

Practica Cassandra

A continuación, se enuncian los puntos a resolver con su respectiva explicación documentación:

- 1. Consumos por origen: se almacenarán consumos (individuales). Los informes se harán mensuales (por cada mes) para cada país y para cada país y director. Debe tenerse en cuenta que estos informes pueden venir limitados por un rango de porcentaje de visualización ('viewing pct').**

Descripción solución:

En mongo se obtiene el país del contrato y se hace un unwind de Movies para obtener la fecha y porcentaje de visualización, así como el nombre del director. Antes de hacer copy en cassandra se debió formatear la fecha debido a que salía el siguiente error:

Failed to import 158 rows: ParseError - Failed to parse 08/01/2016 : time data '08/01/2016' does not match format '%Y-%m-%d', given up without retries

Y esto se debe a que no hay control sobre los formatos de fecha en la instrucción COPY y por tanto se tendría que parchear para agregar la opción. referencia: <https://issues.apache.org/jira/browse/CASSANDRA-8970>

En mongo se crea una colección de preparación para exportar los datos:

```
db.cPractica.aggregate([
  {$project:{_id:0}},
  {$unwind:"$Movies"},
  {$project:{"contract.country":1,
    "viewDate": {$concat: [{ $substr: ["$Movies.Date", 6, 4] }, "-", { $substr: ["$Movies.Date", 3, 2] }, "-", { $substr: ["$Movies.Date", 0, 2] } ]}},
    "Movies.Viewing PCT":1, "Movies.Details.Director.Name":1 }},
  {$out:"cassandraPrep1"}
]);
```

En shell se ejecuta una exportación de datos de mongo:

```
mongoexport --db practicaMongo --
collection cassandraPrep1 /type csv /fields:"_id,contract.country,viewDate,Movies.Viewing PCT,Movies.Details.Director.Name" /out:"cassandraPrep1CSV.txt"
```

resultado en \anexos\cassandraPrep1CSV.txt

Diseño cassandra:

En Cassandra se crea un keyspace con tipo de replicación simple

```
CREATE KEYSPACE myKeyspace WITH replication = {'class': 'SimpleStrategy','replication_factor' : 2};
use myKeyspace;
```

Se crea la tabla que alojara los datos exportados de mongo:

```
create table consumosPorOrigen(  
  id text,  
  contractCountry text,  
  viewDate timestamp,  
  viewingPCT smallint,  
  directorName text,  
  PRIMARY KEY (id)  
);
```

Importación de datos a cassandra:

```
COPY consumosPorOrigen (id,contractCountry,viewDate,viewingPCT,directorName) FROM 'D:\mongo3.4.23\bin\cassandraPrep1CSV.txt' WITH HEADER = 'true';
```

Consulta cassandra:

Ejemplo para el mes de enero con un porcentaje mínimo de visualización del 99%

```
select * from consumosPorOrigen where viewingPCT > 99 and viewDate >'2016-01-27' and viewDate <'2016-02-01' ALLOW FILTERING;
```

- 2. Resumen facturación: para cada cliente (DNI, full name) se almacenarán sus facturas mensuales, con los totales de #taps, tráfico (minutos), y monto de la factura. La consulta principal para esta tabla es el informe anual (totales por año y cliente).**

Descripción solución:

En mongo se obtiene los datos del cliente y la fecha de cargo de la factura y se hace un unwind de Movies para obtener la sumatoria del tráfico de minutos. Antes de hacer copy en cassandra se debió formatear la fecha debido a que salía el siguiente error:

Failed to import 158 rows: ParseError - Failed to parse 08/01/2016 : time data '08/01/2016' does not match format '%Y-%m-%d', given up without retries

Y esto se debe a que no hay control sobre los formatos de fecha en la instrucción COPY y por tanto se tendría que parchear para agregar la opción. referencia: <https://issues.apache.org/jira/browse/CASSANDRA-8970>

En mongo se crea una colección de preparación para exportar los datos:

```
db.cPractica.aggregate([  
  {$project:{_id:0}},  
  {$unwind:"$Movies"},  
  {$project:{"TOTAL":1,  
    "fechaFactura":  
    {$concat: ["20",{ $substr: ["$charge date", 6, 4] }, "- ", { $substr: ["$charge date", 3, 2] }, "- ", { $substr: ["$charge date",0,2] } ]}},  
    "Client.DNI":1,"Client.Name":1,"Client.Surname":1,  
    "traficoMinutos": { $sum : { $multiply: ["$Movies.Details.Duration" ,{$divide:["$Movies.Viewing PCT",100]}}} }  }},  
  {$out:"cassandraPrep2"}  
]);
```

En shell se ejecuta una exportación de datos de mongo:

```
mongoexport --db practicaMongo --  
collection cassandraPrep2 /type csv /fields:"_id,TOTAL,fechaFactura,Client.DNI,Client.Name,Client.Surname,traficoMinutos" /out:"cassandraPrep2CSV.txt"
```

resultado en \anexos\cassandraPrep2CSV.txt

Diseño tabla:

```
create table resumenFacturacion(  
  id text,  
  valorFactura double,  
  chargeDate timestamp,  
  clientDNI text,  
  clientName text,  
  clientSurname text,  
  traficoMinutos double,  
  PRIMARY KEY (id)  
);
```

Importación de datos a cassandra:

```
COPY resumenFacturacion (id,valorFactura,chargeDate,clientDNI,clientName,clientSurname,traficoMinutos) FROM 'D:\mongo3.4.23\bin\cassandraPrep2CSV.txt' WITH HEADER = 'true';
```

Consulta cassandra:

```
select * from resumenFacturacion;
```