

**Un Saas pour le cryptage des données issues d'un
ETL**

Présenté par

Hana Gara Kort

Sous la direction de

**Dr Jalel Akaichi
Maître de conférences**



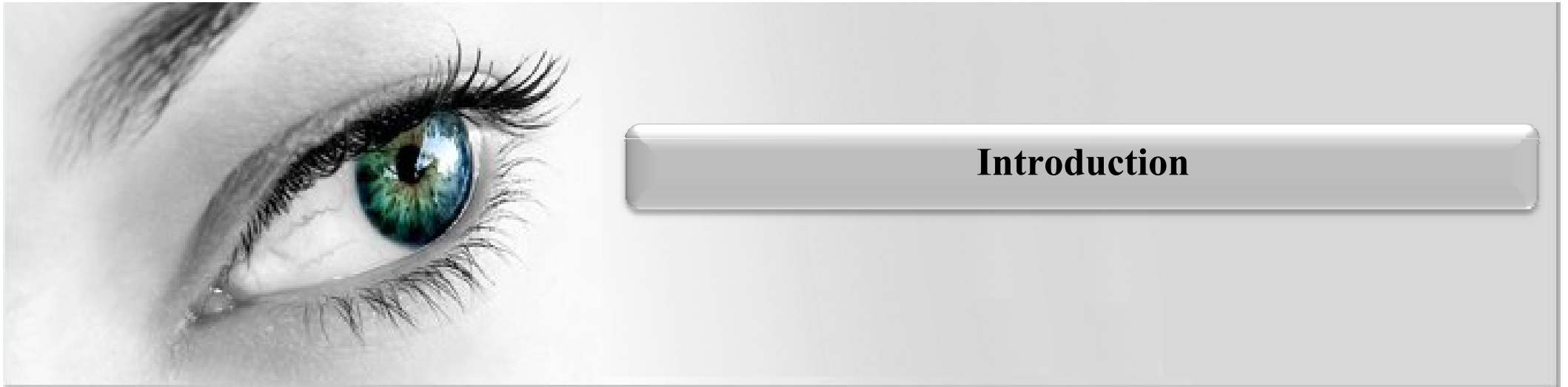
2. Les problèmes de sécurité liés aux architectures des entrepôts de données dans le Cloud

3. Fonctionnalités et scénarios de migration vers le Cloud

5. Mise en œuvre d'un SaaS pour le cryptage

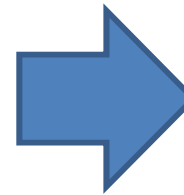
6. Conclusion & Perspectives





Introduction

Introduction



Entrepôt de données

Cloud Computing

Sécurité

Problématique

- ✓ Les entrepôts de données, par leur nature même créent un conflit au niveau de la sécurité.
- ✓ **Augmentation des risques d'attaques et de violations des données** dans ce nouvel environnement.
- ✓ provoquées par des menaces.



Objectifs

- ✓ (1) : Aborder les problèmes de la sécurité liés aux architectures de l'entrepôt de données dans le nuage,
- ✓ (2) : Analyser les différents scénarios possibles pour la migrations des fonctionnalités de l'entrepôt de données vers le Cloud,
- ✓ (3): Proposer une solution SaaS pratique pour le cryptage des données dans l'outil ETL Talend Open Studio.



Définition

Qu'est-ce que le cloud Computing?

Réseau omniprésent: Un accès ubiquitaire au réseau

Service mesuré facturé

élasticité rapide

Service à la demande

Multi Locatères: Une mise en commun des ressources

Caractéristiques

SaaS

PaaS

IaaS

Modèles de service

Publique

Privé

Hybride

Communauté

Modèles de déploiement





**Les problèmes de sécurité liés aux architectures des
entrepôts de données dans le Cloud**

Architectures

1 Data Marts indépendant

2 Data Marts en bus

3 Entrepôt de données centralisé



4 Architecture Hub-and-Spoke

5 Architecture hybride (Fédérée)

6 Architecture trois tiers(niveau)

7 Architecture multi niveaux





Quels sont les problèmes de sécurité liés
aux architectures de l'entrepôt de
données ?

Les problèmes de sécurité

- ✓ 1 La centralisation
- ✓ 2 L'hétérogénéité
- ✓ 3 Le niveau de détail
- ✓ 4 La sécurité de stockage
- ✓ 5 La sécurité de transmission



Quels sont les problèmes de sécurité liés aux architectures de l'entrepôt de données qui peuvent être poser **dans le Cloud** ?

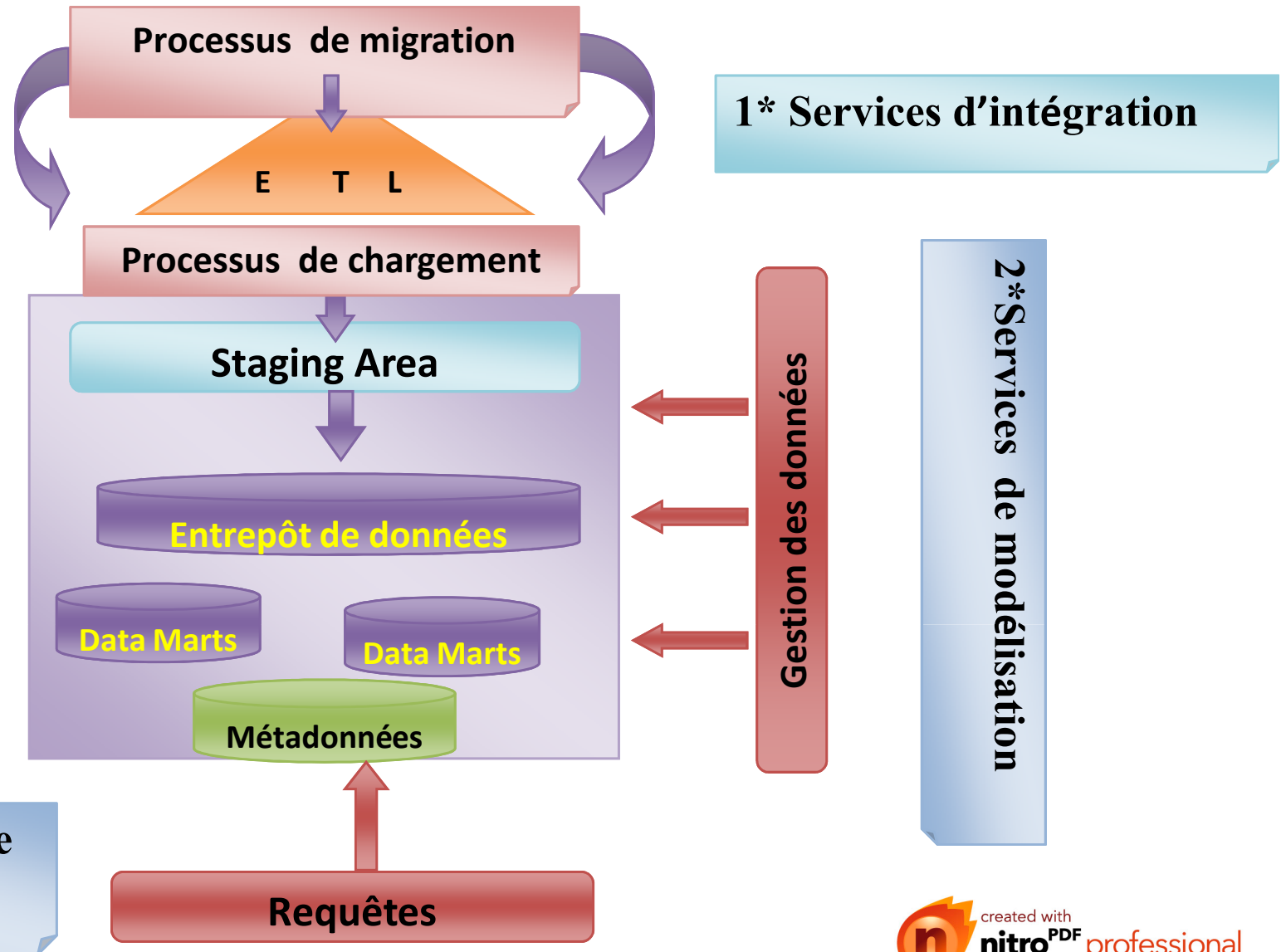
Problèmes liés au Cloud

- ✓ 1 La virtualisation
- ✓ 2 La localité des données
- ✓ 3 L'impact des modèles de déploiement
- ✓ 4 La multi-location
- ✓ 5 La gestion des identités



