

# Proposition d'Offre - Mise en Place Dataplateforme Kubernetes On-Premise

**Client** : CBAO

**Projet** : Déploiement et accompagnement Dataplateforme DaaS

**Date** : 07/12/2025

**Consultant** : Bassirou KA

**Sociétés** : Tecsen France / Back Consulting Sénégal

**Email** : [ka.bassirou@gmail.com](mailto:ka.bassirou@gmail.com)

**Téléphone** : +33 6 34 24 79 25

---

## Sommaire

1. [Contexte et objectifs](#)
  2. [Architecture de référence](#)
  3. [Détail des prestations](#)
  4. [Planification](#)
  5. [Chiffrage](#)
  6. [Modalités d'intervention](#)
  7. [Livrables](#)
  8. [Profil du consultant](#)
  9. [Conditions générales](#)
-

# 1. Contexte et objectifs

## Contexte

La CBAO souhaite mettre en place une dataplateforme moderne et scalable sur infrastructure Kubernetes on-premise pour supporter l'ensemble du cycle de vie des données : ingestion, transformation, stockage, analyse et gouvernance.

## Objectifs

- **Scalabilité** : Architecture élastique basée sur Kubernetes permettant l'adaptation aux besoins croissants
  - **Fiabilité** : Haute disponibilité et résilience pour garantir la continuité de service
  - **Sécurité** : Authentification centralisée (Keycloak) et contrôle d'accès granulaire avec Open Policy Agent (OPA)
  - **Observabilité** : Monitoring complet (Prometheus, Grafana) et traçabilité des données
  - **Gouvernance** : Catalogue de données (OpenMetadata) et lineage pour la traçabilité complète
  - **Datalab** : Environnement collaboratif (Zeppelin) pour l'analyse exploratoire et le développement
  - **Autonomie** : Encadrement de l'équipe Data Engineer pour assurer la pérennité de la plateforme
-

## 2. Architecture de référence

### Vue d'ensemble

La plateforme est basée sur une architecture cloud-native avec séparation des préoccupations et scalabilité indépendante de chaque composant.

### Composants principaux

#### Orchestration et intégration :

- **Apache Airflow** : Orchestration et planification des workflows de données
- **Apache Nifi** : Intégration de données en temps réel et batch

#### Transformation et traitement :

- **DBT** : Transformation SQL avec approche "Data as Code"
- **Apache Spark** : Traitement distribué de grandes volumétries
- **Trino** : Moteur SQL universel pour accès fédéré aux données

#### Stockage et formats :

- **MINIO** : Stockage objet distribué compatible S3
- **Apache Iceberg** : Format de table open-source avec support ACID, time travel et schema evolution
- **PostgreSQL** : Métadonnées et données relationnelles

#### Datalab et gouvernance :

- **Apache Zeppelin** : Notebooks collaboratifs pour analyse exploratoire
- **OpenMetadata** : Catalogue de données avec lineage automatique

#### Sécurité :

- **Keycloak** : Authentification centralisée (SSO) avec OAuth2/OIDC
- **Open Policy Agent (OPA)** : Gestion des politiques d'autorisation

#### Observabilité :

- **Prometheus** : Collecte et stockage des métriques
- **Grafana** : Visualisation et alerting
- **Loki + Promtail** : Agrégation et analyse des logs

#### Backup :

- **Velero** : Backup et restauration Kubernetes

### Organisation Kubernetes

La plateforme est organisée en namespaces dédiés pour une meilleure isolation et gestion :

- **data-platform-core** : Composants centraux (Airflow, Keycloak, OPA)
  - Services critiques nécessitant haute disponibilité
  - Gestion centralisée de l'authentification et des politiques
- **data-platform-storage** : Stockage (MINIO, PostgreSQL, OpenEBS)
  - Données persistantes et métadonnées
  - Volumes persistants gérés par OpenEBS

