

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

Table des matières

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

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
 - **Fiabilité** : Haute disponibilité et résilience
 - **Sécurité** : Authentification centralisée et contrôle d'accès granulaire
 - **Observabilité** : Monitoring complet et traçabilité des données
 - **Gouvernance** : Catalogue de données et lineage pour la traçabilité
 - **Datalab** : Environnement collaboratif pour l'analyse exploratoire et le développement (Zeppelin)
 - **Autonomie** : Encadrement de l'équipe Data Engineer pour assurer la pérennité
-

Périmètre de la prestation

- Revue et validation de l'architecture
 - Encadrement et formation de l'équipe Data Engineer
 - Déploiement des composants sur Kubernetes
 - Configuration et intégration des services
 - Mise en place du monitoring et de l'observabilité
 - Documentation technique et opérationnelle
 - Transfert de compétences
-

Profil du consultant

Consultant : Bassirou KA

Sociétés :

- Tecsen France
- Back Consulting Sénégal

Architecte Data Ops avec :

- 14 ans d'expérience en architecture de données
 - Expertise Kubernetes et cloud-native
 - Expérience sur les technologies de la stack (Airflow, Spark, Trino, Iceberg, etc.)
 - Certifications Kubernetes (CKA) et autres cloud
 - Certification ISO-27001
 - CV en pièce jointe
-

Architecture de référence

Vue d'ensemble

La plateforme est basée sur une architecture cloud-native avec les composants suivants :

Composants principaux

1. **Orchestration** : Apache Airflow
2. **Intégration** : Apache Nifi
3. **Transformation** : DBT + Apache Spark
4. **Stockage** : MINIO (S3-compatible) + Apache Iceberg
5. **Moteur SQL** : Trino
6. **Métadonnées** : PostgreSQL
7. **Datalab** : Apache Zeppelin (notebooks collaboratifs)
8. **Gouvernance** : DataHub ou Apache Atlas
9. **Sécurité** : Keycloak (SSO)
10. **Monitoring** : Prometheus + Grafana
11. **Logs** : Loki + Promtail
12. **Backup** : Velero

Namespaces Kubernetes

- `data-platform-core` : Composants centraux (Airflow, Keycloak)
 - `data-platform-storage` : Stockage (MINIO, PostgreSQL, OpenEBS)
 - `data-platform-integration` : Intégration (Apache Nifi)
 - `data-platform-processing` : Traitement (Spark, DBT, Trino)
 - `data-platform-datalab` : Datalab (Zeppelin)
 - `data-platform-observability` : Monitoring (Prometheus, Grafana, Loki)
 - `data-platform-governance` : Gouvernance (Data Catalog)
-

Détail des prestations

Phase 1 : Revue de l'architecture

Durée : 3 jour-homme

Modalité : 100% présentiel

Activités

1. Analyse de l'existant

- Revue de l'architecture documentée
- Analyse des besoins métier et techniques
- Identification des contraintes et risques
- Validation des choix technologiques

2. Recommandations architecturales

- Validation ou ajustement de l'architecture proposée
- Validation du choix Apache Nifi pour l'intégration
- Optimisations pour l'environnement on-premise
- Stratégie de déploiement par phases

3. Document de synthèse

- Architecture validée avec diagrammes
- Plan de déploiement détaillé
- Matrice de risques et mitigation
- Recommandations de sizing et ressources

Livrables :

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

Phase 2 : Encadrement de l'équipe Data Engineer

Durée : 5 jours-hommes

Modalité : 50% à distance, 50% présentiel

Activités

1. Formation initiale (1.5 JH)

- Architecture de la plateforme
- Principes Kubernetes pour Data Engineers
- Bonnes pratiques Data Ops
- Outils et workflows

2. Co-pilotage technique (2 JH)

- Accompagnement sur les choix techniques
- Code reviews et bonnes pratiques
- Résolution de problèmes complexes

- Optimisation des pipelines

3. Transfert de compétences (1.5 JH)

- Documentation des procédures
- Runbooks opérationnels
- Sessions de questions/réponses
- Formation avancée sur les composants critiques

Livrables :

- Guide d'utilisation de la plateforme
- Runbooks opérationnels
- Documentation des bonnes pratiques
- Sessions de formation enregistrées (si applicable)

Phase 3 : Déploiement des composants sur Kubernetes (Forfait + Régie)

Durée totale : 22 jours-hommes

Modalité : 77% à distance, 23% présentiel

3.1 Infrastructure de base (3 JH - Forfait)

Activités :