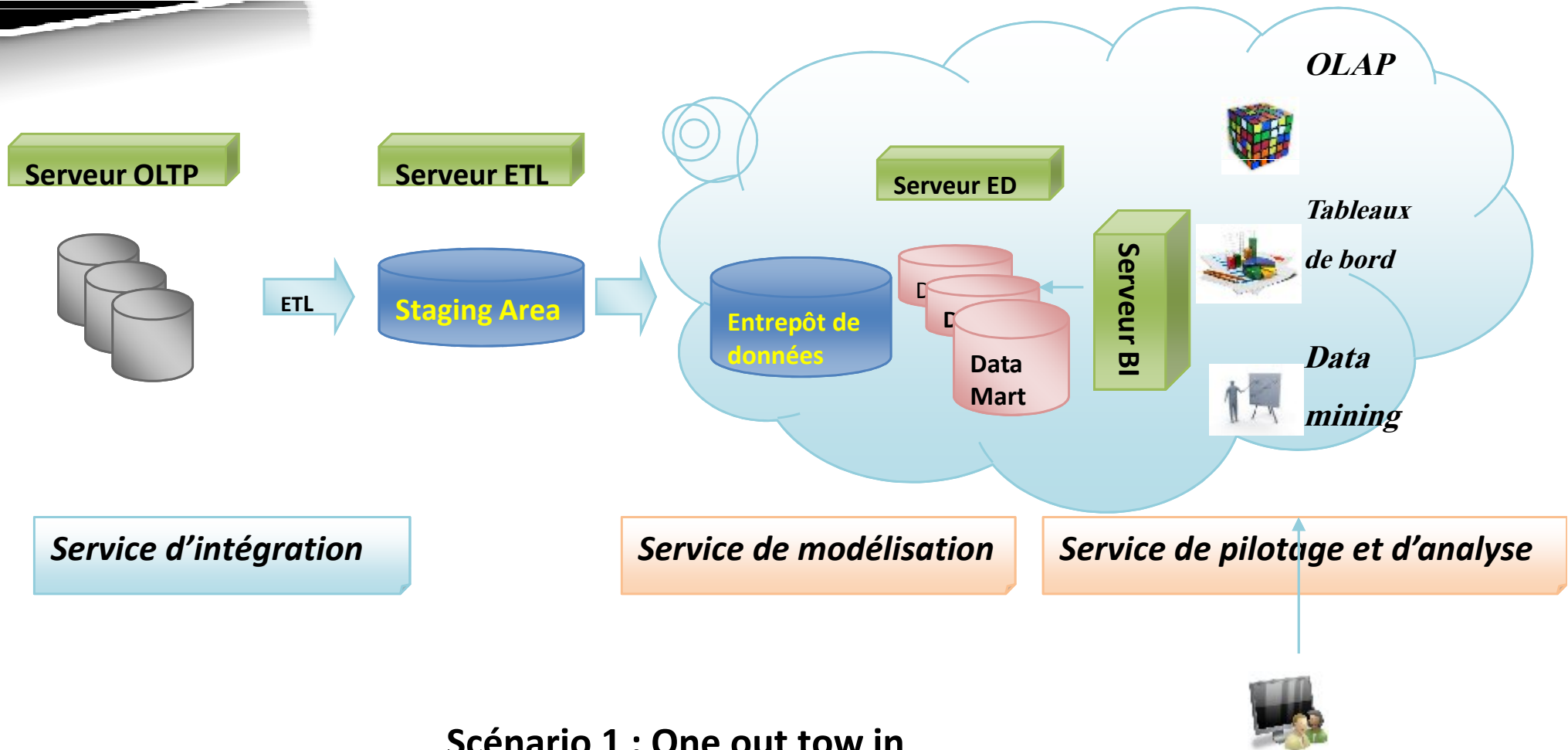
Fonctionnalités de l'entrepôt de données et migration vers le Cloud

