

# DESCRIPTION DES CAS PRATIQUES

Evaluation de la compréhension des UO

## MP 24-35

- **Acheteur:**

Agence de services et de paiement (ASP)  
2, rue du Maupas  
87040 LIMOGES cedex 1

## ***Table des matières***

<b><i>Cas Pratique : Ingestion simple.....</i></b>	<b><i>3</i></b>
<b><i>Cas Pratique : Intégration de Données avec Traitements dans le lac de données .....</i></b>	<b><i>5</i></b>
<b><i>Cas Pratique : Intégration d'un besoin Métier Complet dans la Plateforme LDA .....</i></b>	<b><i>7</i></b>

## Cas Pratique : Ingestion simple

### Description du Cas

L'ASP souhaite intégrer des données provenant de son système Oracle dans un lac de données (Data Lake) basé sur HDFS (Hadoop Distributed File System).

Ce cas pratique implique l'extraction de données à partir de trois tables Oracle, contenant respectivement 10, 15 et 20 champs de types basiques.

L'objectif est de centraliser ces données pour une analyse ultérieure, permettant ainsi une meilleure exploitation des informations au sein de l'organisation.

Ce cas pratique illustre une approche structurée pour intégrer efficacement des données d'un système Oracle vers un lac de données HDFS. En suivant cette méthodologie, l'ASP pourra non seulement centraliser ses informations mais aussi améliorer ses capacités analytiques grâce à une infrastructure moderne et sécurisée.

### Sources de données :

Système Source : Oracle

Tables : 3 tables avec des champs simples

Cible : HDFS (zone brute)

### Attendus

#### Traduction du Cas en Proposition

La proposition d'intégration du candidat doit inclure les éléments suivants :

#### Description Fonctionnelle et Technique :

Extraction des Données : Utilisation d'Apache Sqoop, Python, PySpark, Datalku ou autres outils conçus pour transférer efficacement des données entre une base de données relationnelle comme Oracle et HDFS.

#### Transformation et Chargement :

Les données extraites seront transformées pour répondre aux exigences analytiques. Cela peut inclure la conversion des types de données et le nettoyage des valeurs manquantes. Le chargement dans HDFS se fera via des formats tels que Text, SequenceFiles, Parquet, Avro... adaptés aux besoins d'analyse. Le déversement se fait directement en zone brute du lac.

#### Sécurisation des Données :

Mise en place de mécanismes pour assurer la sécurité des données durant le transfert et le stockage, notamment par l'utilisation de protocoles sécurisés et la gestion des accès.

### Charges, Coûts et Estimation de Planning

Le candidat complète les éléments ci-dessous au sein de l'onglet « SIM. CAS PRATIQUES », pris en compte dans l'analyse au titre du sous-critère prix SCP2 décrit au RC :

#### → Charges en UO P5

Évaluation des ressources nécessaires en termes d'UO.

Les aspects financiers de cette évaluation seront analysés au titre du sous-critère prix SCP2 décrit au RC.

La cohérence des UO mobilisées en termes de niveaux de complexité, de quantité et des profils associés à chaque UO sera analysée au titre du sous-critère technique « Pertinence de la réponse

aux cas pratiques » décrit au RC.

Le candidat complète les éléments ci-dessous au sein du chapitre « Sous critère 2.4 - Pertinence de la réponse au cas pratique » du cadre de réponse technique (CRT) :

→ **Planning :**

Élaboration d'un calendrier prévisionnel détaillant les différentes phases du projet, incluant l'analyse initiale, le développement, les tests et la mise en production.

→ **Description de la réalisation proposée en terme fonctionnel**

*Objectif Principal :*

- Centraliser les données provenant d'Oracle dans un format accessible et exploitable au sein du lac de données.

*Utilisateurs Cibles :*

- Les analystes métiers qui auront besoin d'accéder à ces données pour effectuer des analyses avancées.

→ **Description de la réalisation proposée en terme technique**

*Outils Recommandés :*

- Apache Sqoop pour l'extraction et le chargement des données vers HDFS.
- Apache Hive pour effectuer des transformations supplémentaires sur les données une fois qu'elles sont chargées dans HDFS.
- Hadoop Ecosystem (HDFS) pour le stockage sécurisé et scalable des données.

→ **Processus Technique Détaillé (exemple)**

- Connexion à la base de données Oracle via Sqoop pour extraire les tables ciblées.
- Transformation des données selon les besoins analytiques (normalisation des types de données, gestion des valeurs manquantes).
- Chargement dans HDFS via Sqoop, qui générera un job MapReduce pour garantir l'efficacité du transfert.

## Cas Pratique : Intégration de Données avec Traitements dans le lac de données

### Description du Cas

L'ASP souhaite intégrer des données provenant de son système Oracle dans un lac de données, avec l'ajout de règles de gestion simples et un stockage cible dans Hive.

