

Formation management de projets Big data

Corrigé – TP Big Data

Exercices pratiques

Eddy K. DIABAULT
23/02/2025

Corrigé – TP Big Data

Table des matières

| | |
|--|-----------|
| Exercice 1 : Définition et cadrage du projet Big Data | 2 |
| Exercice 2 : Analyse de cas – Choix d’un projet | 2 |
| Exercice 3 : Rédaction du cahier des charges | 3 |
| Exercice 4 : Conception de l’architecture technique | 4 |
| Exercice 5 : Planification et gestion des ressources | 4 |
| Exercice 6 : Identification et gestion des risques | 5 |
| Exercice 7 : Suivi et contrôle de projet | 6 |
| Exercice 8 : Qualité des données et conformité | 6 |
| Exercice 9 : Communication et gestion du changement | 7 |
| Exercice 10 : Choix de la méthodologie de gestion de projet | 8 |
| Exercice 11 : Simulation de gestion de projet | 8 |
| Exercice 12 : Clôture de projet et rétrospective | 9 |
| Conclusion | 10 |

Voici ci-dessous un corrigé complet, détaillé et commenté des 12 exercices basés sur le projet SmartLogistics. Ce corrigé est conçu pour guider pas à pas des apprenants débutants – y compris ceux en difficulté d'apprentissage – afin qu'ils puissent comprendre et reproduire les notions abordées. Chaque section est expliquée en langage simple et des exemples concrets sont fournis.

Exercice 1 : Définition et cadrage du projet Big Data

Objectif : Comprendre ce qu'est un projet Big Data et identifier ses spécificités.

Corrigé pas à pas :

1. Définir un projet Big Data :

- **Explication :** Un projet Big Data vise à collecter, stocker et analyser d'importantes quantités de données.
- **Notions clés :**
 - **Volume :** La grande quantité de données (ex. : des milliers de transactions, données IoT, feedback clients).
 - **Vélocité :** La rapidité avec laquelle les données arrivent et doivent être traitées (ex. : données en temps réel issues des capteurs des véhicules).
 - **Variété :** La diversité des sources de données (transactions, IoT, retours clients, maintenance).

2. Cadrer le projet SmartLogistics :

- **Contexte :** LogiTech souhaite optimiser sa logistique.
- **Données impliquées :**
 - Transactions, catalogue produits, données IoT, historiques de maintenance, feedback clients.
- **Différence avec un projet informatique classique :**
 - Dans un projet classique, on traite souvent moins de données et avec moins de diversité, alors qu'ici il faut gérer des flux en temps réel et des données hétérogènes.

Commentaire :

Prenez le temps de bien identifier les trois “V” (Volume, Vélocité, Variété). Vous pouvez écrire vos idées sous forme de liste ou schéma pour visualiser le cadrage du projet.

Exercice 2 : Analyse de cas – Choix d'un projet

Objectif : Sélectionner et justifier un projet Big Data pertinent.

Corrigé pas à pas :

Corrigé – TP Big Data

1. Présenter le cas SmartLogistics :

- **Problématique :** La chaîne logistique de LogiTech est complexe et nécessite une optimisation pour réduire les coûts et améliorer la satisfaction client.
- **Objectifs du projet :**
 - Optimiser la gestion des livraisons (planification, suivi en temps réel).
 - Améliorer la maintenance de la flotte (anticiper les pannes).
 - Augmenter la satisfaction client (réduire les retards et analyser les feedbacks).

2. Justifier le choix du projet Big Data :

- **Données massives et variées :** Nécessité d'exploiter des données provenant de différentes sources (transactions, capteurs, feedback).
- **Besoins en temps réel :** Pour réagir rapidement aux imprévus (ex. : panne d'un véhicule détectée via IoT).
- **Prise de décision améliorée :** L'analyse des données permet de prendre des décisions éclairées pour optimiser la logistique.

Commentaire :

Expliquez clairement pourquoi l'utilisation d'un système Big Data est indispensable dans ce cas. Insistez sur le lien entre les besoins réels de l'entreprise et les capacités offertes par le Big Data.

Exercice 3 : Rédaction du cahier des charges

Objectif : Formaliser les besoins fonctionnels et techniques du projet.

Corrigé pas à pas :

1. Introduction et description du projet :

- **Exemple :**
« Le projet SmartLogistics vise à moderniser la chaîne logistique de LogiTech en exploitant les données issues des transactions, des capteurs IoT des véhicules et des retours clients pour optimiser la planification des tournées, la maintenance et la satisfaction client. »

2. Exigences fonctionnelles :

- Suivi en temps réel des livraisons.
- Affichage d'un tableau de bord des indicateurs clés (temps de traitement, taux de satisfaction, etc.).
- Système d'alerte en cas d'anomalies (pannes, retards).

