

**Power
Week**



Université IBM i 2019



22 et 23 mai

IBM Client Center Paris

S50 – IBM i et Modernisation Entrepôts de données

Corinne Baragoin
c_baragoin@fr.ibm.com

Francis Arnaudies
f.arnaudies@fr.ibm.com

IBM France

Plan de la présentation

- Les architectures de données pour l'analytique évoluent
- Architecture de données: point de vue IBM
- Modernisation des entrepôts de données sur IBM i
- Retour d'expérience



**Power
Week**

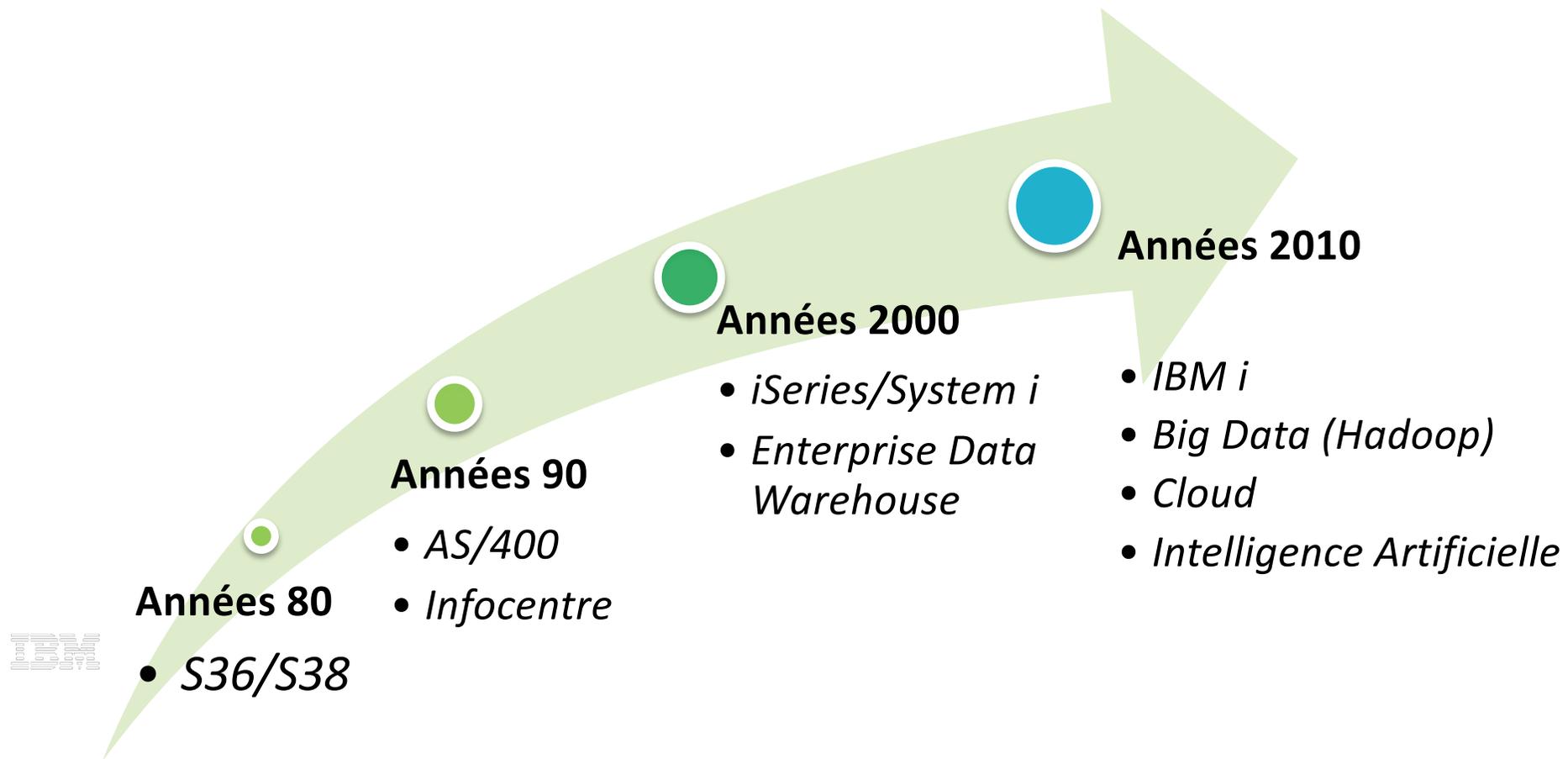
Université IBM i

22 et 23 mai 2019

IBM

Les architectures de données
évoluent

Historique



IBM

Data Warehouse /Big Data

« L'émergence de l'analyse de données volumineuses et l'adoption d'architectures BigData amènent beaucoup de personnes à s'interroger sur l'avenir des entrepôts de données.

