

# Projet M2 BDIA : Données météorologiques et climatiques pour le grand public, sensibilisation au changement climatique

Le projet est calibré pour un groupe de 4 à 6 étudiants.

Encadrants : E. LECLERCQ, A. GILLET

Météo France et d'autres acteurs de la recherche publique mettent à disposition des données météo issues de leurs modèles ainsi que des données climatiques. L'objectif de ce projet est de permettre à des utilisateurs sans connaissances particulières d'avoir accès à ces données, de présenter des indicateurs montrant l'évolution du climat au cours des années mais aussi d'extraire des données.

Le projet représente 150h de travail effectif pour chaque étudiant d'un groupe de 4 à 6. Il consiste à réaliser une application qui met en avant les possibilités offertes par l'exploitation des données de sorties des modèles de Météo France ainsi que les données ouvertes sur le climat. Les principales fonctionnalités sont :

- collecter des données depuis les API Web de Météo France en fonction des besoins utilisateur (par exemple un lieu qu'il a sélectionné) ;
- établir des statistiques générales pour un lieu, une région sur les données météo ;
- calculer des indicateurs statistiques sur données climatiques et les associer aux données météo à différents niveaux de granularité (pays, région, etc.) ;
- visualiser l'évolution des différents indicateurs ;
- fournir une interface permettant de poser des requêtes en langage naturel et d'exploiter un LLM pour générer des requêtes SQL sur la base de données créée ;
- extraire des jeux de données pour une exploitation simple par les usagers.

La collecte des données météo, l'intégration des données météo et climatiques ainsi que leur stockage constituent le socle du projet et de l'application à réaliser. Les données sont récupérables en ligne en utilisant les API Web de Météo France (<https://portail-api.meteofrance.fr/web/fr/api/AROME>) et en téléchargeant les jeux de données depuis <https://www.data.gouv.fr/fr/organizations/meteo-france/> (avec plus de 80 jeux de données disponibles) ou plus généralement <https://www.data.gouv.fr>

## Gestion des données

Dans un premier temps, les données devront être insérées dans une base de données relationnelle qui servira d'entrepôt. Son schéma doit permettre de supporter les différentes fonctionnalités de l'application.

Pour cela, les fichiers bruts devront être étudiés afin d'identifier les différents attributs utiles afin de concevoir un schéma adapté. Un programme de chargement de données devra être développé, il devra être réutilisable et adaptable à différentes sources.

Dans la partie gestion de données, il faudra bien distinguer les données issues de modèles des données issues de l'observation et assurer leur traçabilité.

## Consultation des statistiques générales

Un utilisateur qui possède un compte sur l'application doit pouvoir s'authentifier et consulter des statistiques générales, voire les stations météo proches d'un lieu, consulter des données sur chaque station, visualiser l'évolution.

La problématique principale induite par le volume des données est l'organisation d'une structure de données pour permettre de calculer ou recalculer rapidement les indicateurs, en effet les API Météo France fournissent leurs données régulièrement. Une problématique secondaire concerne la scalabilité c'est-à-dire la capacité du système à absorber les nouvelles données sans compromettre la durée d'exécution du calcul des indicateurs.

En outre, une spécificité de ce projet concerne la partie utilisateur qui ne modifie pas les données mais nécessite des interrogations rapides pour une restitution par l'interface. Il conviendra donc de faire une typologie des requêtes à exécuter pour satisfaire les besoins, d'étudier leur durée d'exécution et de déterminer puis tester une architecture technique et logicielle capable de fournir des données rapidement à la partie *front*.