- **data-platform-integration** : Intégration (Apache Nifi)
    - Pipelines d'ingestion de données
    - Coordination via ZooKeeper pour cluster HA
  - **data-platform-processing** : Traitement (Spark, DBT, Trino)
    - Moteurs de traitement et transformation
    - Scalabilité indépendante selon la charge
  - **data-platform-datalab** : Datalab (Zeppelin)
    - Environnement collaboratif pour analystes et data scientists
    - Notebooks partagés et versionnés
  - **data-platform-observability** : Monitoring (Prometheus, Grafana, Loki)
    - Collecte de métriques et logs
    - Dashboards et alerting
  - **data-platform-governance** : Gouvernance (OpenMetadata)
    - Découverte automatique des métadonnées
    - Lineage et documentation des données
-

### 3. Détail des prestations

#### Phase 1 : Revue de l'architecture (2 JH)

**Objectif :** Valider et affiner l'architecture proposée

**Activités :**

- Analyse de l'existant et des besoins métier
- Validation des choix technologiques
- Recommandations architecturales
- Plan de déploiement détaillé
- Matrice de risques et mitigation

**Livrables :**

- Document de revue d'architecture
- Diagrammes d'architecture mis à jour si besoin
- Plan de déploiement détaillé
- Matrice de risques

#### Phase 2 : Déploiement des composants (15 JH)

**Objectif :** Déployer et configurer tous les composants de la plateforme sur Kubernetes

**Modalité :** 77% à distance, 23% présentiel

**Détail des sous-phases :**

Sous-phase	Description	Durée
3.1	Infrastructure de base (namespaces, quotas, OpenEBS)	2 JH
3.2	Stockage et métadonnées (MINIO, PostgreSQL HA)	2 JH
3.3	Sécurité (Keycloak, OPA, RBAC, cert-manager)	2 JH
3.4	Orchestration et intégration (Airflow, Nifi)	2 JH
3.5	Traitement et transformation (Trino, Spark, DBT)	1 JH
3.6	Datalab (Zeppelin avec intégration Trino/Spark)	1 JH
3.7	Gouvernance (OpenMetadata avec lineage)	1 JH
3.8	Observabilité (Prometheus, Grafana, Loki)	2 JH
3.9	Backup et DR (Velero, scripts PostgreSQL)	2 JH

**Livrables :**

- Manifests Kubernetes / Helm charts pour tous les composants
- Documentation de configuration complète
- Scripts d'installation et de déploiement
- Configuration monitoring avec dashboards Grafana
- Plan de Disaster Recovery

### **Phase 3 : Implémentation d'un cas d'utilisation simple (8 JH)**

**Objectif** : Implémenter un cas d'utilisation concret pour valider la plateforme et démontrer son fonctionnement

**Activités :**

- Analyse et définition du cas d'utilisation avec l'équipe métier
- Ingestion de données sources (via Nifi)
- Transformation des données (via Spark/DBT)
- Stockage dans tables Iceberg
- Création de requêtes SQL (via Trino)
- Analyse exploratoire (via Zeppelin)
- Documentation du cas d'utilisation
- Tests et validation du cas d'utilisation
- Présentation et démonstration

**Livrables :**

- Cas d'utilisation implémenté et opérationnel
- Documentation du cas d'utilisation
- Pipeline de données complet (ingestion → transformation → analyse)
- Notebooks Zeppelin avec visualisations
- Guide de réplication pour autres cas d'utilisation

### **Phase 4 : Support post-déploiement et optimisation (7 JH)**

**Objectif** : Assurer le support et l'optimisation continue

**Activités :**

- Support technique post-production
- Optimisations de performance
- Amélioration des pipelines
- Formation complémentaire à distance
- Documentation des cas d'usage complexes

**Livrables :**

- Rapport d'optimisation
  - Documentation des améliorations
  - Guide d'optimisation
-

## 4. Planification

### Planning global

**Durée totale** : 12 semaines (3 mois)

Période	Activités principales
Semaine 1-2	Phase 1 - Revue de l'architecture
Semaine 3-8	Phase 2 - Déploiement des composants
Semaine 9-10	Phase 3 - Implémentation cas d'utilisation
Semaine 11-12	Phase 4 - Support post-déploiement

