

Rétroplanning Projet Intégré

EthiMask - Framework de Masquage Contextuel & Configuration Apache Atlas

Data Governance & Privacy - Projet Fédérateur 2025-2026

Période : 16 janvier 2026 – 19 janvier 2026

Vue d'ensemble du projet

Contexte

Ce rétroplanning couvre la mise en œuvre complète de deux composants critiques du système de gouvernance des données :

- **EthiMask** : Framework intelligent de masquage adaptatif basé sur un perceptron multi-facteurs
- **Apache Atlas** : Plateforme de catalogage et de lignage des données

Contraintes temporelles

- **Date de début** : Jeudi 16 janvier 2026
- **Date de livraison** : Lundi 19 janvier 2026 (11h00)
- **Durée totale** : 4 jours

1 Rétroplanning détaillé

1.1 16/01/2026 - Jour 1 : Fondations

Tâche	Livrables	Durée
PHASE 1 : Architecture & Configuration initiale		
Architecture Ethimask V0.1 - Définition détaillée du perceptron - Spécification des composantes (R, P, S, H_c, H_f, H_v) - Configuration des poids initiaux ($w_r, w_p, w_s, w_{hc}, w_{hf}, w_{hv}$)	- Schéma d'architecture détaillé - Document de spécification mathématique - Configuration des seuils de masquage	1h30
Installation Apache Atlas - Déploiement environnement (Docker/VM) - Configuration HBase + Solr - Paramétrage Kafka pour les hooks - Test de connectivité	- Instance Atlas opérationnelle - Configuration de base validée - Documentation d'installation	2h00
PHASE 2 : Développement Core		
Implémentation calcul de score - Fonction <code>ComputeRoleScore(u.role)</code> - Fonction <code>ComputePurposeScore(p, allowed)</code> - Fonction <code>GetSensitivity(a)</code> - Fonction sigmoïde $\sigma(x) = \frac{1}{1+e^{-x}}$	- Module <code>score_calculator.py</code> - Tests unitaires des fonctions - Validation mathématique	2h00
Configuration Atlas - Métadonnées - Définition types d'entités personnalisées - Configuration tags PII/SPI - Paramétrage classifications sensibilité - Import métadonnées existantes	- Types Atlas définis - Taxonomie de tags opérationnelle - Script d'import métadonnées	2h00
Développement historique utilisateur - Fonction <code>ComputeCompliance(u, H)</code> - Fonction <code>ComputeFrequency(u, a, H)</code> - Fonction <code>ComputeViolationPenalty(u, H)</code> - Module de logging des accès	- Module <code>user_history.py</code> - Schéma base de données historique - Tests de calcul	1h30

1.2 17/01/2026 - Jour 2 : Masquage & Intégration

Tâche	Livrables	Durée
PHASE 2 (suite) : Techniques de masquage		
Chiffrement Homomorphique (Niveau 1) - Installation TenSEAL/Concrete-ML - Implémentation chiffrement identifiants - Tests calculs sur données chiffrées - Validation performance	- Module <code>homomorphic_encryption.py</code> - Benchmarks performance - Documentation technique HE	2h00
Techniques de masquage (Niv. 2-4) - Niveau 2 : Généralisation (âge→tranche) - Niveau 3 : Privacy Différentielle (ϵ -DP) - Niveau 4 : Suppression (NULL/***) - Tests unitaires par niveau	- Module <code>masking_techniques.py</code> - Configuration paramètres ϵ - Suite de tests complète	1h30
PHASE 3 : Intégration & Lignage		

Orchestration EthiMask (Algorithm 9) - Implémentation flux complet - Intégration tous les modules - Fonction <code>ApplyMasking(value, level)</code> - Logging des décisions	- Module principal <code>ethimask.py</code> - API de masquage fonctionnelle - Base de données d'audit	2h00
Configuration Lignage Atlas - Hooks Airflow/Spark/Hive - Traçabilité transformations de données - Configuration graphe de lignage - Tests de propagation métadonnées	- Hooks configurés et actifs - Graphes de lignage visibles - Documentation traçabilité	2h00
Interface Admin Configuration Poids - US-MASK-03 : Dashboard configuration - Formulaire ajustement poids (w_*) - Validation $\sum w_i = 1$ - Sauvegarde configurations	- Interface web admin - Module de validation poids - Historique configurations	1h30

