

Informazioni sullo sviluppo dei progetti del corso “Basi di Dati 2”

Questo documento fornisce informazioni generali sulla redazione di una proposta di progetto e sul suo svolgimento nell'ambito del corso di Basi di Dati 2. La fase “operativa” del corso ha come obiettivo l'applicazione delle tecniche e degli strumenti trattati durante le lezioni e le esercitazioni. Sono previste due tipologie di progetto, corrispondenti **a due** delle tre macro-aree in cui si suddividono gli argomenti del corso: Database relazionali, **Database No-SQL e Data Warehouse**.

Progetto di tipo No-SQL

I progetti che rientrano in questa categoria devono soddisfare alcuni requisiti minimi. In particolare, poiché durante il corso saranno enfatizzati concetti come le **proprietà BASE**, **il teorema CAP** e il concetto di collection, i progetti orientati al paradigma “No-SQL” devono:

- Prevedere la ricerca online di uno o più dataset pubblici da utilizzare come miniworld del progetto; Utilizzare un DBMS o un approccio No-SQL tra quelli presentati a lezione (es. MongoDB, Neo4j, XML, DOM...);
- Includere almeno due entità (collection) collegate semanticamente attraverso un attributo condiviso (l'equivalente di una foreign key);
- Implementare le operazioni CRUD in modo efficiente e coerente, rispettando le proprietà BASE; Integrare un'operazione di JOIN tra più collection, rispettando almeno due (a scelta) dei tre vincoli del teorema CAP;
- Prevedere lo sviluppo di una interfaccia utente (anche minimale) che consenta in modo interattivo l'esecuzione delle operazioni CRUD e della JOIN.

Un esempio di progetto potrebbe essere l’“*Analisi di film e recensioni da parte degli utenti*” basato su due dataset no-sql (due collection), uno composto da **Film** (titolo, anno, regista, genere, id_film) e uno composto da **Recensioni** (id_review, id_film, utente, voto, testo).

Progetto di tipo Datawarehouse

I progetti che rientrano in questa categoria devono soddisfare alcuni requisiti minimi. In particolare, poiché durante il corso verranno introdotti e approfonditi concetti come l'architettura di un data warehouse, i processi ETL (Extract, Transform, Load), l'uso di strumenti come **Pentaho** e **Kettle**, gli schemi dimensionali (es. a stella o a fiocco di neve) e le interrogazioni OLAP, i progetti orientati al paradigma *Data Warehouse* devono:

- Prevedere la ricerca online di uno o più dataset pubblici contenenti **dati consolidati e storicizzati** (attenzione: non è richiesto un contesto OLTP), da utilizzare come *miniworld* per costruire un contesto informativo significativo;
- Utilizzare uno dei **tool dedicati al Data Warehousing** presentati durante il corso (es. Pentaho);
- Progettare uno **schema dimensionale**, preferibilmente a stella, composto da **una tabella dei fatti e almeno due dimensioni**, con relazioni chiare e l'uso corretto di chiavi primarie e chiavi esterne;
- Implementare un **processo ETL completo** (utilizzando Kettle), che includa: estrazione e pulizia dei dati dai dataset scelti, trasformazione dei dati in tabelle coerenti con il modello progettato, e caricamento nel sistema finale;
- Consentire **interrogazioni OLAP**, come *drill-down*, *roll-up*, *slice* e *dice*, su indicatori quantitativi della tabella dei fatti;
- Prevedere la **pubblicazione dello schema sul server Pentaho**, in modo da permettere all'utente di utilizzare il **cubo OLAP**, ovvero selezionare dimensioni e misure per esplorare i dati tramite grafici, tabelle o viste aggregate.

Un esempio di progetto potrebbe essere lo scaricamento di dataset open source scaricato dal portale ISTAT (ad esempio, contenente dati storici sulle presenze turistiche suddivise per regione, tipo di struttura, provenienza del turista, e periodo) e la sua successiva analisi in termini di business intelligence.

Progetti proposti dal docente

La sezione “Progetti Proposti” sulla piattaforma di elearning del Dipartimento di Informatica presenta proposte di progetto basate su casi d'uso reali. Gli studenti interessati a una delle proposte possono contattare il docente per ottenere maggiori informazioni.

Cosa produrre durante lo svolgimento del progetto

Per i progetti di tipo “NOSQL” e “DATAWAREHOUSE” è necessario produrre:

1. Un documento descrittivo del progetto in formato professionale, contenente almeno le seguenti sezioni: Introduzione, Descrizione del Miniworld, Contesto Applicativo, Soluzione Proposta, Metodologia utilizzata per sviluppare la soluzione;
2. Una demo (proof of concept).

Per i progetti “proposti dal docente”, la descrizione dei task da seguire è disponibile nei file allegati alle proposte.