BANCA D'ITALIA
EUROSISTEMA

10 ESPERTI – PROFILO TECNICO NEL CAMPO DELL'ICT

(Bando del 24 aprile 2026 – lett. A)

Testo n. 3

Due quesiti a scelta – tra sei proposti dalla Commissione – su due differenti ambiti tematici previsti dal programma

COMPUTAZIONE, SOFTWARE, SISTEMI

QUESITO N. 1

Un gruppo editoriale decide di realizzare MSPORT, un sistema per la gestione delle informazioni relative a manifestazioni sportive a squadre (ad esempio, calcio, basket, rugby). Il sistema non sarà accessibile direttamente dagli utenti finali, ma verrà utilizzato dai siti *web* delle varie testate *online* del gruppo.

MSPORT tratterà informazioni come: le varie edizioni delle competizioni, le squadre e gli atleti partecipanti, gli incontri tra due squadre, gli eventi significativi all'interno degli incontri, i punteggi finali e parziali, eventuali classifiche (parziali e finali). MSPORT esporrà delle API REST attraverso cui manipolare le informazioni gestite, le quali saranno utilizzate dai siti *web* e dalle *app* del gruppo, in sola lettura e con interrogazioni fisse e legate alla particolare struttura dei menu di navigazione. Un secondo utilizzo dei dati di MSPORT prevedrà invece l'accesso diretto, per l'esecuzione di interrogazioni estemporanee, non note in fase di *design* del sistema, volte alla determinazione di curiosità e statistiche (ad esempio, riguardanti particolari condizioni nella carriera di un atleta, ricorrenze storiche di eventi rari, valori di prestazioni aggregati).

Una componente di MSPORT, denominata LIVE, verrà utilizzata dai siti *web* del gruppo per la cronaca testuale e grafica degli incontri mentre sono in corso. La cronaca consiste in un'animazione che mostra come si sposta l'azione di gioco sul campo, dandone allo stesso tempo una breve descrizione. LIVE riceverà da un *provider* esterno specializzato un flusso *real-time* di *feed* del tipo:

```
{
   "progressivo": 35624266,
   "evento": "passaggio",
   "squadra": "Italia",
   "da": [54, 32, "Mario Rossi"],
   "a": [70, 66, "Luca Verdi"]
}
```

Nelle triplette "da" e "a", i primi due numeri rappresentano le coordinate dei punti tra cui avviene il passaggio sul campo di gioco, il terzo elemento è un identificativo dell'atleta coinvolto. Il campo "evento" può assumere un insieme noto di valori (ad esempio, "passaggio", "fallo", "meta").

Il flusso di *feed* sarà usato per alimentare la cronaca testuale e grafica *real-time* sui diversi siti delle testate *online* del gruppo, ma anche come una delle sorgenti delle informazioni sugli incontri, una volta terminati.

Considerando lo scenario proposto:

1. si descrivano i principali requisiti non funzionali dei sistemi *software*, eventualmente in relazione a classificazioni o *standard* internazionali. Si indichino, per ciascun requisito, alcune possibili metriche, precisando quali sono le più rilevanti per MSPORT e LIVE. Si fornisca qualche esempio di come, in fase di analisi, si potrebbe arrivare a fissare dei valori obiettivo per le metriche di interesse. Si analizzi criticamente la relazione che intercorre tra le metriche di interesse e il livello di servizio offerto dal *provider*;
2. si illustrino le tecnologie di DBMS relazionale, *document-oriented*, *key-value*, *column-oriented*, *graph-based* e si indichi la più appropriata per i due utilizzi di MSPORT;
3. si descriva una possibile architettura distribuita per il sistema, specificando le componenti costitutive e i *pattern* di integrazione utilizzati. In particolare, l'architettura deve soddisfare le seguenti esigenze:
   - collezionare le informazioni dei *feed* e alimentare le strutture di persistenza selezionate al punto 2, tenendo conto della diversità dei requisiti non funzionali di MSPORT e LIVE;
   - reindirizzare il flusso di *feed real time* verso i siti *web* e le *app* terminali in modo da consentirne la fruibilità;
   - disaccoppiare il flusso proveniente dal *provider*, dai siti *web* e dalle *app* che le consumano, in modo tale da garantire che l'aggiunta o la rimozione di siti o *app* (in seguito, ad esempio, ad acquisizioni o cessioni di testate) non richiedano modifiche strutturali.

