

Basi di Dati 2022-23 – Progetto “Orti Scolastici” (12 CFU)

Dominio applicativo – Specifiche

Si vuole realizzare una base di dati a supporto dell’iniziativa di citizen science rivolta alle scuole “Dalla botanica ai big data”.

L’iniziativa mira a costruire una rete di supporto per le scuole che partecipano a progetti relativi agli orti scolastici. Per ogni scuola si vogliono memorizzare il nome dell’istituto scolastico, il codice meccanografico, la provincia, il ciclo di istruzione (primo o secondo ciclo di istruzione) e se l’istituto beneficia o meno di un finanziamento per partecipare all’iniziativa, in tal caso ne memorizziamo il tipo.

Per ogni scuola c’è almeno una persona di riferimento per l’iniziativa, ma possono essercene diverse. Per ogni persona coinvolta vogliamo memorizzare nome, cognome, indirizzo di email, opzionalmente un contatto telefonico e il ruolo (dirigente, animatore digitale, docente, ...). Nel caso la scuola sia titolare di finanziamento per partecipare all’iniziativa (es. finanziamento per progetto PON EduGreen) si vuole memorizzare se la persona sia il referente e un partecipante al progetto da cui deriva il finanziamento. All’interno della scuola, possono esserci più classi partecipanti all’iniziativa. Per ognuna di esse si vuole memorizzare la classe (es. 4E), l’ordine (es. primaria, secondaria di primo grado) o il tipo di scuola (es. liceo scienze applicate, agrario) e il docente di riferimento per la partecipazione di tale classe.

Ogni scuola ha uno o più orti, identificati da un nome che identifica l’orto all’interno della scuola. Ogni orto può essere in pieno campo o in vaso, ed è caratterizzato da coordinate GPS e una superficie in mq. Si vuole inoltre memorizzare se le condizioni dell’orto lo rendono adatto a fare da controllo per altri istituti (cioè se si trova in un contesto ambientale “pulito” e l’istituto è disposto a collaborare con altri).

Le piante vengono piantate con scopi di biomonitoraggio o fitobonifica. Con biomonitoraggio si intende il monitoraggio dell’inquinamento mediante organismi viventi. Le principali tecniche di biomonitoraggio consistono nell’uso di organismi bioaccumulatori per fornire informazioni sulla situazione ambientale. Fornisce stime sugli effetti combinati di più inquinanti sugli esseri viventi, ha costi di gestione limitati e consente di coprire vaste zone e territori diversificati, consentendo una adeguata mappatura del territorio. Con fitobonifica si intende l’utilizzo delle piante per disinquinare aria, acqua, sedimenti e suoli.

Si considerano un certo numero di specie (vedi allegato 1, da cui si evincono anche le informazioni da memorizzare per ogni specie) per i diversi scopi e per ogni specie vengono utilizzate un certo numero di repliche (cioè esemplari veri e propri delle piante). In particolare, in caso di biomonitoraggio le repliche del gruppo di controllo (“nel pulito”) dovranno essere lo stesso numero di quelle del gruppo per cui vogliamo monitorare lo stress ambientale. Le repliche di controllo potranno essere dislocate in un orto a disposizione dello stesso istituto o in un orto messo a disposizione da altro istituto e andrà mantenuto il collegamento tra gruppo per cui si monitora lo stress ambientale e il corrispondente gruppo di controllo. In particolare, ogni scuola dovrebbe concentrarsi su tre specie e ogni gruppo dovrebbe contenere 20 repliche.

Per ogni specifica pianta messa a dimora, verrà memorizzata la specie, il numero di replica, il gruppo, l’orto, l’esposizione specifica, la data di messa a dimora e la classe che l’ha messa a dimora.

Le rilevazioni (osservazioni) vengono effettuate sulle specifiche piante (repliche) e le informazioni acquisite (in accordo alle schede in Allegato 2) memorizzate con data e ora della rilevazione, data e ora dell’inserimento, responsabile della rilevazione (può essere un individuo o una classe) e responsabile dell’inserimento (se diverso da quello della rilevazione e anche in questo caso può essere un individuo o una classe).