1.3 18/01/2026 - Jour 3 : User Stories & Audit

Tâche	Livrables	Durée
PHASE 3 (suite) : Implémentation User Stories		
US-MASK-01 & US-MASK-02 - Calcul automatique niveau masquage - API accès données par rôle - Tests multi-utilisateurs - Validation règles métier	- Service de masquage automatique - API REST authentifiée - Tests d'intégration	1h30
US-MASK-04 : Dashboard Audit Data Steward - Interface consultation décisions - Filtres par utilisateur/attribut/date - Visualisation scores T' - Export rapports d'audit	- Dashboard Data Steward fonctionnel - Requêtes analytics optimisées - Templates rapports PDF	2h00
PHASE 4 : Recherche & Machine Learning		
Configuration Recherche Atlas - Indexation Solr optimisée - Interface de recherche avancée - Facettes par classification/tags - Auto-complétion métadonnées	- Moteur de recherche opérationnel - Interface utilisateur intuitive - Documentation utilisateur	2h00
US-MASK-06 : Apprentissage patterns - Collecte dataset accès validés - Feature engineering (R, P, S, H) - Entraînement modèle initial - Métriques : précision, rappel, F1	- Dataset d'entraînement structuré - Modèle baseline entraîné - Rapport métriques performance	2h00
Préparation Architecture V1 (NN) - Design sous-réseaux par facteur - Architecture PyTorch/TensorFlow - Fonction de perte $L = \alpha L_{privacy} + (1 - \alpha)L_{utility}$ - Plan de migration V0.1 \rightarrow V1	- Schéma architecture réseau neuronal - Code squelette PyTorch - Document roadmap V1	1h30

1.4 19/01/2026 - Jour 4 : Tests & Livraison

Tâche	Livrables	Durée
PHASE 4 (suite) : Validation & Tests		
Tests d'intégration Ethimask-Atlas - Masquage automatique lors catalogage - Propagation tags via lignage - Cohérence audit Ethimask-Atlas - Tests de charge (100+ req/s)	- Suite tests d'intégration complète - Rapport benchmarks performance - Identification goulots d'étranglement	2h00
Tests scénarios métier - Accès médecin vs chercheur - Escalade privilèges détectée - Audit trail complet - Conformité RGPD validée	- Cas de tests métier documentés - Matrices de traçabilité - Rapport conformité RGPD	1h30
Documentation technique - Architecture détaillée (diagrammes) - Guide déploiement - API documentation (Swagger) - Justifications mathématiques	- Documentation technique complète - README déploiement - Spécification API REST	2h00
Documentation utilisateur - Guide Admin (configuration poids) - Guide Data Steward (audit) - Guide utilisateur final - Tutoriels vidéo (scripts)	- Manuels utilisateur par rôle - FAQ et troubleshooting - Scripts tutoriels	1h30
PHASE 5 : Livraison finale		
Validation & Préparation démo - Tests de non-régression complets - Vérification tous les livrables - Scénarios de démonstration - Support présentation	- Rapport tests de validation - Checklist complétée - Script de démonstration	2h00
Livraison projet (avant 11h00) - Archive complète du projet - Documentation finale - Code source versionné - Démonstration live	- LIVRAISON PROJET - Archive ZIP/Git tag - Documentation PDF - Démo réussie	1h00

2 User Stories - Mapping avec le planning

User Story	Description	Date d'implémentation
US-MASK-01	Calcul automatique niveau de masquage approprié	18/01/2026
US-MASK-02	Accès aux données avec masquage adapté au rôle	18/01/2026
US-MASK-03	Configuration des poids du perceptron (Admin)	17/01/2026
US-MASK-04	Audit des décisions de masquage (Data Steward)	18/01/2026
US-MASK-05	Application chiffrement homomorphique niveau 1	17/01/2026
US-MASK-06	Apprentissage patterns d'accès	18/01/2026

3 Cas d'usage Apache Atlas - Planning

Cas d'usage	Implémentation	Date
Catalogage	Enregistrement datasets et métadonnées	16/01/2026
Lignage	Traçage transformations (pipeline Airflow)	17/01/2026
Classification	Application tags PII/SPI automatiques	16/01/2026
Recherche	Découverte données par Data Stewards	18/01/2026
Audit	Logging modifications métadonnées	17/01/2026