

IBM SolutionsConnect 2013

L'IBM TechSoftware nouvelle génération

28, 29 et 30 août
IBM Client Center Paris



#solconnect13

Transformez vos opportunités en succès

IBM®

IBM SolutionsConnect 2013

IM01P2:

**Le Big Data pour enrichir,
complémenter et travailler en
synergie avec vos Warehouses**

Corinne BARAGOIN

c_baragoin@fr.ibm.com



#solconnect13

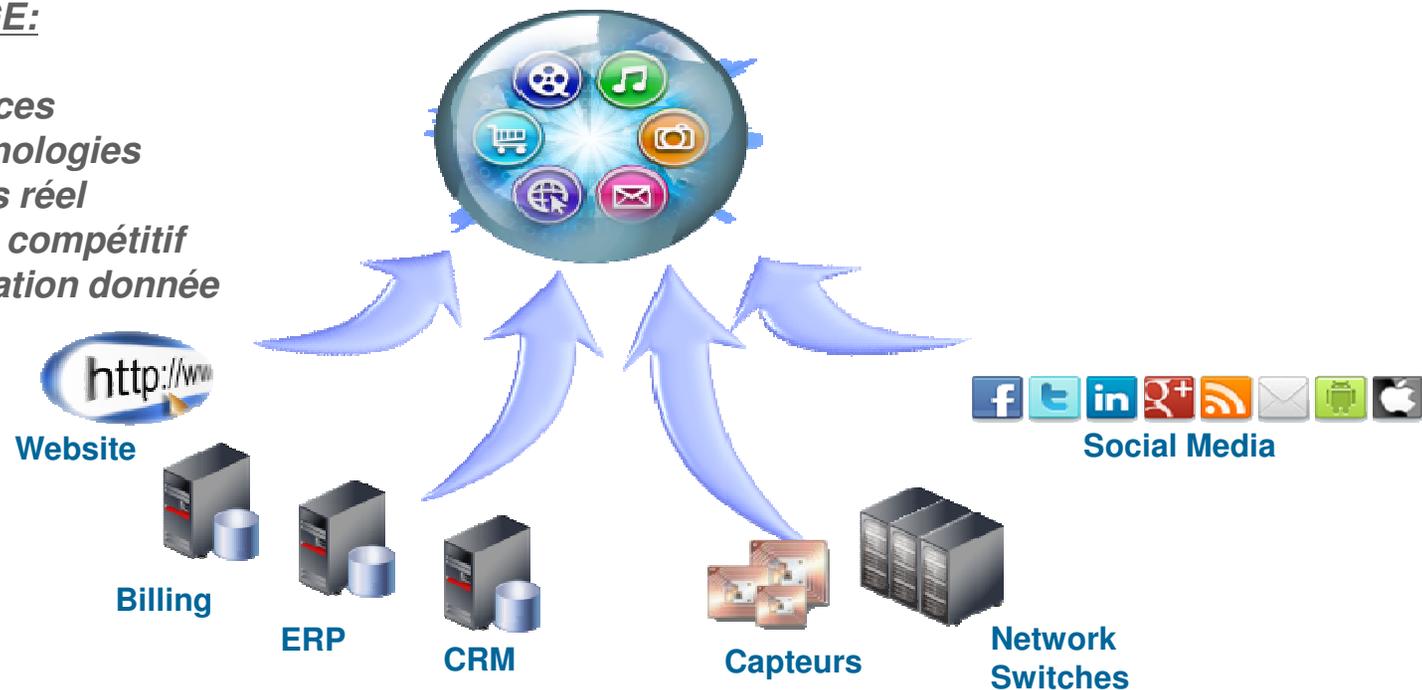


Le succès du Big Data est lié au fait que la technologie rend possible l'analyse de *TOUTES* les données

Créer de la valeur à partir de données sous-utilisées pour décider et agir avec agilité

CE QUI CHANGE:

*Nouvelles sources
Nouvelles technologies
Analyses temps réel
Environnement compétitif
Valeur/Monétisation donnée*



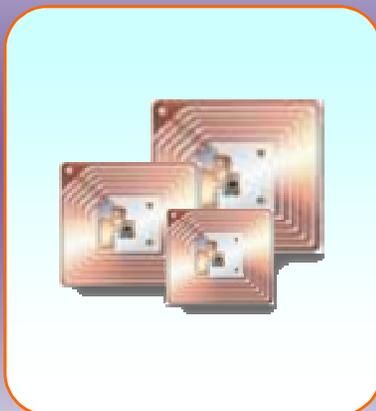
De nouvelles opportunités s'ouvrent au delà des sources traditionnelles de l'entreprise

Transactional & Application Data



- Volume
- Données structurées
- Débit

Machine Data



- Vélocité/Faible latence
- Semi-structurées
- Ingestion

Social Data



- Variété
- Non Structurées
- Véracité

Enterprise Content



- Variété
- Non Structurées
- Volume



#solconnect13



Domaines majeurs d'application du Big Data



Exploration Big Data

Chercher, visualiser et comprendre les données Big Data pour améliorer sa connaissance



Vue 360° améliorée

Développer une véritable vision unifiée, intégrant des sources internes et externes



Extension des systèmes de sécurité

Détecter les fraudes et les risques en temps réel



Analyse opérationnelle

Analyser les données venant des machines pour améliorer les résultats et prévenir en temps réel



Extension de Data Warehouse

Intégrer le big data et les capacités des data warehouse pour augmenter leur efficacité



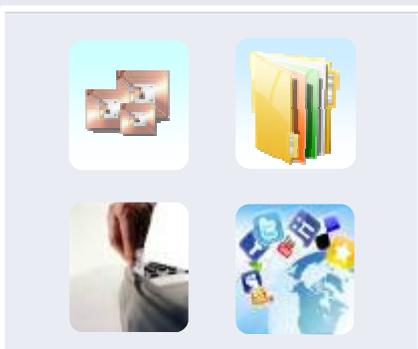
#solconnect13



Extension de Data Warehouse: les besoins

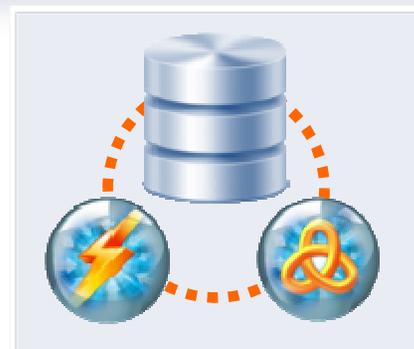


Intégrer le big data et les capacités des data warehouse pour augmenter leur efficacité opérationnelle



Besoin de variété de données

- Données structurées, non structurées, en Streaming nécessaires à des analyses plus avancées
- Besoin de faible latence (heures- et non pas semaines ou mois)
- Accès à l'ensemble des données



Optimiser l'infrastructure du DW

- Optimiser les coûts de stockage et de maintenance en migrant les données rarement utilisées sur Hadoop
- Réduire les coûts de stockage au travers du traitement optimisé des données de type Streaming
- Améliorer la performance des DW en déterminant les données les plus appropriées à stocker sur le DW

Un exemple



Constant Contact a transformé l'efficacité de ses campagnes marketing par email

Besoin

Analyser 35 milliards d'emails annuels (opens, clicks ...) afin de mieux guider leurs clients sur les meilleures dates et périodes pour envoyer les emails pour un taux de réponse maximum.