   Si argomenti come le diverse scelte consentono di realizzare gli obiettivi non funzionali identificati al punto 1, evidenziando eventuali alternative e *trade-off*;
4. si illustrino alcune soluzioni tecniche che potrebbero essere usate per inviare in modalità *push*, ai *browser* aperti sul servizio dei lettori, gli aggiornamenti della cronaca degli incontri. Si discutano quali conseguenze una comunicazione *push* dal *server* ha, in generale, dal punto di vista del consumo di risorse sul *server*, della complessità infrastrutturale e della scalabilità.

QUESITO N. 2

Un'organizzazione sta evolvendo le proprie applicazioni per essere ospitate su un'infrastruttura basata su *container*.

1. Si illustrino le tecniche di virtualizzazione di sistema, confrontando criticamente le strategie di *system virtualization*, *os-level virtualization (o container virtualization)* e *process virtualization* in termini di prestazioni, isolamento, sicurezza e gestione operativa.
2. Con riferimento alla *container virtualization*, si descrivano gli ambiti di applicazione, i relativi benefici e gli eventuali punti di attenzione.
3. Si illustri il ciclo di vita di un *container* dalla costruzione fino alla dismissione, facendo riferimento a livello concettuale agli strumenti e alle componenti a supporto delle diverse fasi del ciclo di vita. Si descriva inoltre una possibile *pipeline* che consente di automatizzare il rilascio delle componenti applicative basate su *container*.
4. Si illustrino i vantaggi tecnici e operativi derivanti dall'adozione di una piattaforma di orchestrazione dei *container*. Si descriva inoltre come tale piattaforma può gestire i seguenti due scenari: i) incrementi repentini del volume di richieste indirizzate a una o più applicazioni basate su *container* e ii) malfunzionamenti di uno o più componenti applicative.

CRITTOGRAFIA, DISTRIBUTED LEDGER TECHNOLOGT (DLT), PRIVACY

QUESITO N. 3

La crittografia a chiave pubblica consente di instaurare canali sicuri su reti non sicure; in tale contesto, i protocolli di *Authenticated Key Exchange* (AKE) permettono a due parti di concordare una chiave di sessione garantendo al tempo stesso autenticazione e riservatezza.

1. Si illustrino i principali requisiti di sicurezza dei protocolli AKE, con riferimento all'autenticazione delle parti, alla segretezza della chiave di sessione, alla resistenza ad attacchi attivi e alla protezione delle sessioni precedenti e successive.
2. Si discutano le modalità di implementazione dell'AKE nello standard *Transport Layer Security* TLS (con eventuale riferimento ad una particolare versione del protocollo), descrivendone l'architettura generale e il modo in cui esso realizza i requisiti di sicurezza individuati nel punto precedente, assumendo il caso di attaccanti classici.
3. Si descriva, in termini generali, come le componenti crittografiche del protocollo TLS possono essere sostituite o affiancate da soluzioni *post-quantum* e si discutano i principali effetti di tale migrazione su compatibilità e prestazioni.
4. Si descrivano i protocolli *Password AKE* (PAKE) come variante degli AKE basata su *password*, illustrandone i requisiti di sicurezza e le principali difficoltà costruttive, con particolare attenzione alla resistenza ad attacchi di dizionario *offline* e all'autenticazione reciproca, e ai possibili ambiti di applicazione.