Fonctionnalités



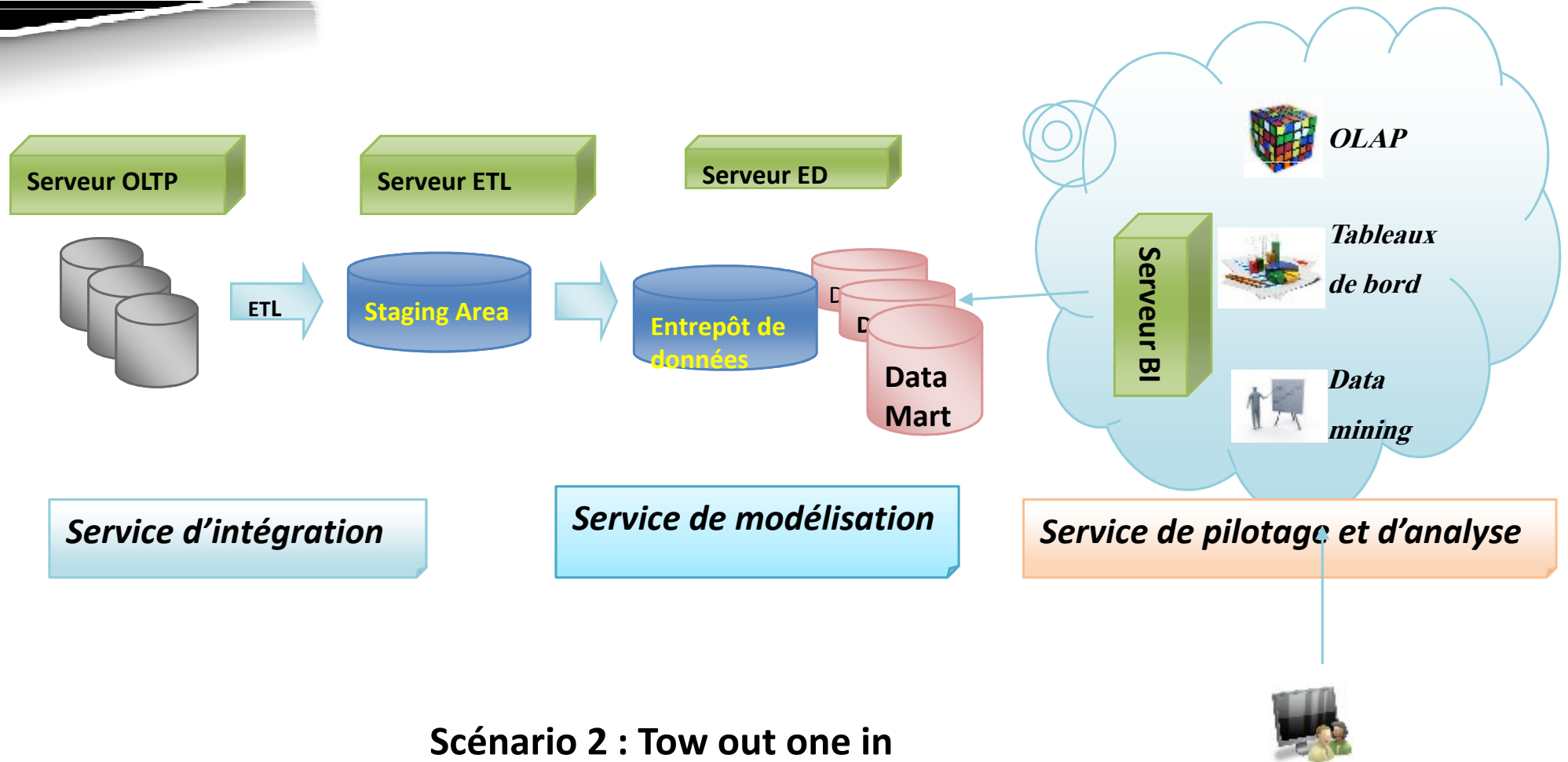
3* Services de pilotage et d'analyse

Scénario 1



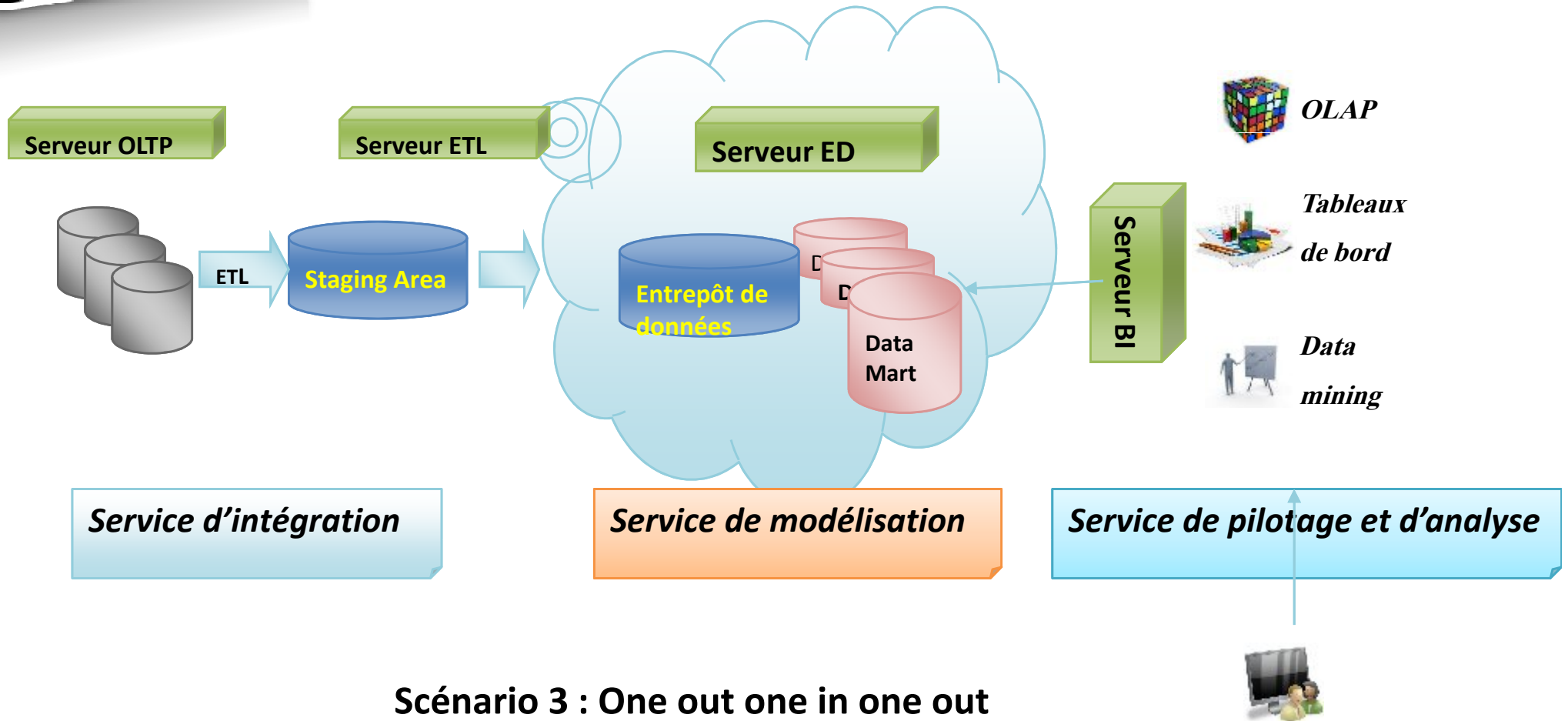
Scénario 1 : One out tow in

Scénario 2



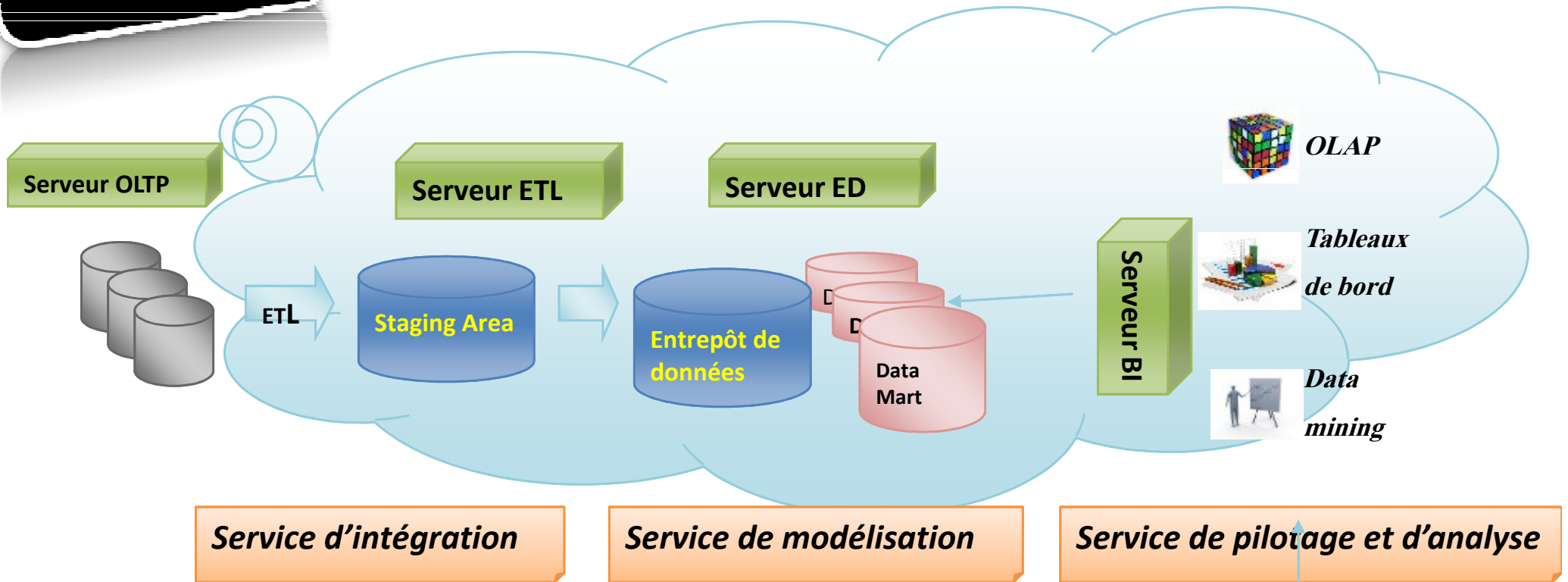
Scénario 2 : Tow out one in

Scénario 3



Scénario 3 : One out one in one out

Scénario 4



Scénario 4 : All in

Constatations

Externaliser le minimum	Déléguer à un prestataire compétant
le grand travail	
un système de sécurité complexe contre les attaques	Solution fiable
source de vulnérabilité et pas très souvent efficace.	
cout >>	cout << et sécurité





Architecture de sécurité

Architecture de sécurité

Cloud : Fournisseur de service : Destination de données

SaaS

PaaS

IaaS

Infrastructure d'analyse

Entreprise cliente :
Source de données

Bases de données



Répertoire FTP



Fichiers (XML,
Texte, Cvs, Excel ..)



