Document de conception technique

Projet : Application de Veille des Marchés Publics

Auteur : logo service

Date : 05/05/2025

## Introduction

Dans le cadre de la facilitation de l’accès aux marchés publiques pour les entreprises du Burkina Faso, et suite à la réalisation du document technique et fonctionnel , ce document intervient en vue d’avoir une vision claire et globale de la façon dont le projet sera développé .

## Rappel des exigences fonctionnelles

Les fonctionnalités principales de L’application de veille des marchés publique sont :

* La gestion des entreprises,
* l’extraction automatique,
* l’analyse intelligente,
* les notifications et alertes,
* l’interface administrateur

## présentation du modèle conceptuel de données

Après avoir analysé les exigences fonctionnelles du projet, nous avons opté pour l’utilisation de **deux types de bases de données** : une base **relationnelle (SQL)** et une base **non relationnelle (NoSQL)**. Ce choix a pour objectif d’**optimiser les performances** du système en s’adaptant à la nature des données traitées.

La base de données SQL sera utilisée pour le **stockage des données structurées** telles que les informations sur les utilisateurs, les entreprises et les marchés. En parallèle, la base NoSQL permettra de **gérer efficacement les données non structurées**, comme les fichiers PDF et les textes extraits de ces documents.

Cette approche hybride permet d’exploiter les **forces complémentaires** des deux technologies : la fiabilité et la cohérence du SQL pour les données bien définies, et la souplesse du NoSQL pour les contenus plus libres et volumineux. Elle garantit ainsi un système de veille **plus performant, évolutif et cohérent**.

La structure globale du stockage des données peut donc être résumée comme suit :

* **Base SQL** : utilisateurs, entreprises, marchés publics
* **Base NoSQL** : fichiers PDF, contenu textuel extrait, logs de traitement, etc.

1.structure de la base de donnée SQl :

1. Table : Entreprise

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique de l'entreprise | 1 |
| nom | VARCHAR(255) | Nom complet de l'entreprise | Fournitures Express SARL |
| siret | VARCHAR(20) | Numéro d'identification (SIRET-like) | BF1234567890 |
| adresse | TEXT | Adresse postale complète | 01 BP 1234 Ouagadougou |
| email | VARCHAR(255) | Adresse email de contact | [contact@fournitures-express.bf](mailto:contact@fournitures-express.bf) |
| telephone | VARCHAR(20) | Numéro de téléphone | +226 25 30 00 00 |
| representantNom | VARCHAR(255) | Nom du représentant légal | OUEDRAOGO |
| representantPrenom | VARCHAR(255) | Prénom du représentant légal | Fatoumata |
| dateInscription | DATETIME | Date d'inscription de l'entreprise sur la plateforme | 2025-04-23 10:00:00 |

1. **Table : Domaine**

*Stocke les domaines d'activité normalisés.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique du domaine | 1 |
| libelle | VARCHAR(255) | Nom du domaine d'activité normalisé | Fournitures de bureau |

**3. Table : Publication**

*Stocke les informations sur les publications (ex: le Quotidien).*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| publicationId | INT | Identifiant unique de la publication | 1 |
| datePublication | DATE | Date de publication | 2025-04-10 |
| titre | VARCHAR(255) | Titre de la publication | Quotidien N° 4115 |
| source\_url | VARCHAR(255) | URL de la page source | <http://www.dgcmef.gov.bf/fr/taxonomy/term/16> |
| reference | VARCHAR(50) | Référence unique de la publication | N° 4115- |
| typePublication | VARCHAR(50) | Type de publication (ex: "Quotidien d'information") | Quotidien d’information |
| fichierPdf | VARCHAR(255) | Chemin vers le fichier PDF téléchargé | /data/pdfs/Quotidien-Ndeg4115\_061307.pdf |
| contenuExtrait | TEXT | Contenu textuel extrait du PDF | (Texte intégral du PDF - voir exemple partiel dans le précédent message, commence par "MINISTERE DE...") |

**4. Table : TypeProcedure**

*Stocke les types de marchés normalisés (ex: "Fournitures", "Travaux", ….).*

*Catégoriser les grandes sections du document qui regroupent différents marchés. Elle sert à structurer le document en grandes parties homogènes.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique du type de marché | 1 |
| libelle | VARCHAR(255) | Libellé du type de marché (ex: "Fournitures", "Travaux", "Résultats...") | Résultats provisoires des ministères, institutions... |