Cependant, des enquêtes récentes montrent que plus de 60% des entreprises exploitent encore entre 2 et 5 entrepôts de données.

Beaucoup de gens ont parlé d'éliminer complètement l'entrepôt de données. Mais la réalité est que l'entrepôt de données offre une valeur que le Data Lake n'aborde pas, et inversement. Le véritable défi consiste à combiner l'entrepôt de données avec une architecture de Data Lake, des pipelines de données modernes et des cas d'utilisation de l'analyse. » *Bloor Group, Data Warehousing in the Age of Data Lakes*



Laurent Mignon, Président de la BPCE (Banque Populaire Caisse d'Epargne), 29 janvier 2019 : <http://www.larevuedudigital.com/la-banque-bpce-en-attente-de-son-data-lake-pour-passer-a-la-vitesse-superieure/>

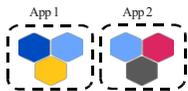
« Tout dépend de la manière dont vous allez utiliser la donnée » reprend le dirigeant. « La technologie vous pouvez l'appeler Machine Learning, Deep Learning, intelligence artificielle, etc. le véritable problème est comment bien utiliser la donnée dont nous disposons pour améliorer notre capacité à mieux servir nos clients. C'est le premier enjeu » insiste-t-il.

Le problème est de bien utiliser la donnée existante

« Si nous réussissons cela, je pense que nous transformerons en profondeur notre entreprise. De manière implicite, l'usage de la donnée était réalisé dans la tête de nos conseillers en agence. Nous devons élargir l'accès à la donnée » dit-il.

Architecture de données: tendances

FASTER



Faster time to deployment
(Containerized Micro-Services)

BIGGER



Scalability

SMARTER



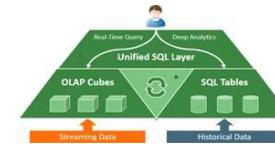
Machine Learning / Deep Learning frameworks (Python, Spark ML, TensorFlow, Caffe)

TRUSTED



Data Swamp->Data Lake

REAL-TIME SQL



One SQL Layer
(Across Historical, Real-time)

HYBRID MULTI-CLOUD



IBM Cloud



La problématique

Chez un client (représentatif...)

6 mois

C'est le temps moyen de mise en œuvre d'une donnée non prioritaire dans la BI.

Durée moyenne : 6 mois / coût moyen : 10 à 15k€

8 interlocuteurs différents en séquentiel

La nécessité de savoir ce que l'on cherche et de le spécifier → contraire à un mode d'expérimentation



Les grands axes d'évolution

- Simplicité et Performances
 - Bases colonne, en mémoire
 - Scalabilité/ performances pour les utilisateurs
 - Maîtrise des coûts
- Evolution vers le temps-réel
 - Intégration de données en temps-réel
 - Analytique en temps-réel
- Evolution de l'architecture de données
 - On-premise
 - Dans le cloud public
 - Hybride
- Montée en puissance du cloud privé
 - Approche Container (Docker)



**Power
Week**

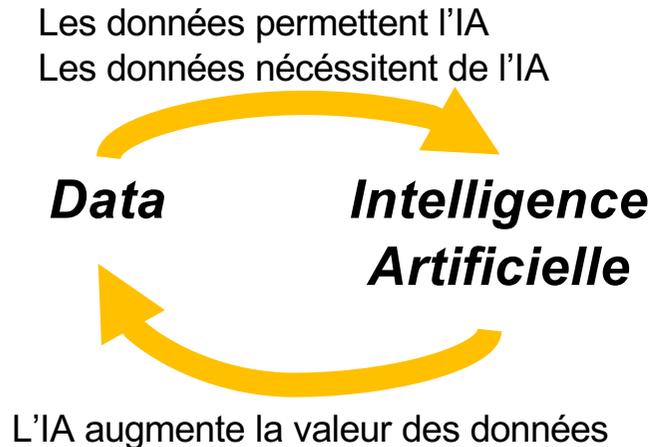
Université IBM i

22 et 23 mai 2019

IBM

Architecture de données: Point de vue IBM

A l'ère de l'Intelligence Artificielle, les données sont essentielles



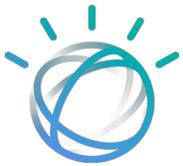
Enjeux liés aux données

- Quelles données pour entraîner les modèles ?
- Les données sont-elles aisément accessibles ?
- Quels sont les volumes ?
- S'agit-il de données structurées? Non-structurées ?