QUESITO N. 4

Una *blockchain permissionless* (*Layer 1*) consente l'esecuzione di *smart contract*, ma può elaborare un numero limitato di transazioni. Per aumentare la capacità del sistema, un *rollup* esegue le transazioni al di fuori del *Layer 1* e pubblica periodicamente sulla *blockchain* almeno un *commitment* al nuovo stato, rendendo inoltre disponibili, secondo il meccanismo previsto dal protocollo, i dati necessari alla verifica e alla ricostruzione dello stato.

Si considerino i modelli di *rollup* ottimistico (in cui il nuovo stato viene accettato provvisoriamente e può essere contestato, entro un periodo di contestazione - *challenge period* - mediante una prova della frode) e di *rollup* con prova di validità, spesso denominato ZK *rollup* (in cui l'aggiornamento viene accettato solo se accompagnato da una prova crittografica della correttezza della transizione di stato).

1. Si spieghi che cos'è un *rollup* e perché l'esecuzione di più transazioni al di fuori del *Layer 1* può aumentare la capacità complessiva del sistema. Si descriva, in termini generali, il ruolo del *sequencer*, dei *batch*, del *commitment* allo stato e dei dati resi disponibili, chiarendo quali funzioni rimangono affidate al *Layer 1* e perché non è necessario considerare il *sequencer* pienamente fidato.
2. Si confronti il processo di creazione e finalizzazione di un *batch* nei due modelli. Si chiarisca come possa essere rilevato e gestito un aggiornamento non valido. Si distinguano la finalizzazione sul *Layer 1* della transazione che pubblica il *batch*, la definitività dell'aggiornamento di stato nel *rollup* e i tempi necessari per utilizzare i fondi all'interno del *rollup* o trasferirli dal *rollup* al *Layer 1*.
3. Si considerino i seguenti scenari:
   a) il *sequencer* propone un *batch* contenente una transazione non valida, ad esempio, un trasferimento superiore al saldo disponibile;

b) il *sequencer* è indisponibile o rifiuta di includere le transazioni di alcuni utenti;
c) il *sequencer* pubblica il *commitment* al nuovo stato, ed eventualmente una prova valida, ma i dati necessari a ricostruire lo stato non sono disponibili.

Per ciascuno scenario e distinguendo i modelli se rilevante, si indichi se possano essere compromesse la correttezza dello stato, la possibilità degli utenti di utilizzare o recuperare i propri fondi oppure la capacità del sistema di progredire. Inoltre, si discuta, quando pertinente, il ruolo delle contestazioni, dell'inclusione forzata delle transazioni, della pubblicazione dei dati e delle procedure di uscita o recupero.

4. Si discutano i principali compromessi tra *rollup* ottimistici e ZK *rollup* rispetto a latenza, costi, complessità crittografica, disponibilità dei dati e dipendenza dal *sequencer*. Con riferimento ad un'applicazione che richiede prelievi rapidi e un'elevata affidabilità del regolamento, si indichi quale modello risulta preferibile e da quali ulteriori condizioni dipende la scelta.

INTELLIGENZA ARTIFICIALE, MACHINE LEARNING, DATA SCIENCE

QUESITO N. 5

Una banca utilizza un modello per classificare i soggetti da essa finanziati in funzione della loro probabilità di *default* ([1]) nelle due classi "*cliente sano*" e "*cliente da attenzionare*". Un cliente viene classificato come "*cliente da attenzionare*" se la relativa probabilità di *default* assume valori al di sopra di una soglia, altrimenti viene classificato come "*cliente sano*".

Il modello utilizzato attualmente è una regressione logistica, addestrato su un *dataset* composto dalle caratteristiche dei soggetti finanziati. Ogni osservazione ha una data di riferimento, che può assumere 5 valori (31/12/2017; 31/12/2018; 31/12/2019; 31/12/2020; 31/12/2021). I soggetti finanziati vengono etichettati come *default* se alla fine dell'anno solare successivo alla data di riferimento non sono stati in grado di adempiere alle proprie obbligazioni creditizie (e.g., non hanno pagato le rate del mutuo per 90 giorni), altrimenti *non in default*. Il 2% dei soggetti finanziati del *dataset* è etichettato come *default*, mentre il restante 98% come *non in default*. Il *team* di modellazione propone di sostituire la regressione logistica con un nuovo modello denominato "*modello challenger*" per migliorare le prestazioni.