**5. Table : Marche**

*Table "mère" contenant les informations communes à tous les marchés.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| marcheId | INT | Identifiant unique du marché (commun aux appels d'offres et résultats) | 1 |
| publicationId | INT | Clé étrangère vers la table Publication | 1 |
| typeProcédureId | INT | Clé étrangère vers la table TypeMarche | 1 |
| ministere | VARCHAR(255) | Ministère concerné | (A extraire du document) |
| region | VARCHAR(100) | Région concernée | (A extraire du document) |
| objet | TEXT | Description de l'objet du marché | Acquisition de fournitures de bureau... |
| budgetMin | DECIMAL(15,2) | Budget minimal (si fourchette) | 76224870.00 |
| budgetMax | DECIMAL(15,2) | Budget maximal (si fourchette) | 81000000.00 |

**6. Table : AppelOffre**

*Stocke les informations spécifiques aux appels d'offres.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique (hérité de Marche) | 2 |
| dateDepot | DATE | Date limite de dépôt des offres | (A extraire du document) |
| referenceDossier | VARCHAR(255) | Référence du dossier d'appel d'offres | (A extraire du document) |
| lieuDepot | TEXT | Lieu de dépôt des offres | (A extraire du document) |
| conditionsParticipation | TEXT | Conditions de participation (éligibilité, documents à fournir, etc.) | (A extraire du document) |
| criteresSelection | TEXT | Critères de sélection des offres (prix, qualité technique, délais, etc.) | (A extraire du document) |
| cautionnement | DECIMAL(15,2) | Montant du cautionnement (si applicable) | (A extraire du document) |
| dureeValiditeOffres | VARCHAR(255) | Durée de validité des offres | (A extraire du document) |

**7. Table : Resultat**

*Stocke les informations spécifiques aux résultats d'attribution.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique (hérité de Marche) | 1 |
| dateAttribution | DATE | Date d'attribution du marché | (A extraire du document) |
| entrepriseId | INT | Clé étrangère vers la table Entreprise (identifiant de l'entreprise attributaire) | 1 |
| montantAttribue | DECIMAL(15,2) | Montant final attribué à l'entreprise | 78000000.00 |
| referenceDecision | VARCHAR(255) | Référence de la décision d'attribution | (A extraire du document) |
| nombreOffresRecues | INT | Nombre d'offres reçues | (A extraire du document) |
| delaiExecution | VARCHAR(255) | Délai d'exécution du marché | (A extraire du document) |
| motifRejetAutresOffres | TEXT | Motif du rejet des autres offres | (A extraire du document, ex: "Offre anormalement basse", voir page 5 pour des exemples de motifs abrégés comme "Non conforme, offre...") |

**8. Table : Lot**

*Stocke les informations sur les lots, avec une relation polymorphique vers AppelOffre ou Resultat.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique du lot | 1 |
| marcheId | INT | Clé étrangère vers la table Marche | 1 |
| numeroLot | INT | Numéro du lot (si applicable) | 2 |
| description | TEXT | Description spécifique du lot | Acquisition de fournitures de bureau... |
| montant | DECIMAL(15,2) | Montant du lot (si disponible) | (A extraire du document) |

**9. Table : Alerte**

*Gère les alertes envoyées aux entreprises.*

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| id | INT | Identifiant unique de l'alerte | 1 |
| entrepriseId | INT | Clé étrangère vers la table Entreprise | 1 |
| publicationId | INT | Clé étrangère vers la table Publication | 1 |
| typeAlerte | VARCHAR(50) | Type d'alerte (ex: "Nouveau marché", "Attribution de marché") | Nouveau marché |
| dateAlerte | DATETIME | Date et heure de l'alerte | 2025-04-23 10:00:00 |
| contenuAlerte | TEXT | Résumé de l'alerte envoyée | Nouveau marché publié : Acquisition de fournitures de bureau... |
| CanalAlerte | VARCHAR(30) | Moyen par lequel l’alerte à été envoyée | Sms,email,whatsapp |

**10. Table: Utilisateur**

***Gère les comptes utilisateurs de la plateforme.***

|  |  |  |  |
| --- | --- | --- | --- |
| **Champ** | **Type** | **Description** | **Exemple** |
| **id** | **INT** | **Identifiant unique de l'utilisateur** | **1** |
| **email** | **VARCHAR(255)** | **Adresse email de l'utilisateur** | [**admin@example.com**](mailto:admin@example.com) |
| **motDePasse** | **VARCHAR(255)** | **Mot de passe (hashé)** | **hashed\_password** |
| **nom** | **VARCHAR(255)** | **Nom de l'utilisateur** | **Admin** |
| **prenom** | **VARCHAR(255)** | **Prénom de l'utilisateur** | **Super** |
| **role** | **VARCHAR(20)** | **Rôle de l'utilisateur (ex: "admin", "entreprise")** | **admin** |
| **entrepriseId** | **INT** | **Clé étrangère vers la table Entreprise (si applicable)** | **(NULL pour un admin)** |