Enjeux liés à l'IA

- Les modèles ont-ils été suffisamment testés ?
- Avec quels jeux d'essai ?
- Choix des algorithmes
- Utilisation dans les applications

L'approche Data & AI d'IBM



AI

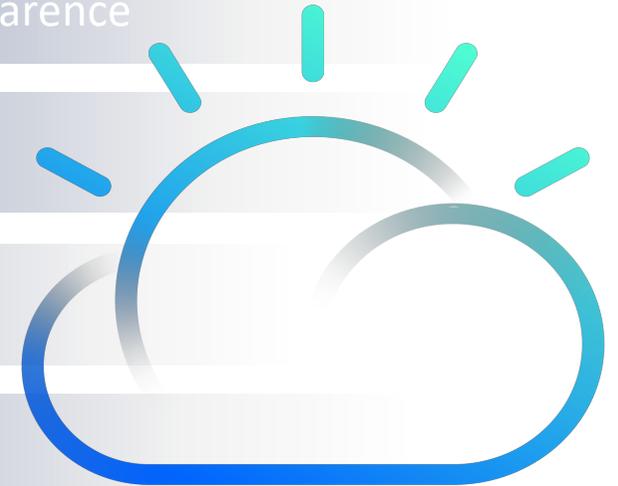
INFUSER – Opérationnaliser l'IA avec confiance et transparence

ANALYSER – Passer à l'échelle avec l'IA et l'Analytics partout

ORGANISER – Créer un socle données de confiance

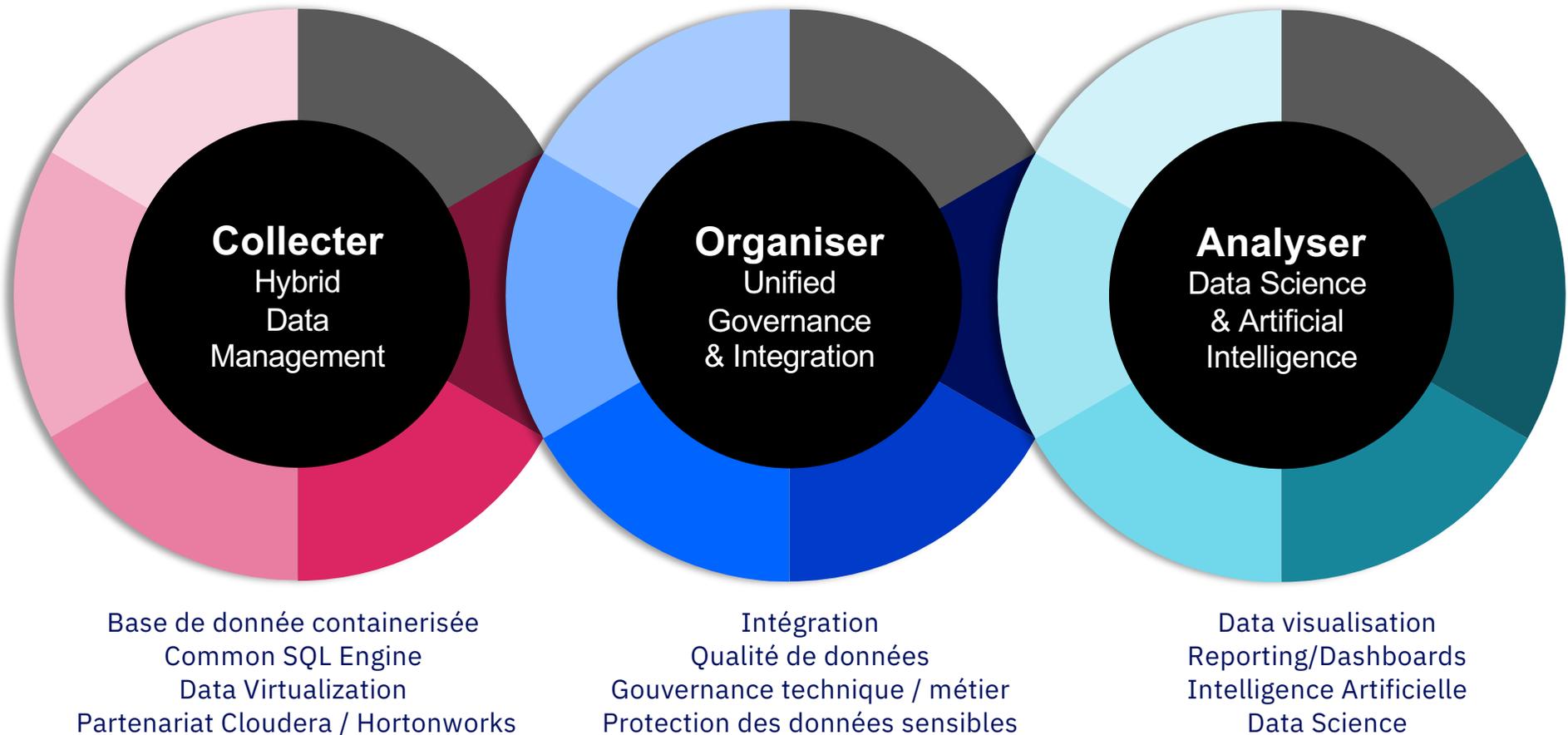
COLLECTER – Rendre la données simple et accessible