### Jalons

- **J1** (Semaine 2) : Validation de l'architecture
- **J2** (Semaine 4) : Infrastructure de base opérationnelle
- **J3** (Semaine 6) : Services principaux déployés
- **J4** (Semaine 8) : Plateforme complète déployée
- **J5** (Semaine 10) : Validation et recette
- **J6** (Semaine 12) : Mise en production

### Déplacements

- **Mi-décembre 2025** : 1 semaine en présentiel à Dakar (5 jours ouvrés)
    - Kick-off et revue d'architecture
    - Déploiement des composants de base
  - **Fin-février 2026** : 2 semaines en présentiel à Dakar (10 jours ouvrés)
    - Déploiements critiques nécessitant accès au cluster
    - Tests de validation
    - *Note : Billets d'avion pris en charge par le client*
  - **Reste** : Travail à distance
-



## 5. Chiffrage

### Répartition des jours-hommes

Phase	Prestation	Durée (JH)
Phase 1	Revue de l'architecture	2
Phase 3.1	Infrastructure de base	2
Phase 3.2	Stockage et métadonnées	2
Phase 3.3	Sécurité et authentification	2
Phase 3.4	Orchestration et intégration	2
Phase 3.5	Traitement et transformation	1
Phase 3.6	Datalab - Apache Zeppelin	1
Phase 3.7	Gouvernance et catalogue	1
Phase 3.8	Observabilité	1
Phase 3.9	Backup et DR	1
Phase 3.10	Support et ajustements	2
Phase 4	Implémentation cas d'utilisation	8
Phase 5	Support post-déploiement	7
<b>TOTAL</b>		<b>32 JH</b>

### Détail du chiffrage

#### Répartition par phase :

- Phase 1 : 700 000 FCFA
- Phase 3.1 à 3.9 : 4 200 000 FCFA
- Phase 3.10 : 700 000 FCFA
- Phase 4 : 2 800 000 FCFA
- Phase 5 : 2 450 000 FCFA

**Frais de déplacement** : 700 000 FCFA (billet d'avion)

### Récapitulatif financier

Poste	Montant HT
Prestations (32 JH)	11 200 000 FCFA
Frais de déplacement (billet d'avion)	700 000 FCFA
<b>TOTAL HT</b>	<b>11 900 000 FCFA</b>
TVA (20%)	2 380 000 FCFA

<b>TOTAL TTC</b>	<b>14 280 000 FCFA</b>
------------------	------------------------

**Tarif unitaire :** 350 000 FCFA / jour-homme

**Conditions de facturation**

- Facturation par phase selon avancement
    - 30% à la commande
    - 40% à la validation de chaque phase
    - 30% à la livraison finale
-

## 6. Modalités d'intervention

### Répartition présentiel / distanciel

Phase	Présentiel	Distanciel	Total
Phase 1 - Revue architecture	2 JH	0 JH	2 JH
Phase 2 - Déploiement	5 JH	10 JH	15 JH
Phase 3 - Implémentation cas d'utilisation	2 JH	6 JH	8 JH
Phase 4 - Support	0 JH	7 JH	7 JH
<b>TOTAL</b>	<b>9 JH (28%)</b>	<b>23 JH (72%)</b>	<b>32 JH</b>

### Organisation du travail

#### Présentiel à Dakar :

- Déploiements critiques nécessitant accès au cluster
- Formation et transfert de compétences
- Tests de validation
- Points d'avancement réguliers

#### Distanciel :

- Préparation et documentation
- Développement des configurations
- Code reviews
- Support et résolution de problèmes
- Suivi et reporting

### Communication

- **Points d'avancement** : Hebdomadaires (1h)
  - **Outils** : Teams/Zoom, Slack, Git, Confluence/Wiki
  - **Reporting** : Rapport hebdomadaire d'avancement
-

## 7. Livrables

### Phase 1 : Revue de l'architecture

- Document de revue d'architecture
- Diagrammes d'architecture mis à jour
- Plan de déploiement détaillé
- Matrice de risques et mitigation