**11. Table: SecteurActivite**

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | exemple |
| **Id** | **INT** | **Identifiant unique du secteur d’activité** | **1** |
| **Libelle** | **VARCHAR(255)** | **Description du secteur d’activité** | **Bâtiment et Travaux Publics (BTP)** |

**12. Tables de jointure (PublicationDomaine et EntrepriseDomaine)**

*Gèrent les relations N-N entre publications/domaines et entreprises/domaines.*

* **Table: PublicationDomaine**

|  |  |
| --- | --- |
| Champ | Type |
| publicationId | INT |
| domaineId | INT |

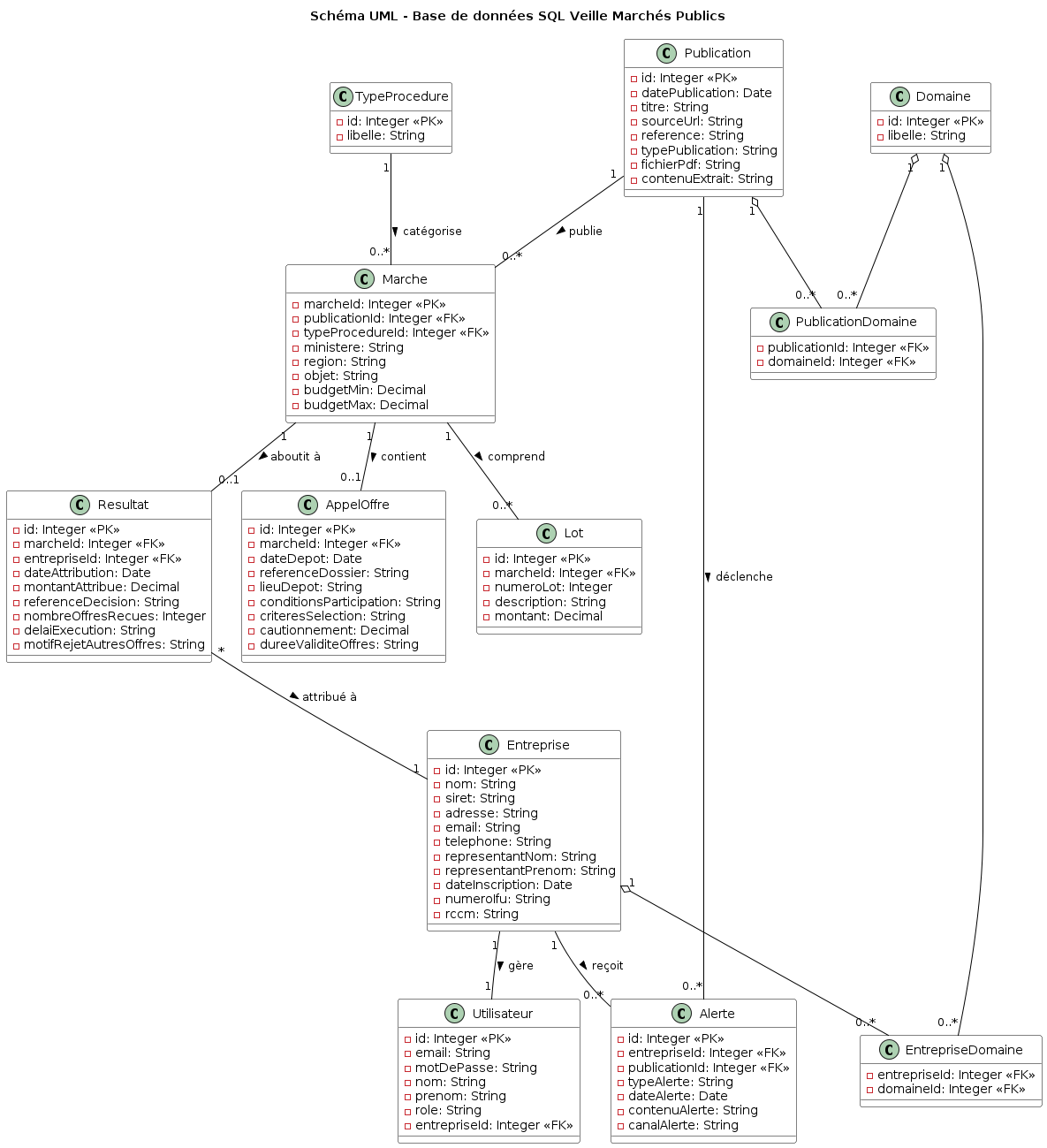
* **Table: EntrepriseDomaine**

|  |  |
| --- | --- |
| Champ | Type |
| entrepriseId | INT |
| domaineId | INT |

