

12 ESPERTI PROFILO TECNICO CON CONOSCENZE NELLE TECNOLOGIE DEI SISTEMI INFORMATICI DISTRIBUITI E A REGISTRO DISTRIBUITO E DELL'INTELLIGENZA ARTIFICIALE
(Bando del 2 dicembre 2020 – Lettera A)

Testo n. 1

Tre quesiti su tre delle quattro materie del programma scelti tra gli otto proposti dalla Commissione (due per ogni materia del programma).

1. Programmazione, algoritmi, strutture e modelli dati

QUESITO 1A

Una coppia invita al ricevimento del suo matrimonio N persone, e vuole distribuirle ai K tavoli disponibili nel modo migliore possibile, tenendo conto della loro reciproca affinità. Assumendo che a ciascun tavolo possano sedersi un numero minimo e massimo di invitati, e che l'affinità reciproca tra due invitati sia nota e sia espressa da un numero intero (anche negativo), la candidata/il candidato:

1. descriva l'algoritmo più appropriato per determinare la disposizione che massimizza l'affinità complessiva, intesa come somma delle affinità presenti allo stesso tavolo;
2. indichi la struttura dati più adeguata ad eseguire l'algoritmo;
3. esprima la complessità dell'algoritmo, motivandola;
4. descriva un algoritmo a complessità inferiore che consenta di determinare una soluzione subottimale in un tempo minore.

QUESITO 1B

Un'azienda di autotrasporti opera sul territorio nazionale con un parco proprio di automezzi. Ciascun automezzo è identificato dalla targa ed è contraddistinto dal peso massimo trasportabile. Ogni trasporto organizzato dalla ditta è caratterizzato dalla località di partenza e di arrivo, dall'automezzo utilizzato, dal peso del carico trasportato, dal cliente committente e dalla data del trasporto. Ciascun cliente ha una ragione sociale ed opera in uno specifico settore di attività economica.

Con riferimento al dominio descritto, la candidata/il candidato:

1. definisca il modello entità-relazione indicando le chiavi, gli attributi e la cardinalità delle relazioni;
2. progetti un possibile *star schema* per il *data warehouse* aziendale che consenta di:
 - a. monitorare con cadenza semestrale e annuale il peso della merce trasportata per settore di attività economica del cliente;
 - b. individuare le rotte tra regione e regione caratterizzate dal maggior numero di trasporti;
 - c. individuare i trasporti effettuati con un carico inferiore al 50% rispetto a quello massimo consentito dall'automezzo;
3. ipotizzando che l'azienda voglia migliorare l'efficienza del processo di manutenzione della flotta tramite l'acquisizione e l'analisi dei dati sullo stato dei veicoli, che sono trasmessi con formati differenti e variabili nel tempo da una pluralità di sensori posti sugli automezzi,

descriva i limiti dei sistemi RDBMS nel soddisfare efficientemente questa esigenza e argomenta una possibile soluzione alternativa.

2. Architetture dei sistemi distribuiti

QUESITO 2A

Un'azienda che gestisce i servizi energetici di una grande città intende realizzare una piattaforma di gestione della propria rete di distribuzione.

Il sistema deve fornire le seguenti funzionalità:

- collezionare informazioni provenienti da sensori remoti e contatori, ubicati in punti strategici della rete e presso le utenze; l'invio dei dati di consumo da parte dei contatori avviene solo nel primo giorno di ogni mese;
- permettere agli utenti del sistema di eseguire interrogazioni in tempo reale sui dati acquisiti;
- generare allarmi automatici (SMS o *email*) al superamento di valori di soglia; gli allarmi generati vengono memorizzati in un *datastore* e sono consultabili dagli utenti del sistema.

Sulla base della descrizione precedente, la candidata/il candidato:

1. proponga una possibile architettura del sistema basata su un modello a *microservizi* evidenziandone i vantaggi rispetto ad una architettura monolitica;
2. descriva i principali *pattern* architetturali usati per la gestione della persistenza e della condivisione dei dati tra *microservizi*, specificandone vantaggi e svantaggi; indichi, poi, quale dei *pattern* descritti è stato utilizzato nella soluzione proposta;
3. proponga una possibile soluzione per gestire il picco di messaggi inviati dai contatori all'inizio di ogni mese che permetta un'ottimizzazione dell'utilizzo delle risorse del sistema.

QUESITO 2B

In un database distribuito di tipo *key/value*, lo spazio delle chiavi viene organizzato in partizioni, che sono distribuite ai nodi di un *cluster* e opzionalmente replicate in più copie per incrementarne la disponibilità e fornire un certo grado di tolleranza ai guasti.

1. La candidata/il candidato descriva l'organizzazione dei dati comunemente utilizzata per le partizioni di un *database* distribuito, e illustri le soluzioni adottate per garantire che i dati non sufficientemente replicati non possano essere utilizzati dai *client* applicativi.

Si consideri poi un *database* distribuito configurato in *strong consistency* e costituito da un *cluster* di 3 nodi, nel quale il *leader* di una partizione con *replication factor* 3 abbia replicato con successo un dato X solo su uno dei *follower*, e abbia ricevuto un errore di *timeout* per il secondo.

2. La candidata/il candidato indichi, con le opportune motivazioni, quali delle seguenti affermazioni sono vere:
 - a. il dato X è replicato su solo 2 dei 3 nodi responsabili di gestire la partizione, e sarà necessario effettuare la replica sul terzo quando questo apparirà di nuovo disponibile;
 - b. qualora il fenomeno si ripettesse nel breve termine, il nodo *leader* dovrebbe mettere *offline* la partizione, rendendola non disponibile per letture e scritture;
 - c. se un *client* fa richiesta del dato X in lettura e la richiesta viene servita dal nodo che non ha ricevuto la replica, il dato letto dal *client* può essere sbagliato.



