# Prérequis outil :

Installer pyspark sur sa machine locale

Vous pourriez vous aider du lien suivant :

<https://sparkbyexamples.com/pyspark/install-pyspark-in-anaconda-jupyter-notebook/>

Après avoir installé le notebook avec PySpark, veuillez ensuite lire l’exercice ci-dessous et développer un notebook en PySpark permettant de fournir les résultats attendus. Bon courage à vous !

# Prérequis métier :

* Un routeur peut avoir plusieurs interfaces
* Un routeur peut être connecté à un autre routeur via une interface composée de plusieurs liens, on appelle ce type de lien un « lag » (Link Agrégation Group)
* La capacité d’un port lag est la somme des capacités des liens unitaires
* Pour les liens lag le nom du lag est le même entre l’équipement et son client

# Exercice :

**Les documents à fournir à l’issu de l’exercice sont :**

* **Notebook du script**
* **Fichiers csv topologie**
* **Fichier csv résultat du trafic**

## Fichiers

Etant donnée les sets de données suivants :

### Fichier 1 (Trafic) :

Fichier contenant les données permettant d’avoir le trafic aperçu sur une interface d’un équipement (routeur) par tranche de 5min

* **Fichier 2 (Topologie) :**

Fichier contenant les connexions entre les différents équipements et via quelles interfaces

* **Fichier 3 (Donnees\_geographique) :**

Fichier contenant l’ensemble des informations géographique des équipements réseau

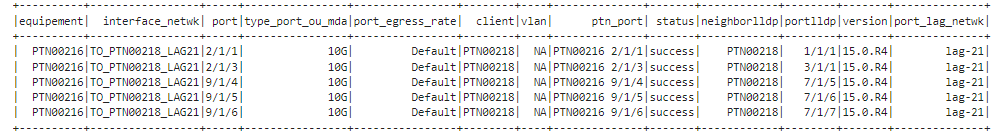
* **Fichier 4 (coûts ISIS) :**

Fichier contenant l’ensemble des coûts isis (permet d’avoir un routage unique entre deux équipements)

## Partie 1 : génération de topologie

### Préambule :

* Les fichiers en annexe contiennent plusieurs types d’équipements : PTN, CSG, FHT… On ne s’intéressera dans ce cas d’étude qu’aux liens entre les équipements PTN.
* Un routeur peut être connecté à un autre routeur via deux types de liens :
* Lien unitaire : se fait entre deux routeurs via des liens de capacite maximale 10G et via des ports du type (1/1/1)
* Lien « lag » : regroupement de plusieurs liens unitaires, ces liens sont peuvent s’appeler par exemple lag-24.

Exemple :   
ci-dessus le lag-21 entre les deux équipements PTN00216 et PTN00218 est constitué de 5 liens unitaires et peuvent être vus sur les colonnes port et portlldp (l’information du lag est contenue dans la colonne port\_lag\_netwk).

### **Travail demandé :**

A partir du fichier 2, fichier 3 et fichier 4 il faudra calculer le fichier de topologie qui permet d’avoir l’ensemble des lags avec tous les ports équipement et client. Pour les cas où le lien est constitué d’un lien unitaire garder le nom du port équipement.

### Filtres :

Voici l’ensemble des filtres à appliquer aux données pour avoir que l’ensemble des équipements cibles :

* + Fichier 2 :
    - Garder que les liens qui relient deux équipements « PTN »
    - Supprimer les lignes avec admin\_state\_port=down
    - Supprimer les lignes avec vlan différent de NA
    - Supprimer les lignes avec type\_port\_ou\_mda différent de 10G
  + Ne garder lors des croisements que les données du fichier 2, le fichier 3 et le fichier 4 viennent compléter les infos du fichier 2.

### Opérations :

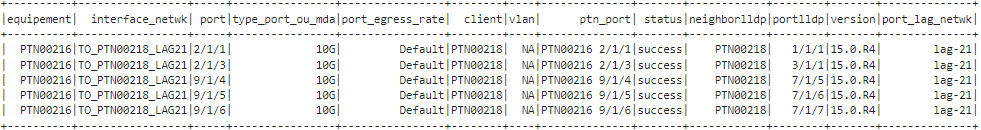
* + Création de la colonne id\_lien = equipement + ‘\_’ + client
  + Création de la colonne port et port\_client qui sont la liste de tous les ports unitaires qui constitue le lag
  + Les colonnes rôle, configuration sont à joindre deux fois pour l’équipement et son client. Mettre le suffixe \_A quand il s’agit de l’équipement et \_B quand il s’agit du client.