Données de tout type,
où qu'elles soient

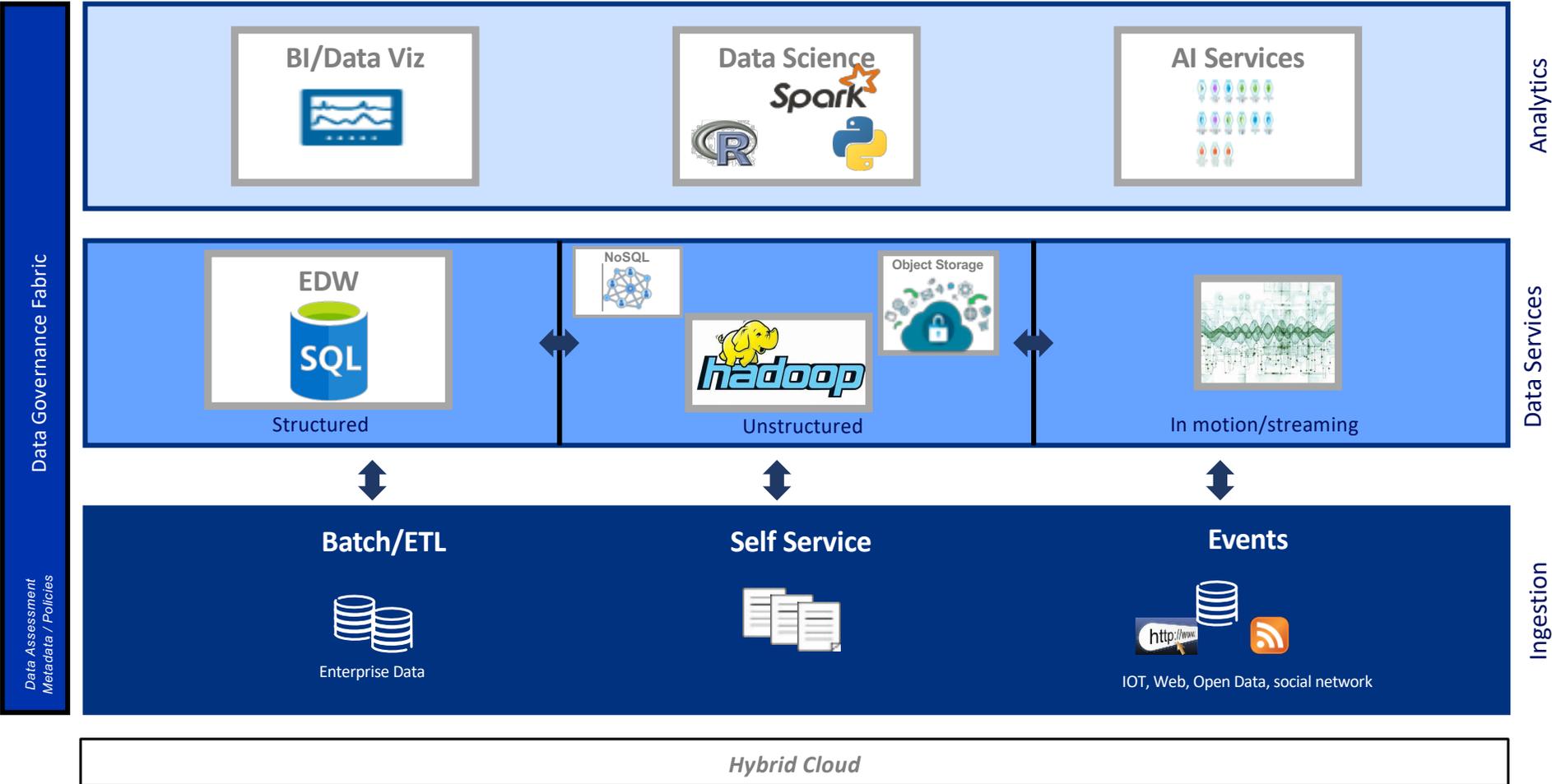


MODERNISER
ses applications
analytiques à l'ère de
l'IA et du multi-cloud

L'approche Data & AI d'IBM: fonctions



Architecture Data & AI



**Power
Week**

Université IBM i

22 et 23 mai 2019

Modernisation des entrepôts de données existants sur IBM i

IBM

Evolution des Entrepôts de données sur IBM i

- Améliorer la satisfaction des utilisateurs et répondre aux nouveaux besoins métier (plus de données cross métiers, externes, plus de granularité, plus d'historiques ...)
 - Bénéficier d'une solution performante et simple de mise en œuvre et d'utilisation, optimisée pour un usage analytique
 - Isoler des systèmes opérationnels
