'incidents "MedWatch" de la FDA sont une source d'informations pour les **fabricants de matériel médical** pour comprendre les problèmes rapportés par les consommateurs sur leurs produits. Ils contiennent des informations **structurées & non structurées**. Un fabricant peut aussi **analyser le contenu interne**, comme les demandes de garantie ou les incidents de support.

Analyse d'incident santé analyse du contenu



Cette vue montre les **Deviations** (ou anomalies) au cours du temps pour la facette sélectionnée – ici, "Generic Device Name"

Analyse d'incident santé détection du matériel en cause



Une fréquence inhabituelle d'incidents concernant "Infusion Pump" en 2008, nécessitant une exploration plus fine.

Analyse d'incident santé investigation des origines

IBM Cognos Content Analytics Collection: (change) Logged in as: anonymous user | Preferences | My Profile | Help | About | Log Out

Help for query syntax
 /*:* /*keyword\$.deviceinformation.genericname/"INFUSION PUMP" /*date\$.month/"200804"
 Search Clear ?
 Search within results

Saved Searches (0) Advanced Search Query Tree

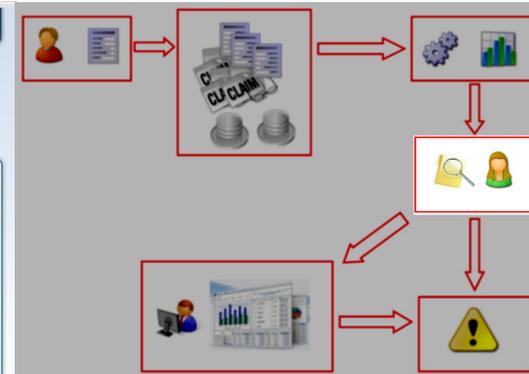
Documents Facets Time Series Deviations Trends Facet Pairs

708 results matched
 You searched for "/*:* /*keyword\$.deviceinformation.genericname/"INFUSION PUMP" /*date\$.month/"200804""

Facet Navigation Default order
 Filter:
 Clear

Part of Speech
 Phrase Constituent
 Noun Phrase
 Predicate Phrase
 Predicate with Adverb
 Noun - Predicate
 Verb - Noun
 Conjunction Phrase
 Named entity
 My Keywords
 Manufacturer Information
 Manufacture Name
 Report to Manufacturer
 Remedial Action
 Device Information
 Brand Name
 Generic Name
 Device Availability
 Device Problem
 Device Age
 Device Report Code
 Device Evaluated by Manufacturer
 Product Problem Flag
 Date Removed Flag

Keywords	Frequency	Correlation
<input type="checkbox"/> report ... issue	397	28.5
<input type="checkbox"/> report ... condition	219	31.1
<input type="checkbox"/> investigate ... under	213	30.2
<input checked="" type="checkbox"/> deplete ... battery	126	63.2
<input type="checkbox"/> be ... report	113	3.2
<input type="checkbox"/> display ... error	108	10.9
<input type="checkbox"/> be ... UNK	104	6.6
<input type="checkbox"/> be ... unknown	98	4.2
<input type="checkbox"/> be ... known	87	12.4
<input type="checkbox"/> circuit ... board	45	34.7
<input type="checkbox"/> be ... defective	38	19.4
<input checked="" type="checkbox"/> damage ... battery	33	54.2
<input type="checkbox"/> replace ... due	32	18.9
<input type="checkbox"/> print ... circuit	32	26.1
<input type="checkbox"/> be ... in	27	1.8
<input type="checkbox"/> confirm ... in	21	13.8
<input type="checkbox"/> level ... were	16	23.3
<input type="checkbox"/> detail ... were	14	7.5
<input type="checkbox"/> obtain ... information	14	11.4
<input type="checkbox"/> report ... failure	13	6.6
<input type="checkbox"/> be ... due	13	2.0
<input type="checkbox"/> will ... be	13	1.4
<input type="checkbox"/> level ... over	10	62.3
<input type="checkbox"/> wake ... UP	8	5.2
<input type="checkbox"/> be ... use	8	6.0



Utilisant les **Facettes** sur les **Composants de phrase**, nous remarquons la référence fréquente à des **problèmes de batteries** pour les incidents "Infusion Pump 2008". Nouvelle exploration plus fine..

Analyse d'incident santé connaissance des causes

IBM Cognos Content Analytics Collection: (change) Logged in as: anonymous user | Preferences | My Profile | Help | About | Log Out

Help for query syntax
 /*:* /*keyword\$.deviceinformation.genericname/"INFUSION PUMP"/"date\$.month"/"200804" (/*keyword\$.phrase.pred_phrase.verb_noun/"deplete ... battery" OR /*keyword\$.phrase.pred_phrase.verb_noun/"damage ... battery" OR /*keyword\$.phrase.pred_phrase.verb_noun/"fail ... battery")

Search Clear ?
 Search within results

Saved Searches (0) Advanced Search Query Tree

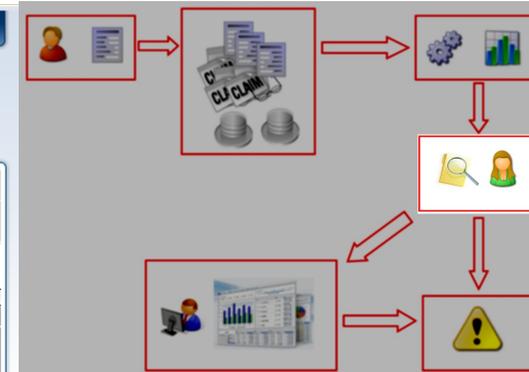
Documents Facets Time Series Deviations Trends Facet Pairs

Results 1-10 of 154 (154 results matched) View by file type: ALL Results per Page 10 1 2 3 4 5 6 7 8 9 10

You searched for "/*:* /*keyword\$.deviceinformation.genericname/"INFUSION PUMP"/"date\$.month"/"200804" (/*keyword\$.phrase.pred_phrase.verb_noun/"deplete ... battery" OR /*keyword\$.phrase.pred_phrase.verb_noun/"damage ... battery" OR /*keyword\$.phrase.pred_phrase.verb_noun/"fail ... battery")"

Facet Navigation	Default order	Source	Relevance	Date	Title	Thumbnail
Filter:		Windows file system	100.00%	4/4/08	MAUDE-1023177.xml	
INFUSION PUMP BAXTER HEALTHCARE PTE. LTD. NA 2M8153 NA NA Device was returned to ... DEVICE EVALUATION WS COMPLETED AND THE FAILED BATTERY TEST REVEALED DEPLETED BATTERIES. THE ...		Windows file system	100.00%	4/3/08	MAUDE-1023097.xml	
INFUSION PUMP BAXTER HEALTHCARE PTE. LTD. NA 2M8151 NA NA Device was returned to ... WAS CONFIRMED (3 DISCHARGES BELOW ALARM THRESHOLD FOUND) DUE TO POTENTIALLY DAMAGED BATTERIES. ... DURING SERVICE TESTING, DEPLETED BATTERIES WERE FOUND. ACCORDING TO THE HOSPITAL REPRESENTATIVE, ...		Windows file system	100.00%	4/3/08	MAUDE-1023105.xml	
INFUSION PUMP BAXTER HEALTHCARE PTE. LTD. NA 2M8163 NA NA Device was returned to ... AND THE DEPLETED BATTERY CONDITION WAS CONFIRMED DUE TO POTENTIALLY DAMAGED BATTERIES. BAXTER ... DURING SERVICE TESTING DEPLETED BATTERIES WERE FOUND. ACCORDING TO THE HOSPITAL REPRESENTATIVE ...		Windows file system	100.00%	4/3/08	MAUDE-1023116.xml	
INFUSION PUMP BAXTER HEALTHCARE PTE. LTD. NA 2M8163 NA NA Device was returned to ... AND THE DEPLETED BATTERY CONDITION WAS CONFIRMED DUE TO POTENTIALLY DAMAGED BATTERIES. BAXTER ... DURING SERVICE TESTING DEPLETED BATTERIES WERE FOUND. ACCORDING TO THE HOSPITAL REPRESENTATIVE ...		Windows file system	100.00%	4/3/08	MAUDE-1023106.xml	

Facet Navigation: Part of Speech, Phrase Constituent, Named entity, My Keywords, Manufacturer Information, Device Information (Brand Name selected), Patient, Report Information



Par la vue Documents, nous pouvons prendre connaissance du **texte source original** sur 154 incidents de batterie sur "Infusion Pump". Le texte significatif est mis en **évidence**.

Analyse d'incident santé rapport des types de matériels

IBM Cognos Content Analytics

Collection: (change) Logged in as: anonymous user | Preferences | My Profile | Help | About | Log Out

Help for query syntax

/*:*/keyword\$.deviceinformation.genericname/"INFUSION PUMP"/"date\$.month"/"200804" (/keyword\$.phrase.pred_phrase.verb_noun/"deplete ... battery" OR /keyword\$.phrase.pred_phrase.verb_noun/"damage ... battery" OR /keyword\$.phrase.pred_phrase.verb_noun/"fail ... battery")

Search Clear ?

Search within results

Saved Searches (0) Advanced Search Query Tree

Documents Facets Time Series Deviations Trends Facet Pairs

154 results matched

You searched for "/*:*/keyword\$.deviceinformation.genericname/"INFUSION PUMP"/"date\$.month"/"200804" (/keyword\$.phrase.pred_phrase.verb_noun/"deplete ... battery" OR /keyword\$.phrase.pred_phrase.verb_noun/"damage ... battery" OR /keyword\$.phrase.pred_phrase.verb_noun/"fail ... battery")"

Facet Navigation Default order

Filter:

Clear

Part of Speech

Phrase Constituent

Named entity

My Keywords

Manufacturer Information

Manufacture Name

Report to Manufacturer

Remedial Action

Device Information

Brand Name

Generic Name

Device Availability

Device Problem

Device Age

Device Report Code

Device Evaluated by Manufacturer

Product Problem Flag

Date Removed Flag

Patient

Patient Treatment

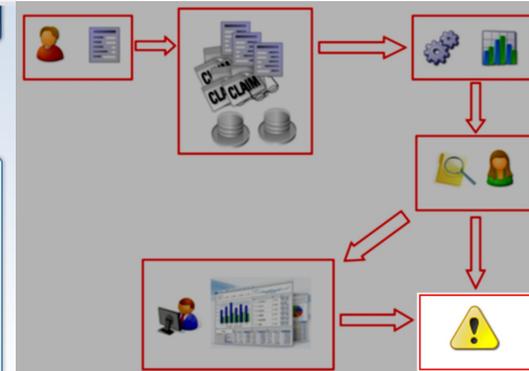
Patient Outcome

Report Information

Date Received

Event Location

Keywords	Frequency	Correlation
<input type="checkbox"/> COLLEAGUE CX VOLUMETRIC INFUSION	56	91.8
<input type="checkbox"/> COLLEAGUE VOLUMETRIC INFUSION PU	41	107.1
<input type="checkbox"/> COLLEAGUE 3CX VOLUMETRIC INFUSIO	22	58.3
<input type="checkbox"/> COLLEAGUE 3 VOLUMETRIC INFUSION	14	60.8
<input type="checkbox"/> COLLEAGUE CX VOLUMETRIC INFUSION	11	1,178.6
<input type="checkbox"/> COLLEAGUE 3CX VOLUMETRIC INFUSIO	7	1,232.1
<input type="checkbox"/> COLLEAGUE TRIPLE CHANNEL VOLUME	2	55.3
<input type="checkbox"/> REFURBISHED COLLEAGUE VOLUMETRI	1	49.2

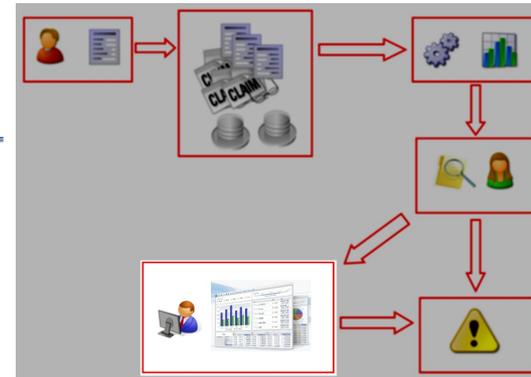


Par la **facette Brand**, la **synthèse, avec fréquence et corrélation**, indique les types de matériels concernés par ces incidents ("COLLEAGUE").