### Exemple :

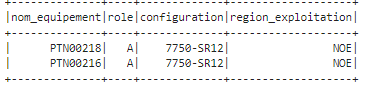
Ci-dessous un exemple de résultat du lien qui sépare les deux équipements PTN00216 et PTN00218

A partir des données relatives aux deux équipements

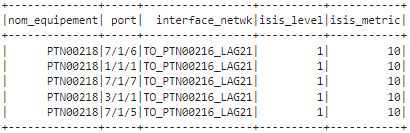
Fichier 2 :



Fichier 3 :



Fichier 4 :



On crée le résultat final pour le lien PTN00216\_PTN00218 (il faut noter que le id\_lien PTN00216\_PTN00218 est le même pour les deux équipements PTN00216 & PTN00218. Et pas de lien PTN00218\_PTN00216)

|  |  |  |
| --- | --- | --- |
| **id\_lien** | PTN00216\_PTN00218 | PTN00216\_PTN00218 |
| **equipement** | PTN00216 | PTN00218 |
| **client** | PTN00218 | PTN00216 |
| **interface\_client** | lag-21 | lag-21 |
| **capacite\_lien** | 50 | 50 |
| **port** | [2/1/1, 2/1/3, 9/1/4, 9/1/5, 9/1/6] | [1/1/1, 3/1/1, 7/1/5, 7/1/6, 7/1/7] |
| **port\_client** | [1/1/1, 3/1/1, 7/1/5, 7/1/6, 7/1/7] | [2/1/1, 2/1/3, 9/1/4, 9/1/5, 9/1/6] |
| **interface\_equipement** | lag-21 | lag-21 |
| **role\_A** | A | A |
| **configuration\_A** | 7750-SR12 | 7750-SR12 |
| **region\_exploitation** | NOE | NOE |
| **role\_B** | A | A |
| **configuration\_B** | 7750-SR12 | 7750-SR12 |
| **isis\_level** | 1 | 1 |
| **isis\_metric** | 10 | 10 |

*NB : l’exemple ci-dessus est transposé pour la facilité de lecture*

Le fichier résultat de cette étape est à fournir dans le résultat final.

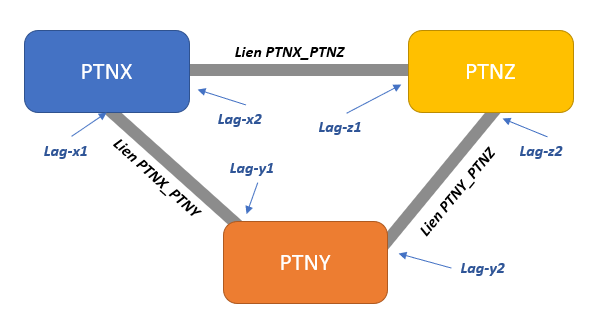
## Partie 2 : Calcul du taux de charge des liens

### Descriptions des données :

Dans le fichier1 des données de trafic on trouve le trafic maximum par heure qui transite pour chaque équipement PTN sur chacun de ses ports.

NB : Les ports de type lag sont aussi présents comme un port à part entière et leurs valeurs présentent la somme de trafic de tous les ports qui les constituent.

### Travail demandé :



Etant donné l’architecture ci-dessus qui inclus :

* 3 équipements PTN X,Y & Z, chacun a deux lag x & y et les 6 ports sont interconnectés entre eux.
* On disposerait dans ce cas-là du trafic (in et out) des 6 lags pour chacun des 3 équipements
* **Le but dans cet exercice est d’identifier le trafic maximum aperçu sur chacun des liens durant toute la journée.**
* Pour identifier le trafic maximum d’un lien il faut effectuer les opérations suivantes :
  + Identifier le maximum du trafic sur cet équipement entre le trafic rentrant et sortant (in & out), on appellera cette variable trafic\_max
  + Identifier le maximum du trafic absolu sur la journée.
  + Identifier le maximum du trafic aperçu entre les deux extrémités du lien.
* Par exemple pour avoir le trafic maximum du lien PTNX\_PTNY il faut regarder le trafic max de (PTNX\_lag-x1, PTNY\_lag-y1) et ainsi de suite.
* Il faudra ensuite calculer la variable TDC.   
   TDC = trafic\_max/capacite\_lien \* 100
* Il faudra ensuite créer une variable avec le nom interval\_charge avec les modalités suivantes :

|  |  |
| --- | --- |
| TDC | Interval\_charge |
| <=70 | Inf 70% |
| <=90 | Sup 70% Inf 90% |
| <100 | Sup 90% |
| >=100 | Erreur |

### Consignes :

* L’unité des variables trafic\_in et trafic\_out est en mégabyte (Mb) alors que l’unité de la variable capacite\_lien etn gigabyte (Gb)
* L’unité des variables trafic\_in, trafic\_out & trafic\_max sont à mettre en Gb dans le résultat final.

### Exemple fichier résultat :

ci-dessus un exemple du résultat attendu sur le lien PTN00216\_PTN00218

|  |  |
| --- | --- |
| **id\_lien** | PTN00216\_PTN00218 |
| **trafic\_max** | 7,03 |
| **trafic\_in** | 7,03 |
| **trafic\_out** | 0,87 |
| **TDC** | 14 |
| **intervalle\_charge** | inf 70% |
| **equipement** | PTN00216 |
| **interface\_equipement** | lag-21 |
| **role\_A** | A |
| **configuration\_A** | 7750-SR12 |
| **client** | PTN00218 |
| **interface\_client** | lag-21 |
| **role\_B** | A |
| **configuration\_B** | 7750-SR12 |
| **capacite\_lien** | 50 |
| **isis\_level** | 1 |
| **isis\_metric** | 10 |
| **region\_exploitation** | NOE |
| **date** | 2022-08-01 |

**Bon courage 😊**