

21/11/2025

DATALAKE CLIENT & RISQUES NATURELS (NOAA + USGS)

Améto Cornelia ADANTO
MIA - IPSSI MONTPELLIER

Année-Scolaire : 2025-2026

SOMMAIRE

- Introduction
- 1- Sources de données
- 2- Architecture du DataLake
- 3- Pipeline d'ingestion
- 4- Traitement et Normalisation (SILVER)
- 5- Détection d'anomalies
- 6- Tendance à long terme
- 7- Analyse géographique
- 8- Dashboard
- Conclusion

Introduction

Ce projet consiste à construire un DataLake capable d'ingérer, stocker, transformer et analyser des données climatiques et de risques naturels provenant de deux sources principales : NOAA (météo, climat) et USGS (tremblements de terre, inondations, feux). L'objectif est de détecter des anomalies climatiques, d'étudier les tendances à long terme et de réaliser une analyse géographique des phénomènes.

1. Sources de données

NOAA

- Données météorologiques : température, vent, précipitations.
- Données climatiques historiques.
- Format : CSV ou JSON.
- Accès : <https://www.ncei.noaa.gov/access>

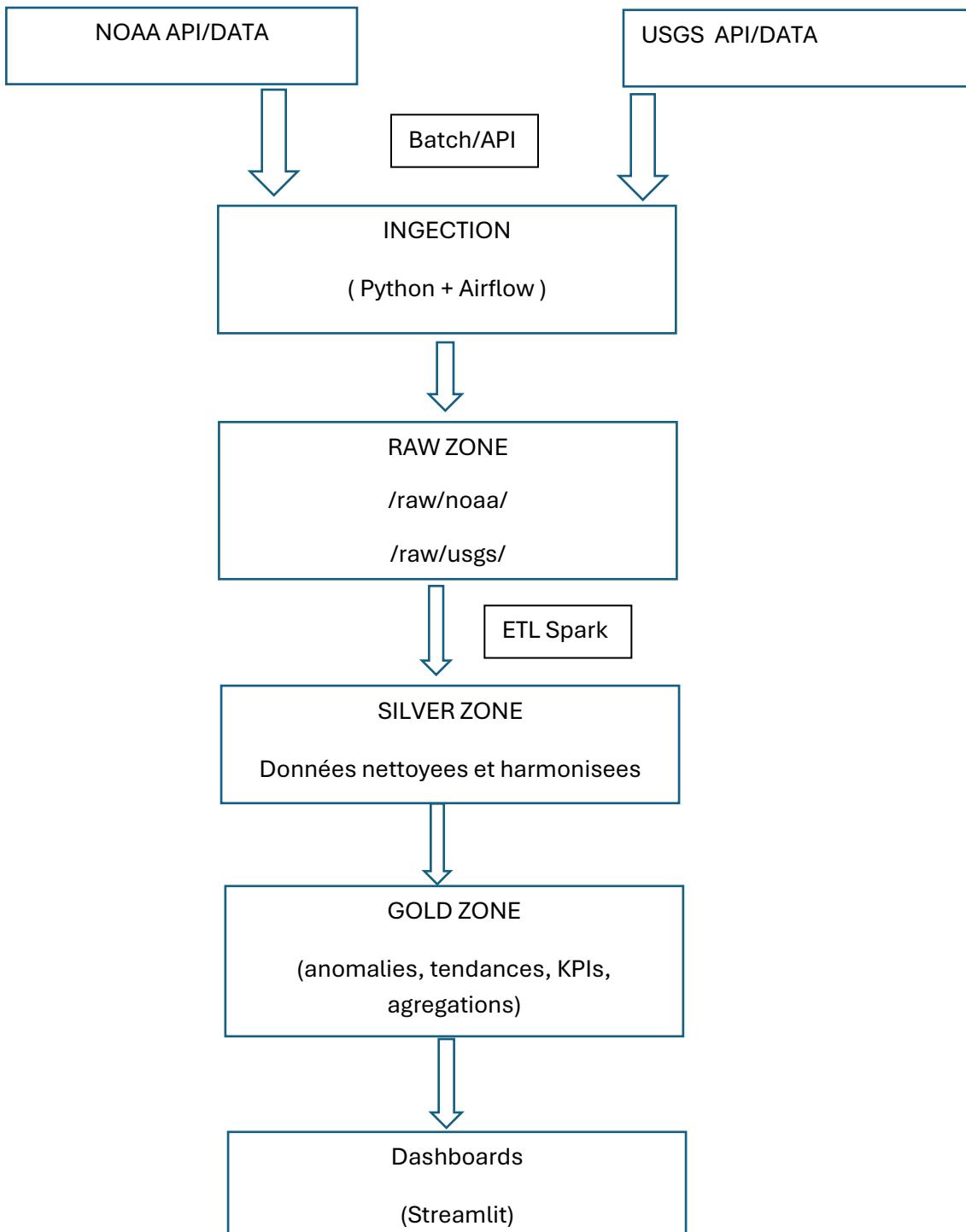
USGS

- Tremblements de terre (flux temps réel).
- Données sur les feux et inondations.
- API : <https://earthquake.usgs.gov/earthquakes/feed/v1.0/>

2. Architecture du DataLake

Le DataLake est organisé en trois zones : RAW, SILVER, GOLD.