- Configuration des namespaces Kubernetes
- Mise en place des Resource Quotas et Limits
- Configuration des Network Policies
- Déploiement d'OpenEBS (storage provisioner)
- Configuration des StorageClasses

Livrables :

- Manifests Kubernetes pour l'infrastructure
- Documentation de configuration

3.2 Stockage et métadonnées (2 JH - Forfait)

Activités :

- Déploiement de MINIO (StatefulSet, HA)
- Configuration des buckets et lifecycle policies
- Déploiement de PostgreSQL (HA avec Patroni)
- Configuration des schémas et extensions
- Tests de performance et résilience

Livrables :

- Helm charts ou manifests pour MINIO et PostgreSQL
- Documentation de configuration
- Procédures de backup/restore

3.3 Sécurité et authentification (2 JH - Forfait)

Activités :

- Déploiement de Keycloak (HA)
- Configuration des realms et utilisateurs

- Intégration OAuth2/OIDC avec les services
- Configuration RBAC Kubernetes
- Mise en place de cert-manager pour TLS

Livrables :

- Configuration Keycloak
- Documentation d'intégration SSO
- Procédures de gestion des utilisateurs

3.4 Orchestration et intégration (2 JH - Forfait)

Activités :

- Déploiement d'Apache Airflow (Helm chart)
- Configuration KubernetesExecutor
- Intégration Keycloak pour authentification
- Déploiement d'Apache Nifi
- Configuration des processors et flows de base
- Configuration ZooKeeper pour coordination cluster
- Tests de flows d'ingestion

Livrables :

- Configuration Airflow opérationnelle
- Configuration Apache Nifi opérationnelle (cluster HA)
- Exemples de DAGs et pipelines
- Exemples de flows Nifi

3.5 Traitement et transformation (2 JH - Forfait)

Activités :

- Déploiement de Trino (coordinateur + workers)
- Configuration des catalogues (Iceberg, PostgreSQL)
- Intégration Keycloak
- Configuration de Spark Operator
- Configuration DBT avec Trino adapter
- Tests d'intégration

Livrables :

- Configuration Trino opérationnelle
- Configuration Spark
- Configuration DBT
- Exemples de transformations

3.6 Datalab - Apache Zeppelin (1 JH - Forfait)

Activités :

- Déploiement d'Apache Zeppelin (StatefulSet)
- Configuration des interpréteurs (Spark, Trino, Python, SQL)
- Intégration Keycloak pour authentification SSO
- Configuration du stockage des notebooks (MINIO ou volumes persistants)
- Configuration des connexions aux services :
 - Connexion à Trino pour requêtes SQL
 - Connexion à Spark pour traitement distribué

- Connexion à MINIO pour accès aux données
- Configuration des permissions et partage de notebooks
- Tests d'intégration avec Trino et Spark
- Documentation d'utilisation pour Data Analysts et Data Scientists

Livrables :

- Zeppelin opérationnel avec SSO
- Configuration des interpréteurs
- Exemples de notebooks (Trino, Spark, visualisations)
- Documentation d'utilisation
- Guide de bonnes pratiques pour notebooks

3.7 Gouvernance et catalogue (1 JH - Forfait)

Activités :

- Déploiement de DataHub ou Apache Atlas
- Configuration des connecteurs de métadonnées
- Intégration avec Airflow, Spark, DBT, Trino, Zeppelin
- Configuration du lineage automatique
- Tests de découverte et lineage

Livrables :

- Data Catalog opérationnel
- Documentation d'utilisation
- Guide de documentation des données

3.8 Observabilité (1 JH - Forfait)

Activités :

- Déploiement de Prometheus Operator
- Configuration des ServiceMonitors
- Déploiement de Grafana avec dashboards
- Intégration Keycloak
- Déploiement de Loki + Promtail
- Configuration des alertes de base

Livrables :

- Stack d'observabilité opérationnelle
- Dashboards Grafana (infrastructure, Airflow, Trino, etc.)
- Configuration AlertManager
- Documentation d'utilisation

3.9 Backup et Disaster Recovery (2 JH - Forfait)

Activités :

- Déploiement de Velero
- Configuration des backups automatiques
- Scripts de backup PostgreSQL
- Configuration de la réplication MINIO
- Documentation des procédures DR
- Test de restauration

Livrables :

- Système de backup opérationnel
- Procédures de restauration documentées
- Plan de Disaster Recovery

3.10 Support et ajustements (Régie - 3 JH)

Activités :

- Résolution de problèmes de déploiement
- Optimisations de configuration
- Ajustements selon retours équipe
- Support technique pendant le déploiement
- Documentation complémentaire

Phase 4 : Tests et validation (Régie)