Le informazioni ambientali relative a pH, umidità e temperatura vengono acquisite mediante sensori o schede Arduino (vedi Allegato 3, da cui si possono dedurre le informazioni da monitorare per i diversi tipi di sensore/scheda), si vogliono memorizzare numero e tipo di sensori presenti in ogni orto (e le repliche associate a quel sensore). Le informazioni possono essere

rilevate tramite app e inserite nella base di dati oppure essere trasmesse direttamente da schede Arduino alla base di dati. Si vuole tenere traccia della **modalità di acquisizione** delle informazioni.

Attività da svolgere

Il progetto consiste nelle seguenti attività (per ognuna di esse il materiale da consegnare è dettagliato nella sezione seguente).

I. **Progettazione logica:** Lo scopo di questa attività è la definizione di uno schema logico relativi a una base di dati relazionale per il dominio sopra descritto, ipotizzando un carico di lavoro a partire dalle specifiche sopra fornite (es. inserimento scuole, orti, repliche, misurazioni, analisi dei dati per attività di biomonitoraggio, ecc). Il progetto dovrà essere completo della specifica di tutti i vincoli di integrità. Si dovrà verificare la qualità dello schema relazionale progettato, applicando i principi della teoria della normalizzazione. In particolare, si dovrà stabilire se lo schema è normalizzato rispetto alla forma normale di Boyce Codd o rispetto alla terza forma normale. Se non è normalizzato rispetto alla terza forma normale, proporre una decomposizione senza perdita e che preserva le dipendenze per almeno una relazione non normalizzata e indicare quale forma normale soddisfa.

II. Realizzazione:

Lo scopo di questa attività è la definizione in PostgreSQL dello schema logico, il suo popolamento e la realizzazione di alcune operazioni di interrogazione sulla base di dati, nonché di alcune routine e trigger.

A. La definizione dello schema logico della base di dati progettato dovrà contenere la specifica di tutti i vincoli individuati in fase di progettazione. Per ogni vincolo andrà individuata e indicata l'opportuna modalità di implementazione (vincoli di tipo CHECK o trigger). È richiesta l'implementazione dei soli vincoli di tipo CHECK. Per quanto riguarda il popolamento, la base di dati creata dovrà essere popolata con dati sufficienti almeno a verificare che i vincoli di dominio espressi siano verificati e che in generale le operazioni di cui si richiede l'implementazione funzionino correttamente.

Su tale basi di dati si richiede di realizzare:

- B. La definizione di una vista che fornisca alcune informazioni riassuntive per ogni attività di biomonitoraggio: per ogni gruppo e per il corrispondente gruppo di controllo mostrare il numero di piante, la specie, l'orto in cui è posizionato il gruppo e, su base mensile, il valore medio dei parametri ambientali e di crescita delle piante (selezionare almeno tre parametri, quelli che si ritengono più significativi).
- C. Le seguenti interrogazioni
 - a. determinare le scuole che, pur avendo un finanziamento per il progetto, non hanno inserito rilevazioni in questo anno scolastico;
 - b. determinare le specie utilizzate in tutti i comuni in cui ci sono scuole aderenti al progetto;
 - c. determinare per ogni scuola l'individuo/la classe della scuola che ha effettuato più rilevazioni.
- D. Le seguenti procedure/funzioni:
 - a. funzione che realizza l'abbinamento tra gruppo e gruppo di controllo nel caso di operazioni di biomonitoraggio;
 - b. funzione che corrisponde alla seguente query parametrica: data una replica con finalità di fitobonifica e due date, determina i valori medi dei parametri rilevati per tale replica nel periodo compreso tra le due date.
- E. I seguenti trigger:
 - a. verifica del vincolo che ogni scuola dovrebbe concentrarsi su tre specie e ogni gruppo dovrebbe contenere 20 repliche;
 - b. generazione di un messaggio (o inserimento di una informazione di warning in qualche tabella) quando viene rilevato un valore decrescente per un parametro di biomassa.

