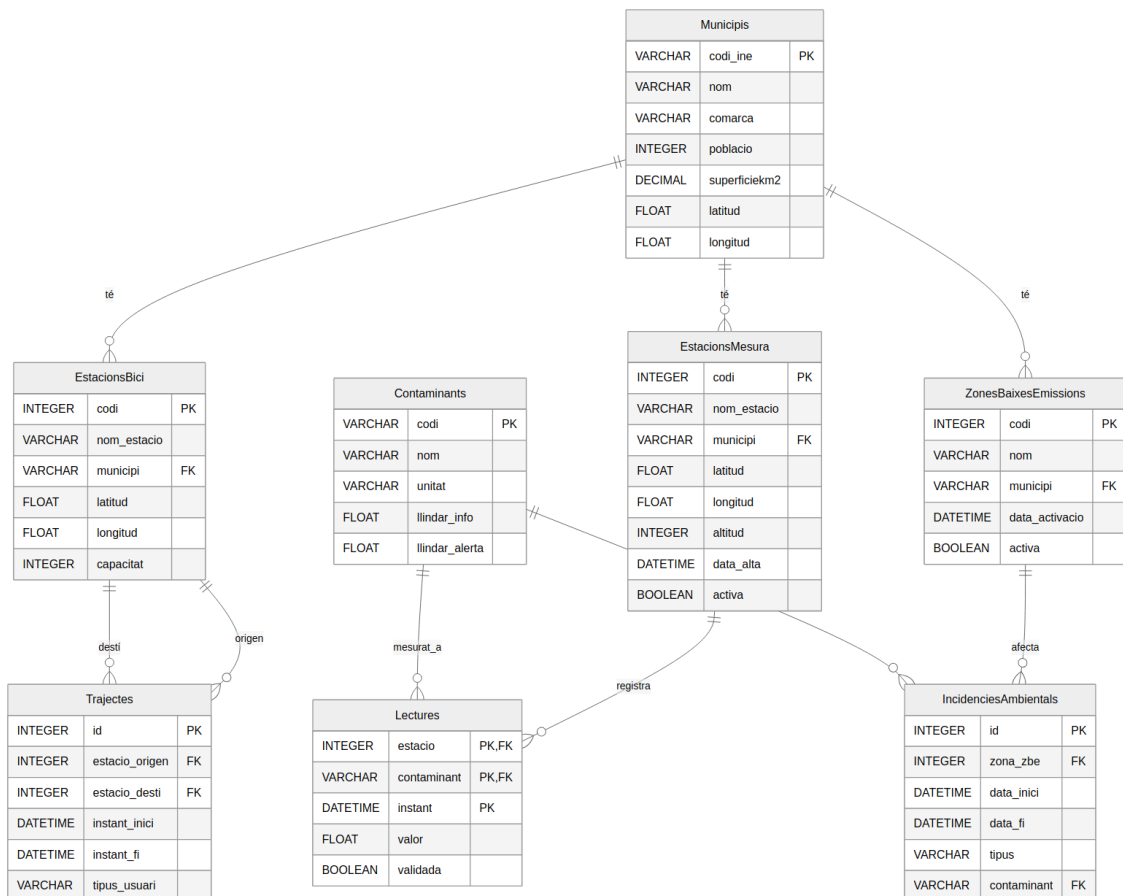


Convocatòria R-02/26-E Tècnic/a superior informàtica - perfil especialitzat en analítica de dades

Durada de la prova: 2 hores

Suposeu un sistema fictici per a la gestió de la qualitat de l'aire i la mobilitat sostenible a l'àrea metropolitana de Barcelona. El sistema manté informació sobre estacions de mesura ambiental, els sensors que operen, les lectures de contaminants i les estacions de bicicleta compartida amb els seus trajectes.

A continuació es presenta l'esquema de la base de dades transaccional, així com les sentències SQL de creació de les taules. Podeu assumir que els instants (moments en el temps) es codifiquen com a valors numèrics *timestamp* i que les coordenades geogràfiques es codifiquen en graus decimals (EPSG:4326, WGS84).