Bénéfices

Grâce à la mise en œuvre d'InfoSphere BigInsights pour l'analyse des emails, en complément d'une solution Pure Data for Analytics et Cognos BI:

- 15-25% d'augmentation des réponses client dans les campagnes marketing par email
- Performances améliorée par un facteur de 40
- Temps d'analyse réduit d'heures en secondes

"Our work analyzing email delivery times has already given our customers a 15-25% lift in their email campaign performance – and that means more customers in their doors and increased revenue." — Jesse Harriott, Chief Analytics Officer





Extension de Data Warehouse

Enjeux Métiers

- Prendre des **décisions** plus rapidement
- Offrir plus de services à ses clients voire une meilleure QoS
- Affiner et rendre plus **pertinente** la prise de décision par :
 - **L'enrichissement** des données du Data Warehouse
 - Des **analyses plus complexes** et des modèles plus riches sans contraintes
 - La prise en compte de toute information utile à mon métier **quelle qu'elle soit, où qu'elle soit**

Enjeux Techniques

- **Intégrer de nouvelles technologies** et faire évoluer son architecture de données
- Historiser les données dans un environnement dédié à **bas coût**
- Intégrer de **nouvelles sources de données** non forcément structurées et mise en qualité
- **Analyser, agréger et corrélér** d'importants volumes de données **structurées** ou **non-structurées, historisés** ou **en temps réel**
- Explorer / fédérer toute information et en offrir une **vue unifiée**

Messages clés

- Étendre le Data Warehouse pour être en capacité de traiter toutes les informations et de prendre les meilleures décisions
- Diminuer le **coût du Data Warehouse**



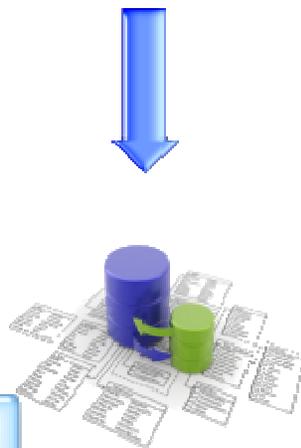
Des approches différentes

Approche traditionnelle *Analyses structurées et répétables*

Métier
détermine quelle
question poser



IT
structure les
données pour
répondre à la
question



- Rapports de vente mensuels
- Analyse de profitabilité
- Sondages clients

Approche Big Data *Analyse itérative & exploratoire*

IT
fournit une
plateforme qui
autorise la
prospective



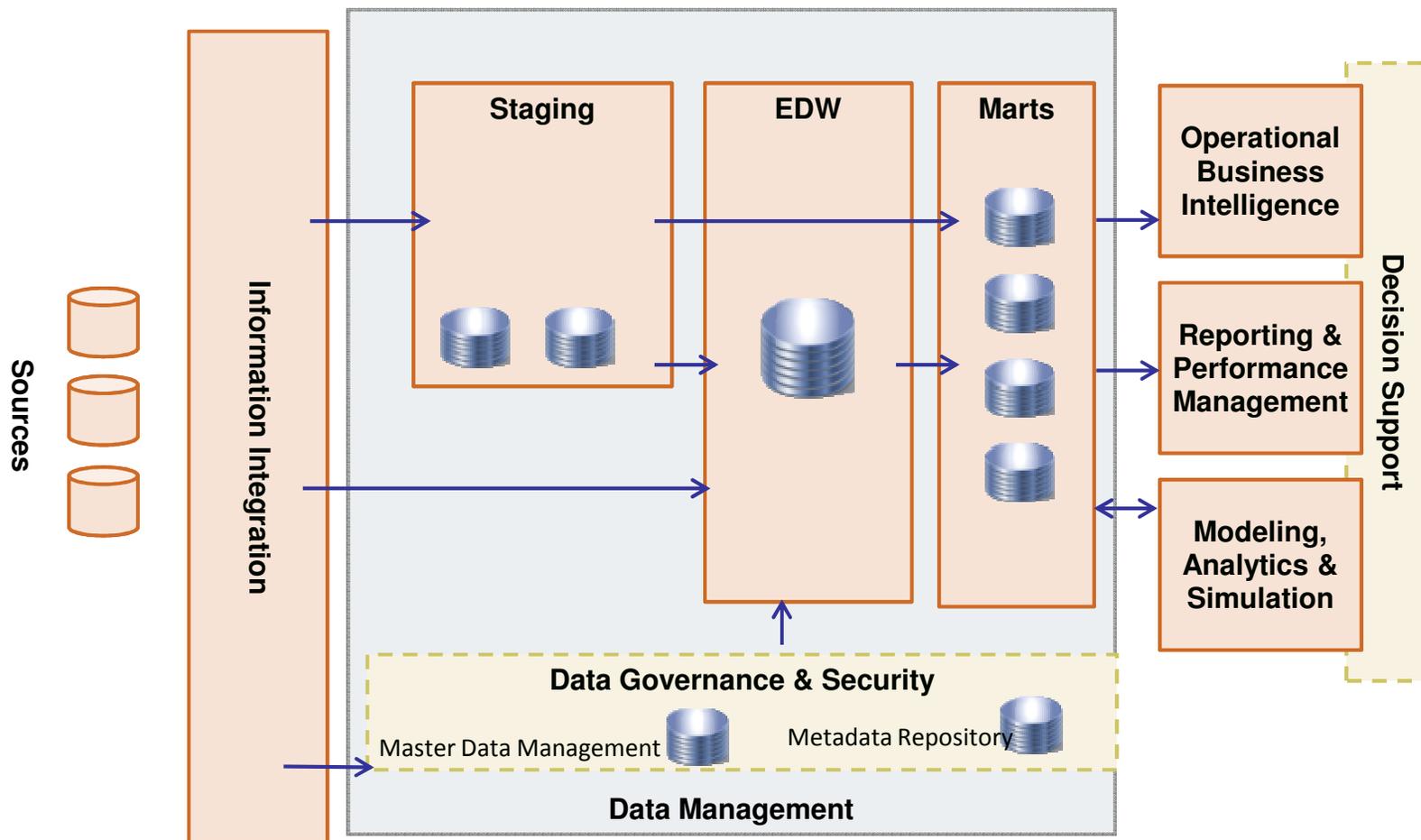
Métier
explore les
questions qui
pourraient être
posées



- Ressenti sur la marque
- Stratégie Produits
- Utilisation maximale des actifs



Exemple d'architecture traditionnelle ...