III: Progettazione fisica, elaborazione delle interrogazioni e controllo dell'accesso

- A. Lo scopo di questa attività è la progettazione fisica della base di dati, la sua creazione in PostgreSQL, un popolamento di dimensioni significative, la verifica dei piani di esecuzione scelti dal sistema per alcune interrogazioni prima e dopo la creazione dello schema fisico. In particolare, si richiede di
- individuare un carico di lavoro composto da tre interrogazioni a vostra scelta, specificandole in SQL e in linguaggio naturale; le interrogazioni devono contenere solo condizioni di selezione e join, almeno una deve includere una selezione con condizione complessa e almeno una deve includere una operazione di join;
 - progettare lo schema fisico a partire dal carico di lavoro individuato, motivando opportunamente, in modo sintetico, le scelte effettuate;
 - definire lo schema fisico progettato in PostgreSQL;
 - per le relazioni coinvolte nel carico di lavoro, inserire un numero di tuple significativo che portino ad occupare, per ciascuna relazione, qualche decina di pagine (in PostgreSQL ogni pagina, per default, ha dimensione 8 kb). Per far ciò, e solo per le relazioni coinvolte nel carico di lavoro, si suggerisce di utilizzare un generatore automatico di dati (ad esempio www.datanamic.com). Per raggiungere più velocemente l'occupazione di spazio disco indicata, si suggerisce di aggiungere solo a queste relazioni un attributo semanticamente inutile (dummy), definito con tipo char(n) o text. Questo permetterà di estendere la dimensione di ciascuna tupla e di arrivare all'occupazione di spazio disco indicata con un numero inferiore di tuple.
 - Individuare ed esaminare il piano di esecuzione scelto dal sistema e il tempo di esecuzione per le interrogazioni nel carico di lavoro **prima e dopo la creazione dello schema fisico**; confrontare i piani e i tempi ottenuti nei due casi, fornendo una giustificazione per i risultati ottenuti.
- B. Si richiede di definire i ruoli **insegnante, gestore globale del progetto, referente di istituto, referente della scuola e studente** (di una delle classi coinvolte nel progetto), individuando una opportuna gerarchia tra di essi. Assegnare ai ruoli sopra descritti i privilegi che ritenete ragionevoli nel dominio applicativo considerato, tenendo in considerazione quanto specificato nella descrizione del dominio, motivando le scelte effettuate. Assegnare i ruoli definiti a 5 utenti a vostra scelta. Specificare in SQL la politica di controllo dell'accesso individuata.

Deliverable (prodotti da consegnare)

Ogni attività al punto precedente dovrà produrre un deliverable (prodotto) di tipo software o di tipo report (documentazione) che illustri il processo di progettazione e le scelte effettuate. In particolare, i prodotti attesi per ogni attività sono i seguenti (S=codice SQL, D=documentazione in pdf):

PARTE I

1. [D] **Requisiti ristrutturati** in modo da eliminare ambiguità.
2. [D] **Progetto concettuale**, articolato in
 - (a) schema ER;
 - (b) documentazione relativa ai domini degli attributi (dizionario dati ed entità);
 - (c) vincoli non esprimibili nel diagramma;
 - (d) specifica dei tipi di gerarchie di generalizzazione.

Per i gruppi che hanno effettuato la peer review va effettuata la consegna completa, includendo anche i punti 1. e 2. finali che devono essere consistenti con i documenti di progettazione relativi ai passi successivi.

3. [D] **Progetto logico**, articolato in
 - (a) schema ER ristrutturato;
 - (b) eventuali modifiche dei domini degli attributi e informazioni sui domini di eventuali attributi introdotti;
 - (c) modifiche all'elenco di vincoli del modello concettuale (nuovi vincoli, eventuali vincoli eliminati o modificati);
 - (d) documentazione relativa alle scelte fatte per eliminare le gerarchie di generalizzazione;
 - (e) schema logico;
 - (f) verifica di qualità dello schema e eventuali ottimizzazioni applicate tenendo in considerazione il carico di lavoro

PARTE II

4. [S] Script SQL per la **creazione dello schema** logico della base di dati in accordo allo schema relazionale ottenuto alla fine della fase di progettazione logica, per la porzione necessaria per i punti successivi (cioè le tabelle coinvolte dalle interrogazioni nel carico di lavoro, nella definizione della vista, nelle interrogazioni, in funzioni, procedure e trigger). Lo schema dovrà essere comprensivo dei vincoli esprimibili con check, e per il **popolamento** di tale base di dati.
5. [D] diagramma che visualizza lo script SQL in forma grafica ottenuto con DataGrip (vedi [Aulabweb](#) per come crearla da 4.).
6. [S] Il codice SQL sviluppato per implementare le **interrogazioni** e la **vista** richieste, inserendo, come commento nello script, la corrispondente richiesta in linguaggio naturale.
7. [S] Il codice PL/pgSQL sviluppato per implementare le **funzioni e procedure** richieste, inserendo, come commento nello script, la specifica relativa a ciascuna funzione o procedura.
8. [S] Il codice PL/pgSQL sviluppato per implementare i **trigger** richiesti, inserendo, come commento nello script, la specifica relativa a ciascun trigger.