- RAW : stockage brut des fichiers NOAA et USGS.
- SILVER : données nettoyées et normalisées.
- GOLD : résultats analytiques (anomalies, tendances, cartes, KPIs).



3. Pipeline d'ingestion

- Ingestion NOAA (Batch) : téléchargement des fichiers météorologiques par période et dépôt dans la zone RAW.

Ingestion des Données

Données NOAA

Télécharger données NOAA

Données USGS

Télécharger données sismiques

364 enregistrements NOAA téléchargés							
	STATION	DATE	LATITUDE	LONGITUDE	ELEVATION	NAME	TE
0	1,001,099,999	2023-01-01	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	1
1	1,001,099,999	2023-01-02	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	2
2	1,001,099,999	2023-01-03	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
3	1,001,099,999	2023-01-04	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
4	1,001,099,999	2023-01-05	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
5	1,001,099,999	2023-01-06	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
6	1,001,099,999	2023-01-07	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
7	1,001,099,999	2023-01-08	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
8	1,001,099,999	2023-01-09	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3
9	1,001,099,999	2023-01-10	70.9333	-8.6667	9	JAN MAYEN NOR NAVY, NO	3

- Ingestion USGS (API) : appel de l'API toutes les X minutes et stockage des résultats dans la zone RAW.

Ingestion des Données

Données NOAA

Télécharger données NOAA

Données USGS

Télécharger données sismiques

229 séismes récents collectés

	magnitude	place	time	longitude	latitude	depth
0	2.8	8 km W of Templeton, CA	2025-11-20 23:54:58	-120.7985	35.5617	5.41
1	2.6488	4 km SSW of Knik-Fairview, Alaska	2025-11-20 23:37:23	-149.6437	61.4778	55.5395
2	0.72	2 km NW of The Geysers, CA	2025-11-20 23:36:09	-122.772	38.7855	2.51
3	1.31	7 km NW of The Geysers, CA	2025-11-20 23:01:56	-122.8075	38.8248	2.52
4	1.5	44 km S of Sandy Valley, Nevada	2025-11-20 23:01:27	-115.5602	35.4225	7.0615
5	2.2359	83 km SSE of Igigig, Alaska	2025-11-20 22:55:18	-155.0822	58.7021	135.9178
6	2.8	6 km WSW of Westbrook, Texas	2025-11-20 22:53:51	-101.082	32.337	7.0557
7	1.33	2 km SW of Anderson Springs, CA	2025-11-20 22:52:24	-122.7117	38.7588	1.32
8	2.2575	45 km S of Glacier View, Alaska	2025-11-20 22:42:44	-147.6191	61.3977	28.0875
9	2.2751	55 km N of Petersville, Alaska	2025-11-20 22:39:11	-150.8106	62.9946	118.9382

4. Traitement et Normalisation (SILVER)

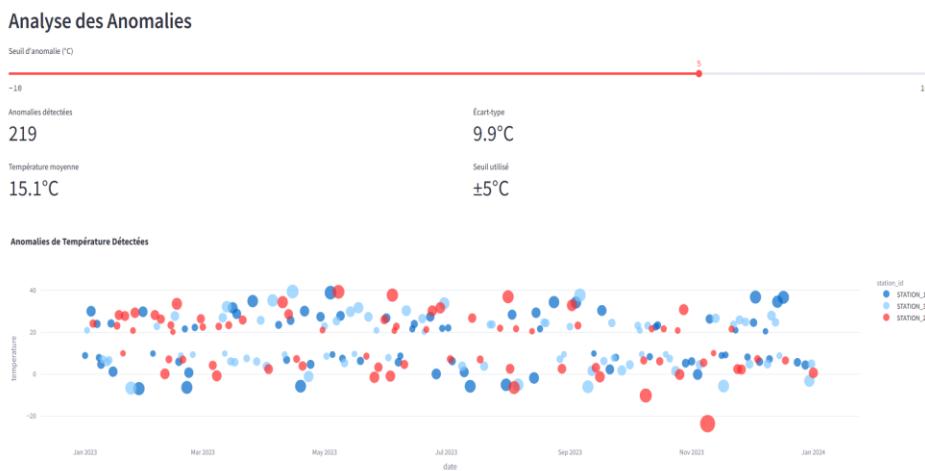
Les données subissent un processus de nettoyage :

- Suppression des valeurs manquantes.
- Harmonisation des colonnes.
- Conversion des unités (ex : $^{\circ}\text{F} \rightarrow ^{\circ}\text{C}$).
- Jointure des différents fichiers sur la date et la localisation.