3. Exigences techniques :

- Capacité à gérer un grand volume de données en temps réel.
- Architecture scalable et résiliente (ex. : utilisation de Hadoop et Spark).
- Sécurité des données (chiffrement, anonymisation pour respecter le RGPD).

4. Contraintes spécifiques :

- Intégration avec les systèmes existants de LogiTech (ERP, CRM).
- Respect des normes de protection des données.

Corrigé – TP Big Data

Commentaire :

Utilisez des titres et des sous-titres pour organiser votre cahier des charges. Chaque section doit être claire et précise, afin que même un débutant puisse comprendre les attentes du projet.

Exercice 4 : Conception de l'architecture technique

Objectif : Imaginer et justifier une architecture technique adaptée aux besoins du projet.

Corrigé pas à pas :

1. Identifier les composants principaux :

- **Collecte des données :** Outils d'ingestion (ex. : Kafka pour les données IoT).
- **Stockage :** Utilisation de systèmes distribués comme Hadoop (HDFS) pour le stockage massif.
- **Traitement des données :** Moteur de traitement en temps réel (Apache Spark).
- **Bases de données NoSQL :** Pour stocker des données hétérogènes (ex. : MongoDB, Cassandra).
- **Visualisation :** Tableau de bord avec des outils comme Tableau ou Kibana.

2. Schématiser l'architecture :

- Dessinez un diagramme simple montrant la chaîne de traitement :
 - **Source de données** (IoT, transactions, feedback) → **Ingestion (Kafka)** → **Stockage (Hadoop HDFS)** → **Traitement (Spark)** → **Base de données NoSQL** → **Visualisation (Dashboard)**

3. Justifier le choix de chaque composant :

- **Kafka :** Pour gérer le flux continu de données IoT.
- **Hadoop :** Pour stocker et gérer de très gros volumes de données de manière distribuée.
- **Spark :** Pour traiter rapidement des données en temps réel, ce qui est crucial pour le suivi des livraisons.
- **NoSQL :** Pour la flexibilité et la rapidité d'accès aux données non structurées.

Commentaire :

Un schéma visuel aide à comprendre l'architecture. Même un dessin à la main numérisé peut être utilisé pour illustrer la chaîne de traitement. N'hésitez pas à ajouter des flèches et des commentaires sur le schéma.

Exercice 5 : Planification et gestion des ressources

Objectif : Décomposer le projet en phases et planifier les ressources nécessaires.

Corrigé pas à pas :

1. Établir les phases du projet :

- **Phase 1 :** Analyse des besoins et définition du cahier des charges.
 - **Phase 2 :** Conception de l'architecture et choix technologiques.
 - **Phase 3 :** Développement et intégration des différents modules (ingestion, traitement, visualisation).
 - **Phase 4 :** Tests et validation (fonctionnels et de performance).
 - **Phase 5 :** Déploiement et formation des utilisateurs.
 - **Phase 6 :** Suivi, maintenance et amélioration continue.
2. **Planification avec un diagramme de Gantt :**
- Par exemple, répartissez les phases sur une période de 6 mois en indiquant les dates de début et de fin de chaque phase.
 - Identifiez les jalons (ex. : validation du cahier des charges, mise en place de l'architecture, premier test fonctionnel).
3. **Gestion des ressources :**
- **Humaines :** Chef de projet, data ingénieur, développeurs, analyste de données, expert sécurité.
 - **Matérielles :** Serveurs, licences logicielles, outils de monitoring et de visualisation.

Commentaire :

La clarté de la planification est primordiale. Même un planning simplifié (tableau ou diagramme) aide à visualiser le déroulement du projet et à comprendre l'enchaînement des tâches.

Exercice 6 : Identification et gestion des risques

Objectif : Repérer les risques et prévoir des mesures d'atténuation.