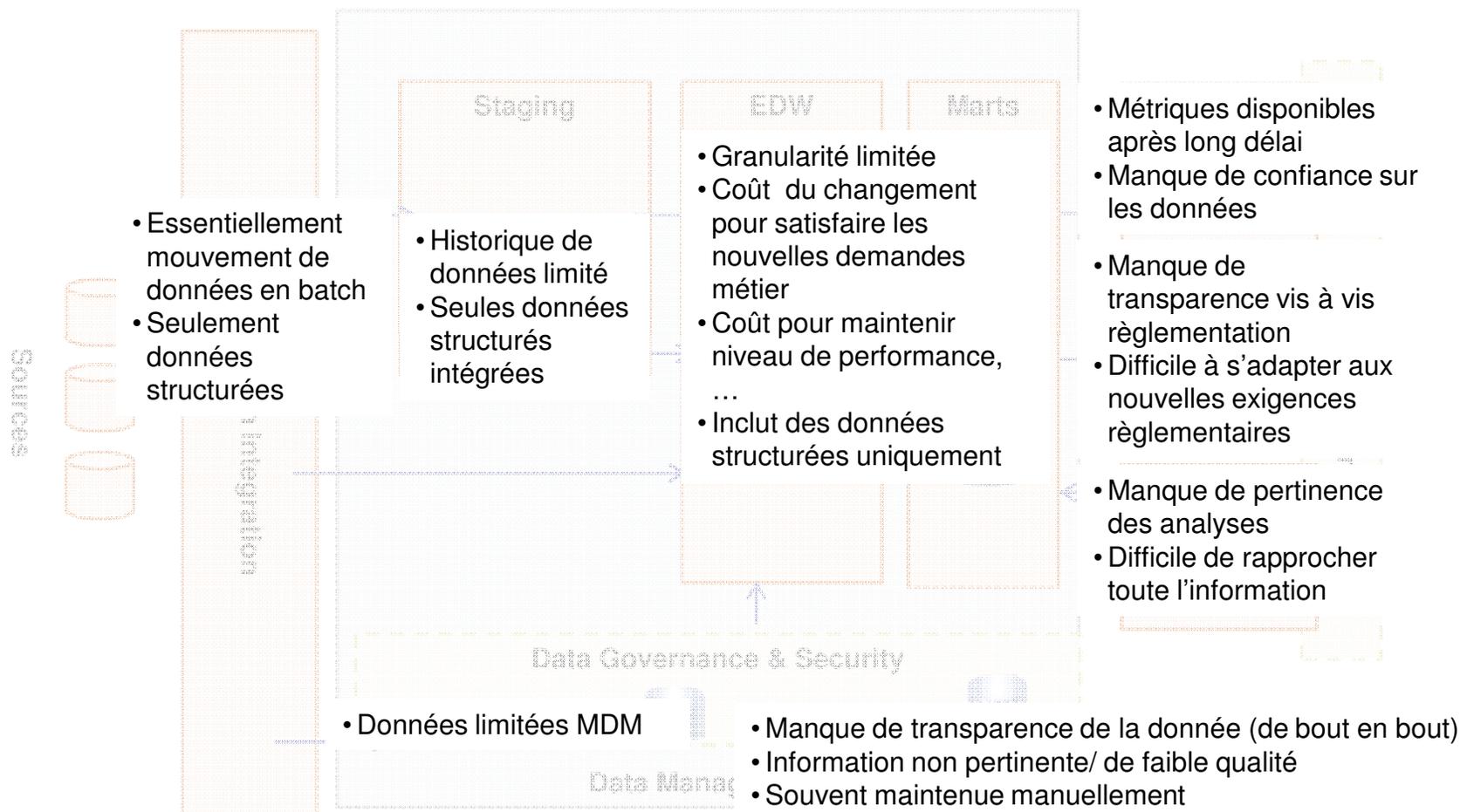
... et ses limitations

Données limitées

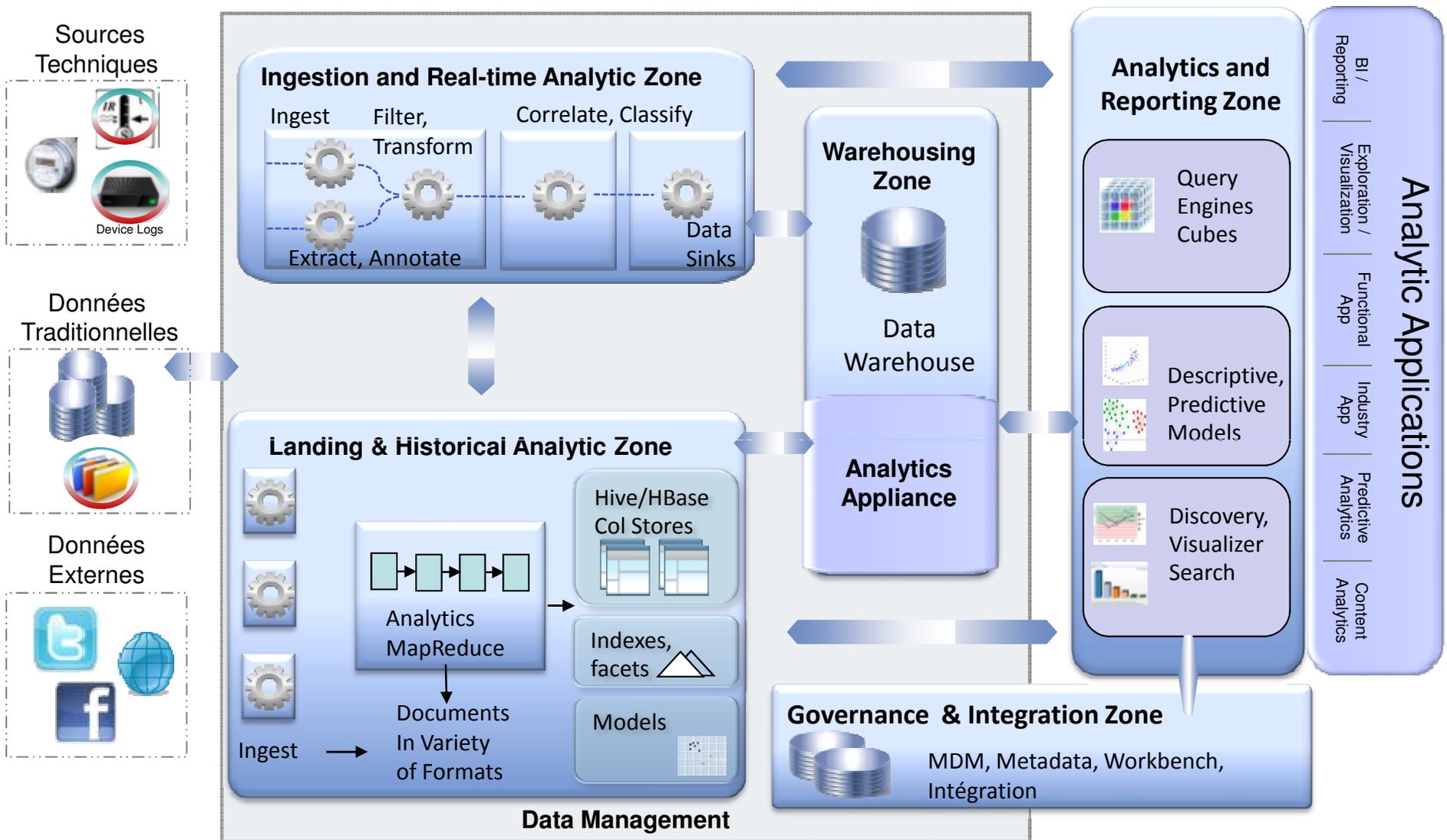
Latence



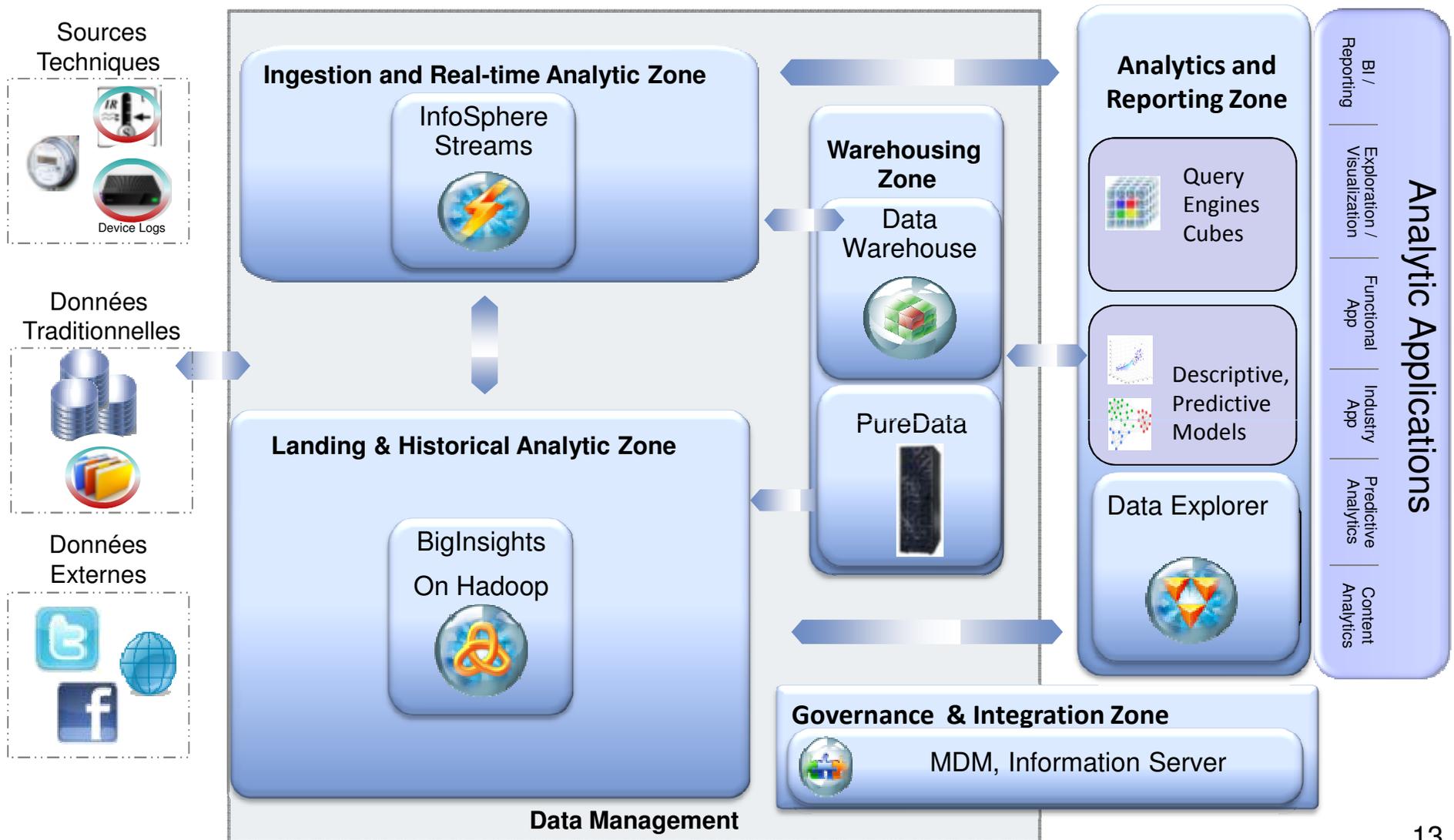
