



Data science e tecnologie per le basi di dati
Politecnico di Torino
Modalità d'esame A.A. 2020/2021
(emergenza COVID-19)

L'esame è costituito *esclusivamente da una parte scritta* svolta tramite PC con l'utilizzo della piattaforma di Ateneo Exam integrata con strumenti di proctoring (Respondus).

- La regolarità di svolgimento della prova mediante Respondus sarà verificata. Comportamenti non regolari saranno sanzionati secondo le regole di Ateneo.

Regole per la prova scritta

- Possono svolgere la prova scritta *esclusivamente* gli studenti regolarmente prenotati all'appello.
- Lo studente deve essere munito di un documento di identificazione con foto.
- **La durata è di 80 minuti.**
- **Non è possibile** utilizzare libri e appunti, dispositivi elettronici di qualsiasi tipo a parte quello utilizzato per l'esame (cellulare, altro PC/laptop/tablet).
- È possibile utilizzare come fogli di brutta 2 fogli bianchi da entrambi i lati, che dovranno essere mostrati davanti alla telecamera prima dell'inizio della prova scritta.
- La scrivania deve essere completamente sgombra, eccetto per la presenza dei 2 fogli bianchi di brutta, di una penna e del dispositivo utilizzato per la prova.
- In caso di problemi di connessione di rete (o altro problema tecnico) si procederà secondo quanto indicato dall'Ateneo nelle "Linee Guida per il sostenimento di esami on line".

Struttura dello scritto

- 4-6 domande a risposta chiusa di teoria (punteggio totale 6/30)
- 1-3 domande a risposta chiusa di progettazione fisica (punteggio totale 5/30). Non è richiesto l'uso di strumento di disegno.
- 1-2 esercizi di progettazione di trigger (punteggio totale 9/30). Gli esercizi sono strutturati come domande aperte da completare digitando il codice SQL all'interno di una casella di testo.
- 1 esercizio su data warehousing che include
 - 1-3 domande a risposta chiusa sulla progettazione di un data warehouse (punteggio totale 3/30). Non è richiesto l'uso di strumento di disegno.

- A partire da uno schema relazionale predefinito, 2-3 interrogazioni SQL (punteggio totale 8/30). Gli esercizi sono strutturati come domande aperte da completare digitando il codice SQL all'interno di una casella di testo.
- Le risposte errate alle domande a risposta chiusa comportano una penalizzazione del 15% del punteggio relativo a ciascuna domanda. Le risposte mancanti valgono zero.

Argomenti delle domande di teoria

1. Preparazione dei dati: discretizzazione, normalizzazione, misure di distanza, ecc.
2. Regole di associazione: algoritmi di estrazione (anche su esempi), tipi di itemset (closed ecc.), indici di qualità
3. Classificazione: algoritmi, validazione di un modello
4. Clustering: algoritmi, indici di qualità (quelli spiegati a lezione)
5. Recovery: algoritmi di ripristino centralizzato e distribuito (anche applicati su esempi di log)
6. Gestione dell'accesso concorrente: descrizione delle tecniche CSR, VSR, 2PL (anche applicate su esempi di schedule)
7. BD distribuite: frammentazione e livelli di trasparenza, 2 phase commit; parallelismo; benchmarking; basi di dati non relazionali

Quaderno

- Ogni esercizio del quaderno consegnato entro la scadenza è valutato 0.5/30.
- I punti conseguiti con la valutazione del quaderno sono ***validi fino alla sessione di settembre 2021*** inclusa.

Verifica del quaderno

- Prima dello scritto di ogni appello saranno comunicati i nomi degli studenti il cui quaderno sarà verificato e la corrispondente data e ora del colloquio di verifica (tipicamente, subito dopo il termine della prova scritta).

Valutazione finale

- In caso di voto dello scritto $\geq 18/30$, voto dello scritto e valutazione dei quaderni sono sommati. ***Il risultato è registrato come voto finale.*** Un voto finale maggiore o uguale a 31/30 senza arrotondamento corrisponde a 30 con lode.
- In caso di esito negativo è registrata la bocciatura.