 - ne pas dégrader/ralentir les temps de traitement des systèmes transactionnels et les temps de réponse utilisateur
 - permettre plus d'exploration des données (requêtes ad hoc)
 - Intégrer facilement et rapidement cette nouvelle solution (Gagner en agilité)
-  • Minimiser l'impact sur l'existant
- Bénéficier d'une solution assurant sécurité, résilience et haute-disponibilité

**Power
Week**

Université IBM i

22 et 23 mai 2019

IBM

Retour d'expérience

Ce client iSeries : 3 ambitions majeures

Refonte Décisionnel

- Satisfaire aux exigences de performances des utilisateurs
- Minimiser l'effort de migration
- Pas de DBA dans l'équipe
- Privilégier l'application ERP ses performances et sa sécurité

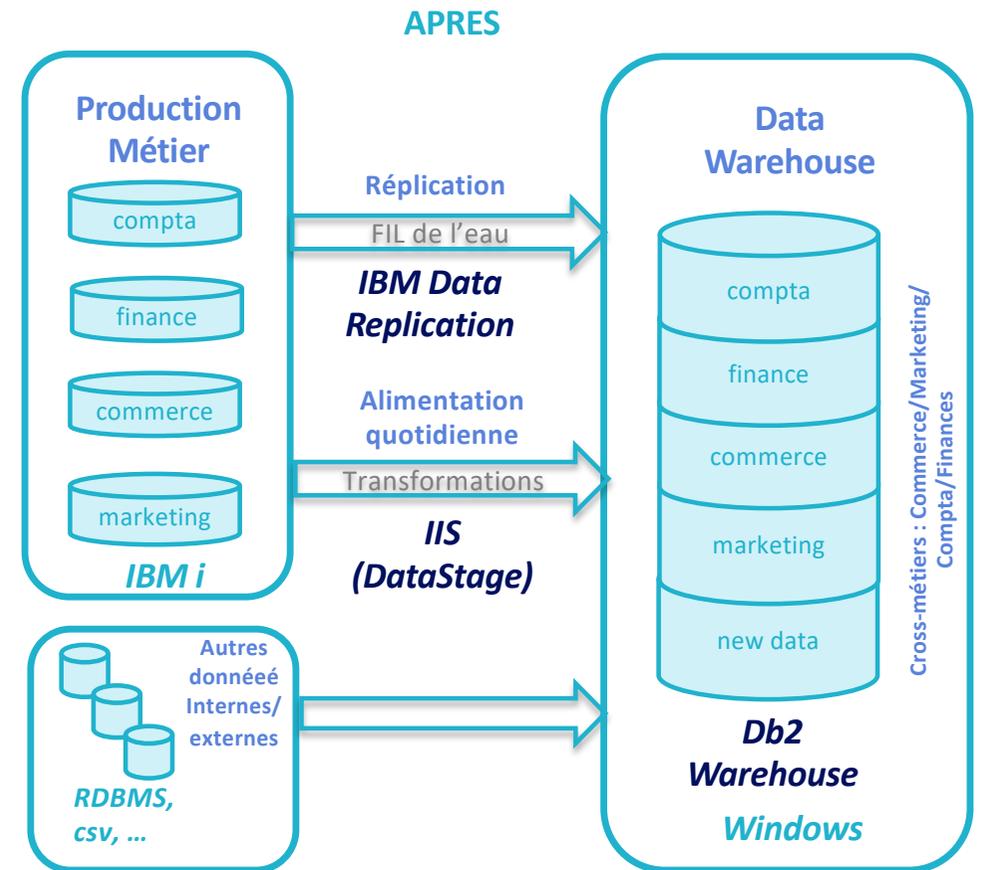
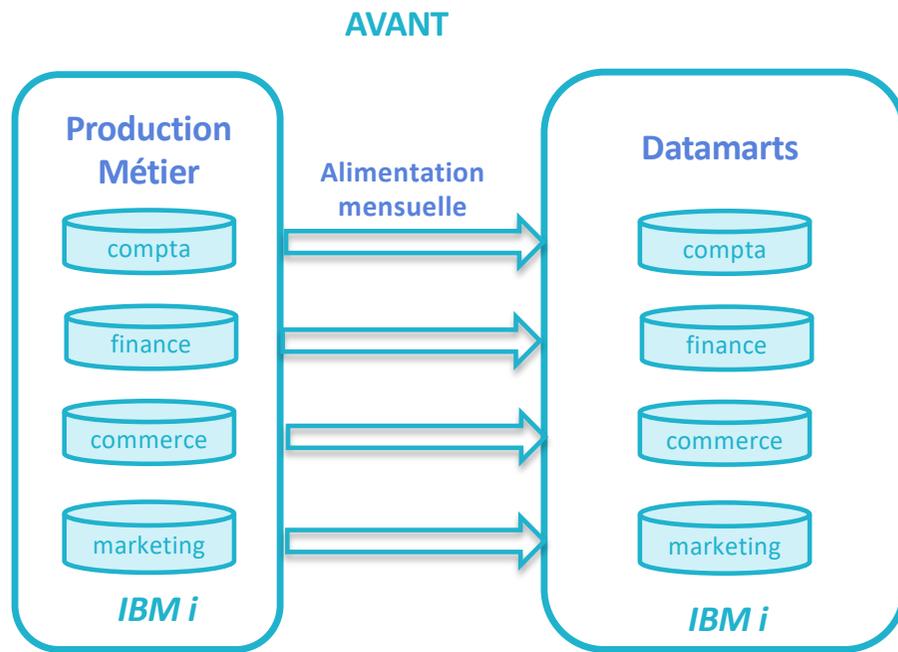
Satisfaire les nouveaux besoins métier