Àrea de Serveis Generals i Transició Digital

Direcció dels Serveis de Recursos Humans

Secció d'Innovació i Projectes Jurídics en Reclutament i Captació del Talent

<pre>CREATE TABLE Municipis (codi_ine VARCHAR(10), nom VARCHAR(255), comarca VARCHAR(255), poblacio INTEGER, superficiekm2 DECIMAL(10,2), latitud FLOAT, longitud FLOAT, PRIMARY KEY (codi_ine)); CREATE TABLE EstacionsMesura (codi INTEGER, nom_estacio VARCHAR(255), municipi VARCHAR(10), latitud FLOAT, longitud FLOAT, altitud INTEGER, data_alta DATETIME, activa BOOLEAN, PRIMARY KEY (codi), FOREIGN KEY (municipi) REFERENCES Municipis(codi_ine)); CREATE TABLE Contaminants (codi VARCHAR(10), nom VARCHAR(255), unitat VARCHAR(50), llindar_info FLOAT, llindar_alerta FLOAT, PRIMARY KEY (codi)); CREATE TABLE Lectures (estacio INTEGER, contaminant VARCHAR(10), instant DATETIME, valor FLOAT, validada BOOLEAN, PRIMARY KEY (estacio, contaminant, instant), FOREIGN KEY (estacio) REFERENCES EstacionsMesura(codi), FOREIGN KEY (contaminant) REFERENCES Contaminants(codi));</pre>	<pre>CREATE TABLE EstacionsBici (codi INTEGER, nom_estacio VARCHAR(255), municipi VARCHAR(10), latitud FLOAT, longitud FLOAT, capacitat INTEGER, PRIMARY KEY (codi), FOREIGN KEY (municipi) REFERENCES Municipis(codi_ine)); CREATE TABLE Trajectes (id INTEGER, estacio_origen INTEGER, estacio_desti INTEGER, instant_inici DATETIME, instant_fi DATETIME, tipus_usuari VARCHAR(50), -- 'abonat', 'ocasional' PRIMARY KEY (id), FOREIGN KEY (estacio_origen) REFERENCES EstacionsBici(codi), FOREIGN KEY (estacio_desti) REFERENCES EstacionsBici(codi)); CREATE TABLE ZonesBaixesEmissions (codi INTEGER, nom VARCHAR(255), municipi VARCHAR(10), data_activacio DATETIME, activa BOOLEAN, PRIMARY KEY (codi), FOREIGN KEY (municipi) REFERENCES Municipis(codi_ine)); CREATE TABLE IncidenciesAmbientals (id INTEGER, zona_zbe INTEGER, data_inici DATETIME, data_fi DATETIME, tipus VARCHAR(100), -- 'avis', 'alerta', 'emergència' contaminant VARCHAR(10), PRIMARY KEY (id), FOREIGN KEY (zona_zbe) REFERENCES ZonesBaixesEmissions(codi), FOREIGN KEY (contaminant) REFERENCES Contaminants(codi));</pre>
---	--

Des de l'Àrea Metropolitana de Barcelona (AMB) es planteja la implementació d'un sistema de suport a la presa de decisions que doni suport a l'anàlisi **descriptiva** i **predictiva**. Per això es considera la ingestió, processament i integració d'informació de les següents fonts de dades externes.

Nota: Les fonts de dades, URLs i exemples proporcionats en aquesta secció són ficticis a efectes de la prova, però recreen l'estructura i el comportament de fonts de dades obertes reals.

Capes geoespacial de municipis i barris

- Descripció: fitxer amb les delimitacions administratives dels municipis de l'AMB i els barris de Barcelona, amb geometries poligonals i atributs demogràfics bàsics

Àrea de Serveis Generals i Transició Digital

Direcció dels Serveis de Recursos Humans

Secció d'Innovació i Projectes Jurídics en Reclutament i Captació del Talent

- Format: GeoJSON (EPSG:4326) i Shapefile (EPSG:25831, UTM)
- Freqüència d'actualització: anual

Prediccions meteorològiques de Catalunya

- Descripció: API REST amb predicció meteorològica municipal horària a 72 hores, incloent-hi temperatura, humitat, velocitat del vent i radiació solar
- Freqüència d'actualització: diària
- Format: JSON
- Exemple:

```
{
  "codiMunicipi": "080193",
  "nomMunicipi": "Barcelona",
  "prediccions": [
    {
      "data": "2026-04-07T08:00:00",
      "temperatura": 16.5,
      "humitat": 72,
      "velocitatVent": 12.3,
      "direccioVent": "SW",
      "radiacioSolar": 450.2,
      "precipitacio": 0.0
    },
    { ... }
  ]
}
```

Intensitat de trànsit a l'AMB

- Descripció: Dades horàries del nombre de vehicles detectats per espies magnètiques en els principals eixos viaris de l'àrea metropolitana
- Freqüència d'actualització: horària
- Format: CSV
- Exemple:

id_tram	data_hora	latitud	longitud	intensitat	velocitat_mitjana
T001	2026-04-07T08:00	41.3925	2.1404	1852	28.5
T001	2026-04-07T09:00	41.3925	2.1404	2340	22.1
T002	2026-04-07T08:00	41.4012	2.1587	987	45.3

L'objectiu d'aquesta prova és el desenvolupament d'un sistema **genèric, dinàmic i escalable**. A continuació es presenta una petita descripció de cada una d'aquestes característiques.

Genèric: S'ha de poder adaptar a tipologies variades de fonts de dades (per exemple: formats, orígens, heterogeneïtat sintàctica o semàntica, dades vectorials i ràster).

Dinàmic: S'ha de poder adaptar a canvis al llarg del temps en les dades actuals (per exemple: canvis en l'esquema, incorporació de nous sensors o contaminants), i ha de poder ingerir informació de noves fonts de dades.

Escalable: Ha de poder processar un gran volum de dades en un entorn distribuït, considerant l'alta freqüència d'ingestió de dades de sensors (*streaming*).