Base tampon+ETL



Codage routine



Clé publique



Fichier java

ED Crypté

Independant
&
homogène



Extraction

Cryptage

Chargement

Collecte



Stockage



Distributio



Un SaaS pour le cryptage des données issues d'un ETL

Chargement des données

The screenshot displays the Talend Open Studio for Data Integration (5.0.1.r74687) interface. The main workspace shows a data pipeline with several components: 'tTableIn', 'tJoin', 'tTableOut', 'tTableIn', 'tTableOut', 'tTableIn', 'tTableOut', 'tTableIn', 'tTableOut', 'tTableIn', and 'tTableOut'. The pipeline is connected by red dashed lines representing data flows.

The bottom panel shows the configuration for the 'tLagRow_2' component. The 'Paramètres simples' section includes the following settings:

- Scénario: Full T1
- Entrez le schéma: Synchroniser les colonnes
- Mode: Jaques
- Talend utilise les colonnes de destination (si définies): Verras (chaque ligne est une liste de données)
- Wagon de données: T1
- Afficher le nom de ce composant en face de chaque ligne de sortie
- Afficher le nom de ce composant en face de chaque ligne de sortie



Configuration du cryptage

The screenshot shows the Talend Open Studio interface for configuring a mapping job. The 'row1' component is selected, and the 'row2' component is being configured. The 'row2' component has a table of columns and expressions, including 'routines.RoutineCryptage...' for encryption.

Expression	Column
routines.RoutineCryptage ...	nomCompte
routines.RoutineCryptage ...	Nom
occ.main.row1.Prenom	Prenom
routines.RoutineCryptage ...	CodeBanque
routines.RoutineCryptage ...	solde



Exécution du Job Trace

The screenshot displays the Talend Open Studio for Data Integration (v5.0.1) interface. The main workspace shows a job execution trace with several data preview windows. The data preview windows show the following information:

row1 (Main)	row2 (Main)	row3 (Main)	row4 (Main)	row5 (Main)	row6 (Main)	row7 (Main)	row8 (Main)
NumLompte	0000000000	Debit	85000	credit	85000.0	NumCrypt	8091
Nom	Maria	Debit	85000.0	credit	85000.0	DebitCrypt	[B@1a80a69
Prenom	Maria	credit	85000.0	Debit	85000.0	Nom	[R@14384c7
Tel	22310909	Nom	Maria	Prenom	Maria	CreditCrypt	[U@1c0e197
CodeAgence	1	Tel	22310909	Tel	22310909	NomCrypt	[R@erh281
Solde	0000000000	CodeAgence	1	CodeAgence	1	PrenomCrypt	[U@1bb60c3
						TelCrypt	[B@u10be
						CodeAgenceCrypt	[B@1a11bb6

The interface also shows a flow diagram with components like 'row1 (Main)', 'row2 (Main)', 'row3 (Main)', 'row4 (Main)', 'row5 (Main)', 'row6 (Main)', 'row7 (Main)', 'row8 (Main)', 'row9 (Main)', 'row10 (Main)', 'row11 (Main)', 'row12 (Main)', 'row13 (Main)', 'row14 (Main)', 'row15 (Main)', 'row16 (Main)', 'row17 (Main)', 'row18 (Main)', 'row19 (Main)', 'row20 (Main)', 'row21 (Main)', 'row22 (Main)', 'row23 (Main)', 'row24 (Main)', 'row25 (Main)', 'row26 (Main)', 'row27 (Main)', 'row28 (Main)', 'row29 (Main)', 'row30 (Main)', 'row31 (Main)', 'row32 (Main)', 'row33 (Main)', 'row34 (Main)', 'row35 (Main)', 'row36 (Main)', 'row37 (Main)', 'row38 (Main)', 'row39 (Main)', 'row40 (Main)', 'row41 (Main)', 'row42 (Main)', 'row43 (Main)', 'row44 (Main)', 'row45 (Main)', 'row46 (Main)', 'row47 (Main)', 'row48 (Main)', 'row49 (Main)', 'row50 (Main)', 'row51 (Main)', 'row52 (Main)', 'row53 (Main)', 'row54 (Main)', 'row55 (Main)', 'row56 (Main)', 'row57 (Main)', 'row58 (Main)', 'row59 (Main)', 'row60 (Main)', 'row61 (Main)', 'row62 (Main)', 'row63 (Main)', 'row64 (Main)', 'row65 (Main)', 'row66 (Main)', 'row67 (Main)', 'row68 (Main)', 'row69 (Main)', 'row70 (Main)', 'row71 (Main)', 'row72 (Main)', 'row73 (Main)', 'row74 (Main)', 'row75 (Main)', 'row76 (Main)', 'row77 (Main)', 'row78 (Main)', 'row79 (Main)', 'row80 (Main)', 'row81 (Main)', 'row82 (Main)', 'row83 (Main)', 'row84 (Main)', 'row85 (Main)', 'row86 (Main)', 'row87 (Main)', 'row88 (Main)', 'row89 (Main)', 'row90 (Main)', 'row91 (Main)', 'row92 (Main)', 'row93 (Main)', 'row94 (Main)', 'row95 (Main)', 'row96 (Main)', 'row97 (Main)', 'row98 (Main)', 'row99 (Main)', 'row100 (Main)'. The status bar at the bottom shows 'démarrer', 'pcd@h111', 'Talend Open Studio...', 'C:\Program Files\...', 'bin\workspace\...', 'FR', and '26/35'.

Schéma de la Table

Édition des données - PostgreSQL 9.1 (localhost:5432) - basecrypte - operationcrypte

Fichier Édition Affichage Outils Aide

Pas de limite

	NumLrypt bytea	DebitLrypt bytea	LreditLrypt bytea	NomLrypt bytea	PrenomLrypt bytea	IelLrypt bytea	LodeAgenceLrypt bytea
1	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
2	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
3	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
4	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
5	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
6	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
7	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
8	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
9	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
10	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
11	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
12	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
13	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair
14	<données binaires>	<données binaires>	<données binair	<données bi	<données bi	<données bi	<données binair



Conclusion & Perspectives

Conclusion

- ✓ L'apparition du Cloud Computing a apporté plusieurs avantages en termes de puissance du calcul, de rapidité d'exécution et de réduction des couts.
- ✓ l'empêchement majeur des entreprises pour l'externalisation de leurs systèmes d'information vers le Cloud est **le manque de confiance accordée** au fournisseur de services
- ✓ L'objectif de cette solution est de **diminuer la vulnérabilité** de transmission et assurer **la confidentialité**.