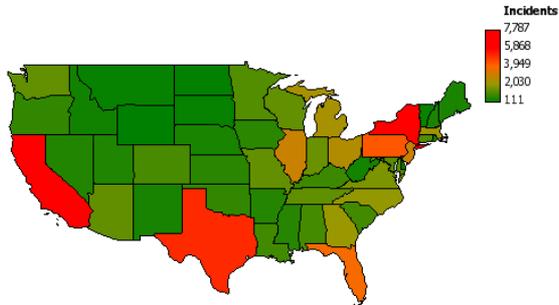
Analyse d'incident santé tableau de bord de suivi



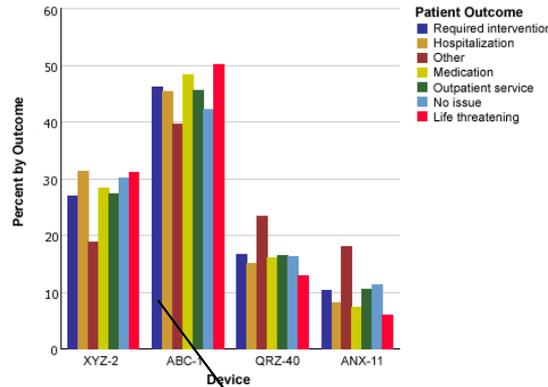
Quality Early Insight



Incidents by State



Patient Outcomes by Device

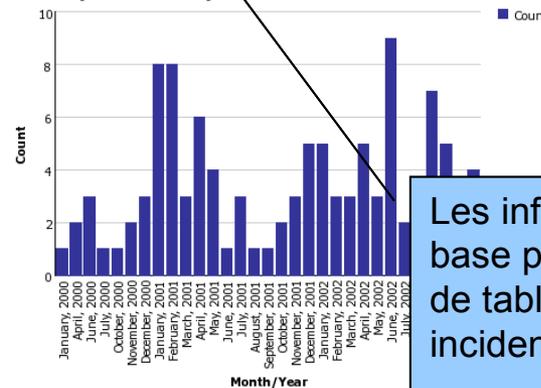


Top correlated medications reported by device

Top medications for XYZ-2	Count
MOTRIN	1,066
TYLENOL	808
IBUPROFEN	452
VICODIN	435
TORADOL	299
KEFLEX	279
BENADRYL	214
LORTAB	207
TETANUS TOXOID	179
LIDOCAINE	177
PERCOCET-5	160
MORPHINE	145
ATIVAN	140
TYLENOL NO. 3	137

Top medications for ABC-1	Count
TYLENOL	425
VICODIN	353
TORADOL	222
TETANUS TOXOID	218
DARVOCET-N	214
MORPHINE	205
DEMEROL	171
MOTRIN	169
TETANUS DIPHTHERIA TOXOID	159
PERCOCET-5	132
LORTAB	131
KEFLEX	130
NORMAL SALINE	100
IBUPROFEN	96

Battery mentions by Date



Les informations peuvent servir de base pour la génération d'états et de tableaux de bords de suivi des incidents..

Analyse d'incident santé suivi des actions correctives

U.S. Department of Health & Human Services

FDA U.S. Food and Drug Administration

Home | Food | Drugs | Medical Devices | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Radiation

News & Events

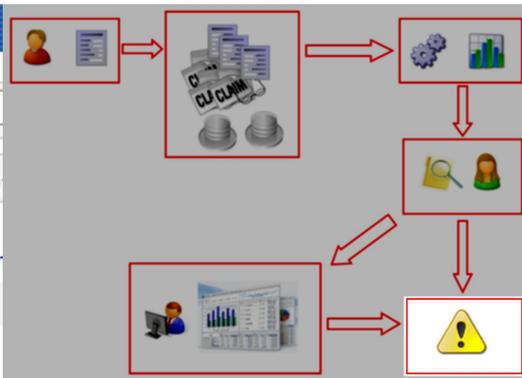
Home > News & Events > Newsroom > Press Announcements

FDA NEWS RELEASE

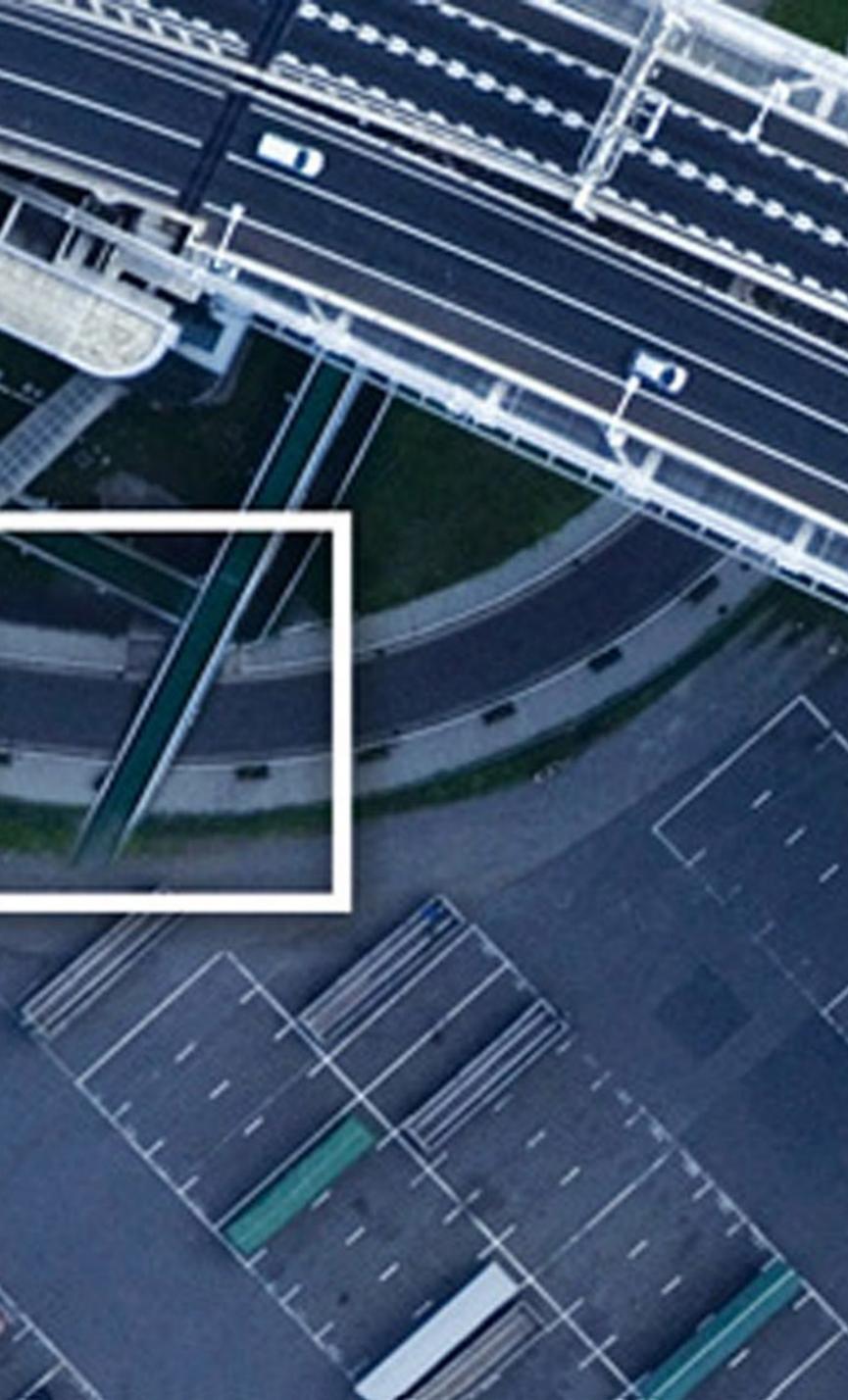
For Immediate Release: May 3, 2010
Media Inquiries: Dick Thompson, 240-753-3049, dick.thompson@fda.hhs.gov
Consumer Inquiries: 888-INFO-FDA

FDA Issues Statement on Baxter's Recall of Colleague Infusion Pumps

The U.S. Food and Drug Administration sent a letter to Baxter Healthcare Corp. on April 30 ordering the company to recall and destroy all of its Colleague Volumetric Infusion Pumps (Colleague pumps) currently in use in the United States. This action is based on a longstanding failure to correct many serious problems with the pumps. The FDA believes there may be as many as 200,000 of those pumps currently in use.



L'organisme officiel de santé, FDA ordonne au fabricant de rappeler les produits défectueux ("COLLEAGUE").



BIG DATA & BIG DATA GOUVERNANCE

DOCUMENTATION

MARS 2012



Renaud Finaz de Villaine

MICROPOL 
Driving Distinction

OLF 



Sommaire

- **L'explosion de l'information**
- **BIG DATA**
- **Big Data & nouveaux concepts BI**
- **Big Data & Gouvernance**

EXPLOSION DE L'INFORMATION



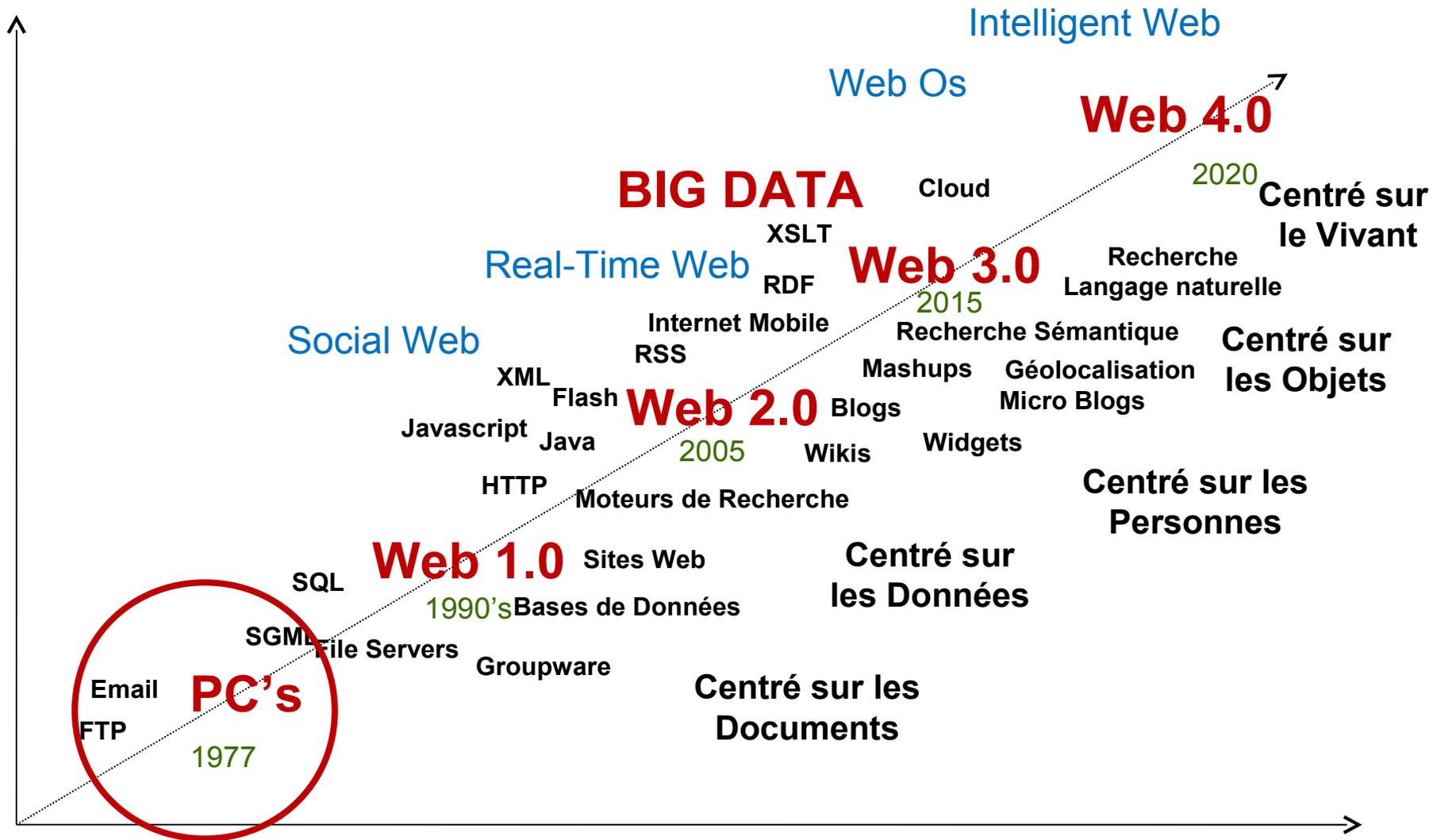
On devrait passer de 1,6 Zettabytes en 2010 à 35 Zettabytes en 2020 !!

- ➔ 30 milliards de données (messages, photos) sont ajoutées sur Facebook tous les mois.
- ➔ Plus de 2 milliards de vidéos ont été visionnées sur Youtube ... hier.
- ➔ 32 milliards de recherches ont été effectuées le mois dernier sur Twitter.