Il *team* di validazione confronta i due modelli sullo stesso campione di test. La regressione logistica ottiene un valore dell'*Area Under the Receiver Operating Characteristic Curve* (AUC-ROC) = 0,78. Il *modello challenger* ottiene AUC-ROC = 0,86 sul test set, ma AUC-ROC = 0,97 sul training set.

Si richiede di:

1. spiegare di che tipo di problema di *machine learning* si tratta, specificandone le caratteristiche e discutendo l'adeguatezza delle metriche comunemente usate, con particolare riferimento all'accuratezza;
2. illustrare il compromesso *bias-variance* e interpretare i dati forniti, indicando cosa segnala lo scarto tra l'AUC-ROC sul training (0,97) e quello sul test (0,86) del *modello challenger*.

Ipotizzando che il *dataset* sia costituito da un numero ridotto di osservazioni (e.g., 200 osservazioni per ogni data di riferimento):

3. si illustri quale strategia di validazione adotterebbe per stimare le prestazioni del modello, selezionando i parametri della strategia di validazione adottata e giustificando la risposta;

[1] Probabilità che un debitore non sia in grado di adempiere alle proprie obbligazioni creditizie entro un determinato orizzonte temporale (per il caso in esame ipotizziamo essere 1 anno solare).

4. si scelga un *modello challenger*, illustrandone il funzionamento e motivando criticamente tale scelta. Si discuta inoltre quali tecniche potrebbero attenuare eventuali problemi di *overfitting*.

QUESITO N. 6

La qualità e l'affidabilità delle informazioni anagrafiche rappresentano elementi fondamentali per il corretto funzionamento dei processi bancari. Una banca intende realizzare una piattaforma centralizzata per la gestione delle informazioni anagrafiche dei soggetti, capace di integrare dati provenienti da fonti interne ed esterne alla banca e di garantire una visione univoca, coerente e aggiornata.

La piattaforma deve consentire l'individuazione e la correzione di duplicati, errori e incongruenze nei dati relativi ai soggetti, assicurando allo stesso tempo elevati livelli di disponibilità, affidabilità, tracciabilità e continuità operativa.

1. Assumendo che i sistemi interni alla banca notificano in tempo reale le modifiche dei soggetti al sistema anagrafico, mentre i sistemi esterni forniscono gli aggiornamenti in modalità differita attraverso flussi dati di grandi dimensioni, si descrivano i principali approcci che si possono utilizzare per l'aggiornamento delle informazioni anagrafiche gestite dalla banca.
2. Si analizzi il processamento dei dati di tipo *batch*, illustrando criticamente i modelli di elaborazione distribuita più diffusi e le soluzioni tecnologiche più utilizzate per questo tipo di elaborazioni.
3. Con riferimento al caso d'uso descritto in premessa, si progetti una *pipeline* di gestione delle informazioni anagrafiche finalizzata alla costruzione di una vista unificata e affidabile dei soggetti, descrivendone le principali fasi, dalla raccolta dei dati provenienti da fonti interne ed esterne fino alla distribuzione delle informazioni ai sistemi utilizzatori. Si discutano inoltre le soluzioni adottate per garantire qualità del dato e la tracciabilità delle modifiche.
4. Si illustri una soluzione, basata su tecniche di *machine learning,* per il problema dell'*entity matching* e l'eliminazione dei duplicati, descrivendo come costruire il *golden record* a partire dalle informazioni provenienti da fonti differenti.

PROVA IN LINGUA INGLESE

If you could take your dream trip anywhere in the world, where would you go, when, and why? Describe your chosen destination and explain what makes it meaningful to you.