Cette approche permet à l'ASP d'intégrer efficacement ses données Oracle dans un environnement Big Data, tout en appliquant des règles de gestion simples et en offrant une visualisation accessible. L'utilisation d'outils open-source comme Hadoop et Hive, combinée à Angular pour la visualisation, offre une solution robuste et économique pour répondre aux besoins de l'entreprise.

### Source

Oracle avec 3 tables (10, 15, 20 champs de types basiques)

### Traitements

2 transformations, 2 agrégats, 1 indicateur calculé

### Cible

HDFS (zone brute) + Table Hive (zone transformée)

### Visualisation

Page Web avec 2 graphiques basiques

### Attendus

#### Traduction du Cas en Proposition

#### Description Fonctionnelle et Technique

##### Extraction des Données

Utilisation d'Apache Sqoop (ou de tout autre outil équivalent) pour extraire efficacement les données des tables Oracle.

##### Transformation et Chargement

Emploi de Hive (ou de tout autre outil équivalent) pour effectuer les transformations et agrégations requises.

##### Stockage

Stockage des données brutes dans HDFS et des données traitées dans des tables Hive.

##### Visualisation

Utilisation d'Angular pour créer une page web interactive avec des graphiques basiques.

#### Charges, Coûts et Estimation de Planning

**Le candidat complète les éléments ci-dessous au sein de l'onglet « SIM. CAS PRATIQUES », pris en compte dans l'analyse au titre du sous-critère prix SCP2 décrit au RC :**

##### → Charges en UO P5

Évaluation des ressources nécessaires en termes d'UO.

Les aspects financiers de cette évaluation seront analysés au titre du sous-critère prix SCP2 décrit au RC.

La cohérence des UO mobilisées en termes de niveaux de complexité, de quantité et des profils associés à chaque UO sera analysée au titre du sous-critère technique « Pertinence de la réponse aux cas pratiques » décrit au RC.

Le candidat complète les éléments ci-dessous au sein du chapitre « Sous critère 2.4 - Pertinence de la réponse au cas pratique » du cadre de réponse technique (CRT) :

→ **Planning**

Élaboration d'un calendrier à estimer, couvrant l'analyse, le développement, les tests et la mise en production.

→ **Description de la réalisation proposée en terme fonctionnel**

**Objectif Principal**

Intégrer les données Oracle dans un lac de données, appliquer des règles de gestion simples, et fournir une visualisation basique.

**Utilisateurs Cibles**

Analystes métiers et décideurs nécessitant un accès rapide aux données traitées et visualisées.

→ **Description de la réalisation proposée en terme technique**

**Outils Recommandés**

- Apache Sqoop pour l'extraction des données Oracle.
- Apache Hive pour le traitement et le stockage des données.
- HDFS pour le stockage des données brutes.
- Site Web Angular pour la création de visualisations web interactives.

→ **Processus Technique Détaillé**

Extraction des données Oracle via Sqoop et stockage dans HDFS.

Création de tables externes Hive pointant vers les données HDFS.

Utilisation de HiveQL pour appliquer les transformations, agrégations et calculs d'indicateurs.

Stockage des résultats dans des tables Hive optimisées (format Parquet).

Développement d'une interface web utilisant Angular pour visualiser les données.

→ **Règles de Gestion**

Deux transformations (ex: normalisation des données, conversion de formats)

Deux agrégats (ex: sommes ou moyennes par catégorie)

Un indicateur calculé (ex: taux de croissance ou ratio)

→ **Visualisation**

Page web interactive avec deux graphiques basiques (histogramme + camembert).

Possibilité de filtrer et d'explorer les données via des contrôles interactifs.

## Cas Pratique : Intégration d'un besoin Métier Complet dans la Plateforme LDA

### Description du Cas

L'ASP souhaite intégrer un besoin métier complet provenant d'une de ses Directions dans sa plateforme LDA (Lac de Données Agricoles).

Ce projet d'intégration complet permettra à l'ASP de centraliser et d'exploiter efficacement les données d'un cas d'usage intégral de l'une de ses Directions au sein de sa plateforme LDA. L'utilisation d'outils puissants comme Spark et MicroStrategy, combinée à une architecture de stockage multi-niveaux (HDFS, Hive, PostgreSQL), offrira une solution robuste et évolutive, capable de gérer des données complexes, y compris géospatiales, tout en fournissant des insights avancés via des modèles de machine learning et des rapports analytiques détaillés.