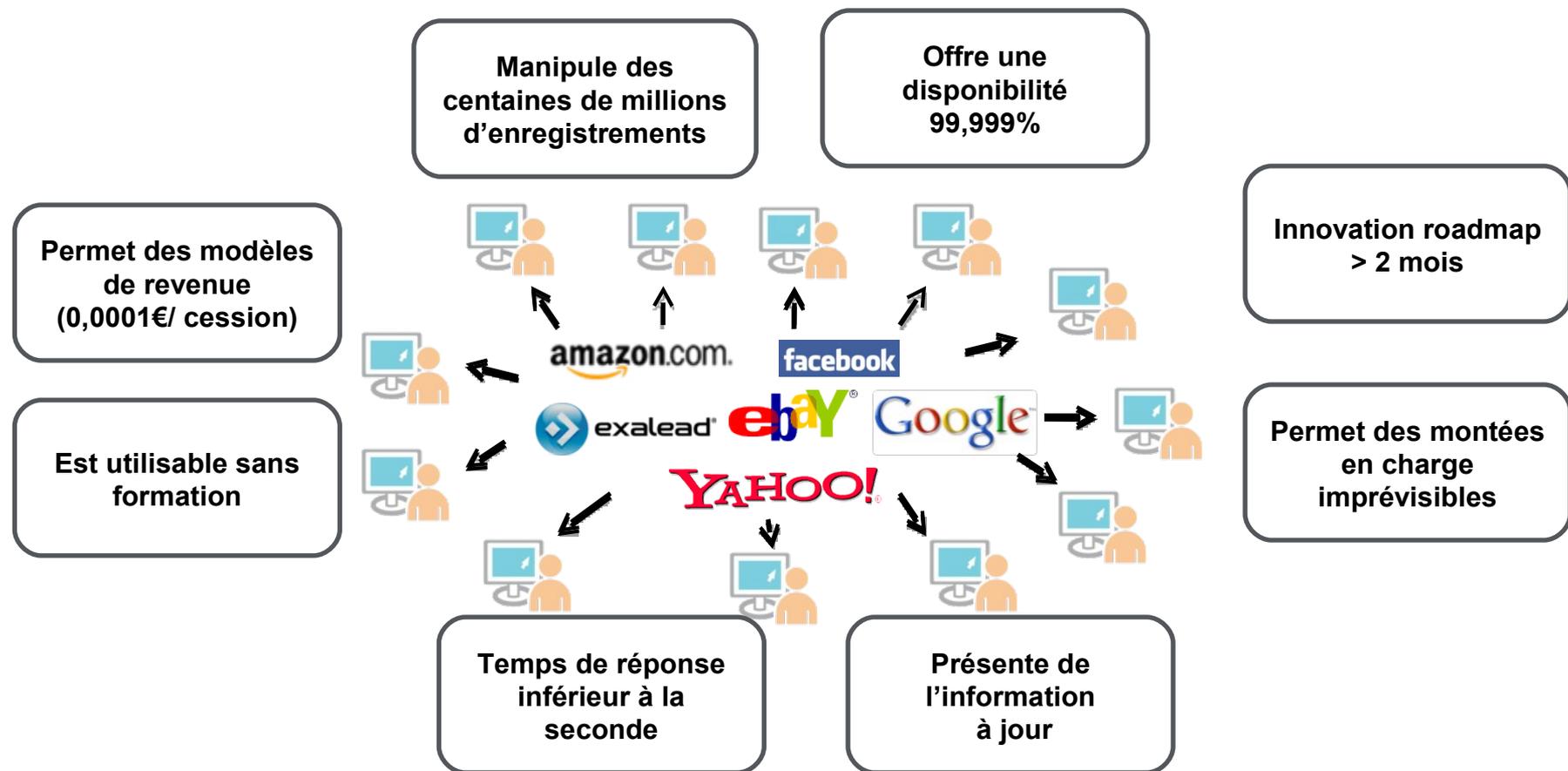
Pour mémoire :

- 1 zettabyte = 1 000 Exabyte
- 1 exabyte = 1 000 Petabyte
- 1 Petabyte = 1 000 Terabyte

QUELLE ÉVOLUTION POUR LE WEB?



LE WEB CHANGE LE JEU !



LES CRÉATEURS DU BIG DATA

Google

- 1 million de serveurs répartis sur 39 Data Centers
- 2,5 milliards de requêtes par jour sur l'ensemble des sites de Google (chiffre juillet 2009)
- 2,5% de la consommation électrique américaine (est. 2009)
- 400 serveurs perdus quotidiennement

Twitter

- Twitter est passé de 2 M/Tweets/jour en janvier 2009 à **250 M/jour** en octobre 2011.

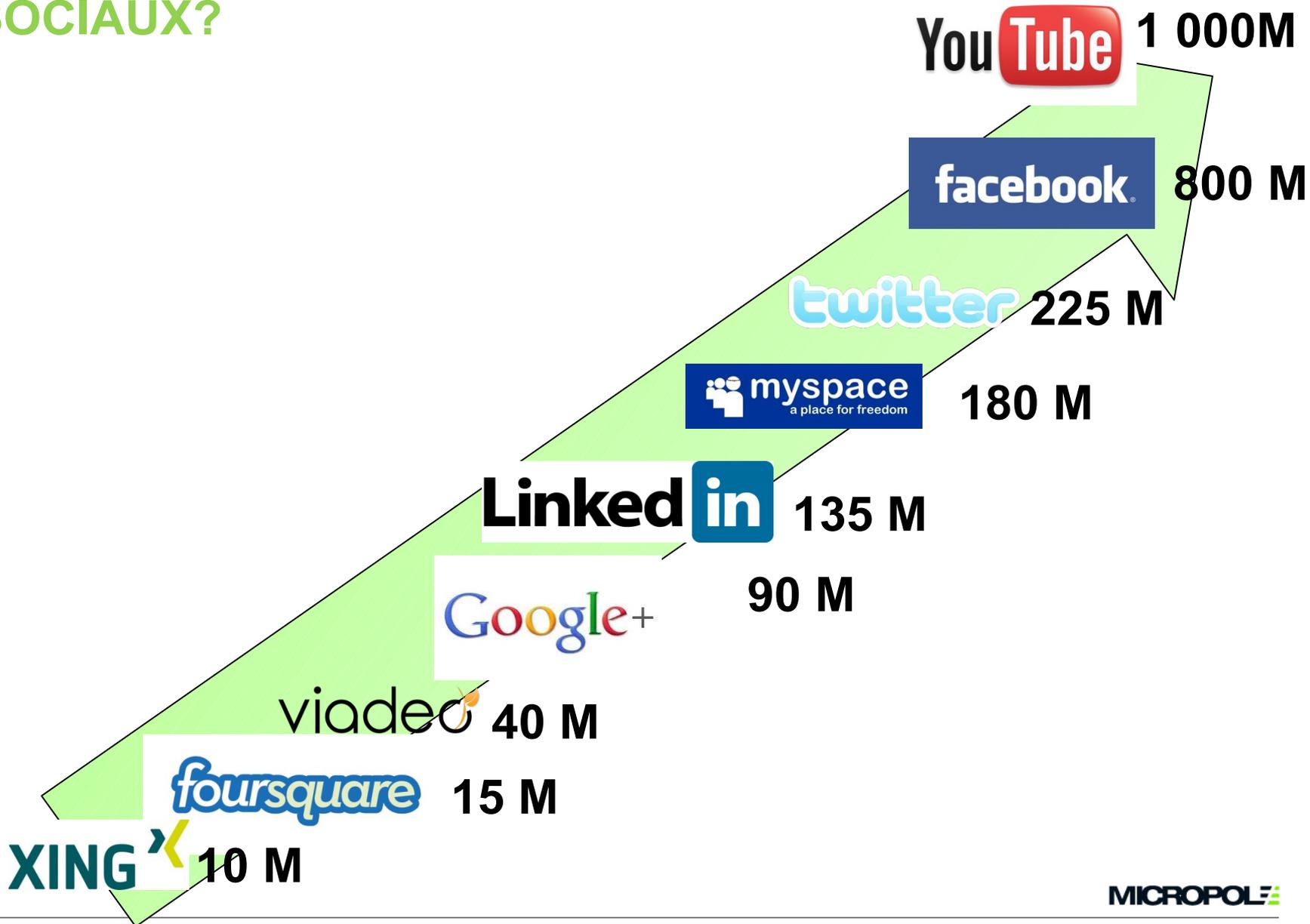
eBay

- 250 M d'utilisateurs enregistrés
- 1 milliard de pages vues au quotidien
- 44 milliards de requêtes /jour
- 3 800 recherches/semaine
- 16 000 serveurs d'applications
- 1,5 Tb de log par jour
- 6h d'indisponibilité par an (99,94%)

Amazon.com

- 60 millions de clients
- > 1 000 services

COMBIEN D'UTILISATEURS SUR LES RÉSEAUX SOCIAUX?





LES PROBLÉMATIQUES POSÉES À L'ENTREPRISE

LES NOUVELLES PROBLEMATIQUES : LES 3 V

→ Volumétrie

鏹 Flux de données structurées, semi et non structurées à :

- Stocker, Analyser, Visualiser, Diffuser

鏹 Ex : Gestion des stocks, des commandes, transport /logistique /RFID, transactions financières, services grand public, géolocalisation, telcos...

→ Variété

鏹 Nouveaux types de requêtes dus à une vision 360°

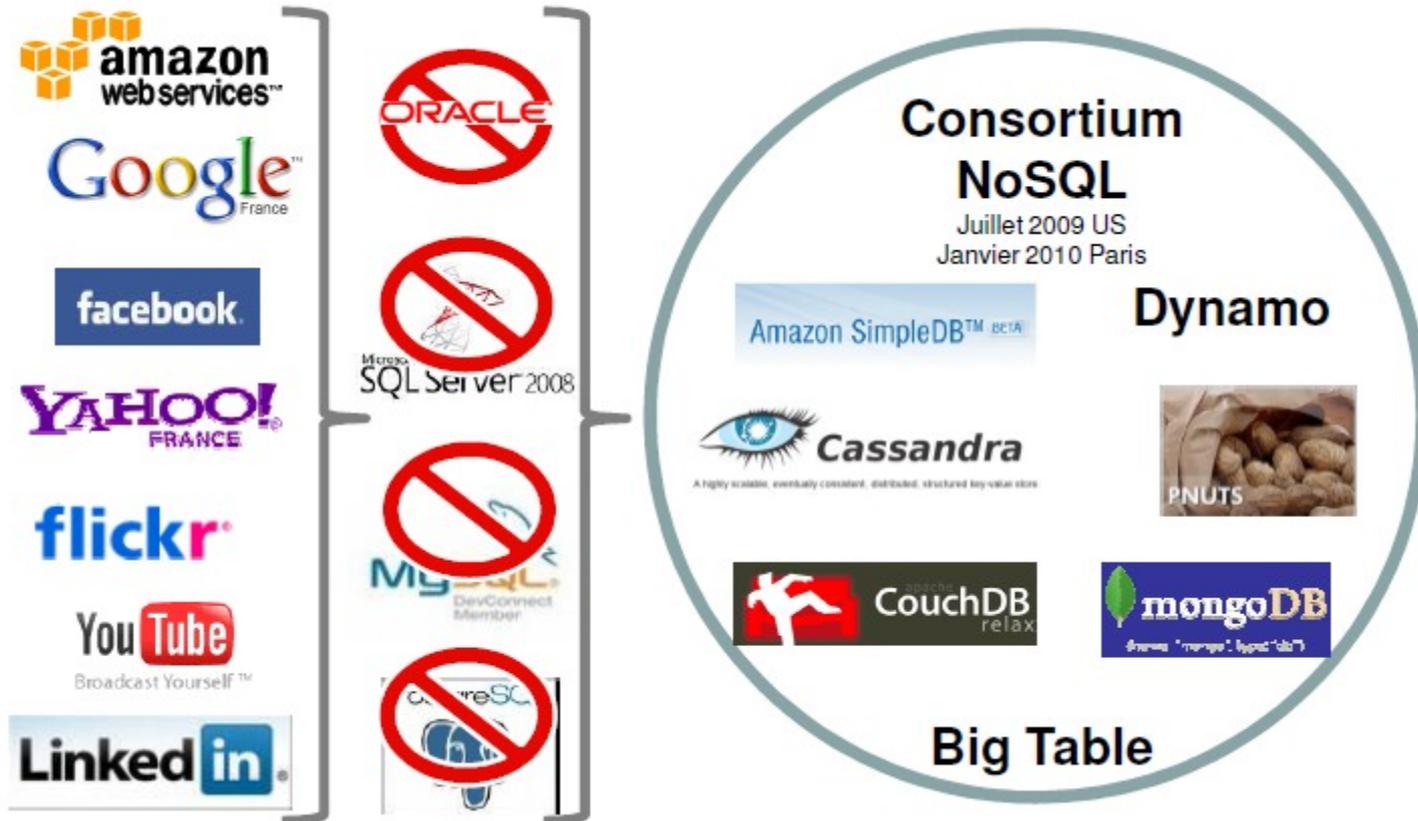
- CRM, BI, ERP, Marketing, SAV, e-reputation, communication financière et clients, données streaming...

→ Vitesse (pas d'attente)

鏹 Gestion des stocks, des commandes, mode, jeu, électronique grand public...

Une remise en cause des standards

REMISE EN CAUSE DES STANDARDS



■ ■ AVEC EN PLUS.....

→ Performance

鏞 Requetes plus rapides

鏞 Grands nombres d'écritures

鏞 Parallélisations des traitements

→ Scalabilité

鏞 N petites machines plutôt qu'une grosse machine !

鏞 Rajout ou suppression aisé de ressources pour répondre aux pics de charge

→ Résilience / Disponibilité

鏞 Sur N machines les panes sont isolées

鏞 Back up restore plus rapide

鏞 Des niveaux d'indisponibilité très élevé : quelques heures/an

→ Approche plateforme

鏞 Pour répondre à des multiples services

QUELLES RÉPONSES ?

→ Augmenter la puissance disponible

鏞 la loi de Moore tient bon

鏞 les fermes de serveurs explosent

鏞 l'empreinte écologique et énergétique aussi

→ Virtualisation & Cloud-computing

→ Le Big data

→ De nouveaux concepts BI

鏞 SBA (Search Based Application)

鏞 BI Search

鏞 BI Social





1. LE BIG DATA ET LES BASES NOSQL

→ Objectifs :

銅 Traiter un volume très important de données aussi bien structurées que non structurées, se trouvant sur des terminaux variés (PC, smartphones, tablettes, objets communicants...), produites ou non en temps réel depuis n'importe quelle zone géographique dans le monde.

→ 3 types de Technologies :

銅 Une accélération matérielle à l'aide de mémoires dynamiques DRAM ou Flash

銅 Le recours à des bases de données massivement parallèles (Massively Parallel Processing)

銅 Des solutions utilisant des formats de bases de données non relationnelles basées sur NoSQL.

QUELQUES ACTEURS DU BIG DATA

EMC²

IBM

SAP

1010
data

GREENPLUM.
A DIVISION OF EMC

Microsoft ORACLE

INFOBRIGHT

kognitio
Competitive advantage from data

PARACCEL™

aster data
— more data. big insights. —

amazon
web services™

COUCHBASE

DataStax

hadoop

HADAPT

KARMASPHERE

pentaho
POWERFUL ANALYTICS MADE EASY™

MAPR™
TECHNOLOGIES
EASY. DEPENDABLE. FAST.