Limitations métier



Architecture logique Big Data



Composants d'une architecture Big Data



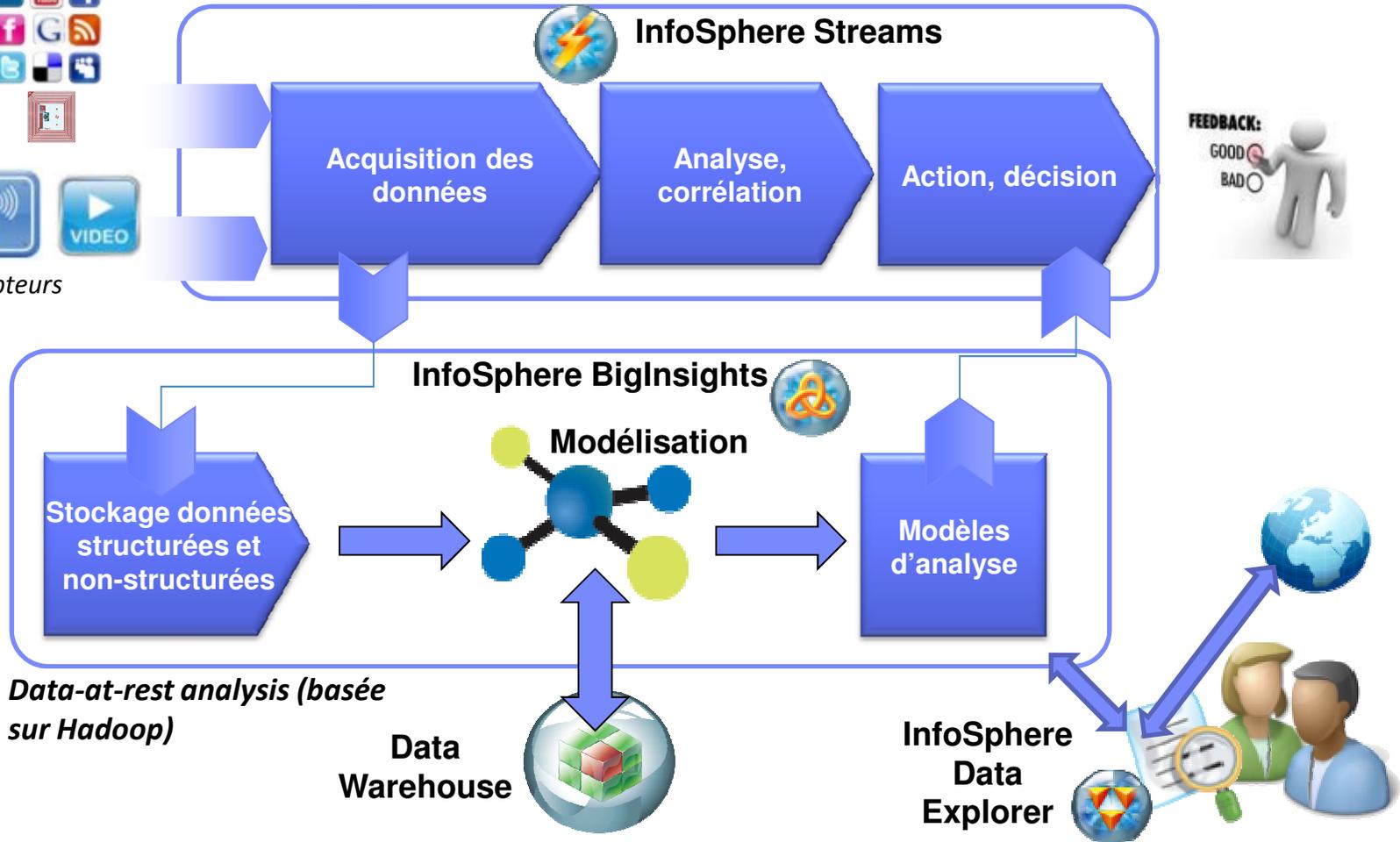
Différentes technologies: le cercle vertueux