Corrigé pas à pas :

1. **Identifier les risques potentiels :**
- **Risque technique :** Panne des systèmes de collecte ou de traitement (ex. : défaillance d'un serveur Hadoop).
 - **Risque de sécurité :** Violation de la confidentialité des données clients (non-conformité RGPD).
 - **Risque organisationnel :** Retards dans la livraison des phases du projet.
 - **Risque lié aux données :** Données de mauvaise qualité ou incomplètes.
2. **Construire une matrice de risques :**
- Pour chaque risque, indiquer :
 - **Description du risque**
 - **Impact (faible, moyen, élevé)**
 - **Probabilité d'occurrence**
 - **Plan d'atténuation**
 - **Exemple :**
 - *Risque : Panne serveur – Impact : élevé, Probabilité : moyenne.*
 - *Mesure : Mise en place d'une redondance et d'un système de sauvegarde automatique.*
3. **Plan d'action :**

Corrigé – TP Big Data

- Précisez les actions à mener en cas de survenance du risque (ex. : procédure d'alerte, plan de reprise après sinistre).

Commentaire :

Utilisez un tableau pour organiser vos risques. Chaque colonne doit être clairement étiquetée afin que même quelqu'un qui débute puisse suivre la logique de gestion des risques.

Exercice 7 : Suivi et contrôle de projet

Objectif : Mettre en place des indicateurs et un tableau de bord pour piloter le projet.

Corrigé pas à pas :

1. Définir des indicateurs clés (KPIs) :

- **Temps de traitement des données :** Par exemple, temps moyen par job.
- **Taux de disponibilité du système :** Pourcentage de fonctionnement sans panne.
- **Taux de satisfaction client :** Basé sur les retours des feedbacks.
- **Respect des délais :** Mesure du retard ou de l'avancement par rapport au planning.

2. Concevoir un tableau de bord :

- Présentez les indicateurs sous forme de graphiques simples (barres, courbes, jauges).
- Exemple d'éléments à inclure :
 - Un graphique montrant le temps de traitement par jour.
 - Un indicateur de disponibilité en temps réel.
 - Un graphique de satisfaction client basé sur les notes.

3. Suivi régulier :

- Proposez des réunions hebdomadaires ou mensuelles pour analyser les données du tableau de bord et ajuster la gestion du projet.

Commentaire :

L'objectif est de permettre à tout le monde, y compris aux débutants, de comprendre rapidement l'état d'avancement du projet grâce à des visuels simples et des indicateurs clairs.

Exercice 8 : Qualité des données et conformité

Objectif : Mettre en place un plan pour assurer la qualité des données et leur conformité.

Corrigé pas à pas :

1. Définir un plan qualité des données :

Corrigé – TP Big Data

- **Procédures de vérification** : Mise en place de scripts ou d'outils automatisés pour détecter les erreurs (ex. : données manquantes, incohérences dans les formats).
 - **Nettoyage des données** : Procédure de suppression ou correction des données erronées.
 - **Validation des données** : Contrôles réguliers pour s'assurer que les données respectent les normes établies.
2. **Assurer la conformité** :
- **Exigences RGPD** : Anonymiser les données personnelles (ex. : masquer les adresses emails ou les identifiants clients).
 - **Sécurité** : Stocker les données dans des environnements sécurisés et mettre en place des protocoles d'accès stricts.
3. **Documentation et audits** :
- Rédiger une documentation détaillée sur les processus de qualité et prévoir des audits réguliers pour vérifier la conformité.

Commentaire :

La clarté dans la procédure de vérification aide à instaurer la confiance dans les analyses. Chaque étape doit être expliquée de façon simple pour permettre à tous de comprendre l'importance de la qualité et de la sécurité des données.

Exercice 9 : Communication et gestion du changement

Objectif : Élaborer un plan de communication et de conduite du changement.

Corrigé pas à pas :

1. **Plan de communication** :
- **Publics cibles** : Équipe projet, direction, utilisateurs finaux.
 - **Messages clés** : Objectifs du projet, bénéfices attendus, étapes importantes.
 - **Canaux de communication** : Réunions régulières, emails, intranet, newsletters internes.
 - **Exemple** :
 - Réunion hebdomadaire pour faire le point sur l'avancement.
 - Newsletter mensuelle pour informer de la mise à jour des fonctionnalités.
2. **Plan de gestion du changement** :
- **Formation** : Organiser des sessions de formation pour les utilisateurs du nouveau système.
 - **Documentation** : Créer des guides pratiques et des tutoriels.
 - **Accompagnement** : Prévoir un support pour répondre aux questions et résoudre les problèmes rencontrés lors de la transition.

Commentaire :

Un plan de communication bien structuré facilite l'adhésion des équipes et réduit les résistances au changement. Expliquez chaque étape en précisant les moyens mis en œuvre pour soutenir les utilisateurs.

Exercice 10 : Choix de la méthodologie de gestion de projet

Objectif : Comparer et choisir entre une approche Agile et Waterfall adaptée au projet SmartLogistics.