PERVASIVE

REVOLUTION
ANALYTICS

splunk >

mahout

cloudera

Bixo

Apache™
Tika

VERTICA

■ LES DÉFIS À RELEVER DANS LES ANNÉES À VENIR

- Fini la prédominance des bases de données relationnelles et du langage SQL !
- Les nouvelles bases sont orientées **colonnes** (Vertica), **graphes** (Neo4J), **documents** (MongoDB). Sans compter celles **en mémoire** (VoltDB) ou spécialisées dans le traitement **d'événements complexes** (Streambase).
- Un autre paradigme du Big Data est l'**architecture distribuée**, laquelle étale les traitements sur un grand nombre de machines. Les technologies **Hadoop et Mapreduce** sont particulièrement sollicitées. On les trouve, par exemple, chez les géants du web : Facebook, eBay, LinkedIn, Twitter et Yahoo.
- Les **techniques de visualisation** constituent un autre défi, qui retient l'attention de beaucoup de chercheurs. En effet, rien ne sert d'extraire la substantifique moelle d'une montagne de données si personne ne peut la comprendre et l'interpréter.



2. LES SBA (Search Based Application)

UNE NOUVELLE APPROCHE : LE SBA

→ Un nouveau concept le **SBA** (Search Based Application)

鑄 Conçues sur la base de technologies de type Index et Web, les SBAs offrent aux entreprises les moyens de concevoir rapidement des services d'accès à l'information innovants pour leurs collaborateurs, leurs clients et leur partenaires sans bouleverser les architectures en place.

鑄 Il s'agit donc de construire des applications qui vont reposer non plus sur une base de données mais sur **l'index d'un moteur de recherche**. Les écrans de l'application se basant sur un index, sont à créer *from scratch*, en fonction du besoin métier (BI, CRM, ...) auquel on souhaite répondre

→ Trois acteurs français

鑄 Exalead, leader dans le domaine du SBA, plateforme d'intégration de données multimédia sémantique non intrusive

鑄 Sinequa, qui commence sa mutation du search d'entreprise vers le SBA !

鑄 Polyspot

Zone de recherche « langage naturel » Exalead

Composant outil BI

Fonctions de recherche basées sur des sources internes inexploitées

Zone de paramètres personnalisés

exalead | (Suivi du lancement produit X) | Recherche | Recherche avancée

Dashboard Lancement produit

[Historique des recherches](#)

[Vos paramètres](#)
 Articles récents
 Mails
 Mots clés concurrents
 Société xxxxx
 Société yyyy
 Période d'analyse
 3 dernières semaines

[Modifiez votre profil](#)

Suivi du volume de ventes / canaux

Canal	Semaine 1	Semaine 2	Semaine 3	Semaine 4	Semaine 5
Internet	1.0	1.0	1.0	1.0	1.0
Librairie spécialisée	2.0	2.0	2.0	2.0	2.0
Distributeur 1	2.0	2.0	2.0	2.0	2.0
Distributeur 2	4.0	4.0	4.0	4.0	4.0
Distributeur 3	4.0	4.0	4.0	4.0	4.0
Distributeur 4	5.0	5.0	5.0	5.0	5.0

Informations canaux de ventes

Alertes Internes

Thèmes	Source	Joindre ce correspondant	Pertinence
Accueil positif qualité	Reporting interne mail	Pierre dupond Alert Tran	75% 80%
Trop cher	outil crm	xxxx xxxx	100% 90%
Arrivée tardive	outil crm	yyy yyy	72% 20%

Sur le net aujourd'hui

Comparatif de notoriété (parution sur le web)

	semaine 10	semaine 11	semaine 12
Produit X	5	12	17
pdt concurrent 1	14	10	8
pdt concurrent 2	8	10	16
pdt concurrent 3	22	18	24

Focus semaine 12

Produit	Notoriété
Produit X	37%
pdt concurrent 1	26%
pdt concurrent 2	12%
pdt concurrent 3	25%

Informations canaux de ventes

Derniers mail échangés « lancement »

Ceux qui ont acheté le produit X achètent aussi

Indicateurs construits par le moteur de recherche Exalead et consolidés par un outil de BI

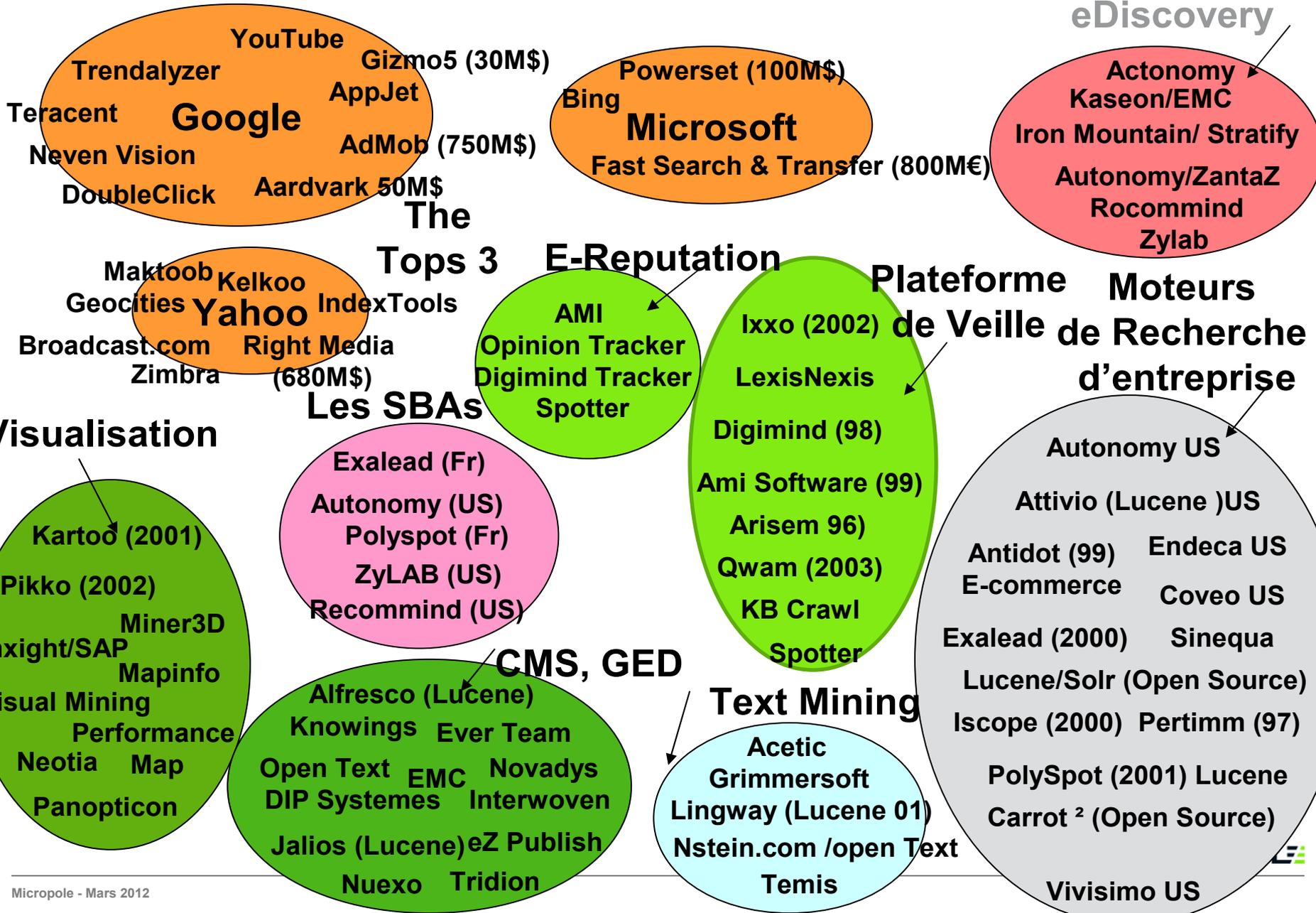
Associations par des fonctions classiques de recherche

Exploitation des données structurées

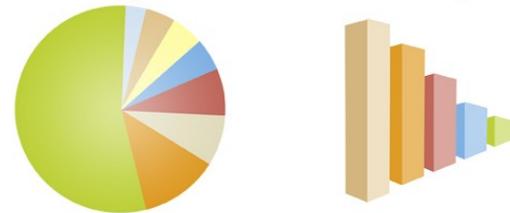


3. LE SEARCH BI

LES MOTEURS DE RECHERCHE 2010'S



APPROCHE N°3 : LE SEARCH BI

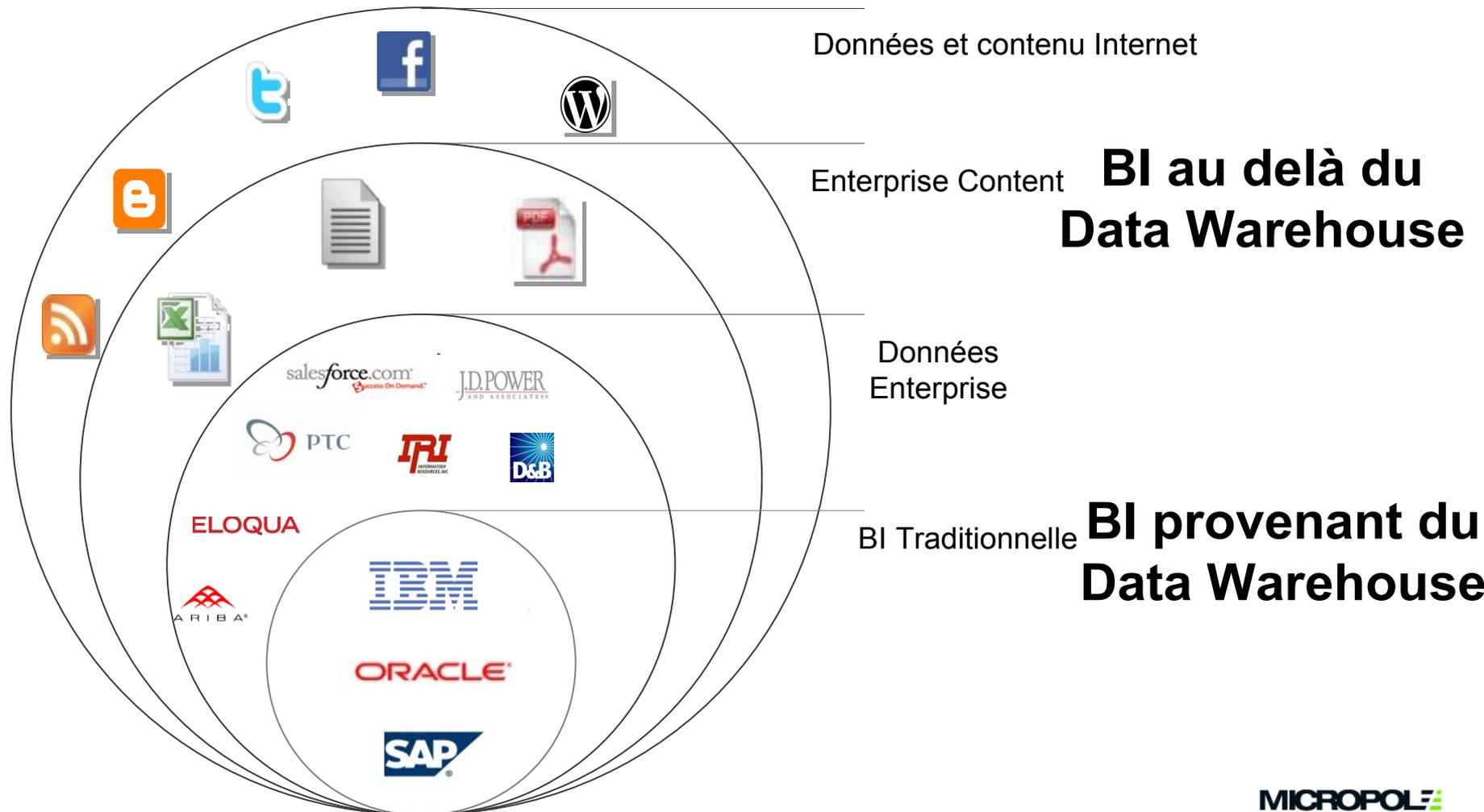


Simplicité du search

Puissance de la BI

*Aider les utilisateurs non avertis, à trouver ce qu'ils cherchent,
et à comprendre ce qu'ils trouvent*