3. Distributed Ledger Technology e cryptoasset

QUESITO 3A

La piattaforma *Bitcoin* è una delle più conosciute implementazioni della *blockchain*.

La candidata/il candidato, prendendo come riferimento il sistema *Bitcoin*:

1. descriva sinteticamente le caratteristiche di decentralizzazione, trasparenza e disintermediazione della *blockchain*, il ruolo del *miner* e l'algoritmo di consenso;
2. descriva, inoltre, l'utilizzo del *Merkle tree* nel contesto del sistema e, dopo una sintetica definizione dello stesso, ne sottolinei l'importanza e il suo utilizzo;
3. infine:
 - a. descriva con uno schema la creazione di un albero di *Merkle* per un blocco di cinque transazioni (A, B, C, D, E);
 - b. descriva schematicamente il processo di verifica di inclusione della transazione B nell'albero.

QUESITO 3B

Una società che opera nel settore della ristorazione emette buoni pasto elettronici in forma di *crypto-token*, che possono essere spesi negli esercizi (bar, supermercati, etc.) che li accettano. I *token* sono disponibili in diversi tagli di importo fisso.

La società espone un servizio *online* per acquistare i *token*, e consente agli esercizi di versare i *token* ricevuti da un loro cliente accreditando un proprio conto tramite un bonifico bancario.

L'autenticità dei *token* e la spendibilità di questi da parte del cliente che li ha acquistati vengono garantite dall'uso di tecniche basate sulla crittografia a chiave pubblica. A ciascun *token* è associata una chiave pubblica che lo identifica.

Rispetto a quanto descritto sopra, la candidata / il candidato:

1. descriva una possibile soluzione al problema di rendere i *token* spendibili solo da parte di chi li ha acquistati presso la società;
2. indichi in che modo gli esercizi possono ricevere, durante lo scambio, un'evidenza del fatto che i *token* sono ceduti in cambio di quanto consumato o acquistato;
3. spieghi se è possibile, in assenza di *blockchain*, impedire il fenomeno del *double spending*, e in caso affermativo indichi una possibile soluzione;
4. con riferimento al punto precedente, confronti la soluzione proposta con *blockchain permissionless* con consenso di tipo *Proof-of-Work*, con particolare riguardo ai tempi necessari per la chiusura della transazione.

4. Intelligenza artificiale

QUESITO 4A

Si considerino le seguenti affermazioni:

- Se la porta è aperta e la finestra è aperta, allora c'è corrente d'aria nella stanza
- Se piove, allora la finestra è chiusa e il davanzale è bagnato
- Se non piove, allora la finestra è aperta
- La porta è aperta
- Il davanzale non è bagnato



La candidata/il candidato:

1. descriva sinteticamente un formalismo o sistema di rappresentazione della conoscenza che permetta di esprimere le precedenti affermazioni, evidenziandone le principali caratteristiche;
2. rappresenti la conoscenza espressa dalle precedenti affermazioni utilizzando il formalismo o sistema per la rappresentazione della conoscenza descritto al punto 1.;
3. nel formalismo o sistema di rappresentazione della conoscenza descritto al punto 1., e data la conoscenza espressa al punto 2., esprima la domanda "C'è corrente d'aria nella stanza?", e dica qual è la risposta a questa domanda in tale formalismo o sistema, spiegando il perché.

QUESITO 4B

Un'azienda attiva nella rilevazione e fornitura di dati meteorologici di un'area metropolitana vuole utilizzare la sua base dati per fornire previsioni giornaliere sulle precipitazioni attese per il giorno successivo. Allo scopo, estrae i seguenti dati dalle serie storiche della sua base dati su cui addestrare un modello di *machine learning*:

- Timestamp (es. 2018-01-01, 2018-01-02, ...)
- TemperaturaMinimaC (es. 5, 7, 9, ...)
- TemperaturaMassimaC (es. 12, 17, 22, ...)
- VelocitàMediaVentoKMH (es. 2.56, 8.02, 11.34, ...)
- NeveCM (es. 0.00, 0.50, 1.34, ...)
- PrecipitazioniCM (es. 0.06, 0.52, 2.97, ...)

Il 20% della base dati, estratto uniformemente, viene accantonato per essere utilizzato come *validation set* X_V . Selezionando sottoinsiemi X_{Tm} di dimensione m dai dati restanti, su ciascuno di essi viene addestrato un modello calcolando l'errore complessivo (*loss*) su *training set* X_{Tm} e *validation set* X_V al fine di studiarne il comportamento al variare di m .

Nel contesto dello scenario descritto, la candidata/il candidato:

1. descriva brevemente le principali tecniche di normalizzazione dati e *feature scaling* valutandone l'applicazione al caso in esame motivando la risposta;
2. valuti possibili effetti collaterali connessi ai criteri adottati per la composizione degli insiemi di validazione e addestramento X_V e X_{Tm} , suggerendo eventualmente un approccio alternativo;
3. definisca i concetti di *bias* e varianza nell'ambito degli algoritmi di apprendimento automatico e le loro implicazioni sul comportamento di un modello addestrato;
4. suggerisca possibili strade percorribili per migliorare il funzionamento del modello qualora il comportamento osservato esibisca elevato *bias* o varianza.

Prova in lingua inglese

Social networking represents a significant part of everyday life. Describe pros and cons of being connected.