## Travail à réaliser sur la première partie d'octobre à décembre

Le travail se concentrera principalement sur le stockage, la consultation des données et le calcul d'indicateurs. Il s'agit de construire le socle technique de l'application qui sera utilisé ensuite pour les autres fonctionnalités du projet. Pour cela, les principales tâches à réaliser sont les suivantes :

- réaliser les spécifications fonctionnelles ;
- concevoir le schéma de stockage des données, nécessitant une phase d'étude des données disponibles ainsi que des données nécessaires pour l'application ;
- développer un composant ETL (*Extract Transform Load*) pour alimenter l'entrepôt de données. Ce module doit pouvoir être réutilisé facilement pour l'ajout de nouvelles sources de données ;
- spécifier l'architecture technique, les différentes technologies (intergiciels, langages et frameworks) et valider les choix par des expérimentations ;
- développer la partie serveur, qui permettra aux utilisateurs d'accéder aux différentes fonctionnalités de l'application via une interface Web.

## Travail à réaliser sur la seconde partie de janvier à mars

La seconde partie se concentrera sur les fonctionnalités utilisateur avec :

- la réalisation de Web services pour les grandes fonctionnalités ;
- la réalisation de l'interface utilisateur (Web, Mobile ou autre) ;
- l'expérimentation et la réalisation de l'interface d'interrogation en langage naturel.

Les fonctionnalités listées précédemment ne sont pas séquentielles. Vous devrez découper le projet en tâches, définir les rôles de chacun, etc.

**Contrainte :** tous les développements doivent se faire avec des logiciels libres ou *open source*.

## Liens et documentation utiles

- API Météo France : <https://portail-api.meteofrance.fr/web/fr/>
- Exemple d'explorateur de données : <https://www.data.gouv.fr/fr/reuses/explorateur-de-donnees-climatiques/>
- Sur la génération SQL avec d'un LLM : <https://github.com/Sinaptik-AI/pandas-ai>

## Organisation du travail autour du projet

Le projet a pour objectif de faire appliquer conjointement les concepts, les connaissances et les compétences acquises dans les différents cours du Master. Il se concentre sur :

- la mise en œuvre des techniques ;
- l'apprentissage du travail en groupe, la manière dont le travail se répartit lorsque différentes personnes occupent différents rôles ;
- la réalisation d'un **travail collaboratif** ;
- l'application des principes et des bonnes pratiques du génie logiciel.

## Organisation du groupe

Les membres du groupes doivent choisir un ou plusieurs rôles. Tous les membres auront forcément le rôle développeur.

**Responsable de groupe** : l'étudiant assurant ce rôle est en charge de coordonner le groupe, et est également le point de contact avec les encadrants du projet.

**Devops** : l'étudiant devops sera responsable de l'installation des différents logiciels/paquets sur les VM mises à disposition. Il devra également assurer la cohérence des différents choix techniques.

**Référent données** : l'étudiant assurant ce rôle est chargé de s'assurer de la cohérence des données de bout en bout de l'application.

**Référent développeur** : le référent développeur est en charge d'assurer la cohérence globale des différents développements réalisés pour chaque fonctionnalité.

**Référent qualité** : chargé de garantir la cohérence des tests de validation par rapport aux fonctionnalités définies.

**Développeur** : tous les étudiants ont également le rôle de développeur et doivent participer activement à la réalisation des fonctionnalités.

D'autres rôles seront à assurer pour chaque réunion. Ces rôles peuvent être assurés par différents étudiants pour chaque réunion.

**Animateur de réunion** : doit veiller à la bonne répartition de la parole et au suivi de l'ordre du jour (défini préalablement par tous, envoyé par email aux membres du groupe et gardé comme trace dans le wiki du projet).

**Gardien du temps** : doit s'assurer que le temps passé pour chaque partie de la réunion est en accord avec ce qui était prévu.

