



Tendances Logicielles
Hiver 2006





***Découverte des données et
retour d'expérience clients***

Steven Haddad

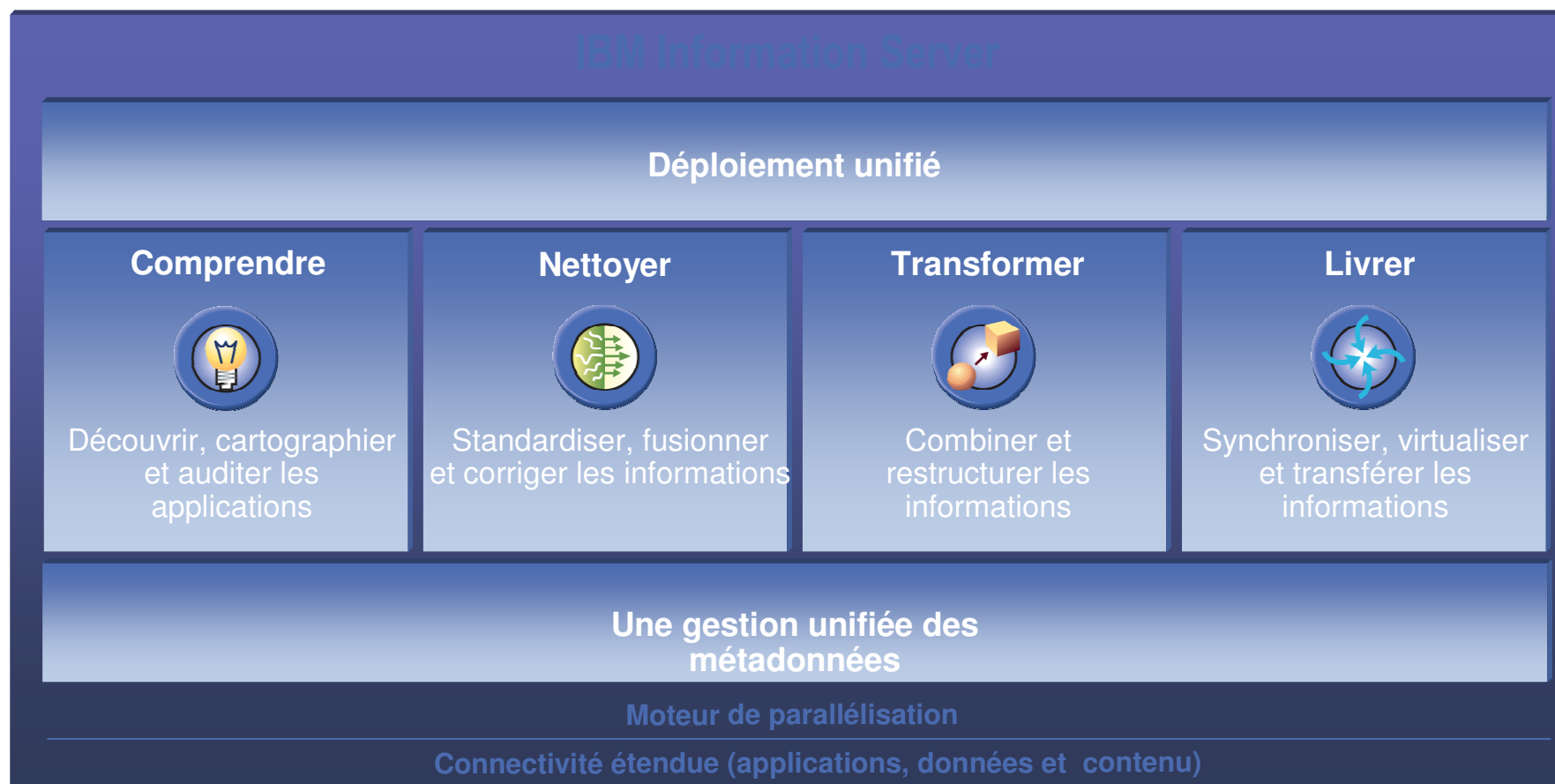
Eric Messin



Agenda

- **Découverte des données**
 - Démarche et outils
- **Cas clients & retours sur investissement**
- **Partenaires IBM associés**

IBM Information Server : *Pour un accès à une information fiable*



Exemples de services d'information

Services de validation de données

Validation des données par rapport aux règles métiers

Services de nettoyage de données

Harmonisation et dédoublonnage des données

Services de transformation de données

Transformation des données en provenance de différentes sources

Services de données partenaires

Intégration de données Électroniques partenaires
SWIFT, EDI, HIPAA

Services de données opérationnelles

Accès unifié aux données opérationnelles

Services de données analytiques

Accès aux données analytiques en temps-réel

Services de données non structurées

Accès aux données non structurées en relation avec des données structurées

Services de données de référence

Consultation et intégration des données de référence

Les étapes de la démarche découverte des données à l'échelle de l'entreprise

Mini DQA

Livrables
Rapports d'anomalies
Durée 5 à 10 jours

DQA Data Quality Assessment Niveau Projet

Livrables
Pertinence & cohérence
des données
Harmonisation optimale
Durée 1 à 3 mois

CCID Centre de Compétences Intégration de Données

Livrables
Mise en œuvre de
Méthodologie qualité des
données
Durée 3 à 9 mois

Agenda

- **Découverte des données**
 - Démarche et outils
- **Cas clients & retours sur investissement**
- **Partenaires IBM associés**

Découverte sur systèmes d'accès

• Les outils fonctionnent !