Exercici 1 (7 punts)

El primer exercici té com a objectiu el disseny del model analític i la definició dels indicadors clau per a l'anàlisi descriptiva. Es plantegen els següents indicadors:

- Índex de qualitat de l'aire (IQA) per municipi i franja horària, representat sobre mapa.
- Evolució temporal de la correlació entre intensitat de trànsit i nivells de NO₂ per zona.
- Comparativa de l'ús de bicicleta compartida (nombre de trajectes i durada mitjana) en funció de les condicions meteorològiques.

1.1) (4 punts)

Proposa l'esquema conceptual del *data warehouse* o *lakehouse* que ha de donar suport a la implementació dels indicadors descrits. Inclou les taules de fets i dimensions, indicant claus, mesures i jerarquies. Justifica el tipus d'esquema triat (estrella, floc de neu o constel·lació).

Resposta esperada: aproximadament mitja pàgina. Pots utilitzar un diagrama esquemàtic o una llista estructurada.

1.2) (3 punts)

Per a cada un dels 3 indicadors anteriors, indica:

- Quin tipus de visualització o gràfic utilitzaries i per què (l'IQA per municipi ha d'estar representat sobre un mapa).
- Quins filtres interactius proposes per fer el quadre de comandament interactiu (mínim 3 filtres globals).

Resposta esperada: una taula o llista breu. No cal esbós gràfic.

Exercici 2 (7 punts)

En aquest exercici es planteja la definició de l'arquitectura del sistema des de la ingestió de dades al *data lake* fins a les capes d'explotació. **L'arquitectura ha de preveure tant el processament *batch* com el processament en temps real (*near-real-time*) de les dades de sensors.**

2.1) (5 punts)

Proposa l'arquitectura del sistema completant la taula següent. Per a cada component que identifiquis, indica el seu rol, el format físic (si escau) de les dades que gestiona, la tecnologia concreta proposada i una justificació. L'arquitectura ha d'identificar clarament les diferents capes (ingestió, emmagatzematge, processament, servei) i seguir un patró reconegut (Medallion/Lakehouse, Lambda, Kappa o equivalent), que has d'indicar abans de la taula.

Component	Rol/Capa	Format físic	Tecnologia	Justificació breu

Si necessites més files, pots afegir-ne. S'esperen entre 5 i 8 components com a referència.

2.2) (2 punts)

A partir de la teva proposta tecnològica de l'apartat 2.1, discuteix els pros i contres de l'enfocament triat respecte a:

- Ús de serveis gestionats al núvol (BigQuery, Synapse, Databricks...) versus solucions autogestionades o *on-premise*.

- Cost, governança, escalabilitat i interoperabilitat.
- Reutilització de la informació al sector públic: com encaixaria la publicació de dades obertes derivades dins de l'arquitectura proposada.

Resposta esperada: aproximadament mitja pàgina.

Exercici 3 (6 punts)

L'últim exercici té com a objectiu plantejar els processos de DataOps i MLOps per al sistema.

3.1) (4 punts)

Describeu els processos d'ETL/ELT a implementar per carregar les dades de les fonts origen als models destí. Presenta un diagrama a alt nivell on es descriguin les operacions a dur a terme, diferenciant entre fluxos *batch* i fluxos *streaming/near-real-time*. Per a cada flux, indica la tecnologia emprada (Airflow, DBT, Spark Structured Streaming, Kafka, etc.) i justifica les decisions.

Adicionalment, es vol implementar un model predictiu que, donades les condicions meteorològiques previstes i la intensitat de trànsit observada, predigui el nivell de qualitat de l'aire per a les properes 24 hores a cada municipi. **Describeu breument:**

- Quin tipus d'algorisme de *machine learning* proposaries (regressió, sèries temporals, *gradient boosting*...) i per què.
- Quines variables (*features*) utilitzaries per entrenar el model.

Resposta esperada: el diagrama pot ser un esquema senzill amb caixes i fletxes; el text de suport ha de ser d'aproximadament mitja pàgina.

3.2) (2 punts)

Describeu breument l'estratègia operativa del sistema en dos àmbits:

(a) DataOps - per als *pipelines* de dades de l'apartat 3.1:

- Com garantiries la qualitat de les dades al llarg del *pipeline* (proves de qualitat, validació d'esquemes, detecció d'anomalies)? Esmenta eines concretes.
- Com automatitzaries el desplegament dels *pipelines* (CI/CD) i quins controls de qualitat aplicaries abans de promocionar canvis a producció?

(b) MLOps - per al model predictiu de qualitat de l'aire:

Àrea de Serveis Generals i Transició Digital

Direcció dels Serveis de Recursos Humans

Secció d'Innovació i Projectes Jurídics en Reclutament i Captació del Talent

- Quina estratègia seguiries per detectar la degradació (*drift*) del model: quins tipus de *drift* consideres rellevants (*data drift*, *concept drift*), quines mètriques monitoraries i amb quina freqüència?
- Quina estratègia de reentrenament proposes (periòdica, basada en *triggers*, contínua) i com desplegaries el nou model en producció (*A/B testing*, *canary*, *blue-green*, etc.)?

Resposta esperada: aproximadament mitja pàgina total per àmbit. No cal entrar en detalls de configuració.