PARTE III

9. [D] **Progetto fisico e sua validazione**, articolato in
 - a) Descrizione in linguaggio naturale e il codice SQL sviluppato per implementare le **interrogazioni del carico di lavoro**.
 - b) **progetto fisico**, contenente l'elenco degli indici che si intendono creare per le interrogazioni contenute nel carico di lavoro (specificando relazione di riferimento e chiave di ricerca), il loro tipo (ordinato/hash, clusterizzato/non clusterizzato) e la motivazione che ha portato alla loro creazione.
 - c) Una tabella che, per ogni tabella coinvolta nelle operazioni del carico di lavoro, riporti il numero di **tuple inserite e la dimensione in blocchi** della tabella.
 - d) Descrizione dei **piani di esecuzione** scelti dal sistema (prima e dopo la creazione dello schema fisico) per le interrogazioni contenute nel carico di lavoro, riportando la **visualizzazione del piano di esecuzione** prodotta da PostgreSQL e i tempi di esecuzione (prima e dopo la creazione dello schema fisico); confrontare i piani e i tempi ottenuti nei due casi, fornendo una giustificazione per i risultati ottenuti.
10. [D] Descrizione della politica di **controllo dell'accesso** scelta, motivando le scelte effettuate. Si suggerisce di riassumere la politica di controllo dell'accesso con una tabella contenente una riga per ogni tabella e una colonna per ogni ruolo. Ogni cella (i,j) dovrà contenere i privilegi che si intendono assegnare all'utente j sulla tabella i.
11. [S] Script SQL per la **creazione dello schema fisico** della base di dati e per la specifica delle **interrogazioni** contenute nel **carico di lavoro**, inserendo, come commento nello script, la corrispondente richiesta in linguaggio naturale.
12. [S] Script SQL per l'implementazione della politica di **controllo dell'accesso**.

Modalità di svolgimento e consegna

Svolgimento

Il progetto deve essere svolto **a gruppi**. Nel caso di studenti che hanno aderito al patto d'aula il gruppo è quello **assegnato, per gli altri** la composizione dei gruppi (tre persone) va dichiarata su AulaWeb (sezione Progetto) al momento dell'inizio del lavoro e **al più tardi una settimana prima dell'appello in cui si consegna il progetto**.

Scadenze per la consegna

Sono previste consegne **in corrispondenza di ogni appello di esame**, con consegna da effettuarsi **via AulaWeb entro le ore 9.00 della data dell'esame scritto**.

Modalità di consegna su AulaWeb

Si deve consegnare, utilizzando il compito Consegna progetto su AulaWeb, un singolo file **zip** con nome = GruppoXX con XX numero del gruppo (sarà impostata la consegna di gruppo) contenente la documentazione e il codice richiesti ai punti precedenti, opportunamente organizzata in file.

Il file zip dovrà contenere

- un unico documento pdf per la parte I, organizzato in sezioni, il titolo di ciascuna sezione dovrà corrispondere al deliverable corrispondente
- un documento pdf contenente il diagramma DataGrip dello schema logico (prodotto II.5)
- un file SQL per la parte II, organizzato in sezioni (indicate come commenti nel file), inserendo come commento la richiesta a cui si riferisce il codice SQL (e riportando per ogni interrogazione/vista la specifica come commento). I file SQL devono essere file SQL ... cioè formato solo testo e apribili ed eseguibili in pgAdmin;
- un unico documento pdf per la parte III, organizzato in sezioni, il titolo di ciascuna sezione dovrà corrispondere al deliverable corrispondente (si noti che questo documento, oltre a spiegazioni in linguaggio naturale dovrà anche riportare del codice SQL, come da descrizione deliverable PARTE III);
- un file SQL per la parte III, organizzato in sezioni (indicate come commenti nel file), inserendo come commento la richiesta a cui si riferisce il codice SQL (e riportando per ogni interrogazione/vista la specifica come commento). Includere in un file separato il popolamento di dimensioni significative generato con datanamic per la verifica dei piani di esecuzione. I file SQL devono essere file SQL ... cioè formato solo testo e apribili ed eseguibili in pgAdmin.