### Phase 2 : Déploiement

- Manifests Kubernetes / Helm charts pour tous les composants
- Documentation de configuration de chaque composant
- Scripts d'installation et de déploiement
- Configuration Zeppelin avec intégration Trino/Spark
- Exemples de notebooks Zeppelin
- Configuration OPA avec intégration Keycloak
- Procédures de backup/restore
- Configuration de monitoring (dashboards Grafana)
- Documentation d'intégration SSO
- Plan de Disaster Recovery

### Phase 3 : Implémentation cas d'utilisation

- Cas d'utilisation implémenté et opérationnel
- Pipeline de données complet
- Notebooks Zeppelin avec visualisations
- Documentation du cas d'utilisation
- Guide de réplication

### Phase 4 : Support post-déploiement

- Rapport d'optimisation
- Documentation des améliorations
- Guide d'optimisation

### Livrables transversaux

- Documentation technique complète
  - Architecture as Code (Terraform/Helm/Kustomize)
  - Procédures de maintenance
  - Guide de troubleshooting
-

## 8. Profil du consultant

**Consultant** : Bassirou KA

**Sociétés** : Tecsen France / Back Consulting Sénégal

**Architecte Data Ops** avec 14 ans d'expérience en architecture de données et dataplateformes.

### Expérience

- 14 ans d'expérience en architecture de données et data engineering
- Expertise en conception et déploiement de dataplateformes cloud-native
- Spécialisation en architectures Kubernetes on-premise et cloud
- Expérience sur projets grands comptes en France et en Afrique
- Accompagnement d'équipes Data Engineering et transfert de compétences

### Compétences techniques

**Orchestration et traitement** : Apache Airflow, Apache Spark, Apache Nifi, DBT, Trino

**Stockage et formats** : Apache Iceberg, Delta Lake, MINIO, S3-compatible storage, PostgreSQL

**Infrastructure et DevOps** : Kubernetes, Helm, Kustomize, Terraform, CI/CD, Prometheus, Grafana, Loki

**Sécurité et gouvernance** : Keycloak, OAuth2/OIDC, Open Policy Agent (OPA), RBAC, Data Catalog (OpenMetadata)

**Cloud** : AWS, Azure

### Certifications

- **CKA** : Certified Kubernetes Administrator
- **ISO-27001** : Information Security Management Systems
- Certifications cloud (AWS/Azure)

### Domaines d'expertise

- Architecture de dataplateformes modernes
  - Migration vers architectures cloud-native
  - Mise en place de stratégies Data Ops
  - Optimisation de performances et coûts
  - Sécurisation et gouvernance des données
  - Formation et accompagnement d'équipes
-

## 9. Conditions générales

### Prérequis

#### Infrastructure :

- Cluster Kubernetes opérationnel
- Accès administrateur au cluster
- Stockage persistant disponible (OpenEBS ou équivalent)
- Réseau configuré (ingress controller)
- Accès internet pour téléchargement d'images Docker

#### Équipe :

- Équipe Data Engineer disponible (2-3 personnes)
- Accès aux environnements de développement/staging/production
- Accès aux sources de données pour tests

#### Outils :

- Accès Git pour versioning
- Outils de communication (Teams/Slack)
- Accès à la documentation existante
- Accès VPN

### Gestion des risques

#### Risques identifiés :

##### 1. Complexité de l'infrastructure Kubernetes

- *Mitigation* : Expérience confirmée, documentation détaillée

##### 2. Compatibilité des composants

- *Mitigation* : Tests en environnement de développement, versions validées

##### 3. Performance et sizing

- *Mitigation* : Recommandations de sizing, tests de charge

##### 4. Disponibilité de l'équipe

- *Mitigation* : Planning flexible, communication régulière

### Propriété intellectuelle

- Les livrables (code, configurations, documentation) sont la propriété de la CBAO
- Le consultant conserve le droit d'utiliser les connaissances acquises (sans divulguer d'informations confidentielles)

### Confidentialité

- Engagement de confidentialité sur les informations de la CBAO
  - Respect de la réglementation applicable sur les données
-