5. Détection d'anomalies

Les principales anomalies calculées sont :

- Température extrêmement haute : $\text{temp_today} > \text{mean} + 2 \times \text{std}$,
- Température extrêmement basse : $\text{temp_today} < \text{mean} - 2 \times \text{std}$,
- Variation brutale de température ($\geq 10^{\circ}\text{C}$ en 24h),
- Vent extrêmement fort : $\text{wind_today} > \text{mean} + 2 \times \text{std}$,
- Rafale anormale : $\text{gust} > \text{historical_max}$,
- Température incohérente pour la latitude,
- Phénomène extrême dans une zone historiquement stable.



6. Tendances à long terme

Analyse statistique sur plusieurs années :

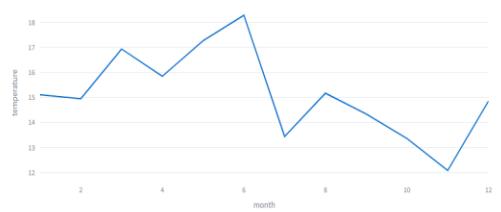
- Évolution des températures moyennes.
- Evolution des précipitations mensuelles.
- Variation de la vitesse moyenne du vent.
- Détection des épisodes extrêmes récurrents.

Tendances

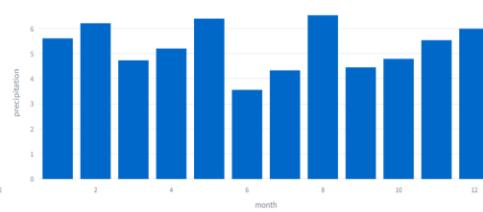
📊 Analyse statistique sur plusieurs années :

- Évolution des températures moyennes
- Évolution des précipitations mensuelles
- Variation de la vitesse moyenne du vent
- Détection des épisodes extrêmes récurrents

Tendance Saisonnière des Températures



Évolution des Précipitations Mensuelles



Analyse des tendances à long terme

Variation de la Vitesse Moyenne du Vent



Détection des Épisodes Extrêmes

Événements extrêmes détectés

54

Seuil température extrême

31.7°C

Seuil vent extrême

35.5 km/h

date	temperature	wind_speed	precipitation	region
13	2023-01-14 00:00:00	24.1511	35.8826	0.6192
17	2023-01-18 00:00:00	28.1556	14.8345	17.47
25	2023-01-26 00:00:00	29.2631	17.5865	29.965
30	2023-01-31 00:00:00	18.151	36.3565	7.2518
35	2023-02-05 00:00:00	28.1698	33.085	26.1908
42	2023-02-12 00:00:00	7.034	36.6846	4.8824
46	2023-02-16 00:00:00	33.5168	34.8364	1.8074
49	2023-02-19 00:00:00	6.8961	36.166	1.9359
61	2023-03-03 00:00:00	10.6366	17.8333	17.6389
71	2023-03-13 00:00:00	32.1358	12.9122	3.4729

7. Analyse géographique

L'analyse géographique inclut :

- Cartographie des anomalies climatiques.
- Analyse des zones à forte activité sismique.
- Comparaison des régions selon latitude, température et risques.
- Détection des zones historiquement stables ayant des événements extrêmes.

Carte Géographique des Séismes



8. Dashboard (Streamlit)

Un tableau de bord est réalisé via Streamlit pour :

- Visualiser les anomalies.
- Afficher les tendances climatiques.

DataLake Climat & Risques Naturels

NOAA (Météo & Climat) + USGS (Risques Naturels)

[Dashboard](#) [Ingestion](#) [Analyse Anomalies](#) [Tendances](#)

Tableau de Bord Principal

Température Moyenne
15.1°C

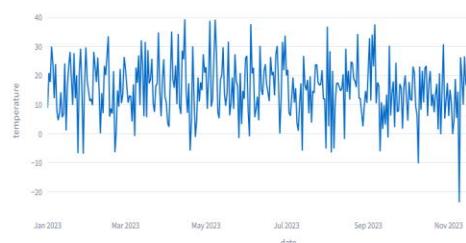
Vent Max
51.1 km/h

Séismes
500

Stations
3

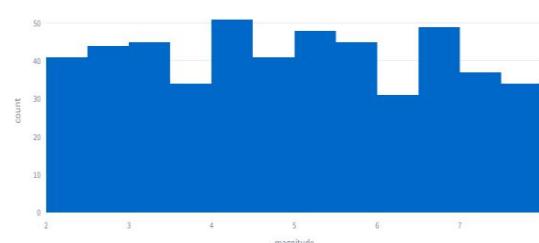
Évolution des Températures

Températures Quotidiennes



Distribution des Séismes

Magnitude des Séismes



- Projeter les données USGS sur une carte.
- Explorer les KPIs (températures moyennes, vents, épisodes extrêmes).

Conclusion

Ce projet permet de construire un DataLake complet, capable de centraliser et analyser des données météorologiques et géophysiques. Il offre une vision détaillée des risques naturels et de l'évolution du climat dans une approche Big Data professionnelle.