Durée : 10 jours-hommes

Modalité : 20% présentiel, 80% distanciel

Activités

1. Tests d'intégration

- Tests end-to-end des pipelines
- Tests de performance
- Tests de résilience (HA, failover)
- Tests de sécurité
- Tests Zeppelin (notebooks, connexions Trino/Spark)

2. Validation fonctionnelle

- Validation avec l'équipe métier
- Tests de charge
- Validation des SLAs

3. Documentation finale

- Mise à jour de la documentation
- Procédures opérationnelles
- Guide utilisateur

Livrables :

- Rapport de tests détaillé
- Documentation finale complète
- Procédures opérationnelles validées
- Guide utilisateur final

Phase 5 : Support post-déploiement et optimisation (Régie)

Durée : 10 jours-hommes

Modalité : 100% distanciel

Activités

1. Support post-déploiement

- Résolution de problèmes post-production
- Support technique et assistance
- Optimisations de performance
- Ajustements de configuration

2. Optimisation continue

- Analyse des performances
- Recommandations d'optimisation
- Amélioration des pipelines
- Fine-tuning des configurations

3. Formation complémentaire

- Sessions de formation avancée à distance
- Support aux équipes utilisatrices
- Documentation des cas d'usage complexes

Livrables :

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

Planification

Planning global (12 semaines)

Semaine 1-2 : Phase 1 - Revue de l'architecture
Semaine 3-4 : Phase 2 - Encadrement équipe (en parallèle du déploiement)
Semaine 3-8 : Phase 3 - Déploiement des composants
Semaine 9-10 : Phase 4 - Tests et validation
Semaine 11-12 : Support et ajustements finaux

Planning détaillé Phase 3

Semaine	Composants	Durée
3	Infrastructure de base + Stockage	2 semaines
4	Sécurité + Orchestration	2 semaines
5-6	Intégration + Traitement	2 semaines
7	Datalab (Zeppelin) + Gouvernance + Observabilité	1 semaine
8	Backup + Support	1 semaine

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
-

Chiffrage

Répartition des jours-hommes

Phase	Prestation	Durée (JH)	Type
Phase 1	Revue de l'architecture	3	Forfait
Phase 2	Encadrement équipe Data Engineer	5	Forfait
Phase 3.1	Infrastructure de base	3	Forfait
Phase 3.2	Stockage et métadonnées	2	Forfait
Phase 3.3	Sécurité et authentification	2	Forfait
Phase 3.4	Orchestration et intégration	2	Forfait
Phase 3.5	Traitements et transformation	2	Forfait
Phase 3.6	Datalab - Apache Zeppelin	1	Forfait
Phase 3.7	Gouvernance et catalogue	1	Forfait
Phase 3.8	Observabilité	1	Forfait
Phase 3.9	Backup et DR	2	Forfait
Phase 3.10	Support et ajustements	3	Régie
Phase 4	Tests et validation	10	Régie
Phase 5	Support post-déploiement et optimisation	10	Régie
TOTAL		50 JH	

Détail du chiffrage

Forfait (24 JH)

Phase	Description	JH	Tarif unitaire	Montant HT
Phase 1	Revue de l'architecture	3	350 000 FCFA	1 050 000 FCFA
Phase 2	Encadrement équipe	5	350 000 FCFA	1 750 000 FCFA
Phase 3.1	Infrastructure de base	3	350 000 FCFA	1 050 000 FCFA
Phase 3.2	Stockage et métadonnées	2	350 000 FCFA	700 000 FCFA
Phase 3.3	Sécurité et authentification	2	350 000 FCFA	700 000 FCFA

Phase 3.4	Orchestration et intégration	2	350 000 FCFA	700 000 FCFA
Phase 3.5	Traitement et transformation	2	350 000 FCFA	700 000 FCFA
Phase 3.6	Datalab - Apache Zeppelin	1	350 000 FCFA	350 000 FCFA
Phase 3.7	Gouvernance et catalogue	1	350 000 FCFA	350 000 FCFA
Phase 3.8	Observabilité	1	350 000 FCFA	350 000 FCFA
Phase 3.9	Backup et DR	2	350 000 FCFA	700 000 FCFA
SOUS-TOTAL FORFAIT		24		8 400 000 FCFA

Régie (26 JH)

Phase	Description	JH	Tarif unitaire	Montant HT
Phase 3.10	Support et ajustements	3	350 000 FCFA	1 050 000 FCFA
Phase 4	Tests et validation	10	350 000 FCFA	3 500 000 FCFA
Phase 5	Support post-déploiement et optimisation	10	350 000 FCFA	3 500 000 FCFA
SOUS-TOTAL RÉGIE		26		9 100 000 FCFA