- Nombreuses incohérences dans l'annuaire
 - ▶ Champs très souvent vides
 - ▶ Deux clefs non uniques, attributs dupliqués
 - ▶ Format incohérents
 - ▶ Données fausses
 - ▶ Données contradictoires
- Normalisation des adresses PAC
- Identification facile de nombreux doublons

Business phone : "No tel number"
Company name : 000049
First name : ++ mit, ???, 1
Office 2nd address : +£bersetzg.
Personal title key : SOGERMA, FR
Country code : AF, FJ
Company : UNDISCLOSED, NOT ASCII

Entry date : 9999-12-01

Entry date > Leaving date

40	XA2	14,20	ALPHACOMS
40	DA2	8,24	ALPHACOM
42	XA2	14,20	IFRS
42	CP2	1,06	IFR
43	XA2	14,20	BREVALEX
43	CP2	0,88	BREVELEX
58	XA2	14,20	JARDIN DES GOURMETS
58	DA2	7,01	LE JARDIN DES GOURMETS
74	XA2	14,20	LE CALVE
74	DA2	8,24	LE CALVEZ

• Premières impressions

- La mise en place est rapide (3 jours de préparation, 2 de développement, 2 de restitution)
- Mais la **méthode** est la clef, bien plus que l'outil...

Découverte sur données clients CRM & SAP

- L'un des leaders mondiaux en France dans le domaine de l'alimentation rapide (snack, aliments pour animaux, plats cuisinés)
- > 10 millions de clients référencés
- CA > 15 milliards €

Objectifs et enjeux	Solution	Résultats
<ul style="list-style-type: none"> ▪ Mise en œuvre d'une solution CRM, afin de construire une vue unique des clients, d'optimiser les campagnes marketing et de mieux segmenter et analyser le marché (10M prospects) ▪ Réduire les coûts de gestion des NPAI ▪ Étendre les capacités du « Global Data Expert Center » aux problèmes de qualité de données 	<ul style="list-style-type: none"> ▪ Mise en œuvre d'une plateforme commune de gestion de qualité de données ▪ Validation d'adresses sur 15 pays ▪ Développement de « Best Practices » et d'un jeu de services de nettoyage (validation et standardisation et de-doublonnage) ▪ La solution utilisée QualityStage 	<ul style="list-style-type: none"> ▪ La plateforme d'intégration gère (standardise, dé-doublonne, vérifie l'intégrité, valide les adresses) plus de <u>100 000 données consommateurs par heure</u> ▪ Ce client a réduit de façon considérable ses coûts opérationnels d'intégration d'information dans les systèmes (CRM : Epiphany, ERP : SAP, etc...) ▪ 5% doublons supplémentaires détectés comparé à une démarche SQL, soit 75k€ d'économie par mois (base 3 à 10 € par courrier, 500k mails/mois)

Découverte sur données financières: Crédit Revolving

- Echantillon de 5000 lignes sur un total fichier de 5 millions de cartes
 - Chiffre d'affaires moyen par carte: 1000 euros
- Résultat sur l'échantillon:
- 26 cartes sans porteurs (20 + 6)

<NULL>	20	0	16446
00000	6	0	90000
002155318	5	0	90000
000570357	4	0	90000
053283414	4	0	90000
050845885	4	0	90000
000324810	4	0	90000
900261517	3	0	90000
900288509	3	0	90000
002551191	3	0	90000
900238793	3	0	90000
001131924	3	0	90000
006793683	3	0	90000

-ROI:

**26 000 euros sur l'échantillon, à extrapoler sur 1000 fois plus de cartes dans le fichier de référence

**Coût de fraudes sur ces cartes sans porteurs

Découverte sur sources hétérogènes

- Effort manuel d'un client du secteur public: – profiling de données sur 5 sources pour un total de 88 tables. Scripts SQL ont été écrits et exécutés en 120 heures, soit 15 jours. Il a fallu ensuite ajouter le temps d'un Expert en données pour valider les résultats.

Ecriture de scripts à la main et exécution pour obtenir les résultats cartographiés

Analyse

4 semaines

- Effort automatisé en utilisant l'outil de cartographie: 5 jours ont suffi pour auditer et présenter les résultats à l'expert en données.

Automated Data
Profiling Tool

Analysis

2 semaines

Découverte des données à l'échelle de l'entreprise exemples de projets qui conduisent au CCID

Mini DQA

Livrables
Rapports d'anomalies
Sur badges d'accès
entreprise

DQA Data Quality Assessment Niveau Projet

Livrables
Pertinence & cohérence
des données
Habilitation applicative et
Confidentialité des
données accessibles

CCID Centre de Compétences Intégration de Données

Livrables
Mise en œuvre de
Méthodologie qualité des
données
Systèmes de cohérence sur
données SAP et non SAP

Agenda

- **Découverte des données**
 - Démarche et outils
- **Cas clients & retours sur investissement**
- **Partenaires IBM associés**

Les investissements des partenaires IBM pour la découverte de vos données

- **Cycle de formation**
 - Produits IBM
 - Proof of Technology IBM
- **Consultants spécialisés**
 - Découverte de données
- **Evénements communs**

Pour plus d'informations

- **Pour plus d'information sur la découverte de vos données**
 - Steven Haddad
 - steven.haddad@fr.ibm.com
 - Eric Messin
 - eric.messin@fr.ibm.com