**Secrétaire** : chargé de réaliser le compte-rendu de la réunion.

## Livrables

### Rapport 1 : Définition des rôles, élaboration des spécifications techniques et fonctionnelles

Dans la première partie du projet, le groupe doit organiser les rôles de chaque étudiant sachant qu'un étudiant peut avoir plusieurs rôles. Par exemple un administrateur de BD, un administrateur système, un référent technique, un responsable du projet, des développeurs. Les fonctionnalités qui seront implémentées doivent faire l'objet d'un rapport, qui indiquera la manière dont les fonctionnalités sont vue par le groupe, quelles seront les personnes concernées pour chaque fonctionnalité en fonction de leur rôle (qui participera activement au développement, qui devra être consulté, quelles sont les dépendances vers les autres fonctionnalités), ainsi que les technologies envisagées pour implémenter la fonctionnalité (elles auront donc été préalablement testées et validées par le groupe).

Ce rapport ne sera pas noté mais fera l'objet d'une évaluation formative afin de rectifier au besoin la compréhension du groupe sur le fond du projet et ses fonctionnalités (l'encadrant fera un retour au groupe). Il permet aussi de guider les étudiants sur la mise en forme correspond aux attendus de l'évaluation finale.

### Contenu du livrable (rapport 1) à fournir pour le 15 novembre :

- organisation de l'équipe et description des rôles ;
- synthèse des fonctionnalités envisagées pour la première phase ;
- choix techniques associés à chaque fonctionnalité.

## Réalisation d'un Proof of Concept

À l'issue de la partie préparatoire, le groupe réalisera un Proof of Concept à partir de ce qu'il aura défini dans le document précédent. Les spécifications techniques de chaque fonctionnalité devront être données, ainsi que les divergences (et leur justification) entre ce qui était prévu et ce qui a été réalisé.

L'organisation du groupe ainsi que les interactions entre les étudiants devront être décrites, tout comme la répartition du travail global et sur chaque fonctionnalité.

**Contenu du livrable (rapport 2 + code en archive annexe) à fournir pour 10 janvier :**

- spécifications techniques de chaque fonctionnalité ;
- justification des choix réalisés ;
- répartition effective du travail.

**Rapport final et présentation**

Ce livrable est attendu en mars à la fin du projet. Un application complètement fonctionnelle remplacera le Proof of Concept. Les consignes pour les rapport et la présentation seront envoyées en décembre.

**Outils pour le travail collaboratif**

Divers outils sont mis à disposition afin que disposiez de l'environnement technique nécessaire pour mener à bien le projet. Tous les outils sont disponibles depuis les salles machines ou depuis l'extérieur de l'université avec une connexion au VPN IEM.

**Machines virtuelles Proxmox**

Le groupe disposera de 2 machines virtuelles Debian sur le serveur Proxmox :

- une VM de développement, afin de tester les développements réalisés par le groupe dans un environnement commun ;
- une VM de production, servant à héberger une version du projet validée sur la VM de développement.

Le groupe disposera des droits root sur les VM, et peut donc réaliser les installations nécessaires à la réalisation du projet, sous la responsabilité de l'étudiant ayant le rôle devops. Le Proxmox est accessible à l'adresse <https://proxmox.iem:8006>, à partir de laquelle il est possible de démarrer/arrêter les VM. Les identifiants de connexion au serveur Proxmox sont les mêmes que pour les machines des salles IEM.

L'accès aux VM peut se faire via ssh, x2go (la partie client pouvant s'installer sur la plupart des distributions <https://wiki.x2go.org/doku.php/doc:installation:x2goclient>) ou via la console de l'interface Proxmox. L'accès via ssh et x2go est à privilégier.

**GitLab**

Le GitLab <https://gitlabvigan.iem> sert de dépôt de code mais aussi de plateforme pour centraliser la documentation et les compte-rendus de réunions grâce au wiki intégré.

Il est configuré pour supporter les *pipelines* CI/CD, et peut donc réaliser automatiquement certaines actions lors de la mise à jour du dépôt.

**Openproject**

Openproject sert à organiser le travail de groupe. Il permet de définir les tâches, les affecter aux membres du groupes, et suivre l'avancée des tâches. Il est disponible à l'adresse <https://op.iem>.