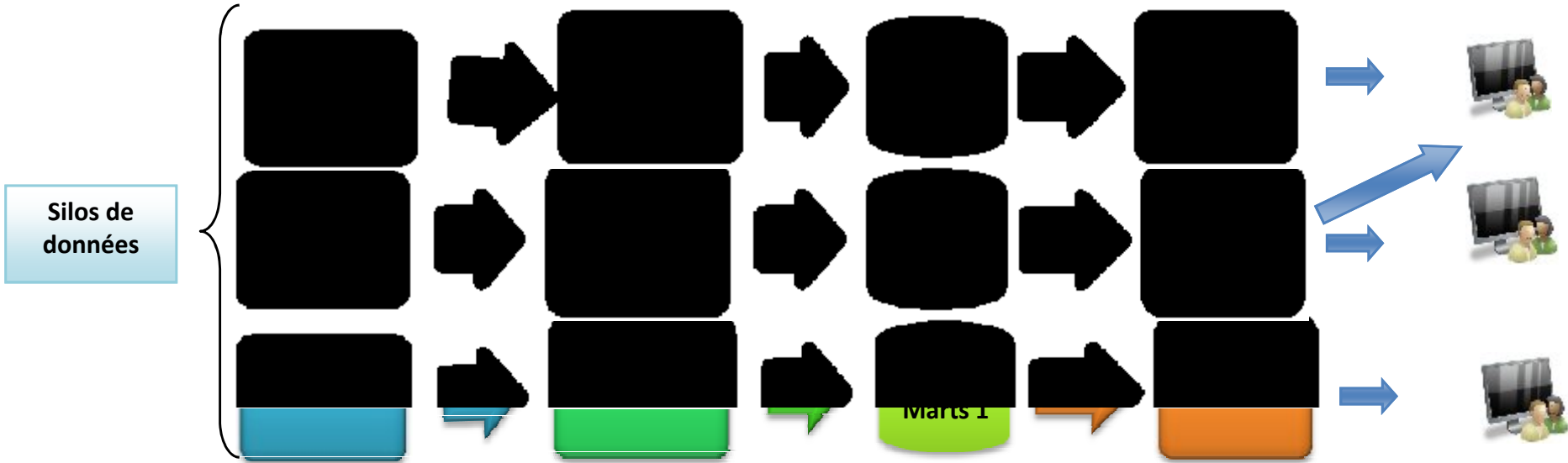
Perspectives

- ✓ Une étude de la possibilité d'intégrer la cryptographie pour la sécurité de **tous** le processus décisionnel y compris le processus ETL.
- ✓ Une étude d'un audit des systèmes décisionnels dans le cloud avec intégration d'une étude de cout en profit des entreprises clientes.
- ✓ Une étude du meilleur algorithme de cryptage des données en Cloud.

