



## Big Data Aplicat. Caso d'us Hive. Desplaçaments de terra.

### BigData Aplicat

#### Cas d'us Hive

##### Desplaçaments de tierra.

- Disposem d'un Dataset de desplaçaments de terra descarregat de la web de la NASA.
- Ens connectem des del nodo2 con beeline (o si no funcionarà directament des de el nodo1). Carreguem una nova taula amb els desplaçaments del fitxer csv.

```
create table deslizamientos (  
  id bigint, fecha string ,  
  hora string ,  
  country string , nearest_places string ,  
  hazard_type string , landslide_type string ,  
  motivo string , storm_name string , fatalities bigint ,  
  injuries string , source_name string , source_link string ,  
  location_description string , location_accuracy string ,  
  landslide_size string , photos_link string ,  
  cat_src string , cat_id bigint ,  
  
  countryname string , near string , distance double , adminname1 string , adminname2 string ,  
  population bigint , countrycode string , continentcode string , key string , version string ,  
  tstamp string , changeset_id string , latitude double , longitude double , geolocation string )  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
```

- Carreguem en una nova taula “países” el fitxer “csv”. Prèviament haurem copiat aquest fitxer “deslizamientos.csv” a la carpeta temporal a “/tmp”.

```
load data local inpath '/tmp/deslizamientos.csv' into table deslizamientos;
```

- Cal analitzar les dades i a més, entregar el següent:
  - Descriure la taula per entendre que conté
  - Construir una consulta Hive SQL per (entregar la consulta i el resultat):
    - contar quants desplaçaments hi ha
    - En què país ha hagut mes de 100 víctimes (fatalities)
    - Nombre de desplaçaments per tipus.
    - Quins son els 10 primers països que tenen mes moviments registrats.



### Big Data Aplicat. Caso d'us Hive. Desplaçaments de terra.

- Crearem la següent taula de països a partir del fitxer "países.csv" que tenim que deixar emmagatzemat en el directori tmp (igual que en el anterior pas).

```
create table paises ( nombre string, cod string) ROW FORMAT DELIMITED FIELDS  
TERMINATED BY ',';
```

```
load data local inpath '/tmp/countries.csv' into table paises;
```

- Cal analitzar les dades i a més, entregar el següent:
  - Construir una consulta Hive SQL per (entregar la consulta i el resultat):  
Quantitat de desplaçaments i motiu per països (nom i codi)
- Guardem el resultat de l'anterior consulta en un fitxer per un futur poder tractar-lo amb ferramentes externes de reporting (p.e. Excel, PowerBI...). Ho guardarà segurament com un fitxer 00000\_0

```
insert overwrite local directory '/tmp/datos' row format delimited fields terminated  
by ',' select ...;
```