

**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. 4**

*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**

Un'azienda di aste *online*, che mette a disposizione una piattaforma digitale a un elevato numero di iscritti, ha necessità di effettuare un'analisi *ex-post* delle transazioni giornaliere eseguite negli ultimi cinque anni. L'analisi ha l'obiettivo di calcolare per ciascun iscritto (rappresentato da un identificativo univoco) il flusso totale mensile di transazioni in euro (inteso come la differenza tra *totale vendite* e *totale acquisti* nel mese).

I dati relativi alle transazioni storiche sono contenute in unico *dataset* di *log*, di cui si riporta un estratto:

DATA	ID_ACQUIRENTE	ID_VENDITORE	IMPORTO_EURO
.....	.....	.....	.....
2020-12-10	Iscritto 1	Iscritto 2	1500
2020-12-10	Iscritto 3	Iscritto 4	820
2020-12-11	Iscritto 4	Iscritto 5	2300
.....	.....	.....	.....

Di seguito un estratto del *dataset* di *output* atteso:

ANNO_MESE	ID	FLUSSO
.....	.....	.....
2020-12	Iscritto 1	-1500
2020-12	Iscritto 2	1500
2020-12	Iscritto 3	-820
2020-12	Iscritto 4	-1.480
2020-12	Iscritto 5	2.300
.....	.....	.....

La candidata/il candidato:

1. descriva sinteticamente il modello di programmazione *MapReduce* evidenziando il ruolo della fase di *shuffle and sort*;
2. fornisca mediante pseudo-codice una soluzione secondo il paradigma *MapReduce* per il calcolo dei flussi totali mensili per iscritto;
3. individui le modifiche necessarie alla soluzione del punto precedente, qualora si volesse ottenere il flusso medio mensile per iscritto (flusso totale mensile del soggetto diviso per il

numero corrispondente di transazioni in ingresso e uscita).



### QUESITO 1B

Si vuole progettare il *software* della centralina di un sistema di videosorveglianza, che pilota attraverso una linea interna le videocamere di un centro commerciale. Attraverso istruzioni ricevute da remoto, la centralina può inviare comandi alle singole videocamere, a quelle di un gruppo, o a tutte.

Ciascuna videocamera appartiene a un solo gruppo, deciso in fase di installazione secondo la sua collocazione. Le videocamere non hanno capacità di elaborazione ma si limitano ad eseguire i comandi ricevuti, pertanto il *software* della centralina ne ha il controllo completo.

Supponendo che la centralina sia programmata con un linguaggio orientato agli oggetti, la candidata/il candidato:

1. indichi quale *pattern* è più adeguato alla gestione dei comandi inviati dalla centralina alle videocamere;
2. proponga e commenti il diagramma delle classi di una possibile soluzione che consenta di applicare tale *pattern* al caso in esame;
3. spieghi perché in un contesto come quello descritto non è opportuna l'applicazione del *pattern Publish/Subscribe* in alternativa a quello proposto.

## 2. Architetture dei sistemi distribuiti

### QUESITO 2A

Un grande rivenditore al dettaglio di *gadget* tecnologici ha implementato il proprio portale di vendita su Internet attraverso un'architettura a microservizi. Per far fronte ai picchi di carico, l'architettura impiega meccanismi di *autoscaling* per le componenti a microservizi, mentre il *database*, esterno alla piattaforma, è unico e dimensionato per il picco di carico.

Si richiede alla candidata/al candidato di:

1. indicare e descrivere le principali caratteristiche necessarie alle componenti applicative a microservizi per essere compatibili con il meccanismo di *autoscaling*;
2. proporre e descrivere una soluzione in grado di garantire le prestazioni desiderate dettagliando i componenti architetturali necessari e illustrando come il sistema reagisce a fronte di un picco di carico;
3. illustrare tecniche e strumenti per il *troubleshooting* delle prestazioni di un'operazione che coinvolge più di un microservizio e una serie di accessi al *database*.

### QUESITO 2B

Una catena di supermercati intende velocizzare gli acquisti dotando ciascun carrello di un *tablet* con un lettore di codice a barre in grado di scansionare e comunicare l'identificativo dei prodotti via API REST ai sistemi informativi.