Corrigé pas à pas :

1. Méthodologie Waterfall :

- **Caractéristiques :** Processus linéaire et séquentiel, chaque phase doit être terminée avant de passer à la suivante.
- **Avantages :** Planification détaillée en amont, bonne visibilité sur les étapes.
- **Limites :** Peu flexible, difficile à adapter en cas de changement.

2. Méthodologie Agile :

- **Caractéristiques :** Approche itérative avec des cycles de développement courts (sprints).
- **Avantages :** Grande flexibilité, adaptation rapide aux changements et retours utilisateurs.
- **Limites :** Moins de planification initiale, peut être difficile pour ceux qui préfèrent un cadre strict.

3. Choix pour SmartLogistics :

- **Recommandation :** Une approche Agile (ou hybride) est conseillée car le projet implique :
 - Des données en temps réel nécessitant des ajustements fréquents.
 - Des retours d'expérience clients et des imprévus (pannes, changements dans la chaîne logistique) qui demandent une réactivité.
- **Justification :** La flexibilité d'Agile permet d'intégrer progressivement les évolutions du projet et d'ajuster le planning en fonction des résultats obtenus.

Commentaire :

Présentez vos arguments sous forme de tableau comparatif si nécessaire. L'objectif est d'aider les apprenants à comprendre que le choix méthodologique doit être adapté aux spécificités du projet.

Exercice 11 : Simulation de gestion de projet

Objectif : Mettre en pratique la gestion de projet en situation réelle avec des imprévus.

Corrigé pas à pas :

1. Déclenchement d'un scénario imprévu :

- **Exemple :** Une panne détectée sur un véhicule de livraison via les capteurs IoT (alerte de température anormale ou baisse soudaine de niveau de carburant).

2. Réaction immédiate :

- **Analyse :** Vérifier les données en temps réel pour confirmer la panne.

- **Communication** : Alerter l'équipe de maintenance et informer le responsable logistique.
- **Réaffectation des ressources** : Rediriger une autre livraison ou planifier une intervention d'urgence.
- 3. **Mise à jour du planning** :
 - Adapter le diagramme de Gantt en intégrant un délai supplémentaire pour la réparation et la réorganisation des tournées.
- 4. **Rédaction d'un rapport de synthèse** :
 - **Contenu du rapport** :
 - Description de l'imprévu.
 - Actions mises en place pour y remédier.
 - Impact sur le planning initial.
 - Leçons apprises et pistes d'amélioration (ex. : renforcer le monitoring ou prévoir une réserve de véhicules).

Commentaire :

L'exercice de simulation doit encourager la réactivité et la flexibilité. L'important est d'expliquer clairement chaque action prise et d'analyser les impacts pour améliorer la gestion future.

Exercice 12 : Clôture de projet et rétrospective

Objectif : Clôturer le projet en formalisant les résultats et en tirant des enseignements.

Corrigé pas à pas :

1. **Rédiger un rapport de clôture** :
 - **Résumé du projet** : Rappel des objectifs initiaux et des résultats obtenus.
 - **Analyse des écarts** : Comparer le planning prévu aux résultats réels et identifier les causes des écarts (retards, imprévus techniques, etc.).
2. **Rétrospective** :
 - **Points positifs** : Identifier ce qui a bien fonctionné (ex. : flexibilité de l'approche Agile, qualité des outils de monitoring).
 - **Axes d'amélioration** : Noter les difficultés rencontrées et proposer des recommandations (ex. : meilleure planification des imprévus, renforcement de la communication).
3. **Leçons apprises** :
 - Rédiger une section « recommandations » pour guider les futurs projets en s'appuyant sur l'expérience acquise.
 - Par exemple : « Il est essentiel d'intégrer des marges de manœuvre dans le planning pour gérer les imprévus techniques. »

Commentaire :

Le rapport de clôture doit être clair et structuré. Il s'agit d'un outil d'apprentissage qui permettra à l'équipe de capitaliser sur l'expérience acquise et d'améliorer les projets futurs.

Conclusion

Ce corrigé détaillé présente, pour chaque exercice, les étapes clés et les justifications nécessaires pour comprendre le management de projets Big Data via le projet SmartLogistics. Chaque section a été pensée pour être accessible aux débutants et aux personnes en difficulté, avec des explications claires, des exemples concrets et des commentaires pour guider la réflexion.

N'hésitez pas à utiliser ce corrigé comme base d'étude et à le compléter avec des travaux pratiques ou des discussions en groupe pour renforcer votre compréhension.