Web, réseau sociaux



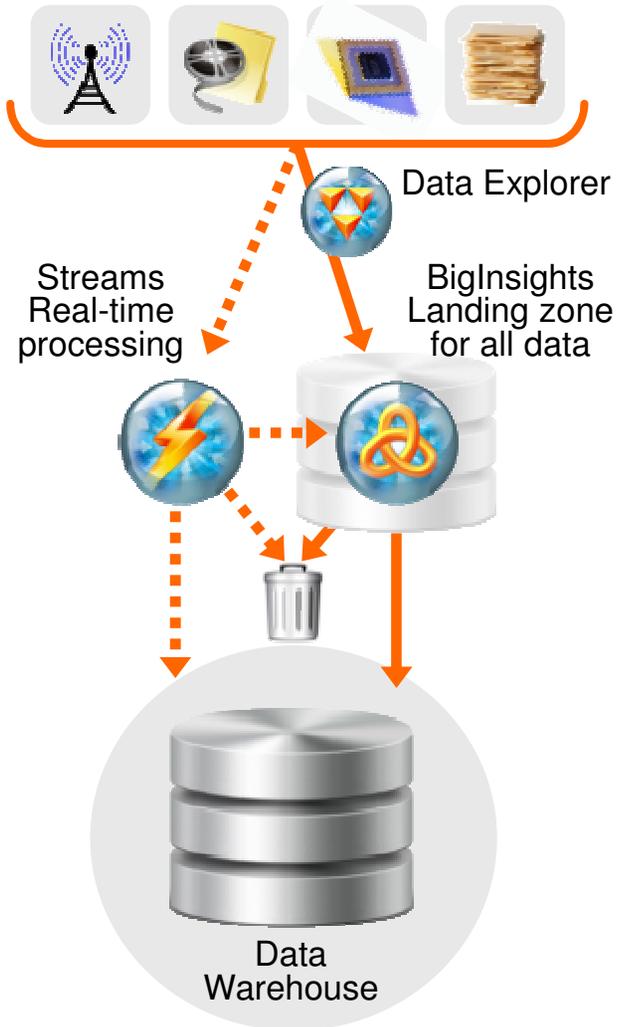
Capteurs

Flux en temps-réel : Analyse de la donnée en mouvement

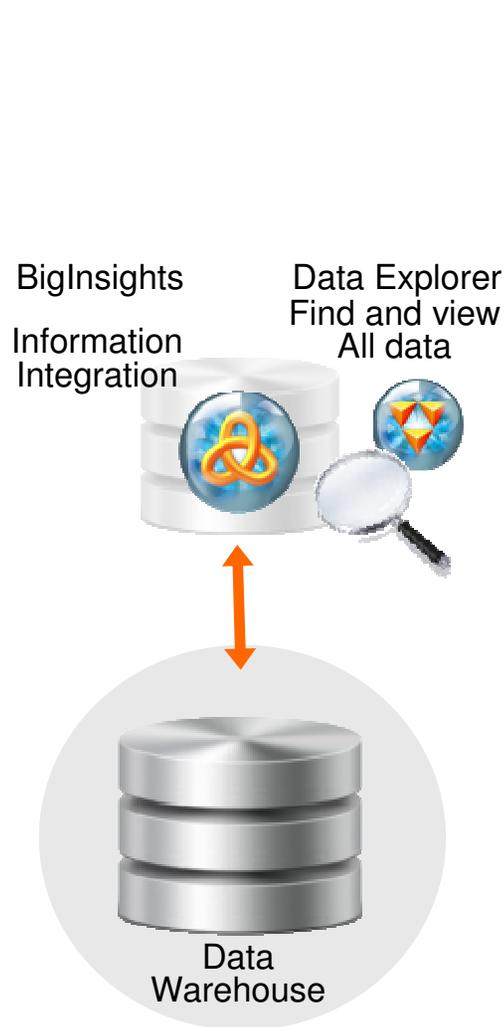


Extension de Data Warehouse : 3 exemples

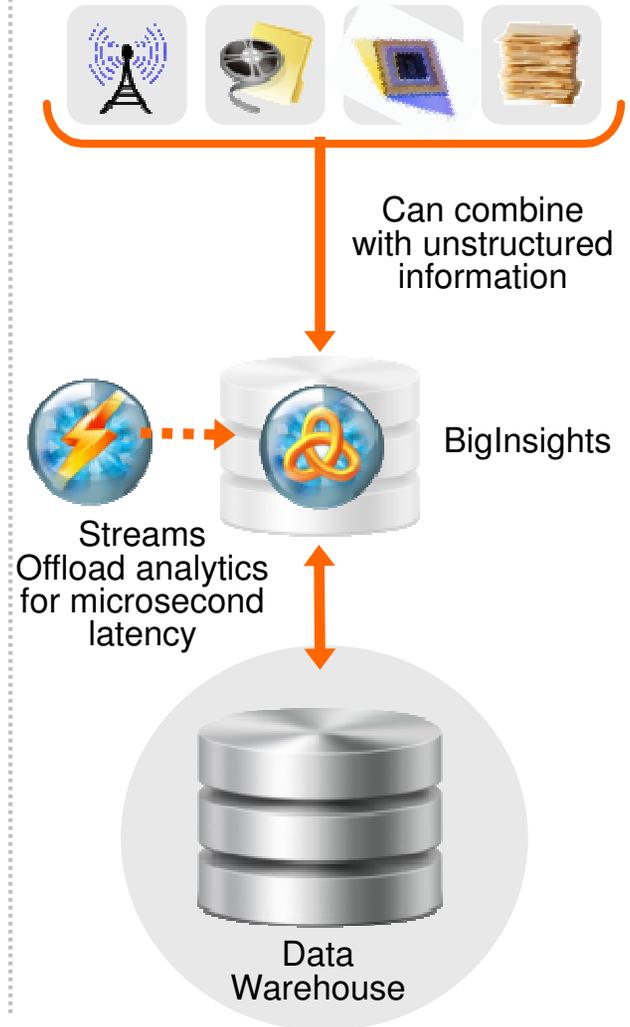
1 Pre-Processing Hub



2 Query-able Archive



3 Exploratory analysis



La plate-forme IBM Big Data

Découvrir, comprendre, rechercher et naviguer dans les sources de données Big Data