### Source

Oracle avec 5 tables (500 champs, incluant des données géospatiales sensibles comme les coordonnées GPS des parcelles)

### Traitements

Mise en qualité des données, règles de gestion complexes, modélisation avancée

### Cible

HDFS (zone brute), Tables Hive (zone transformée), Tables PostgreSQL (zone modélisée)

### Visualisation

Rapport MicroStrategy (10 pages, 50 indicateurs)

### Attendus

#### Traduction du Cas en Proposition

#### Description Fonctionnelle et Technique :

##### Extraction des Données

Utilisation de Sqoop (ou de tout autre outil équivalent) pour l'extraction des données Oracle, y compris les données géospatiales.

##### Traitement et Transformation

Emploi d'Apache Spark (ou de tout autre outil équivalent) pour le traitement massif des données et l'application des règles de gestion.

##### Stockage

Utilisation de HDFS (ou de tout autre support équivalent) pour le stockage brut, Hive (ou de tout autre support équivalent) pour les données traitées, et PostgreSQL (ou de tout autre support équivalent) pour les données optimisées pour la visualisation.

##### Modélisation

Utilisation de Spark MLlib (ou de tout autre outil équivalent) pour les modèles de clustering et de segmentation.

##### Visualisation

MicroStrategy (imposé) pour la création de rapports analytiques avancés.

### Sécurisation des Données :

Mise en place de mécanismes pour assurer la sécurité des données durant le transfert et le stockage, notamment par l'utilisation de protocoles sécurisés et la gestion des accès.

Mise en place de la sécurité d'accès à l'information exposée sur le portail de visualisation.

### Charges, Coûts et Estimation de Planning :

---

Le candidat complète les éléments ci-dessous au sein de l'onglet « SIM. CAS PRATIQUES » », pris en compte dans l'analyse au titre du sous-critère prix SCP2 décrit au RC :

#### → Charges en UO P5

Évaluation des ressources nécessaires en termes d'UO.

Les aspects financiers de cette évaluation seront analysés au titre du sous-critère prix SCP2 décrit au RC.

La cohérence des UO mobilisées en termes de niveaux de complexité, de quantité et des profils associés à chaque UO sera analysée au titre du sous-critère technique « Pertinence de la réponse aux cas pratiques » décrit au RC.

Le candidat complète les éléments ci-dessous au sein du chapitre « Sous critère 2.4 - Pertinence de la réponse au cas pratique » du cadre de réponse technique (CRT) :

#### → Planning

Projet à estimer, incluant les phases d'analyse, développement, tests, et mise en production.

#### → Description de la réalisation proposée en terme fonctionnel

##### Objectif Principal

Intégrer le SI métier dans la plateforme LDA, en assurant la qualité des données, l'application de règles métier complexes, et la production d'insights via des modèles analytiques avancés.

##### Utilisateurs Cibles

Analystes métier, décideurs, et data scientists de la Direction concernée.

#### → Description de la réalisation proposée en terme technique

##### Outils Recommandés

- Sqoop (ou tout autre outil équivalent) pour l'extraction des données Oracle.
- Apache Spark (ou tout autre outil équivalent) pour le traitement massif et les transformations.
- HDFS (ou tout autre support équivalent), Hive (ou tout autre support équivalent), et PostgreSQL (ou tout autre support équivalent) pour le stockage dans les différentes zones.
- Spark MLlib (ou tout autre outil équivalent) pour les modèles de machine learning.
- MicroStrategy (imposé) pour la visualisation et les rapports.

#### → Processus Technique Détaillé (exemple)

- Extraction des données Oracle, y compris les données géospatiales, via Sqoop.
- Stockage des données brutes dans HDFS.
- Traitement de mise en qualité et application des règles de gestion via Spark.
- Stockage des données traitées dans Hive.
- Application des modèles ML (clustering et segmentation) avec Spark MLlib.
- Chargement des données finales et des résultats des modèles dans PostgreSQL.
- Création des rapports MicroStrategy basés sur les données PostgreSQL.
- Mise en place d'un groupe de sécurité spécifique pour accéder aux données relatives au cas d'usage dans l'outil de restitution.



- Traitement des Données Géospatiales
  - Utilisation des fonctionnalités spatiales d'Oracle pour l'extraction initiale.
  - Transformation et analyse des données géospatiales via des bibliothèques Spark dédiées.
  - Stockage optimisé dans PostgreSQL avec l'extension PostGIS pour les requêtes spatiales avancées.
- Modélisation et Analyse Avancée
  - Mise en place de 50 transformations et 50 agrégats pour préparer les données.
  - Création de 50 indicateurs calculés de type BI pour l'analyse métier.
  - Développement de modèles de clustering et segmentation avec Spark MLlib pour des insights avancés.
- Rapport MicroStrategy
  - Conception de 10 pages de rapport intégrant 50 indicateurs clés.
  - Inclusion de visualisations interactives et de tableaux de bord dynamiques.
  - Intégration de filtres et de drill-down pour une exploration approfondie des données.