- Fournir des données plus fraîches, avec une granularité plus fine et plus d'historiques cross métiers
- Répondre aux besoins d'analyse complexes (ad-hoc)
- Accéder à données autres qu'ERP, externes

Exigence intégration

- Rapidité de mise à disposition de l'information (au fil de l'eau)
- Fenêtre batch réduite (Europe, WW)
- Précision et qualité des données

Architecture adoptée



Pourquoi Db2 Warehouse est le plus adapté au monde IBM i ?

- Famille Db2 aussi robuste
 - Facilité d'intégration avec le monde IBM i
 - Solution optimisée pour workloads décisionnels/ hub de données structurées
- Simplicité et Performances du moteur Db2 Warehouse
 - Bases colonne, en mémoire, Compression
 - Scalabilité/ performances pour les utilisateurs (Parallélisation)



■ Approche container (Docker)

- Disponible sous plusieurs modes: Appliance/ On Premise/ Multicloud

Db2 Warehouse: points forts



Simple

Fully-managed data warehouse- As A Service combiné à simplicité "load and go". IBM gère l'installation, la configuration, le tuning, et les opérations de haute disponibilité.

Performant

In memory, columnar database. Mécanismes de compression, data skipping. Scalabilité verticale par ajout de processeurs et horizontale par extension en architecture MPP.

Souple

Lors d'évolution des besoins, possible de déplacer facilement des workloads analytics d'un environnement à un autre dans le Cloud . Possibilité de data virtualisation.

Ouvert

Support et intégration d'analyses avancées: R, Spark au plus près des données . Compatibilité Oracle. Support Natif XML, JSON