- **InfoSphere Data Explorer**

Analyser des petabytes de données multistructurées à bas coût

- **InfoSphere BigInsights** --la version plus d'Hadoop pour le déploiement et l'intégration en entreprise

- **InfoSphere Suite** -- intégrer, sécuriser, maîtriser protéger et archiver Big Data



Traiter les flux de données, les analyser en temps réel

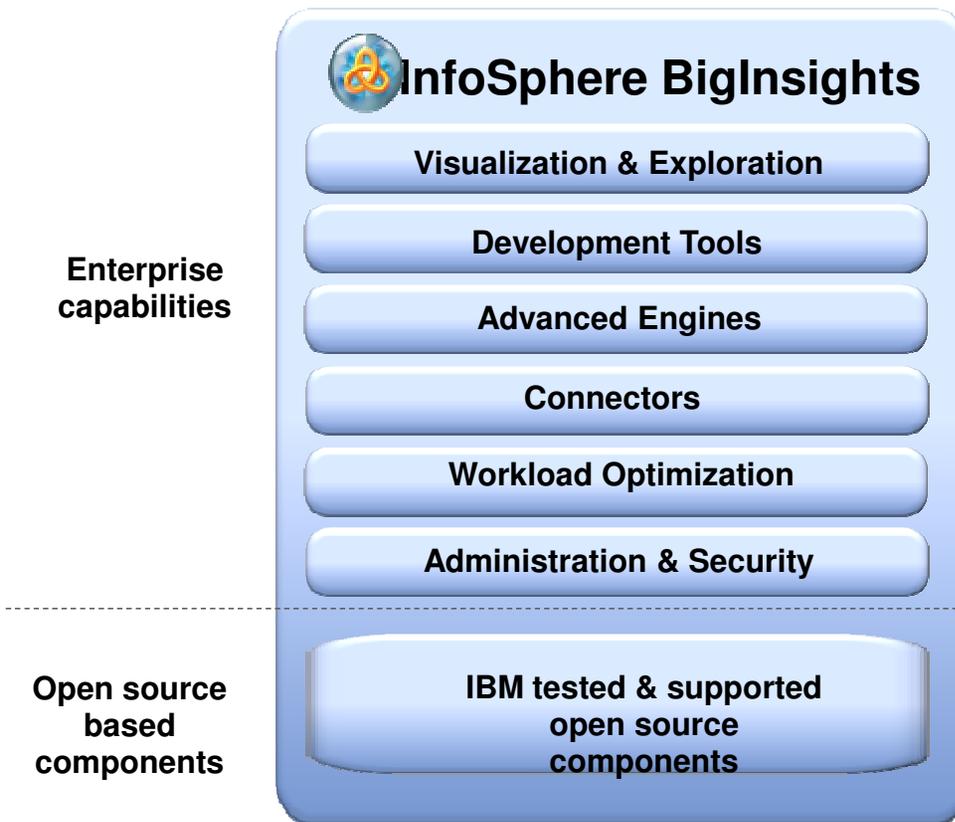
- **InfoSphere Streams** pour l'analyse en continu de flux de données avec des temps de réponse de l'ordre de la sous-milliseconde

Délivrer de manière optimisée de l'analyse avancée

- **PureData System** -- système expert intégré (analyses simplifiées et plus rapides)
- **Accélérateur DB2 BLU** -- in memory database accelerator



Extension de Data Warehouse: BigInsights apporte plus de fonctionnalités que l'Open Source Hadoop



Inclus dans BigInsights

Enterprise Edition:

Real-time Analytics

InfoSphere Streams*

Indexing & Exploration

InfoSphere Data Explorer*

Bi Reporting

Cognos BI Server*



Accélérateurs

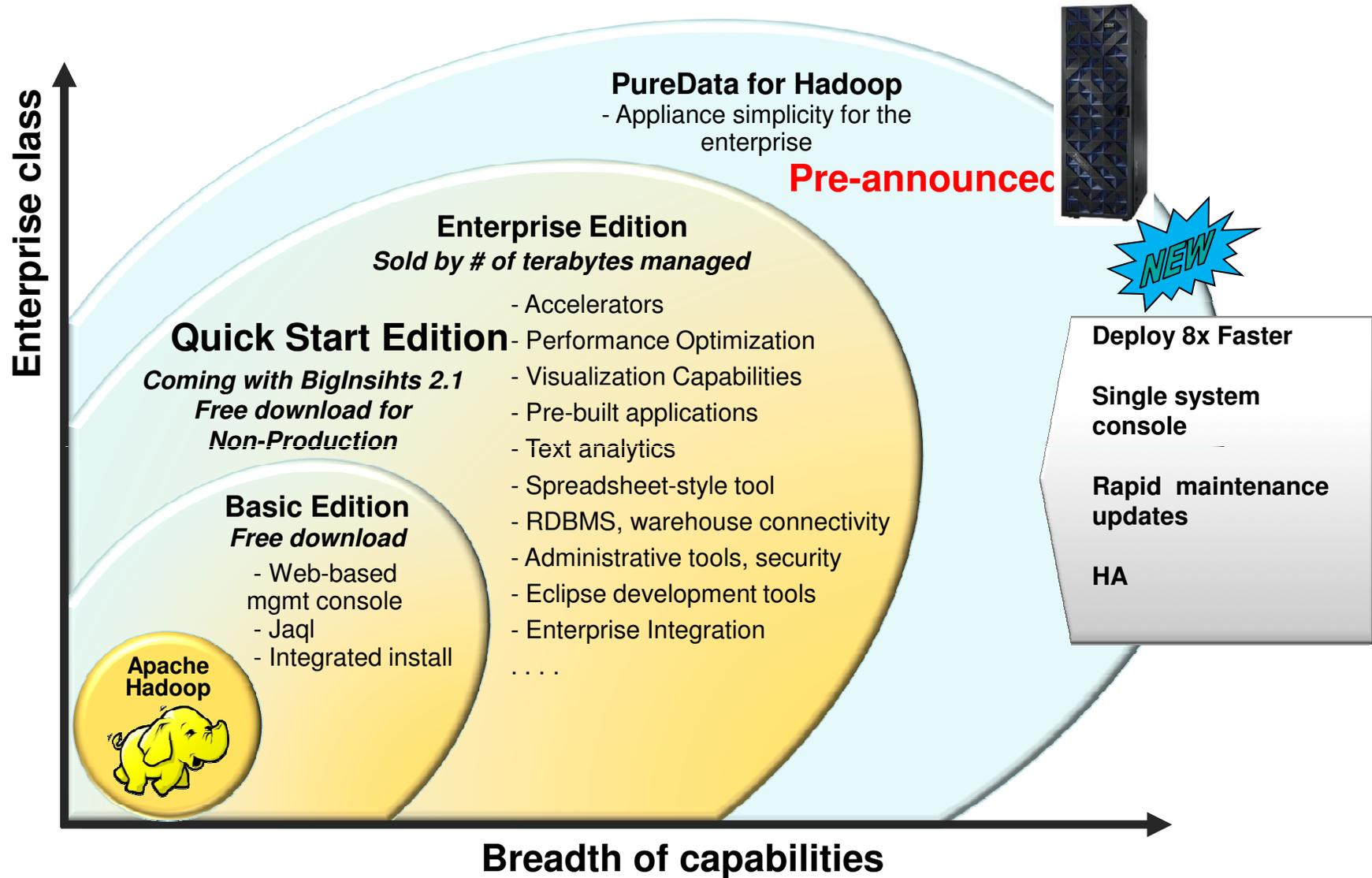


Social Data Analytics Accelerator

Machine Data Analytics Accelerator

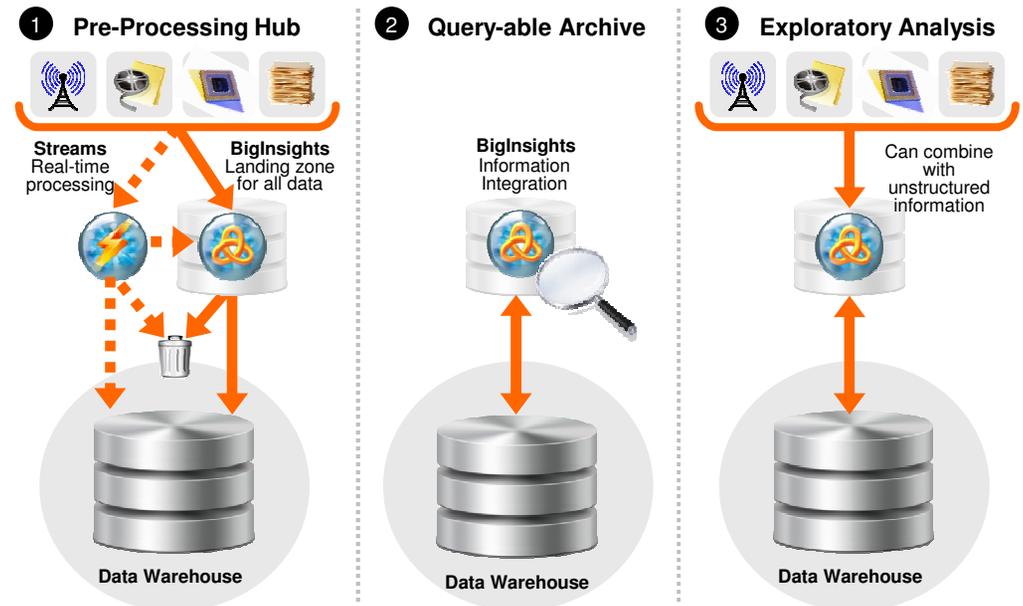
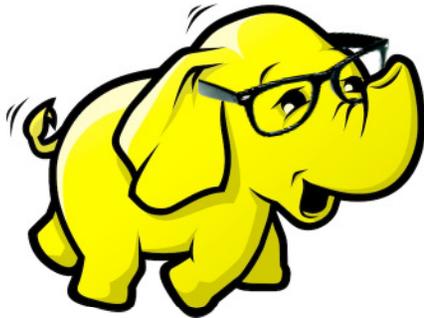
*Limited use license bundles

InfoSphere BigInsights: déclinaisons



Accès SQL pour Hadoop: Pourquoi?

- L'extension du DW est un cas d'usage majeur pour Hadoop

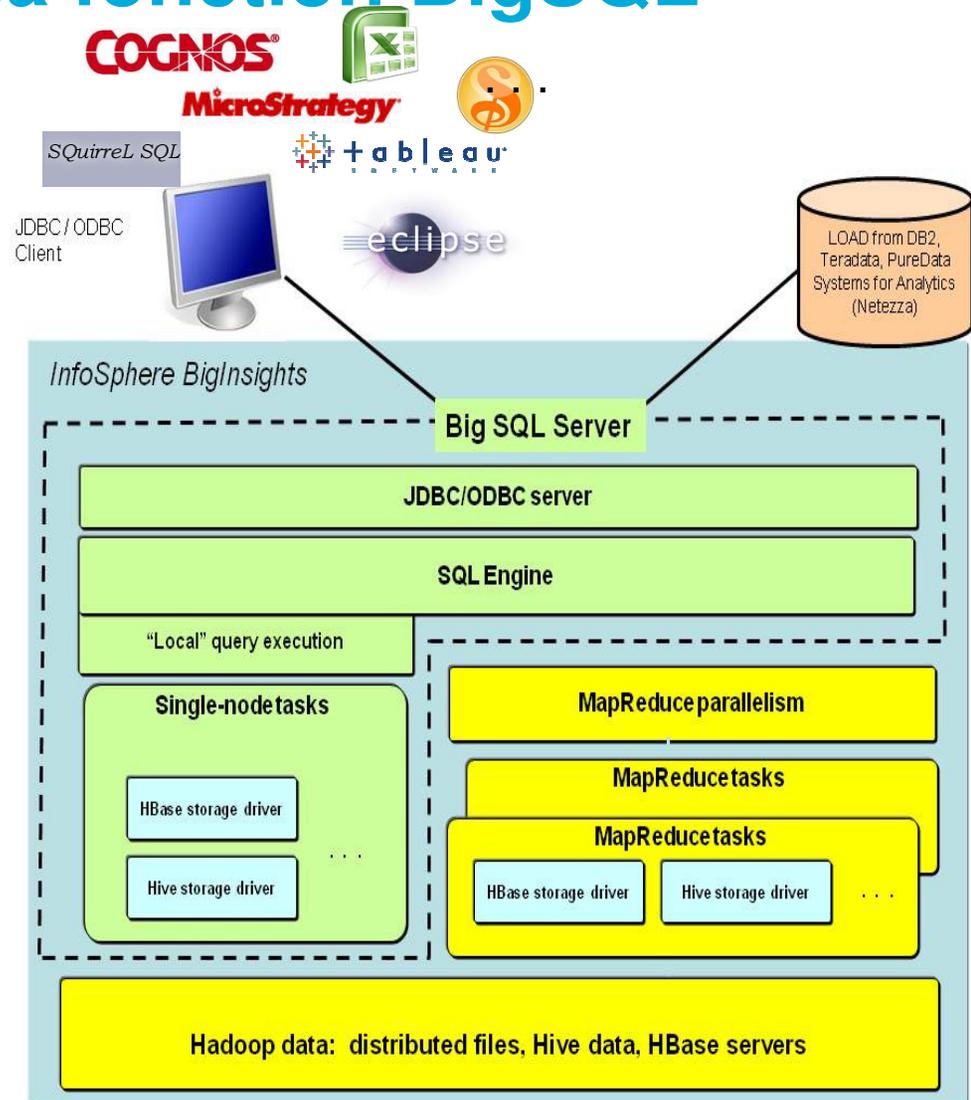


- Hadoop souvent perçu comme complexe:
 - L'API Java MapReduce nécessite une expertise en programmation
 - Langages non familiers (ex Pig) nécessitent des compétences spécifiques
- Le support BigSQL de Biginsights ouvre les données à une plus large audience
 - Syntaxe familière, largement connue
 - Catalogue commun pour identifier données et structures