INTERACTIVE ANALYTICS POUR LA BI AU DELA DU DATA WAREHOUSE



AUCUNE DONNEE OUBLIEE : LES SCHEMAS SONT ISSUS DES DONNEES

Original system



Drug Description

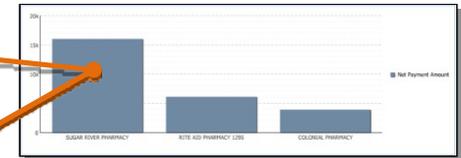
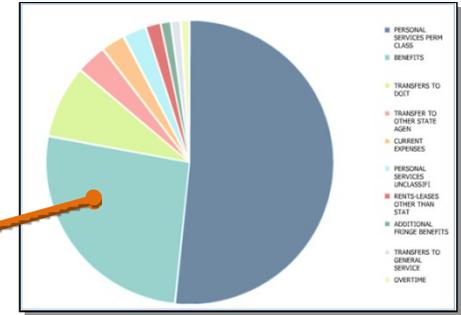
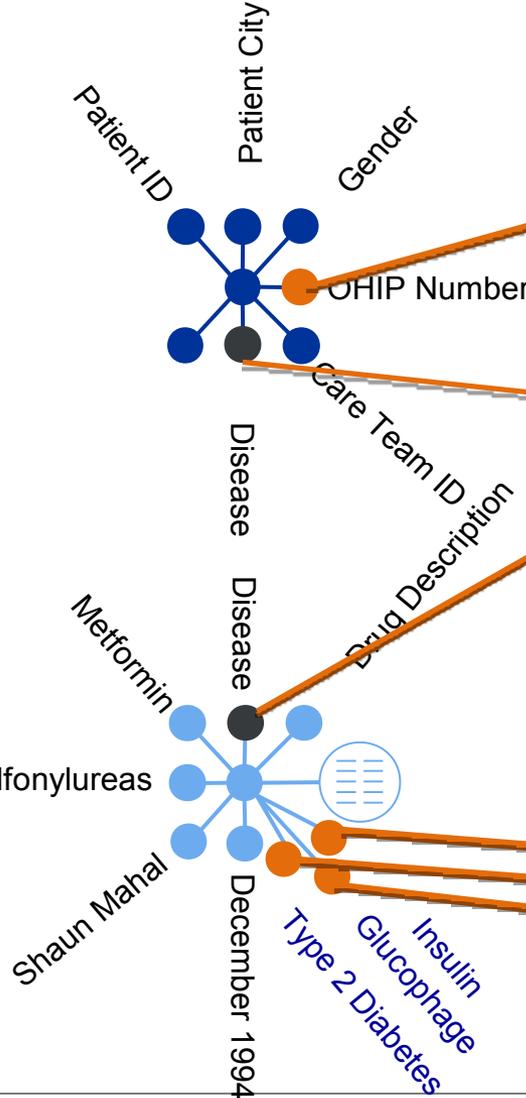
Shaun Mahal | June 20, 2010



Metformin

Metformin was approved for use in the U.S. for treatment of **type 2 diabetes** in **December, 1994**. It is sold under the brand name **Glucophage** and is also available generically. Metformin is approved for treatment with **sulfonylureas**, or with insulin, or as monotherapy (by itself). Glucophage XR Extended Release tablets, a once daily version of metformin, is available. Also, metformin is available...

Endeca index



Range Filters

Medicaid Percentile Ranking

Between 1 - 100 Submit

Guided Navigation

Records

Record Type

- Professional Claim (6859728)
- Pharmacy Claim (1499446)
- Claim (632658)
- Facility Claim (598796)
- Foodstamp (535111)
- Other (244403)
- Cash (190701)

Medicaid Groupings

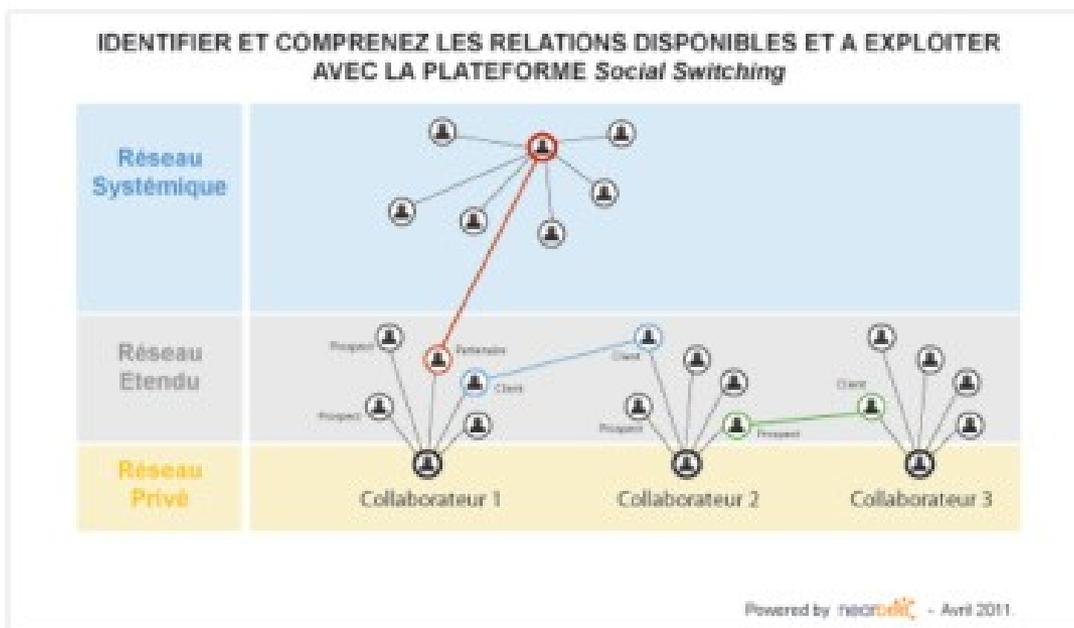
Age Group Service

- 19-64 (4465075)
- <= 18 (3211203)
- >=65 (1281692)



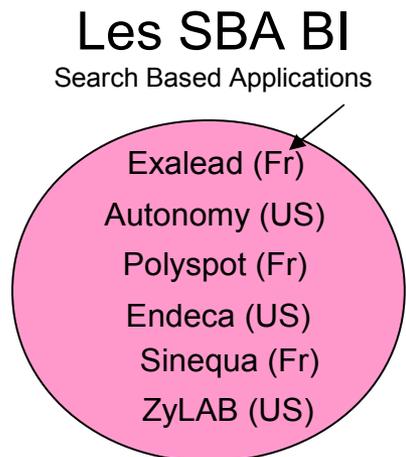
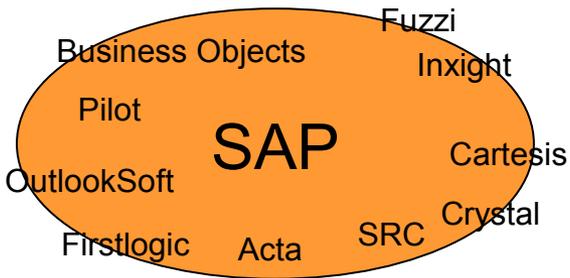
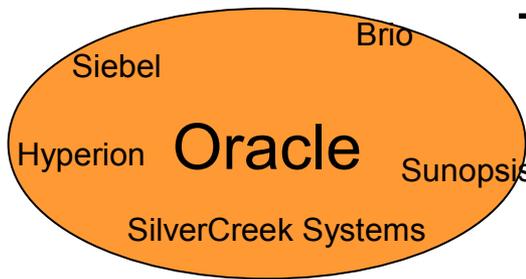
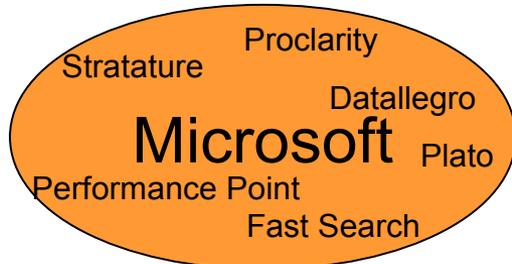
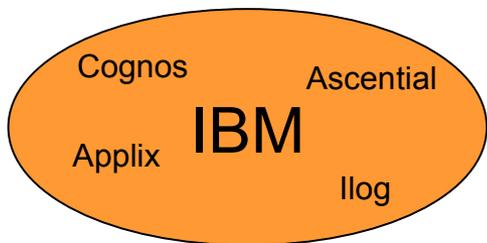
4. LE SOCIAL BI

Le «*Social BI*», se consacre à la valorisation des données sociales, donc humaines, des différents réseaux sociaux de l'entreprise (Privé, Etendu, Systémique) afin de proposer en temps réel les **meilleures mises en relation**, pour **accroître les ventes, identifier un futur partenaire** ou encore **localiser une expertise**

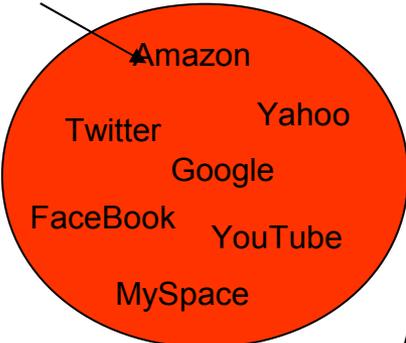


BI, LA CONCENTRATION DES ACTEURS CONTINUE

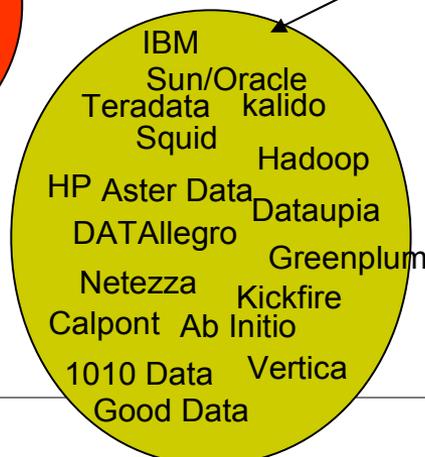
The 4 Tops



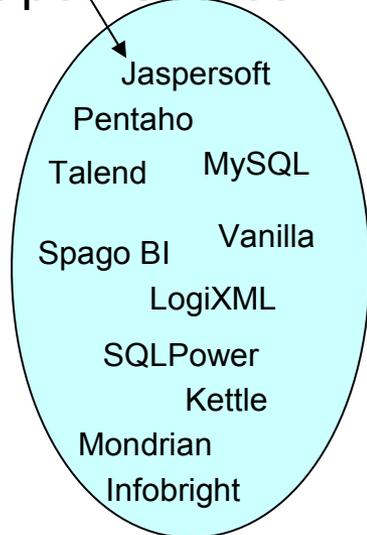
Les troubles Fêtes NoSQL DW



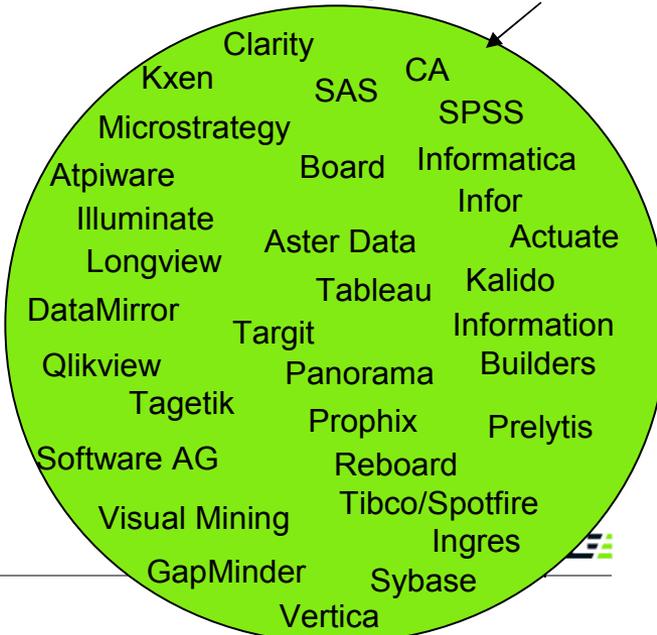
Data Appliances & BigData



Les nouveaux de l'Open Source



Les derniers indépendants



Source : R. FdV
Micropole 2012



GOUVERNANCE DE L'INFORMATION

LA GOUVERNANCE DE L'INFORMATION CONSISTE À

- S'assurer que l'information est identifiée, précise, authentique, collectée, classée, stockée
- S'assurer que ce classement correspond aux exigences de conformité et de règles en vigueur pour chaque domaine d'activité
- Permettre l'utilisation la plus rapide possible de cette information en vue d'alimenter les circuits de décision
- Garantir une collecte et un stockage au moindre coût
- Faire correspondre l'organisation de l'information ainsi stockée à l'organisation de l'entreprise

LE BUT :

- Améliorer l'efficacité des process et de l'entreprise
- Augmenter les revenus

COMMENT ? AGIR SUR PLUSIEURS LEVIERS

Revenu

Up-selling/Cross-selling
Segmentation Marketing
Satisfaction Client
Vue 360°

Efficacité

Supply Chain
Réutilisation des données
Analytics
Dashboard/ Rapports
Les projets qui marchent

Compliance

Sarbanes-Oxley
Bale II/Bale III
Réglementations locales
ou Internationales

Règles sur les données

Process d'entreprise
Outils unifiés
Coordination des data workflow
Data Ownership

Collaboratif

Communication
sur la Data Strategy
Meeting Business/IT
Blogs....

Formation

Partage des valeurs
Incentives
Formation
Reporting KPI

Data Analytics

BI
Reporting
Big Data Analytics
DW / Data Marts

Business Apps

CRM
ERP
Supply Chain
Facturation
RSE/Collaboratif

MDM

Master data
Qualité de données
Version unique de la vérité
Catalogues produits
CIM
HR

LA DIFFÉRENCE ENTRE DATA GOVERNANCE ET BIG DATA GOVERNANCE ?

→ Le Big data Governance doit inclure en plus :

鏞 Les données provenant des médias sociaux

- Le flux Twitter
- Les Blogs,
- Les pages Facebook
- Les profils LinkedIn...

鏞 Les données des smartphones

鏞 Des capteurs de données

鏞 Données sur le climat et le temps

鏞 Données de vidéo surveillance....

鏞 Toutes données de streaming

CE QU'IL FAUT PARTICULIÈREMENT SURVEILLER AVEC LE BIG DATA

- Les coûts d'acquisition (spécialement des meta données)
- Les coûts de stockage
- Le coût de traitement de la donnée pour lui donner du sens
- Le cycle de vie des données et de l'information (Information Life cycle management)
- Trouver le bon compromis entre les 3 V (volume, variété, vélocité) & le niveau suffisant de qualité de données
- Une gestion efficace du référentiel des données pour ne pas acheter le même fichier nommé différemment dans 2 repositories
- Avoir les bons outils vu la complexité et la diversité des données
- Cadre juridique d'usage ou de détention de tout nouveau type de donnée. Ex tout ce qui est du domaine du privé

Attention au traitement des données sociales!