Frais de déplacement

Hypothèses :

- Mi-décembre 2025** : 1 semaine de présence sur site à Dakar (5 jours ouvrés) - Billet à la charge du client
- Mi-février 2026** : 2 semaines de présence sur site à Dakar (10 jours ouvrés)
- Déplacements depuis Paris vers Dakar
- Seul le billet d'avion de fin février est pris en charge

Poste	Quantité	Tarif unitaire	Montant HT
Billet d'avion Paris-Dakar (aller-retour) - Février 2026	1 aller-retour	600 000 FCFA	600 000 FCFA
TOTAL	1 aller-retour		600 000 FCFA

Note : Les frais de déplacement incluent uniquement le billet d'avion Paris-Dakar (aller-retour) de fin février 2026. Le billet de décembre 2025, l'hébergement et les repas sont à la charge du client ou pris en charge séparément.

Récapitulatif financier

Poste	Montant HT
Forfait (24 JH)	8 400 000 FCFA
Régie (26 JH)	9 100 000 FCFA
Frais de déplacement (billet d'avion février)	600 000 FCFA
TOTAL HT	18 100 000 FCFA
TVA (18%)	3 258 000 FCFA
TOTAL TTC	21 358 000 FCFA

Conditions de facturation

- **Forfait** : Facturation par phase selon avancement
 - 30% à la commande
 - 40% à la validation de chaque phase
 - 30% à la livraison finale
 - **Régie** : Facturation mensuelle sur justificatifs
 - **Frais de déplacement** : Facturation mensuelle sur justificatifs
-

Modalités d'intervention

Répartition présentiel / distanciel

Planning des déplacements :

- **Mi-décembre 2025** : 1 semaine en présentiel à Dakar (5 jours ouvrés) - Billet à la charge du client
- **Mi-février 2026** : 2 semaines en présentiel à Dakar (10 jours ouvrés) - Billet pris en charge
- **Reste** : Travail à distance

Phase	Présentiel	Distanciel	Total
Phase 1 - Revue architecture	3 JH	0 JH	3 JH
Phase 2 - Encadrement équipe	5 JH	0 JH	5 JH
Phase 3 - Déploiement	5 JH	17 JH	22 JH
Phase 4 - Tests	2 JH	8 JH	10 JH
Phase 5 - Support post-déploiement	0 JH	10 JH	10 JH
TOTAL	15 JH	35 JH	50 JH

Note : Les 23 JH de régie (Phase 3.10, Phase 4 et Phase 5) sont répartis principalement en distanciel, avec 2 JH de présence sur site pour les tests de validation.

Organisation du travail

Présentiel à Dakar :

- **Mi-décembre 2025 (1 semaine)** :
 - Kick-off et revue d'architecture
 - Formation initiale de l'équipe
 - Déploiement des composants de base
 - Note : Billet d'avion à la charge du client
- **Mi-février 2026 (2 semaines)** :
 - Déploiements critiques nécessitant accès au cluster
 - Formation avancée et transfert de compétences
 - Tests de validation
 - Points d'avancement et ajustements
 - Note : Billet d'avion pris en charge

Distanciel :

- Préparation et documentation
- Développement des configurations
- Code reviews
- Support et résolution de problèmes
- Suivi et reporting
- Travail entre les périodes de présence sur site

Communication

- **Points d'avancement** : Hebdomadaires (1h)

- **Outils** : Teams/Zoom, Slack, Git, Confluence/Wiki
 - **Reporting** : Rapport hebdomadaire d'avancement
-

Livrables

Phase 1 : Revue de l'architecture

- Document de revue d'architecture (15-20 pages)
- Diagrammes d'architecture mis à jour
- Plan de déploiement détaillé
- Matrice de risques et mitigation
- Recommandations de sizing

Phase 2 : Encadrement équipe

- Guide d'utilisation de la plateforme
- Runbooks opérationnels
- Documentation des bonnes pratiques
- Sessions de formation (slides + enregistrements si applicable)

Phase 3 : 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
- Procédures de backup/restore
- Configuration de monitoring (dashboards Grafana)
- Documentation d'intégration SSO
- Plan de Disaster Recovery

Phase 4 : Tests et validation

- Rapport de tests d'intégration
- Rapport de tests de performance
- Documentation finale consolidée
- Procédures opérationnelles validées
- Guide utilisateur final

Livrables transversaux

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

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 du client
- Le consultant conserve le droit d'utiliser les connaissances acquises (sans divulguer d'informations confidentielles)

Confidentialité

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