Covid-19 Cases on World map (Batch processing)

Youtube Ref Link:

[2 Beginner Level Projects For Aspiring Data Engineers ⌛️Covid-19 Data | Real Time Data Processing 👈 - YouTube](https://www.youtube.com/watch?v=GmR2migiu6A)

**Steps:**

1. Covid-19 Cases on World map (Batch processing)

* Batch pipeline where you will read 1 year of Covid-19 data (confirmed, recovered, and deaths) for each country date wise, process the data and aggregate it, store processed data in MySQL/NoSQL database, visualize data on world map.

1. Process to follow:

* Download file from git repo and save it in your local.
* Setup Hadoop in local, here you need to focus on only two main components of Hadoop (HDFS and YARN)
* HDFS is distributed file system so now copy that file from local to HDFS.
* Setup spark in your local.
* Write a spark application which reads file from HDFS and creates a DataFrame.
* Aggregate data in spark application.
* How many total cases so far?
* How many recovered cases so far?
* How many deaths so far?
* Write aggregated data in output table using transactional database like MySQL/Postgres or a NoSQL database like Cassandra/HBase. Prefer MySQL for starter or beginner.
* Use any dashboarding tool like tableau, Power-BI, Grafana, Kibana. Recommended Grafana.
* 1. Setup Grafana in local
* 2. Install plugin for world map
* 3. Connect Grafana to MySQL
* Write query to visualize aggregated data stored in MySQL table on world map.