Guardar la pregunta o la resposta

Datasets virtuals, `CREATE TABLE ... AS` i una API

Noms _____

Superset, base de dades `Dades classe`, esquema `public`, taules `pokemon` i `generations`. L'analogia d'avui és una cuina: les taules són els ingredients. Per parelles: un tecleja, l'altre escriu aquí; canvieu a A2.

Referència

	Recepta	Carmanyola
Guardeu	La pregunta (el SQL).	La resposta (les files).
A Superset	Un **dataset virtual**: una consulta desada.	Una taula nova, feta amb `CREATE TABLE ... AS`.
On viu	A Superset: només el text de la consulta.	A PostgreSQL: una taula real, al costat de `pokemon`.
Quan l'useu	Es torna a cuinar cada vegada.	Se serveix tal com es va guardar.

Un **dataset** és allò a partir del qual Superset pot construir gràfics. Un **dataset físic** apunta a una taula real: una etiqueta en un pot, no una còpia.

Part A · Guardar la pregunta, guardar la resposta (20 min)

A1 · La recepta. SQL → SQL Lab. Escriviu i **Run** (és el JOIN de dimecres):

```sql
SELECT p.id, p.pokemon, p.type_1, p.type_2, p.hp, p.attack, p.defense, p.speed,
       g.generation_id, g.region, g.release_year
FROM pokemon p
JOIN generations g ON p.generation_id = g.generation_id
```

Files _____ `pokemon` en té 949. Per què no 949? _____

Fletxa al costat de **Save** → **Save dataset** → **Save as new**, nom `pokemon_with_region` → **Save & Explore**. El creador de gràfics s'obre en una pestanya nova del navegador: tanqueu-la i obriu **Datasets**.

Tipus de `pokemon_with_region` _____ Tipus de `pokemon` i `generations` _____

A2 · La carmanyola. Pestanya nova de SQL Lab. Primer executeu només el `SELECT` (des de la línia 2). Després afegiu la línia 1 i executeu-ho tot.

```sql
CREATE TABLE pokemon_by_generation AS
SELECT g.generation_id, g.region, g.release_year,
       COUNT(p.id) AS n_pokemon,
       ROUND(AVG(p.attack)) AS avg_attack,
       MAX(p.speed) AS max_speed
FROM generations g
LEFT JOIN pokemon p ON p.generation_id = g.generation_id
GROUP BY g.generation_id, g.region, g.release_year;
```

Només el `SELECT`: files _____ Amb la línia 1: què mostra SQL Lab? _____

On han anat les files? _____

Ara executeu `SELECT * FROM pokemon_by_generation ORDER BY generation_id;` Files _____

Més Pokémon: _____ amb _____ `n_pokemon = 0`: _____ El seu `avg_attack` diu _____

Per què aquesta cel·la no és 0? _____

A3 · Etiqueteu el pot. Superset encara no té cap dataset per a la taula nova. **Datasets → + Dataset**: Database `Dades classe`, Schema `public`, Table `pokemon_by_generation`. Fletxa al costat de **Create and explore dataset** → **Create dataset**.

Tipus de `pokemon_by_generation` a **Datasets** _____

Quin es queda desfasat? (amb el professor, 5 min)

El professor afegeix una novena regió (Paldea, 2022) a `generations` al projector. No ho executeu vosaltres: les vostres taules s'han de quedar com estan. Escriviu la vostra predicció **abans** que s'executi.

Quantes regions mostrarà?	Predicció	Ha mostrat
La **taula** `pokemon_by_generation` (la carmanyola)		
El `SELECT` d'A2 executat de nou com a **consulta** (la recepta)		

Per què? _____

Part B · D'una API a un dataset (16 min)

Una **API** és una adreça a la qual pregunteu i que respon amb dades, no amb una pàgina web. Respon en **JSON**: text fet de noms i valors, on un valor pot contenir més valors. Superset no pot preguntar a una API: **extraieu** i després **carregueu**.

B1 Obriu `https://pokeapi.co/api/v2/pokemon?limit=20` al navegador. Sense clau, sense inici de sessió.

`count` diu _____ Pokémon mostrats _____ Quin camp conté l'adreça de la resta? _____

B2 Obriu `https://pokeapi.co/api/v2/pokemon/charizard`. A SQL Lab executeu `SELECT * FROM pokemon WHERE pokemon = 'charizard'`.

Alçada: API _____ taula _____ Pes: API _____ taula _____ Per què? _____

Com guarda cadascuna els dos tipus? _____

B3 **Extreure, aplanar, carregar. Aplanar** és convertir un JSON imbricat en files i columnes. Descarregueu `api_to_csv.py` de la pàgina Recursos del lloc web del curs. En un terminal, en aquella carpeta: `python3 api_to_csv.py`. Obriu el `pokeapi_sample.csv` que escriu. Després, a Superset: **+ → Data → Upload CSV to database**, base de dades `Dades classe`, esquema `public`, nom de taula `pokeapi_sample` → **Upload**. A SQL Lab: `SELECT type_1, COUNT(*) AS n FROM pokeapi_sample GROUP BY type_1 ORDER BY n DESC`

CSV: files _____ columnes _____ Els dos primers tipus: _____ Tipus a **Datasets** _____

Demà surt un Pokémon nou. Com el feu arribar a aquesta taula? _____

B4 **La llista de comprovació d'una API.** Sis preguntes abans de construir-hi res. Ompliu el que heu vist.

Pregunta	Què cal esbrinar	PokéAPI
Adreça	La URL i els seus **paràmetres** després del `?`.	
Clau	Cal un compte o una **clau d'API**? Una clau és una contrasenya: no l'enganxeu mai enlloc.	
Límits	Quantes peticions podeu fer?	
Paginació	Una resposta ho conté tot? Quin camp és el següent?	
Llicència	Podeu reutilitzar i publicar les dades? Qui heu de citar?	
Instantània	La taula és una còpia d'un moment. Apunteu-ne la data.	

Heu acabat quan Tots els espais estan plens · **Datasets** llista un **Virtual** i almenys tres **Physical**. **Parella ràpida?** Repetiu A1 i A2 amb el vostre conjunt de dades obertes.

Encallats?

- **`relation "pokemon_by_generation" already exists`**: l'heu executada dues vegades. Bona notícia: la taula hi és.
- **`does not allow for DDL/DML`**: no és culpa vostra, aviseu el professor. **`Table already exists` en carregar el CSV**: poseu **If table already exists** a **Replace**. **`pokeapi.co` no s'obre**: demaneu al professor el `pokeapi_sample.csv` ja fet i carregueu-lo.