

```

/data/
  .snapshot/*
  frwiki/
    raw/
      frwiki-20200201-stub-meta-history.xml
      frwiki-20200201-pagelinks.sql

    frwiki-20200201/
      master/
        pageshistory.avsc
        pagelink.avsc
        full/
          frwiki-20200201-pagelinks-*.avro
          frwiki-20200201-stub-meta-history-*.avro
        test/
          *.avro

```

/data/ : dossier principal qui contient tous nos données

/data/.snapshot/s20200720-163848.488/* : contient le snapshot de tous nos données

/data/frwiki/ : contient tout ce qui provient de Wikipedia

/data/frwiki/raw/ : contient tous nos données brutes

/data/frwiki/raw/frwiki-20200201-pagelinks.sql : fichier brute concernant les liens entre les pages (11.04 GB)

/data/frwiki/raw/frwiki-20200201-stub-meta-history.xml : fichier brute concernant toutes les informations et l'historique des pages (70.72 GB)

/data/frwiki/frwiki-20200201/ : contient tout ce qui provient des données brutes du 1er février 2020

/data/frwiki/frwiki-20200201/master/ : contient tous les jeux de données provenant des données brutes wikipédia du 1er février 2020

/data/frwiki/frwiki-20200201/master/pagelink.avsc : schéma avro utilisé pour sérialiser les liens des pages

/data/frwiki/frwiki-20200201/master/pageshistory.avsc : schéma avro utilisé pour sérialiser l'historique et les informations concernant les pages

/data/frwiki/frwiki-20200201/master/full/ : contient tous les données sérialisé (complet)

/data/frwiki/frwiki-20200201/master/full/frwiki-20200201-pagelinks-*.avro: pages links sérialisés

/data/frwiki/frwiki-20200201/master/full/frwiki-20200201-stub-meta-history-*.avro : history sérialisé

/data/frwiki/frwiki-20200201/master/test/ : contient des échantillons de jeux de données sérialisés ==> pour faire des tests de traitement ou d'observer des de nos données