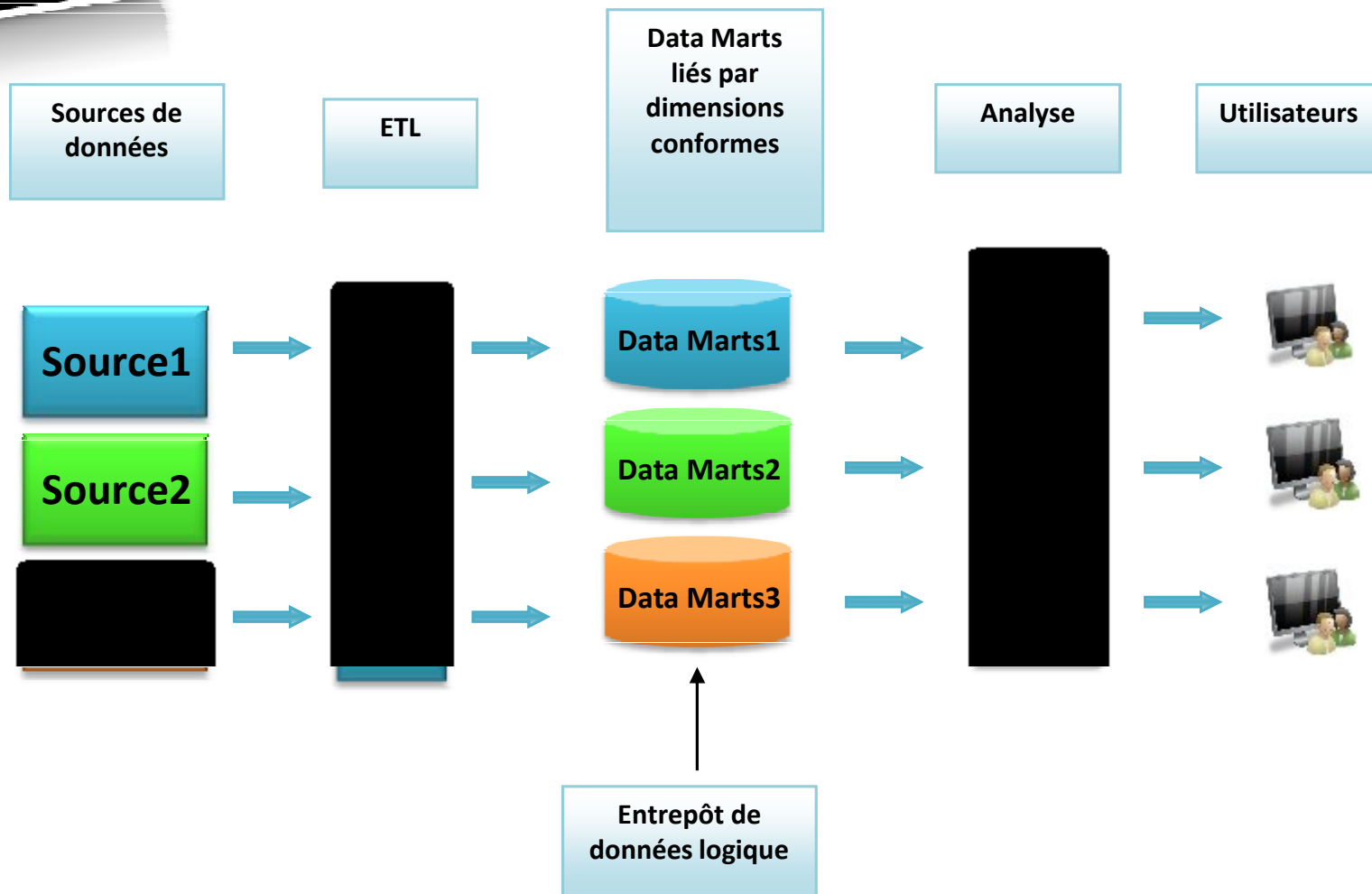
Merci pour votre attention

Architectures

1 Data Marts indépendant

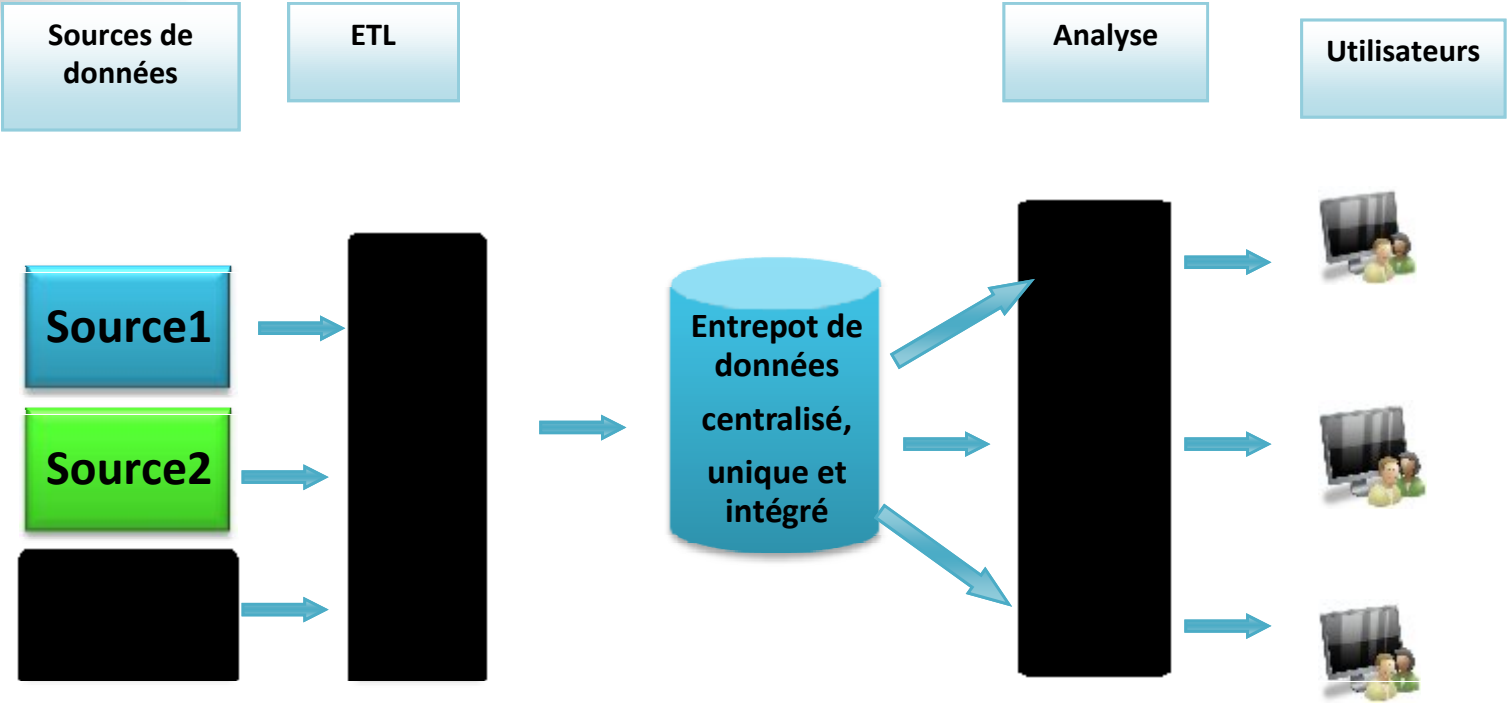


2 Architecture en bus de Data Marts



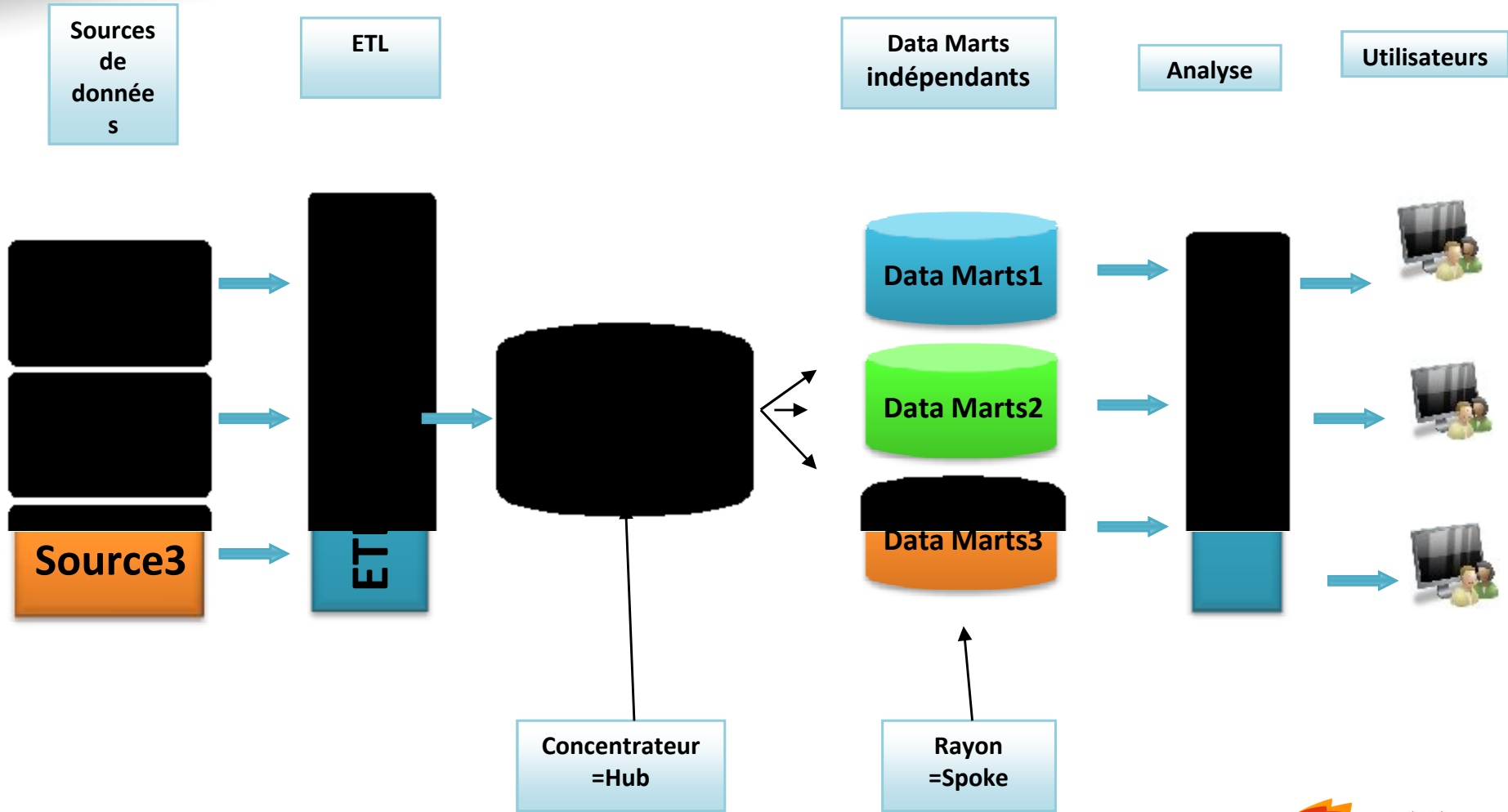
Architectures

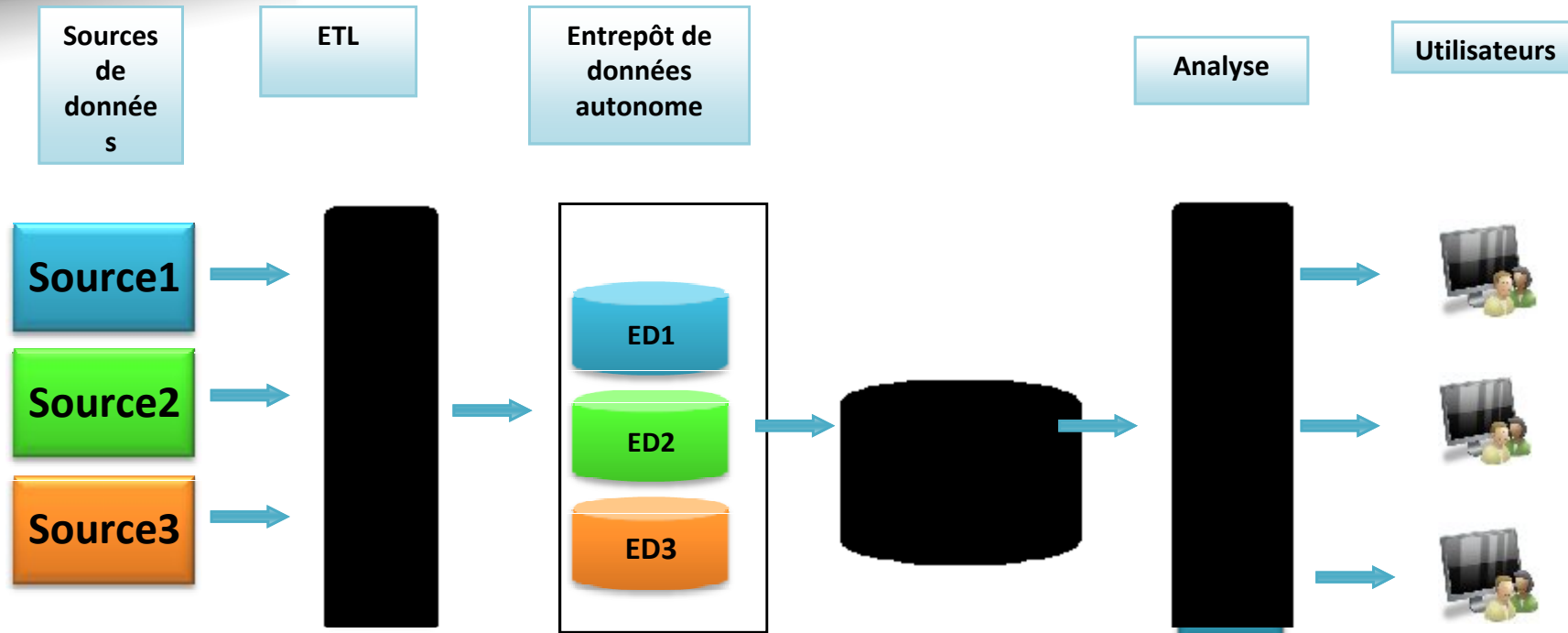