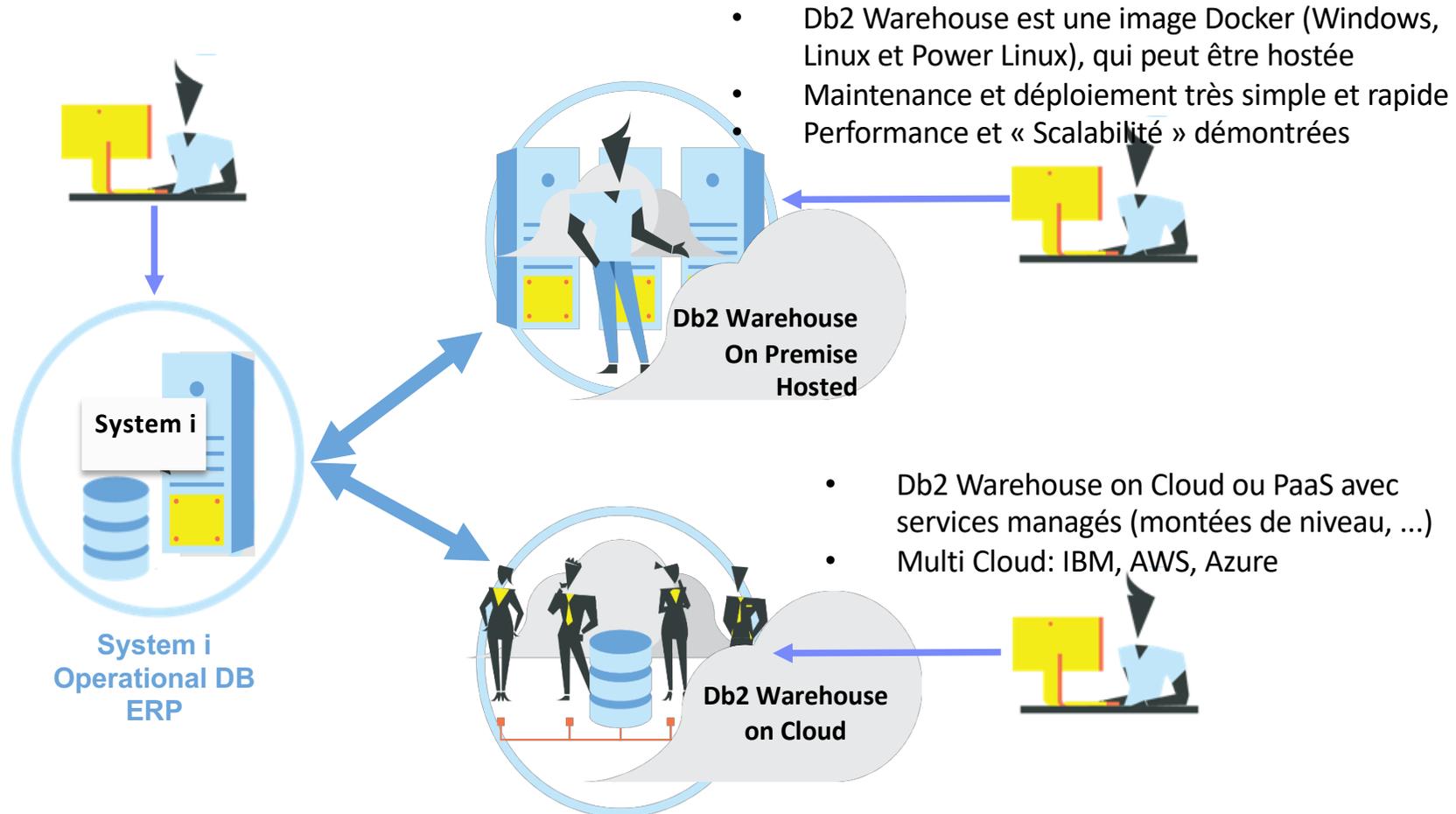


28 Novembre 2016

20

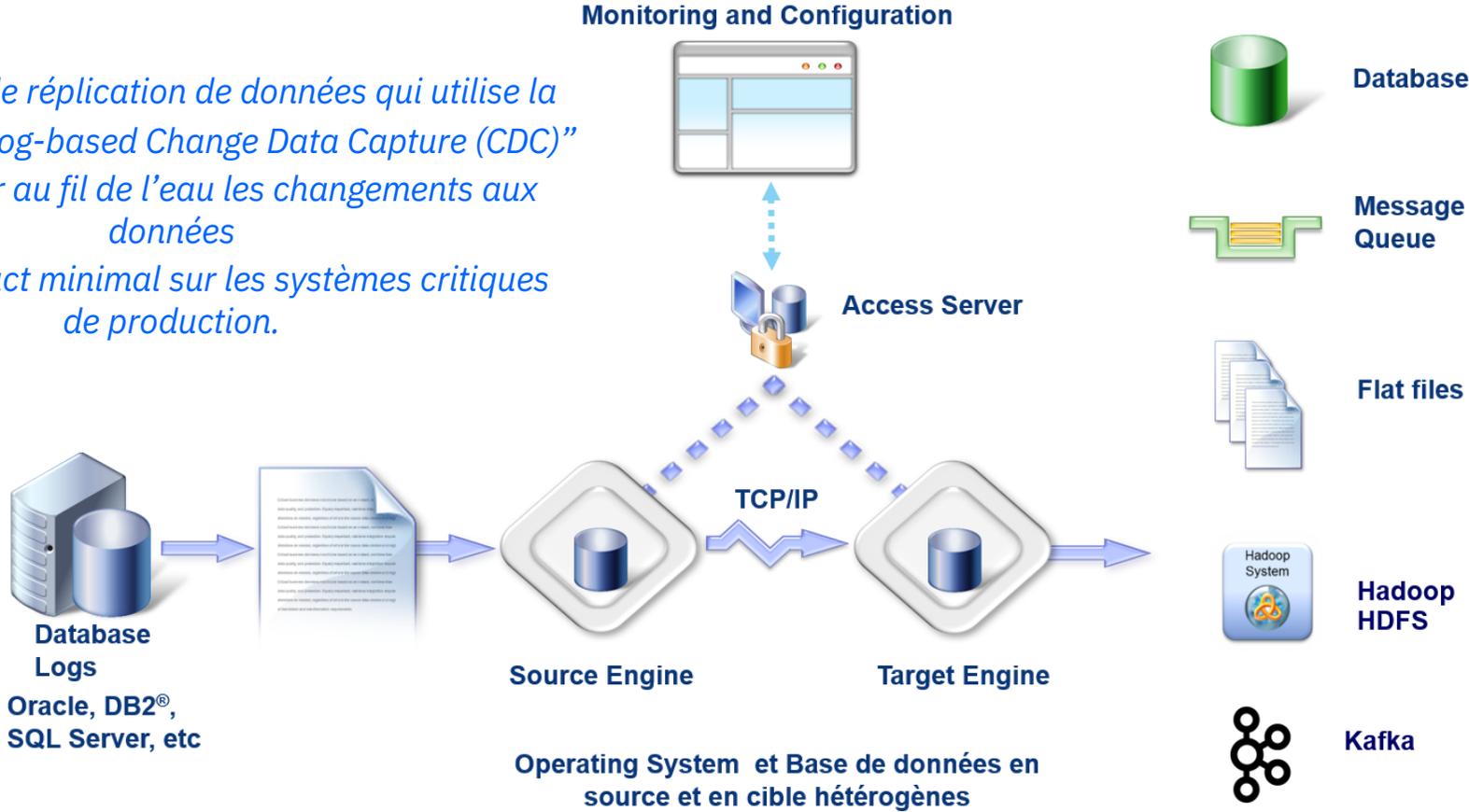
Db2 Warehouse:

base mémoire, portable dans le Cloud



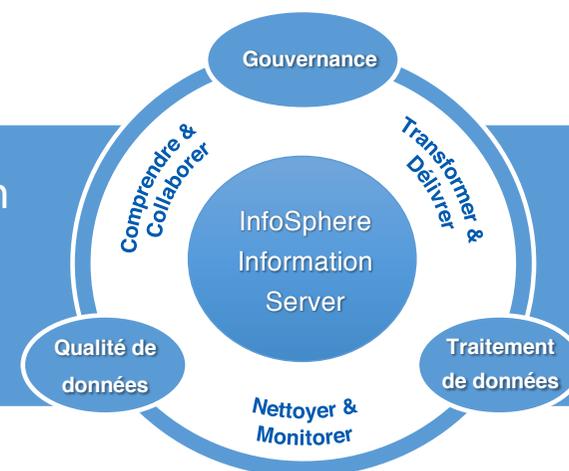
Fast Data avec IBM Data Replication

Une solution de réplication de données qui utilise la technologie “Log-based Change Data Capture (CDC)” pour délivrer au fil de l’eau les changements aux données avec un impact minimal sur les systèmes critiques de production.



InfoSphere Information Server

Intégrer et transformer les données pour délivrer une information cohérente, précise et complète, au sein d'une unique plateforme unifiée par une couche commune de métadonnées



Gouvernance

Comprendre & Collaborer

- Cartographier les données du SI alignées avec le glossaire métier
- Gérer la traçabilité des données
- Gérer les politiques de gouvernance des données



Qualité de données

Nettoyer & Monitorer

- Analyser, classifier & valider les données
- Définir, gérer & monitorer les règles de qualité & exceptions
- Nettoyer & standardiser



Traitement de données

Transformer & Délivrer

- Scalabilité massive
- Puissance de calcul pour tout type de complexité
- Délivrer en mode batch ou temps réel

- Une connectivité commune
- une couche commune de métadonnées
- un moteur d'exécution commun
- un déploiement flexible: On-premise, Cloud ou Hadoop nativement