Il progetto consegnato deve essere completo (rispetto alle specifiche) e funzionante!

Attenzione: sono permesse consegne ripetute, ma farà fede solo l'ultima e, se sufficiente, verrà valutato un solo progetto (quindi, se avete ambizioni di voto, consegnate solo se siete convinti di aver svolto il progetto in maniera per voi soddisfacente, perché non potrete rifarlo per alzarvi il voto). In caso di progetto con valutazione insufficiente, verrà richiesto di consegnare un nuovo progetto (una sola volta per ogni anno accademico), che dovrà tenere conto dei commenti/correzioni ricevuti sul progetto insufficiente.

Criteri di valutazione

- Correttezza e completezza dei deliverable (per esempio, lo schema concettuale, lo schema ristrutturato e lo schema logico devono essere singolarmente corretti).
- Consistenza complessiva dei deliverable e corretta applicazione delle metodologie (per esempio, lo schema concettuale, lo schema ristrutturato e lo schema logico devono essere tra di loro coerenti).
- Qualità della documentazione in termini di chiarezza e coerenza con le richieste.

Allegato 2



Allegato 2

Legenda: * Si intende la misurazione lato*lato della foglia. | ** Da rilevare a fine sperimentazione o a intervalli regolari. | *** Si intende il peso della chioma completamente essicata.

Rilevamento dei parametri utili nelle specie utilizzate ai fini di fitobonifica				 Competenze Digitali nella Scuola competenzedigitali.unige.it	
Nome Istituto Scolastico:					
Data:					
Specie monitorata:					
Tipo di coltivazione: in vaso / in pieno campo					
Tipologia di substrato: terriccio da rinvaso / suolo pre-esistente					
Cosa sto monitorando: SUOLO / ARIA					
Cosa guardare	Domande guida	Cosa misurare	Misure su piante di controllo (la/e pianta/e messa/e in ambiente di riferimento/pulito)	Misure su piante trattate (la/e pianta/e esposta a stress)	Note
Biomassa e struttura	Rilevate delle alterazioni delle dimensioni della pianta rispetto alle piante usate come controllo? Esempi: piante ad accrescimento ridotto, ridotta dimensione delle foglie, ridotto apparato radicale, alterazioni della struttura (es. ramificazioni), diminuzione del peso.	larghezza chioma/foglie (cm)*			
		lunghezza chioma/foglie (cm)			
		peso fresco chioma/foglie (g)**			
		peso secco chioma/foglie (g)***			
		altezza pianta (cm)			
		lunghezza radice/i (cm)			
Alterazioni di fioritura e fruttificazione	Rilevate delle alterazioni nella produzione e/o nella dimensione di fiori e frutti? Esempi: fiori/frutti tardivi o assenti, più piccoli, in minore quantità rispetto al solito.	peso fresco radici (g)			
		peso secco radici (g)			
		n. fiori			
		n. frutti			
Altri danni evidenti	La pianta risulta danneggiata? Potete ad esempio vedere foglie avvizzite, con bruciature o altri segni evidenti? La pianta è più esposta all'attacco agenti patogeni (funghi, virus, batteri)?	n. di foglie danneggiate			
		% di superficie danneggiata per foglia			
Parametri del suolo	Indicare i dati rilevati in campo	pH			
		umidità			
		temperatura			

Legenda: * Si intende la misurazione lato* lato della foglia. | ** Da rilevare a fine sperimentazione o a intervalli regolari. | *** Si intende il peso della chioma completamente essicata.

Allegato 3

Esempi di sensori:

- https://www.amazon.it/WANFEI-Intelligent-Bluetooth-automaticamente-Temperatura/dp/B08C9Q3656/ref=sr_1_2?keywords=xiaomi+flower&qid=1662329490&sr=8-2
- https://www.amazon.it/dp/B096ZBQ1CG/ref=cm_sw_r_ap_1_R1NNGS8QGN263RMVH1S9_0?psc=1

Esempi di schede Arduino:

- <https://docs.arduino.cc/hardware/mkr-env-shield>
- <https://store.arduino.cc/products/arduino-mkr-env-shield-rev2>