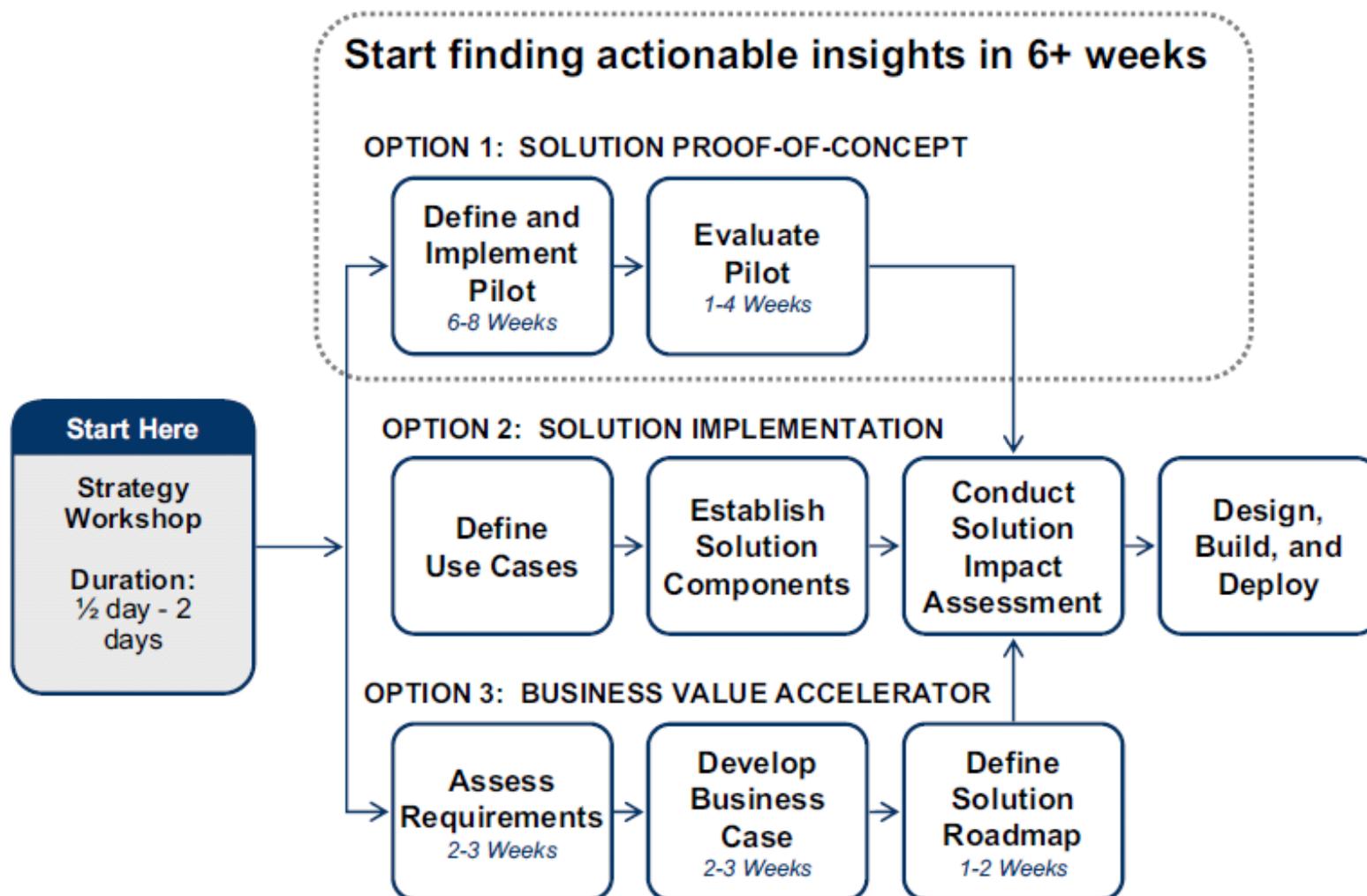


IBM BigInsights et sa fonction BigSQL

- Standard SQL syntax and data types
 - Joins, unions, aggregates . . .
 - VARCHAR, decimal, TIMESTAMP, . . .
- JDBC/ODBC drivers
- Optimization
 - MapReduce parallelism or...
 - “Local” access for low-latency queries
- Varied storage mechanisms appropriate for Hadoop ecosystem
- Integration
 - Eclipse tools
 - DB2, Netezza, Teradata (via LOAD)
 - Cognos Business Intelligence
 - , , ,



Comment engager un projet Big Data



IBM SolutionsConnect 2013

Quelques liens utiles:



Data Warehouse Augmentation

With Vijay Ramaiah
Product Manager IBM Big Data

<http://www.youtube.com/watch?v=JcCWvVIAAMA>

The Hadoop Data Warehouse - Part 1

With James Kobielus
IBM Big Data Evangelist

<http://www.youtube.com/watch?v=Rcxsj-5WjMQ>

The Hadoop Data Warehouse - Part 2

With James Kobielus
IBM Big Data Evangelist

<http://www.youtube.com/watch?v=O8QokzzwbLE>



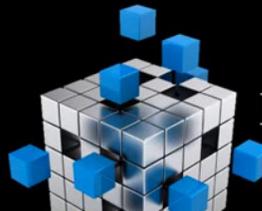
IBM Big Data

- Reference Architectures -
Data Warehouse Augmentation - Part 1

With Bill O'Connell
IBM Distinguished Engineer

August 2013

http://www.youtube.com/watch?v=L_s-x1HAI5k



IBM Big Data

- Reference Architectures -
Enhanced 360 Degree View of the Customer

With Mandy Chessell
IBM Distinguished Engineer

May 2013

<http://www.youtube.com/watch?v=iwvpR2d6rWw>



#solconnect13



IBM SolutionsConnect 2013

Tak (Danois)
شكراً (Arabe)
आभार (Hindi)
고맙습니다 (Coréen)
תודה רבה (Hébreu)
धन्यवाद (Chinois traditionnel)
Spasibo (Russe)
Gracias (Espagnol)
Obrigado (Brésilien portuguais)
Thank You (Anglais)
Trugarez (Breton)
Grazie (Italien)
Dankon (Esperanto)
 Danke (Allemand)
고맙습니다 (Chinois)
Merci (Français)
Tack så mycket (Suedois)
go raibh maith agat (Gaélique)
Данк у (Hollandais)
go raibh maith agat (Gaélique)
இவ்வாயு (Tamil)
Dekujeme Vam (Tchèque)
ขอบคุณ (Thaï)



#solconnect13