* **Table: EntrepriseSecteur**

|  |  |
| --- | --- |
| Champ | Type |
| entrepriseId | INT |
| secteurActiviteId | INT |

1.1Représentation schématique de la base de donnée SQL :

****

2.structure de la base de donnée NOSQL

## **Collection: documentsBruts**

| **Nom du champ** | **Type de données** | **Description** | **Exemple** |
| --- | --- | --- | --- |
| publicationId | NumberInt | Identifiant unique de la publication (Liaison SQL↔NoSQL) | 122 |
| nomFichier | String | Nom du fichier téléchargé | "appel\_offres\_pont\_2025.pdf" |
| contenuFichier | Binary | Contenu binaire ou texte du fichier | BinData(0, "JVBERi0xLjUKJdP...") |
| tailleFichier | NumberInt | Taille du fichier en octets | 2457600 |
| typeContenu | String | Type MIME du fichier | "application/pdf" |
| dateTelechargement | Date | Date et heure du téléchargement | 2025-04-28T14:32:11.000Z |
| urlSource | String | URL d'où provient le fichier | "https://www.marches-publics.gouv.fr/doc123" |
| referenceQuotidien | String | Référence à une publication officielle | "JOUE-2025-042-A12" |
| empreinteMd5 | String | Empreinte MD5 pour vérifier l'intégrité | "a1b2c3d4e5f6g7h8i9j0k1l2m3n4o5p6" |

## **Collection: donneesExtraites**

| **Nom du champ** | **Type de données** | **Description** | **Exemple** |
| --- | --- | --- | --- |
| publicationId | NumberInt | Lien vers la publication d'origine | ObjectId("6437f21a8b9a12345c789012") |
| idDocumentBrut | ObjectId | Référence au \_id de document brut | ObjectId("6437f21a8b9a12345c789013") |
| texteComplet | String | Texte intégral extrait du document | "Le présent marché a pour objet la..." |
| dateExtraction | Date | Date et heure de l'extraction | 2025-04-28T15:20:45.000Z |
| etatTraitement | String | État du processus de traitement | "terminé" |
| qualiteExtraction | Double | Indicateur de qualité de l'extraction | 0.95 |
| sections | Array<Document> | Sections identifiées dans le document | [{"titre": "Spécifications", "contenu": "..."}] |
| entites | Array<Document> | Entités nommées extraites | [{"type": "organisation", "nom": "Mairie de Paris"}] |
| domaines | Array<String> | Domaines d'activité identifiés | ["construction", "infrastructure"] |
| motsCles | Array<String> | Mots-clés extraits du document | ["pont", "rénovation", "acier"] |
| classification | Document | Catégorisation du document | {"catégorie": "travaux", "priorité": "haute"} |

## **Collection: lots**

| **Nom du champ** | **Type de données** | **Description** | **Exemple** |
| --- | --- | --- | --- |
| idDonneeExtraite | ObjectId | Référence au contenu extrait associé | ObjectId("6437f21a8b9a12345c789014") |
| marcheId | NumberInt | Identifiant du marché public | 108 |
| numeroLot | NumberInt | Numéro du lot dans le marché | 3 |
| description | String | Description du lot | "Fourniture et pose des structures métalliques" |
| montantEstime | Double | Montant estimé du lot | 450000.00 |
| montantAttribue | Double | Montant final attribué | 423500.00 |
| entrepriseAttributaire | Document | Entreprise à qui le lot a été attribué | {"nom": "Constructions Métalliques SA", "siret": "12345678901234"} |
| page | NumberInt | Numéro de page de l'information | 12 |
| contexte | String | Contexte textuel entourant l'information | "Le lot 3 concernant les structures..." |