MERCI DE VOTRE ATTENTION



Renaud Finaz de Villaine

CMO

+33 (0) 1 74 18 74 94

rfinaz@micropole.com

Headquarters Corporate : 91-95, rue Carnot 92300 Levallois-Perret France

Tel : +33 1 7418 7418

www.micropole.com

Approche IBM sur la Gouvernance et cas client

*Solange Mathieu-Lartigue,
Ingénieur Avant-Vente,
IBM ECM, IBM France*



Gouvernance de l'information

Pourcentage du budget IT destiné à la conservation des données

17%

Croissance attendue des volumes d'information



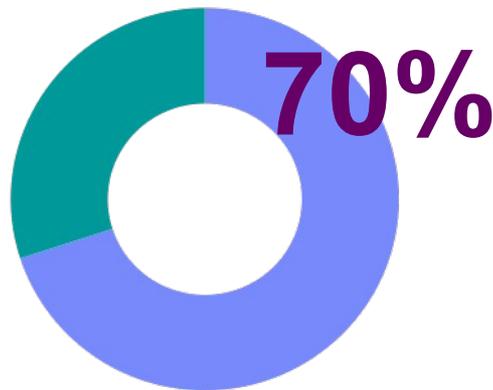
* Zettabyte = 1 trillion gigabytes

- Sources:
- 3. Information Week Survey, December 2009
 - 4. IDC Digital Universe Study, May 2010

Coût moyen constaté pour collecter les informations dans un contexte légal

2,5M €

Données conservées sans nécessité

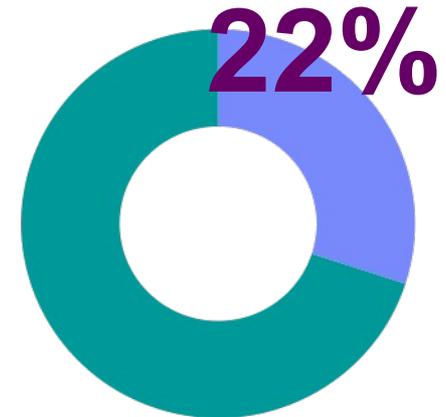


- Sources:
- 1 [Litigation Cost Survey of Major Companies, 2010](#) (from [Conference on Civil Litigation](#), Duke Law School, May 2010)
 - 2 Industry estimates

Entreprises considérant comme pertinente une politique de gouvernance des données

98%

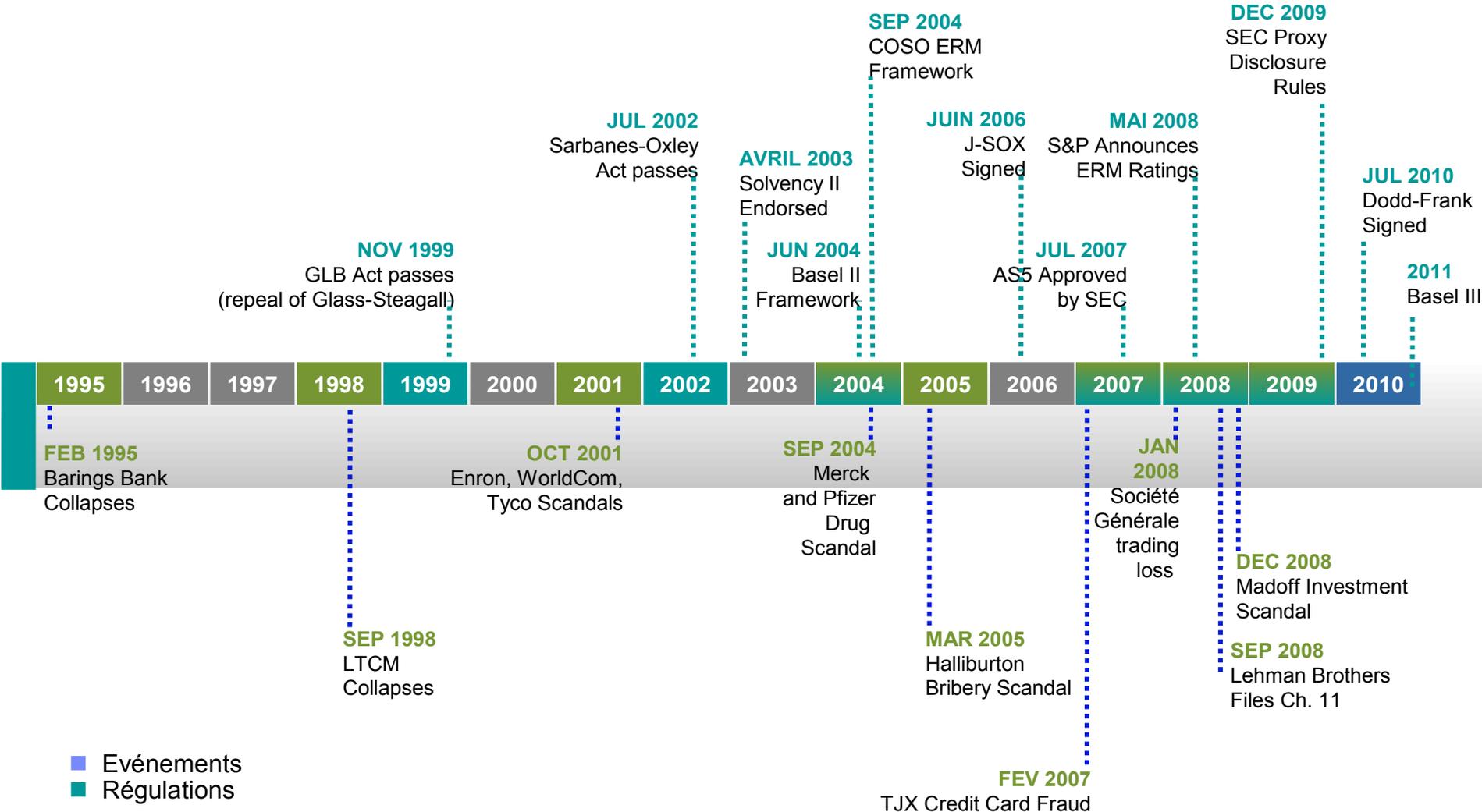
Entreprises la mettant en œuvre et l'appliquant



Source: [CGOC \(Compliance, Governance and Oversight Council\) Benchmark Report on Information Governance](#), October 2010

Un challenge majeur dans l'environnement actuel

...nouvelles réglementations, globalisation, risques accrus et vélocité métier, explosion des informations...



Si vous ne savez pas déterminer la valeur métier ou légal de l'information alors vous devez :

Gérer toutes les informations comme si elles avaient de la valeur

Supposer que vous êtes en conformité

Passer trop de temps à gérer de l'information inutile

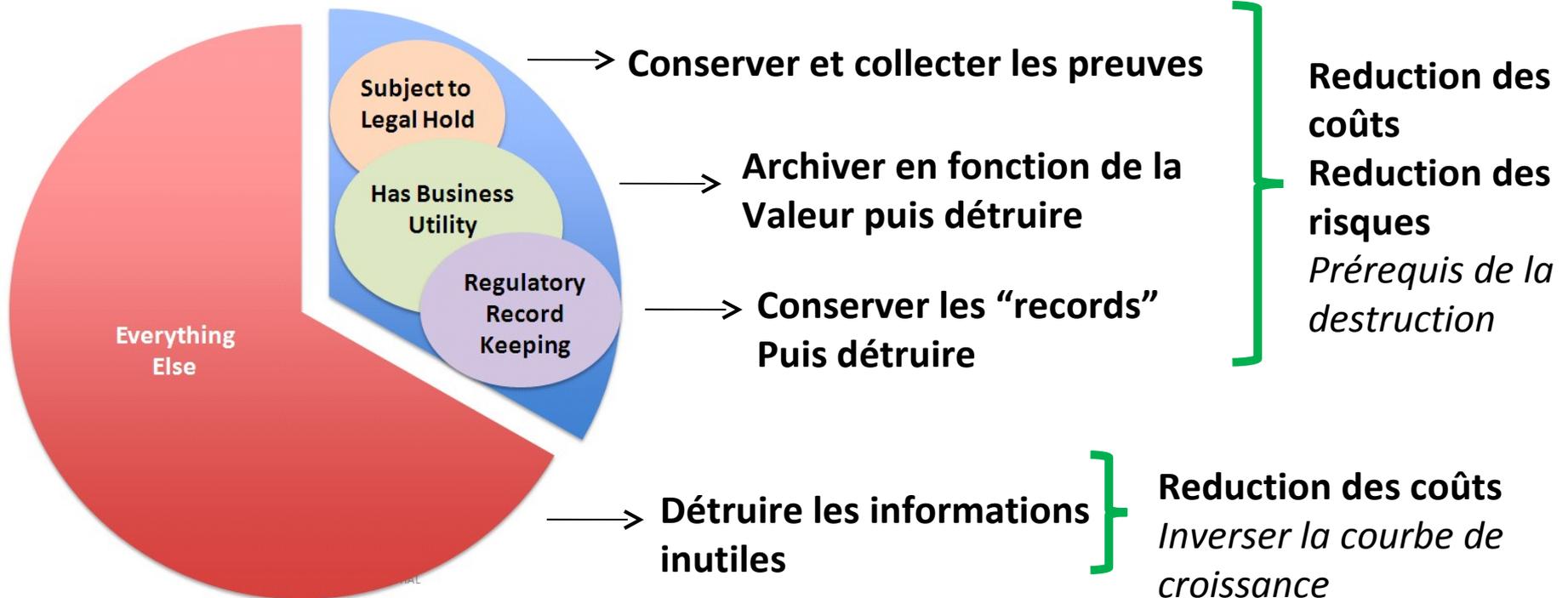
Passer trop de temps à conserver et retrouver l'information

Mais ...

Si vous savez quoi conserver pour être conforme

Et si vous savez quoi conserver pour vos besoins métier (et pour combien de temps)

→ Alors vous pouvez détruire tout le reste





- Comprend les obligations réglementaires
- Responsable des recherches, gels, et autres étapes dans les processus de eDiscovery
- **Souhaite améliorer les processus de gestion des litiges tout en maîtrisant les coûts et la complexité des processus de eDiscovery**



- Peut ne pas maîtriser les obligations légales
- Responsable du système d'information
- **Souhaite maîtriser les coûts IT, améliorer les performances applicatives, les temps de backup mais est confronté à l'explosion des informations.**



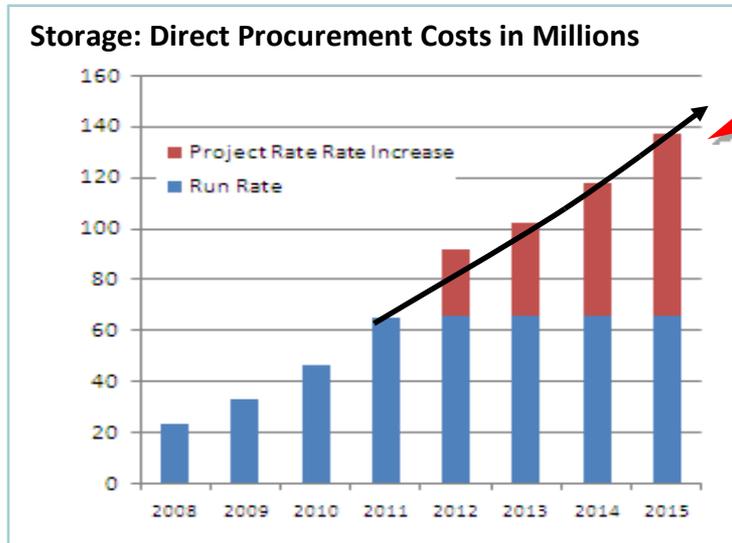
- Comprend les obligations réglementaires
- Responsable d'établir les règles, la taxonomie, les procédures et tous les services liés aux "records".
- **Souhaite une approche globale de rétention et destruction au niveau de l'entreprise, mais est confronté à l'explosion de l'information en volume, variété, ainsi qu'à la complexité et à l'évolution de l'environnement réglementaire.**

Le métier connaît la valeur de l'information mais ...

