

# Mini-Projet: Traitement de données météorologiques massives NOAA

## 1. Introduction

Le but de ce projet est de trouver la température maximale (minimale) enregistrée pour chaque année des données NCDC pour une station donnée.

Les données d'entrée de ce projet sont des données météorologiques enregistrées dans des fichiers séparés (compressés) pour chaque année. Lesdites données sont collectées par le National Climatic Data Center - NCDC (qui dépend de la National Oceanic and Atmospheric Administration NOAA) à partir de capteurs météorologiques du monde entier. Les données météorologiques pour chaque station et par année sont disponibles sur le lien <https://www1.ncdc.noaa.gov/pub/data/noaa>. La liste des stations est disponible sur le fichier <https://www1.ncdc.noaa.gov/pub/data/noaa/isd-history.txt>

chaque station dispose d'un identifiant (code USAF), par exemple celui du port IBN BATOUTA est 601010. Toutes les stations n'ont pas des relevés depuis **1901**, il y en avait très peu gérées à cette époque. Par exemple, les données du port IBN Batouta n'ont commencé à être collectées qu'en **1957**. (voir [isd-history.txt](#))

## 2. Description des fichiers

les fichiers sont généralement nommés station-99999-année.gz. chaque fichier compressé contient des entrées qui ressemblent à ceci :

```
0148601010999992018010100004+35733-005900FM-  
12+002199999V0203401N00461004801CN006000199+01561+01251103471ADDGA1051+  
004801081GE19AGL
```

**601010** : représente l'identifiant de la station météo de l'USAF

**20180101** : représente la date de l'observation

**01561** : représente la température de l'air en degrés Celsius multipliée par dix. Ainsi, la lecture de 01561 équivaut à 15,6 degrés Celsius.

## 3. Travail à faire

1. Chaque étudiant choisira une station marocaine de son choix
2. La première étape consiste à télécharger tous les relevés météo de votre station (un script sera une bonne idée)
3. Placer les fichiers dans HDFS à travers hadoop-master
4. Ecrire les classes MeteoMap, MeteoReduce et MeteoMain
5. Construire un fichier jar correspondant
6. Exécuter le jar sur hadoop-master

7. Rapporter les résultats retrouvés sous forme d'un rapport en utilisant diverse représentations graphiques adéquates. N'oublier pas l'analyse et l'interprétation des données

### **Partie 2 :**

- Stocker les données dans une base de données nosql/fichier plat (csv)
- Utiliser PowerBI en le connectant à la base de données.
  - Créer un rapport statistique descriptive de la station année/mois/saison
  - Créer un dashboard pour apporter des réponse et faire des comparaison entre les années, etc.