## **Collection: journauxExtraction**

| **Nom du champ** | **Type de données** | **Description** | **Exemple** |
| --- | --- | --- | --- |
| idDocumentBrut | ObjectId | Référence au \_id du document PDF stocké dans documentsBruts | ObjectId("6437f21a8b9a12345c789012") |
| horodatage | Date | Date et heure de l'événement | ISODate("2025-04-28T15:10:23.000Z") |
| niveau | String | Niveau de journal | "info" |
| message | String | Message du journal | "Début de l'extraction du document" |
| composant | String | Composant logiciel qui a généré le journal | "ExtractionPDF" |
| details | Document | Informations détaillées sur l'événement | {"pages": 42, "durée": "00:01:23"} |
| traceErreur | String | Trace d'exécution en cas d'erreur | "Error: Impossible de parser la page 5..." |

**Collection : fichierAlerte :**

| **Champ** | **Type (NoSQL)** | **Description** | **Exemple** |
| --- | --- | --- | --- |
| entrepriseId | NumberInt | Identifiant de l'entreprise liée à l'alerte | 501 |
| publicationId | NumberInt | Identifiant de la publication concernée | 1 |
| contenuExtraitId | NumberInt | Identifiant de l'extrait de contenu lié | 101 |
| priorite | String | Niveau de priorité de l'alerte | "Haute" |
| dateCreation | Date | Date de création de l'alerte | 2025-05-01T08:30:00Z |
| contenu | String | Contenu textuel informatif de l'alerte | "Nouvelle publication dans votre domaine d'activité." |
| CanalAlerte | String | Moyen par lequel l’alerte sera transmise | Sms,email,message WhatsApp |

**Collection : indexRecherche :**

|  |  |  |  |
| --- | --- | --- | --- |
| Champ | Type | Description | Exemple |
| publicationId | NumberInt | Identifiant unique de la publication | 1 |
| contenuExtraitId | NumberInt | Identifiant de l'extrait de contenu indexé | 101 |
| texteIndexe | String | Texte indexé pour la recherche | Marché public pour la voirie |
| datePublication | date | Date de publication de l'avis | 2025-04-01 |
| motsCles | |  | | --- | |  |  |  | | --- | | Array<String> | | Liste de mots-clés associés à l'avis | ["voirie", "routes"] |
| domaines | |  | | --- | |  |  |  | | --- | | Array<String> | | Domaines concernés par le marché | ["Travaux Publics"] |
| ministeres | Array<String> | Ministères responsables | ["Min. Transports"] |
| regions | Array<String> | Régions géographiques concernées | ["Île-de-France"] |
| typeMarche | String | Type de marché | Travaux |
| typeProcedure | String | Type de procédure de passation du marché | Appel d'offres |
| derniereMiseAJour | date | Date de la dernière mise à jour de l'information | 2025-04-02 |

3. **Intégration entre les Bases de Données SQL et NoSQL**

Pour répondre à deux types de besoins, on utilise :

* **PostgreSQL** pour les données structurées (utilisateurs, entreprises, marchés).
* **MongoDB** pour les données non structurées (PDF, texte extrait).

**Flux de données étape par étape**

1. **Stockage du PDF**
   * Le scraper télécharge le fichier et l’enregistre dans MongoDB (documentsBruts) en lui ajoutant le même publicationId que dans la table SQL Publication.
2. **Extraction du texte**
   * Un service lit le PDF et place le texte (et ses métadonnées) dans MongoDB (donneesExtraites), toujours en conservant publicationId.
3. **Chargement des informations clés en SQL**
   * À partir du texte extrait, on identifie les données métier (lots, montants, dates…).
   * On met à jour ou on insère ces données dans PostgreSQL (Marche, Lot, etc.) en réutilisant publicationId, marcheId, etc.
4. **Gestion des alertes**
   * MongoDB repère automatiquement les alertes et les stocke provisoirement dans fichierAlerte.
   * Après validation, on récupère ces alertes et on les injecte dans la table SQL Alerte, toujours avec les mêmes identifiants.
5. **Recherche plein texte**
   * MongoDB construit un index (indexRecherche) pour rechercher rapidement dans le contenu extrait.
   * Les résultats renvoient ensuite aux enregistrements SQL correspondants grâce aux mêmes clés (publicationId).

**Pourquoi cette double approche ?**

* **Cohérence et relations complexes**  
  PostgreSQL gère les transactions et les jointures entre tables de façon fiable.
* **Flexibilité et volume**  
  MongoDB stocke facilement de gros documents et du texte libre, sans schéma fixe.
* **Correspondance simple**  
  On conserve exactement les mêmes identifiants (publicationId, marcheId, entrepriseId) des deux côtés, pour relier sans confusion les données structurées et non structurées.