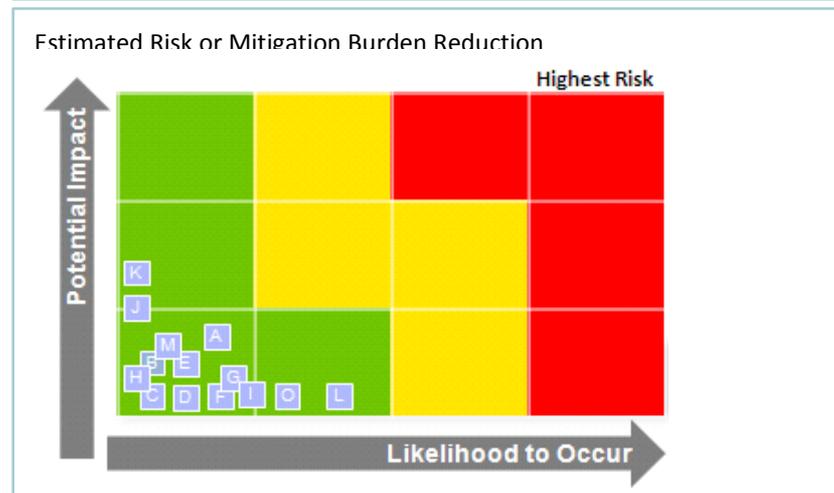
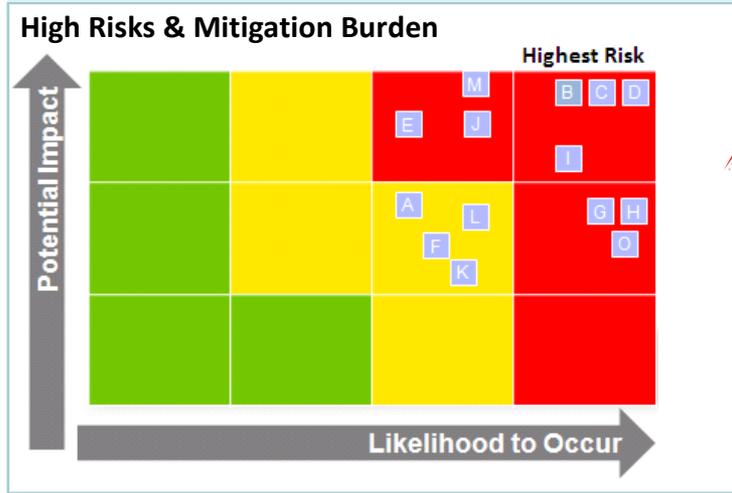
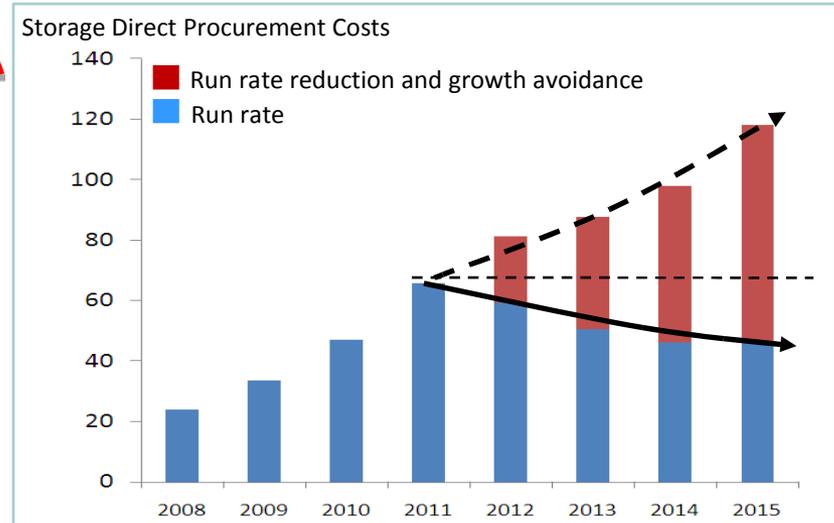


... ne dispose pas des outils adéquats pour transmettre cette information à l' IT ou au juridique.

Run rate costs rise continuously, consume more of budget

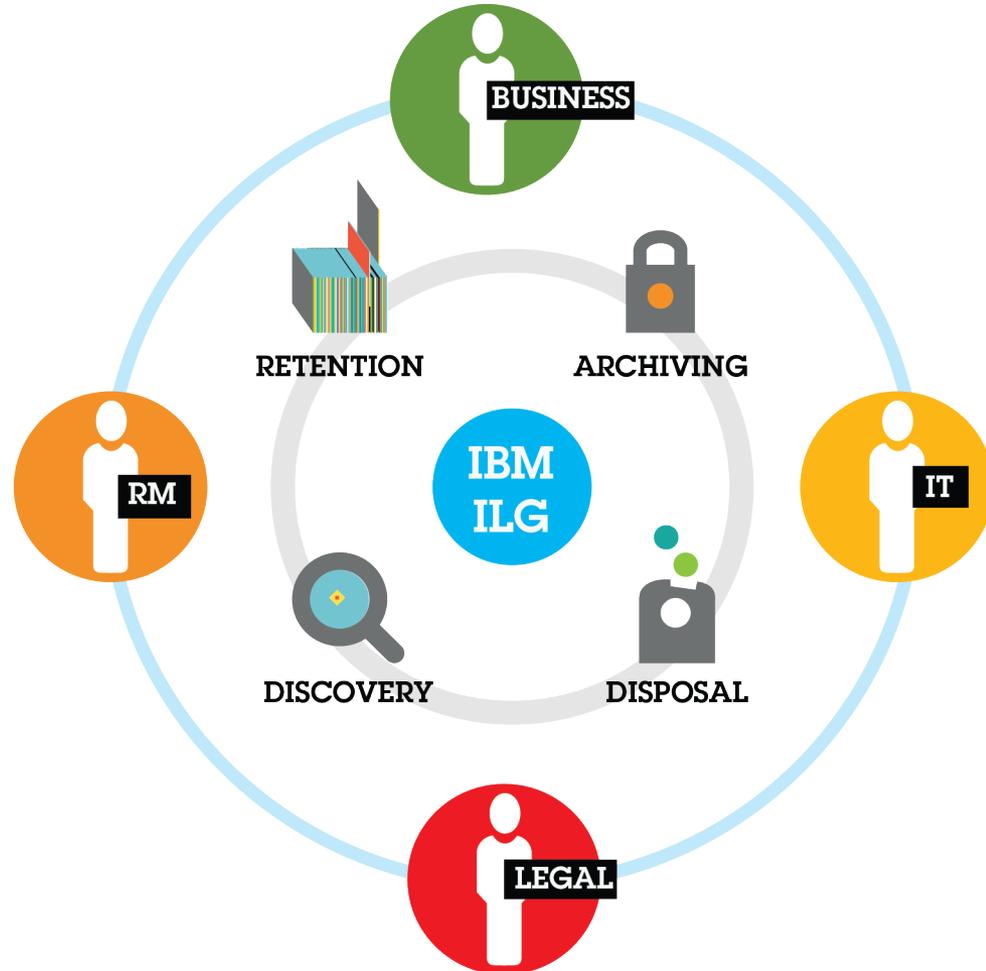


Curbs storage growth, lowers run rate permanently



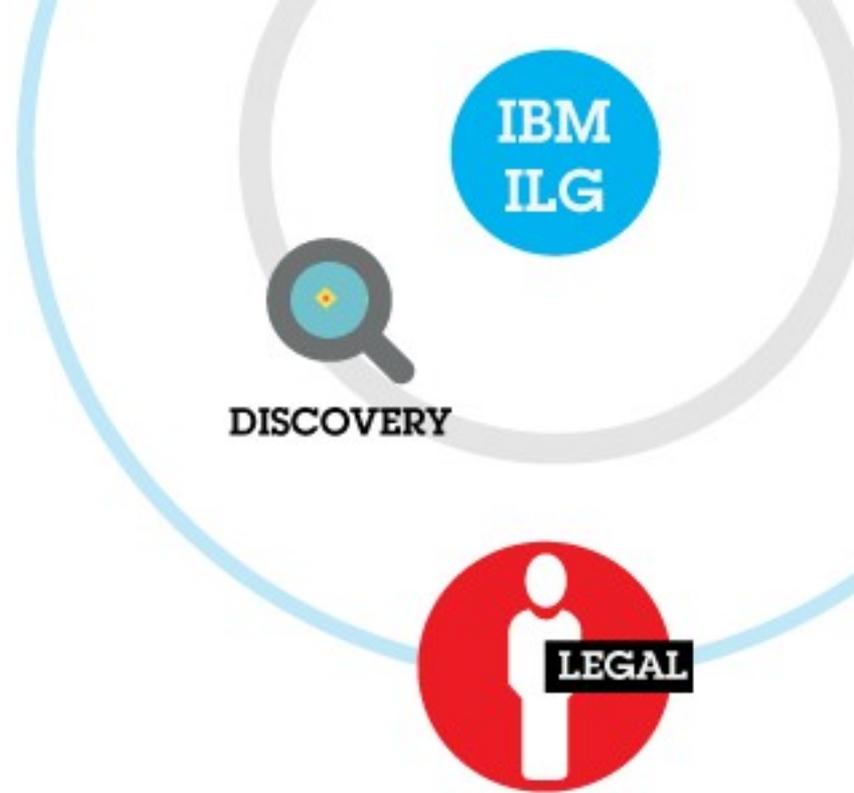
ILG : Information Life Cycle Governance

ILG fournit une solution pour rapprocher le juridique, le record manager, les métiers, et l'informatique.



Un processus eDiscovery Rigoureux et efficace

- Crée et automatise des demandes de gel avec une plus grande précision sur différentes sources de données
- Collecte, sélectionne les éléments de preuve, et maîtrise les coûts associés.
- Gère les obligations des responsables des sources de données en coordination avec l'informatique.



Gérer les Records et leur rétention

- Gère la taxonomie générale et les planning de rétention pour toutes les informations.
- Assure la coordination des programmes de rétention entre les métiers et le juridique.
- Gère les “records” des différents systèmes de façon centralisée.
- Publie les planning de rétention sur les informations structurée et non structurée.

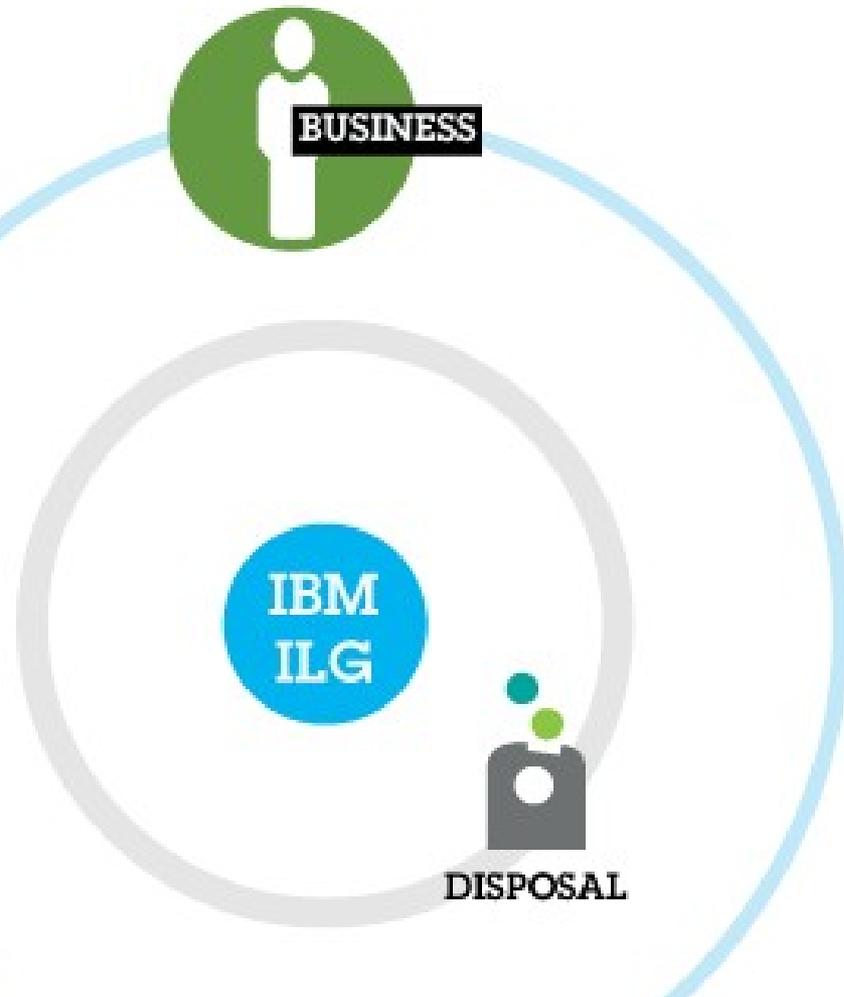


Archiver en fonction de la valeur de l'information

- Archive les fichiers, les email, les contenus applicatifs, les réseaux sociaux en fonction d'une valeur métier.
- Conserve, gele, détruit de façon certaine et efficace.
- Gère le structuré et le non structuré avec les mêmes outils.
- De-duplique et compresse l'information pour réduire les coûts de stockage.