3 Entrepôt de données centralisé



Architectures

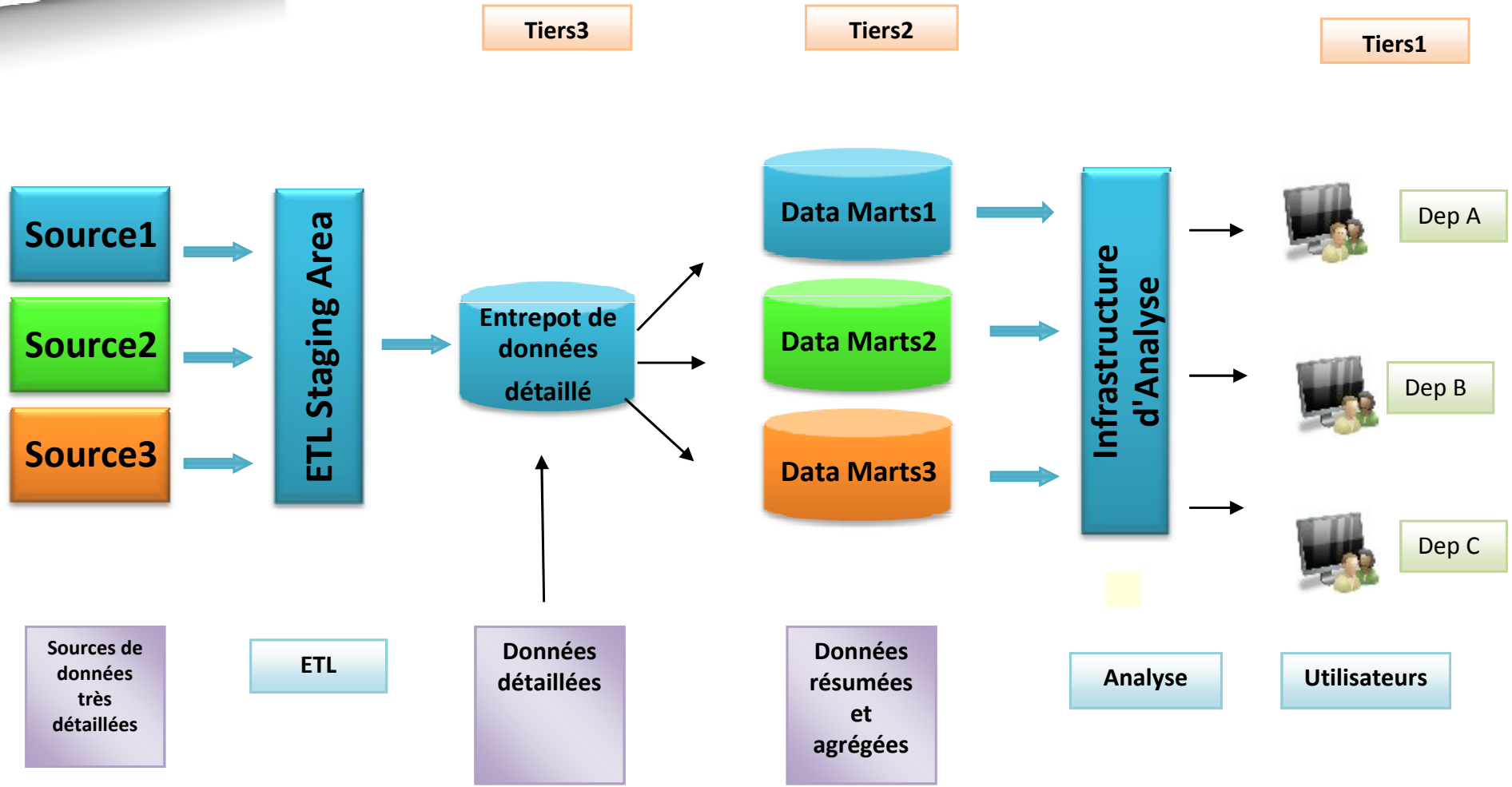
4 Architecture Hub-and-Spoke





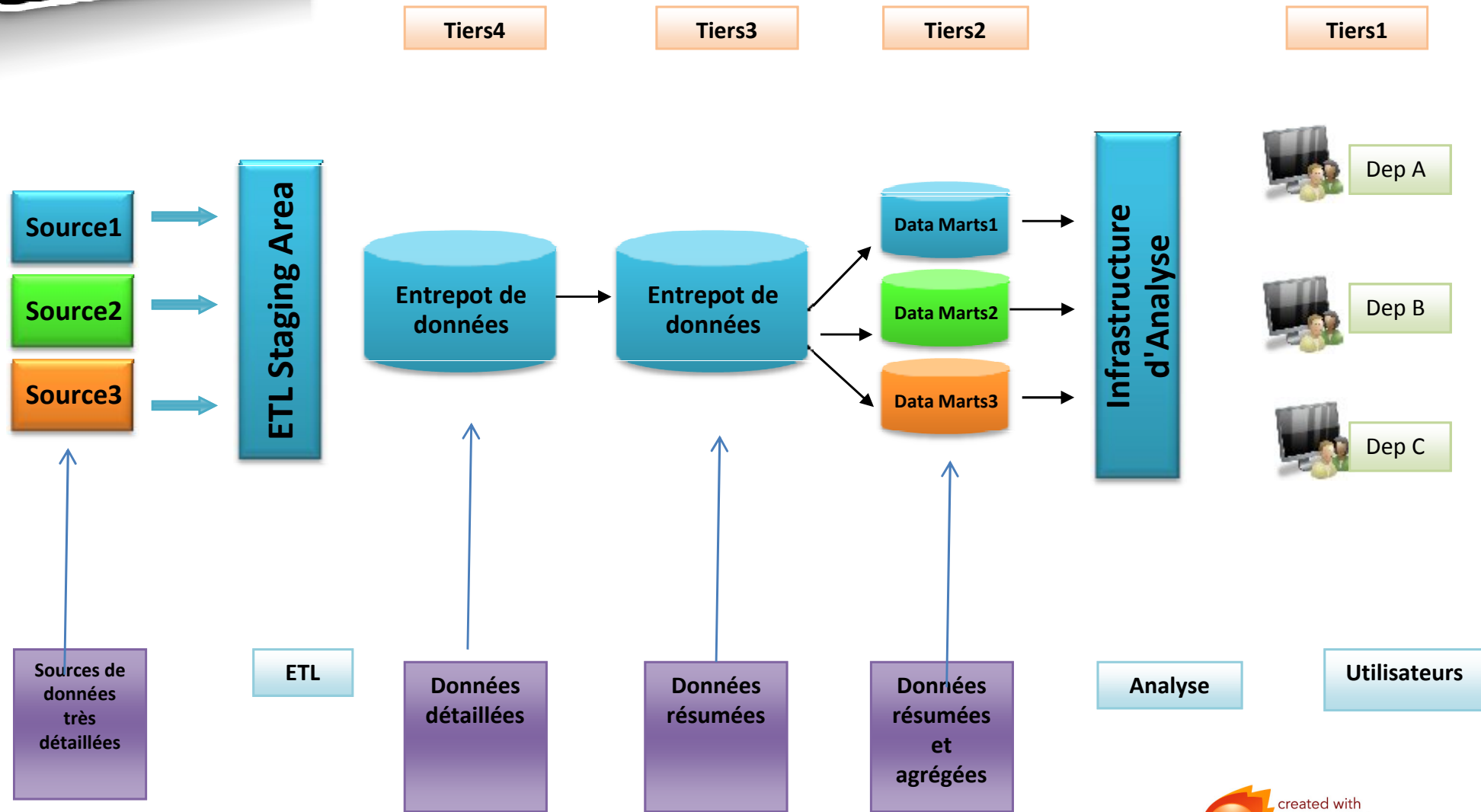
Architectures

6 Architecture trois tiers

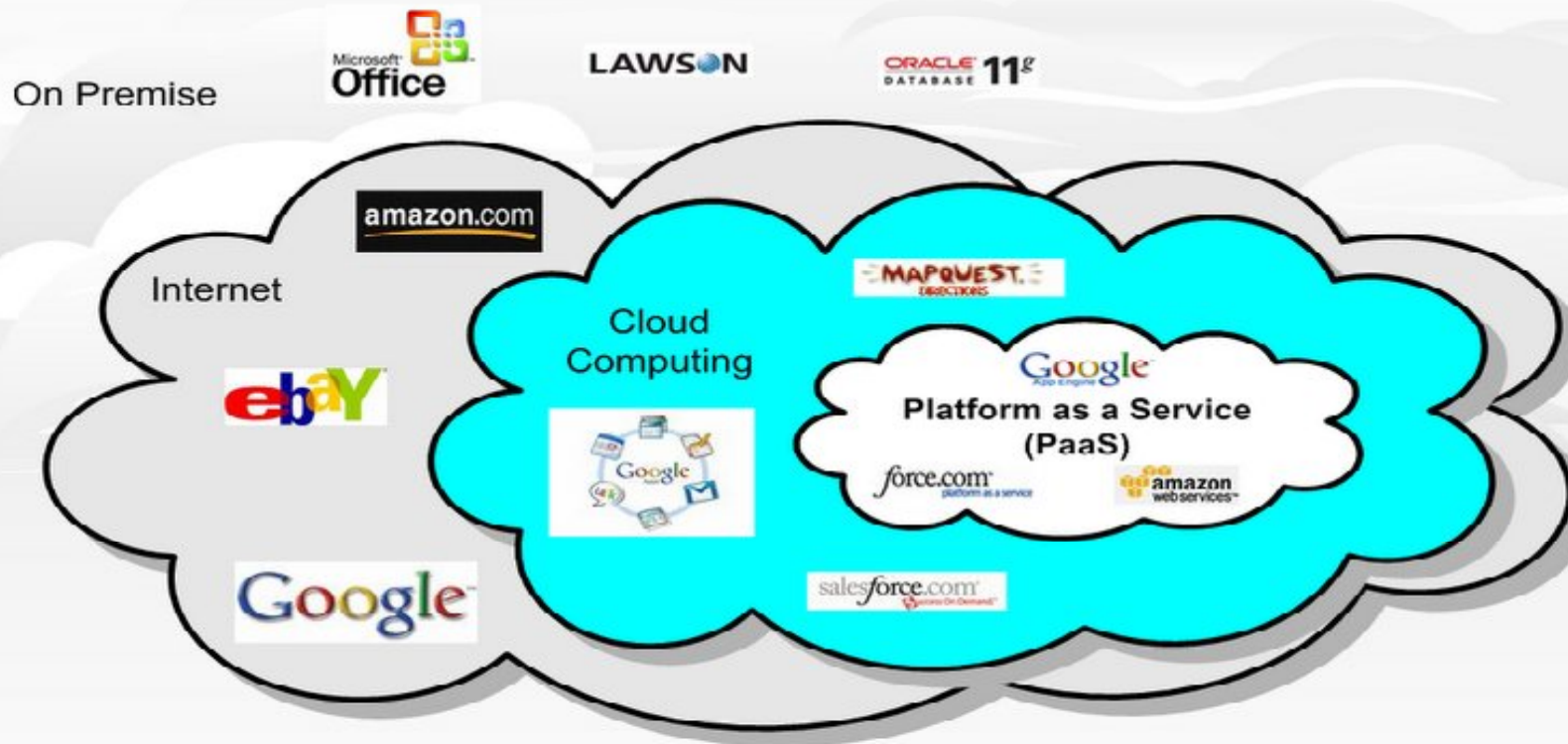