Gouverner et éliminer

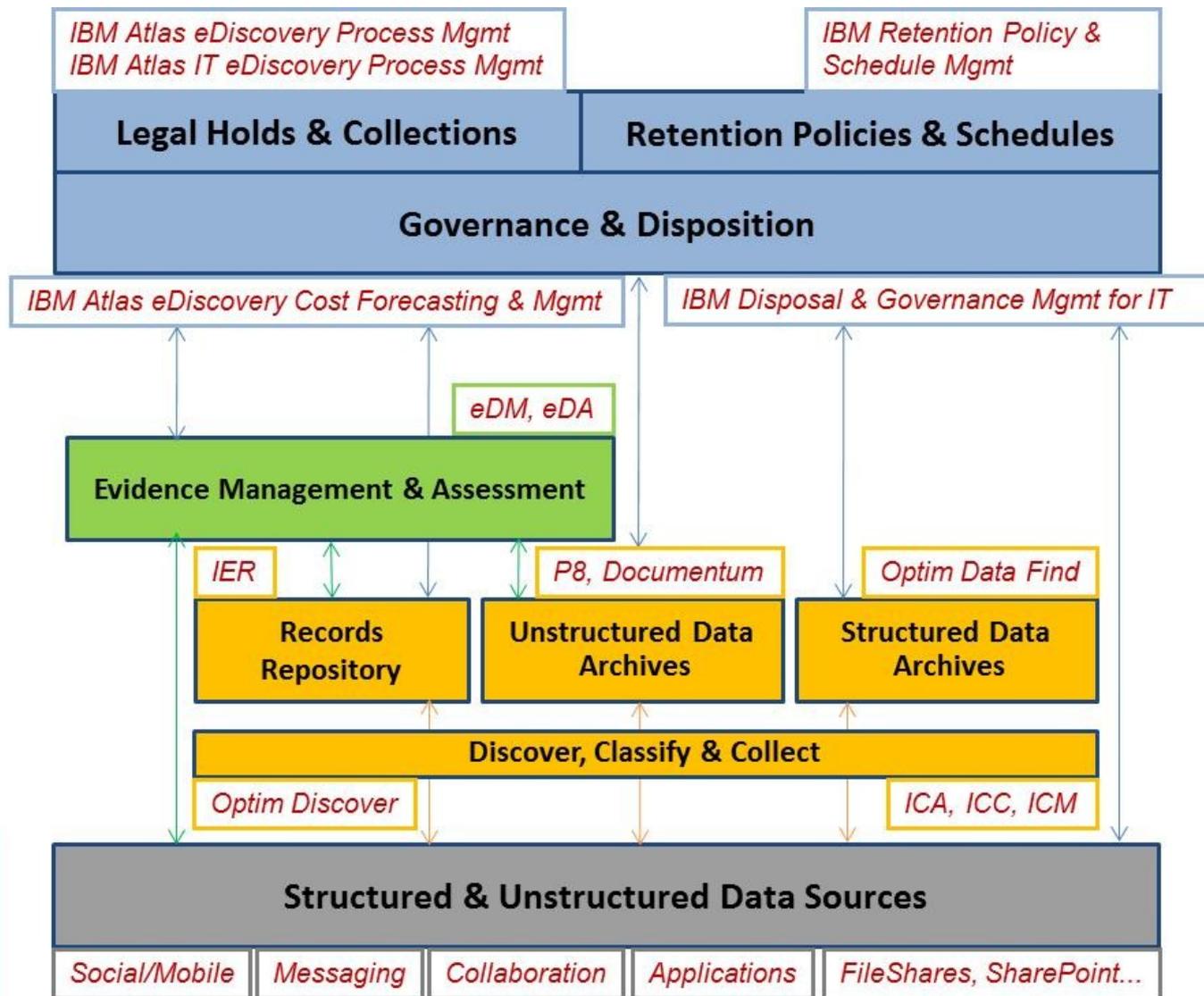


- Accède grâce à un catalogue centralisé aux obligations légales ainsi qu'aux règles appliquées à leurs données.
- Permet en appliquant des politiques de rétention et de gel systématiques de se débarrasser des données inutiles.
- Autorise la destruction des données inutiles
- Permet de freiner la croissance de stockage et gère l'information par sa valeur.

Decisions

Data

Infrastructure



Software Key:

- Decision
- Evidence Management
- Repository / Collection
- Information Sources

Cas client :

Un grand groupe pharmaceutique

Société pharmaceutique mondiale

Environnement



Besoins des patients

Traitements novateurs

Prévention

Des traitements abordables

Système immunitaire

Prise en charge globale

portfolio

Produits pharmaceutiques

Soins ophtalmiques

Gamme "génériques"

Consumer Health (OTC and Animal Health)

Vaccins et diagnostics

Situation de départ pour les programmes de ediscovery, retention et suppression

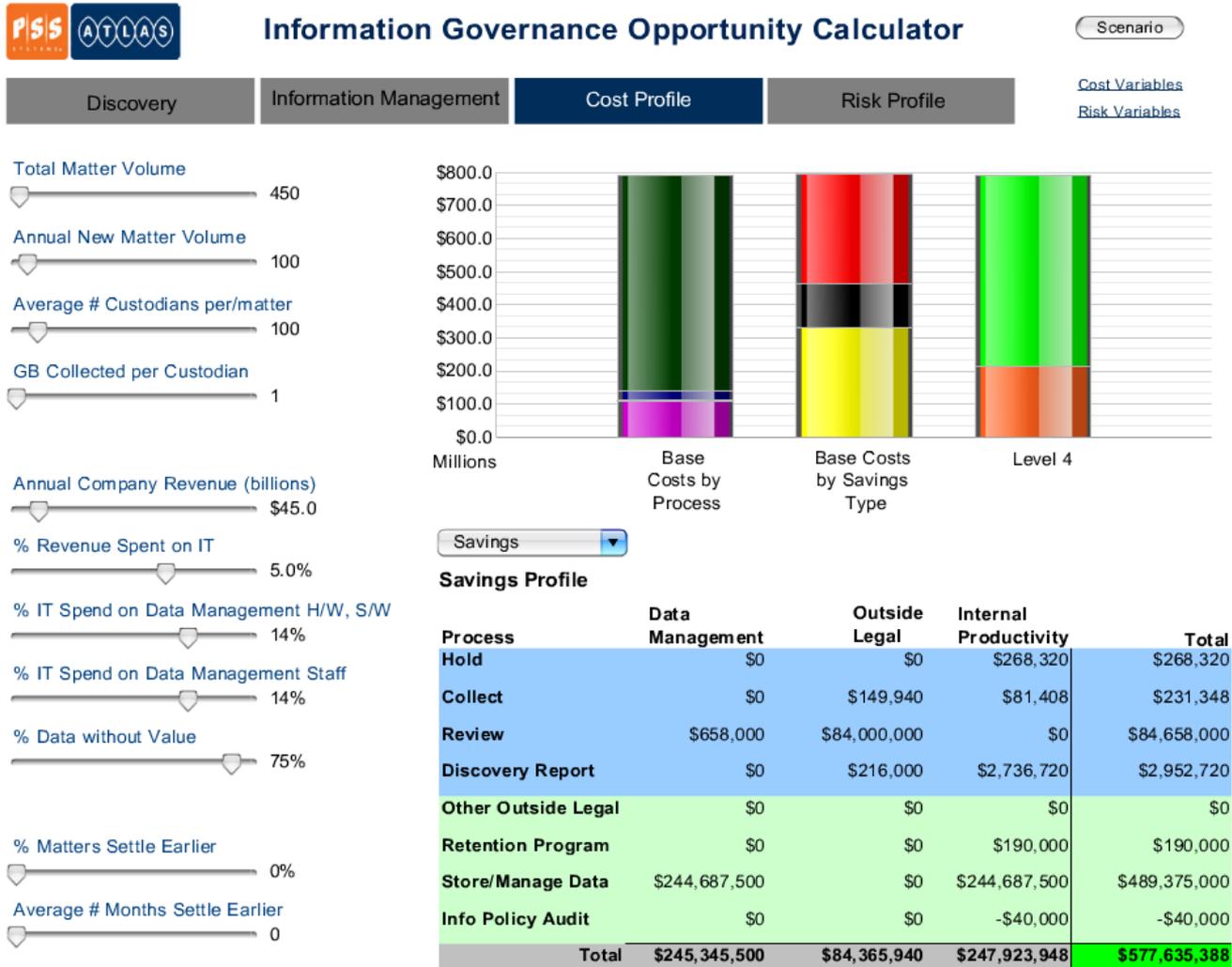
- eDiscovery – Processus manuel et coûteux
 - A l'ouverture du dossier, conservation de l'ensemble du disque
 - Notification des responsables par email des besoins de conservation, mais confirmations renvoyées par fax au service juridique
 - Arriérés importants dans l'enregistrement manuel des réponses des différents responsables
 - Des outils différents de eDiscovery :
 - US
 - Europe

Situation de départ pour les programmes de ediscovery, retention et suppression

- Conservation, archivage et suppression
 - Démarche habituelle (centrée sur le papier depuis 20 ans)
 - 1800 coordinateurs (Records Managers)
 - 256 entités “corporate”
 - 100 pays
 - 7000 contraintes réglementaires
 - 83 pages décrivant les plannings de conservation
 - Plus d’une centaine de plannings et de langues différents
 - Absence de politique unifiée au sein de l’entreprise

- **Implications du programme de conservation**
 - Difficultés à satisfaire aux besoins d’audits actuels et aux nouveaux standards
 - Risques matériels et coûts directement liés à une conservation trop longue des archives

Technique de quantification



Bénéfices du programme

Un programme global pour mieux gérer l'information électronique :

1. Détruire les informations inutiles le plus rapidement possible
2. Stocker et retrouver rapidement et facilement l'information pour répondre à des besoins métiers
3. Identifier et stocker de façon cohérente les archives de l'entreprise
4. Satisfaire aux réglementations en vigueur

Qui requiert des politiques, des procédures et systèmes pour :

1. Solution de verrouillage des données et de gestion des dossiers de litiges permettant de réduire significativement les contraintes métiers et opérationnelles nécessaires pour le respect des obligations réglementaires.
2. Mettre en place des politiques de confidentialité, de gestion sécurisée de l'information et de Records Management
3. Améliorer l'efficacité de la conservation des archives et de l'accès aux données par les métiers
4. Rationaliser les tâches quotidiennes dans la gestion des archives de l'entreprise.

\$300 millions d'économie

EDM Program avoids \$3,000/employee thru 2015



Investissement sur la solution

\$5 million sur Atlas Software

\$10 million sur les logiciels IBM ECM

\$4 million en Services

Prix d'entrée

\$500K en logiciel

\$250K en service

Le projet – Phase 1 – Franchir les obstacles

Quelles sont les bonnes réponses ?

- Solution IBM ou développement interne ?

- Présence de plusieurs centaines de planning de conservation et de contraintes réglementaires.

 - ✓ Ne peut-on pas simplement interpréter cette information?
 - ✓ Plusieurs employés à temps plein sont dédiés à la collecte des besoins liés à l'archivage.

- Scepticisme en interne sur la capacité d'une telle solution à réduire les coûts et les risques, et à être acceptée par les différents utilisateurs.

Le projet – Phase 2 – Réalisation d'une maquette

Démontrer la bonne adéquation aux besoins, adhésion des utilisateurs

- Accompagnement par un partenaire spécialiste des problématiques de gouvernance
- Déploiement, configuration et import des plannings “maîtres” et des demandes de recherches juridiques
 - Plus de 400 lois et réglementations
 - Textes de loi américains, Allemands et suisses
 - Traduction en français, allemand et italien
 - 4 domaines fonctionnels
- Déroulement d'ateliers de travail avec différents responsables fonctionnels.
- Mise à disposition des documents de spécification en adéquation avec les normes qualité de l'entreprise.

Le projet – Phase 3 – Le programme “Pilote”

Le bon timing pour répondre à une urgence: un succès

- Pilote pour le département Propriété Intellectuelle (PI)
 - Suite aux conclusions négatives d’un audit, mise en place d’un environnement de production pour la division “PI” pour répondre à leurs besoins.
 - Le partenaire a mis en place la solution en prenant en compte les différentes réglementations américaines, européennes et suisse ainsi que les traités mondiaux.
 - La division “PI” a réussi son audit

Le projet – Phase 4 – Processus d’approbation au niveau du comité exécutif

- Première tentative
 - Aucune décision prise ; report

- Effort complémentaire = réunion individuelle avec chacune des DSI impliqués dans le projet et analyse de ROI propre à chaque département
 - Assistance du partenaire pour les faits juridiques et réglementaires

- Décision finale = Approbation à l’unanimité par les membres du Board

Le projet – Mise en œuvre du déploiement global

Approche par étape

- Approche “Multi-phase” pour permettre au client de déployer la solution au niveau global, en s’appuyant sur une fondation commune pour chaque entité.
- Couvre tous les aspects de la solution, en incluant :
 - Intégration complète des données RH provenant des sources appropriées, basée sur les besoins définis
 - Définition et intégration avec les déploiements existants aux Etats-Unis et en Suisse.
 - Définition et chargement des plannings de conservation pour les pays identifiés dans le périmètre préparé par le partenaire
 - Génération des rapports basés sur les besoins métiers
- Réutilisation du travail réalisé durant le “projet pilote”

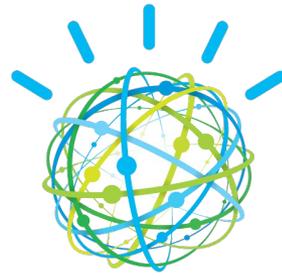
Le projet – Mise en œuvre du déploiement global

Approche par étape

- Effectuer les recherches à caractère juridique aux Etats-Unis, en Suisse, au Japon ainsi que dans de nombreux autres pays identifiés
- Gérer les citations, les plannings de conservation globaux et locaux
- Fournir l'expertise relative aux programmes, domaines et applications permettant de faciliter les activités de gestion des plannings locaux et de gouvernance des archives
- Développer des communications “sur mesure”, questionnaires et enquêtes pour recueillir efficacement les informations
- Développer des connecteurs et les méthodes d'export pour distribuer les plannings de conservation vers les sources de données stratégiques
- Déployer et configurer le module de Gouvernance de l'information pour les équipes informatiques

Principaux résultats obtenus

- ✓ Système unique de conservation pour l'ensemble des directions métiers
- ✓ Mise en œuvre d'un planning de conservation robuste et opposable répondant aux besoins juridiques, métiers et opérationnels en adéquation avec des fonctions de recherche juridiques
- ✓ Association aux sources de données et capacité à s'adapter dans le temps
- ✓ Capacité à gérer les litiges mettant en relation personnes, données et enregistrements
- ✓ Alignement des politiques, des procédures et de la pratique



Grazie धन्यवाद *Merci* ありがとうございます *Obrigado* 多谢
ITALIAN HINDI FRENCH JAPANESE BRAZILIAN PORTUGUESE SIMPLIFIED CHINESE

Thank You

多謝 Gracias Спасибо நன்றி ชอบคุณ Danke شكراً
TRADITIONAL CHINESE SPANISH RUSSIAN TAMIL THAI GERMAN ARABIC

Questions ?