Architectures

7 Architecture multi niveaux

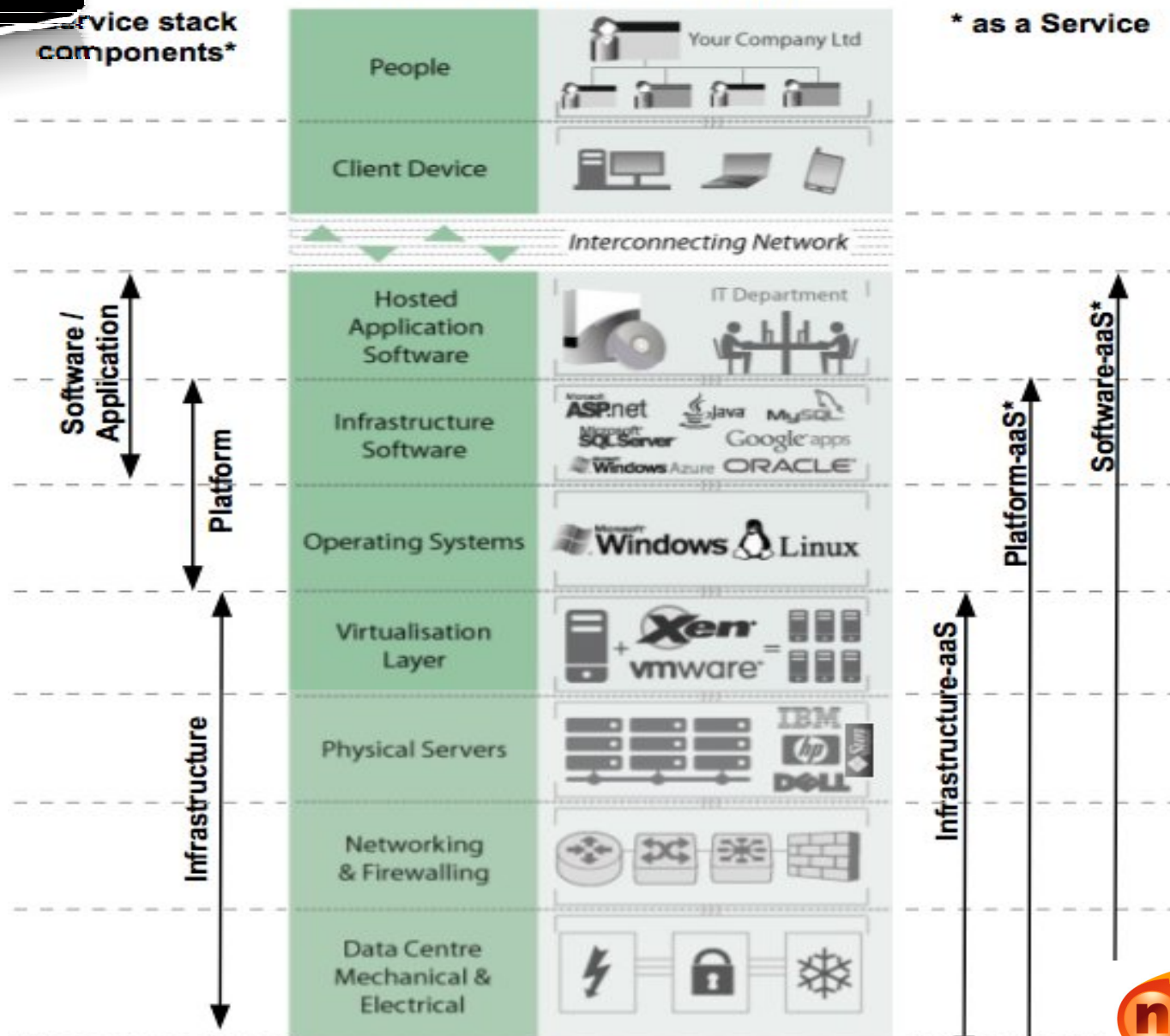


Acteurs du domaine



Domaines de services

Service Layers Definition



Notes:

Brand names for illustrative / example purposes only, and examples are not exhaustive.



SaaS Paas IaaS