Ciascun cliente all'inizio dello *shopping* attiva un "carrello virtuale" mediante la lettura del codice a barre della propria "Cartà Fedeltà"; successivamente scansiona i prodotti che aggiunge al carrello. È possibile rimuovere un prodotto dal carrello premendo un apposito pulsante e scansionandolo nuovamente. Sul *tablet* viene mostrato in tempo reale il contenuto del carrello, con le quantità di ciascuna tipologia di prodotto. Il conto viene calcolato scansionando un apposito codice a barre



presente sul terminale di pagamento.

La candidata/il candidato:

1. descriva lo stile architetturale REST, evidenziandone i tratti salienti;
2. proponga un modello dati idoneo a supportare gli scenari descritti, dando indicazione delle cardinalità minime e massime nelle relazioni tra entità e degli attributi più rilevanti per i casi d'uso descritti;
3. proponga una API REST in grado di supportare gli scenari illustrati, individuando (i) le risorse e le operazioni necessarie e (ii) motivando la scelta dei metodi HTTP e degli *status code* per ciascuna operazione;
4. descriva il *constraint* HATEOAS indicandone finalità, vantaggi e controindicazioni.

### **3. Distributed Ledger Technology e cryptoasset**

#### **QUESITO 3A**

Uno degli aspetti cruciali di una *blockchain* pubblica è la scelta di un adeguato algoritmo di consenso. Questo componente non solo abilita il funzionamento di una *blockchain* come sistema distribuito, ma ne può caratterizzare molti altri aspetti con ripercussioni ad esempio su prestazioni e costi.

Prendendo a riferimento l'architettura *Bitcoin*, la candidata/il candidato:

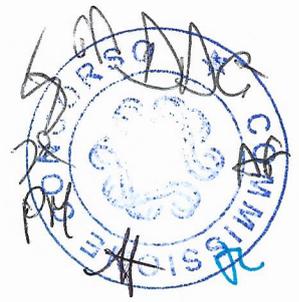
1. descriva sinteticamente il funzionamento dell'algoritmo di consenso impiegato valutando la relazione tra la velocità di creazione dei blocchi e le risorse computazionali allocate nell'operazione di *mining*;
2. valuti il *trade-off* da considerare nell'ipotesi di voler ridurre la latenza media di conferma delle transazioni, con particolare riferimento al fenomeno di creazione di blocchi orfani/estinti;
3. discuta se e in che modo il fenomeno evidenziato al punto precedente viene affrontato in *Ethereum*, anche in relazione all'aspetto di remunerazione dell'attività di *mining*.

#### **QUESITO 3B**

Da un punto di vista tecnico, una transazione in una *blockchain* vede coinvolti principalmente tre attori: il pagante, il ricevente, il *miner*.

La candidata/il candidato:

1. ordini cronologicamente le azioni di seguito elencate, fornendo per ogni azione una descrizione sintetica e l'attore responsabile:
  - a. decodifica dell'indirizzo in *hash* della chiave pubblica;
  - b. inserimento nella *blockchain* della transazione;
  - c. invio dell'indirizzo;
  - d. codifica dell'*hash* della chiave pubblica in indirizzo;
  - e. trasmissione in rete della transazione;
  - f. creazione di una coppia chiave pubblica / chiave privata;
  - g. creazione della transazione;
  - h. *hash* della chiave pubblica;
2. descriva l'algoritmo di consenso "*Proof-of-Work*" evidenziandone i punti di forza e le eventuali criticità;
3. data una *blockchain* con un *total hash rate* pari a 86 *Terahash/s* generato dalle seguenti *mining pools*:
  - MP1: 24,3 TH/s



- MP2: 14,34 TH/s
- MP3: 17,32 TH/s
- MP4: 15 TH/s
- MP5: 15,04 TH/s

descriva quali condizioni siano necessarie per attuare un “attacco 51%” e fornisca un esempio di *double spending*.

#### 4. Intelligenza artificiale

##### QUESITO 4A

Una grande azienda di *e-commerce* intende migliorare il processo di validazione degli acquisti introducendo una funzione che consenta di individuare e bloccare le transazioni riconducibili a sospette frodi. Per raggiungere l’obiettivo intende ricorrere a tecniche di *machine learning* e utilizzare i dati storici delle transazioni di acquisto presenti in forma strutturata negli archivi aziendali. Si segnala che circa l’1% delle transazioni presenti nello storico risulta etichettato come “fraudolento” a seguito di un processo di analisi manuale.

Con riferimento allo scenario indicato, la candidata/il candidato:

1. identifichi il tipo di problema di *machine learning* e descriva sinteticamente un modello idoneo all’obiettivo sopra enunciato;
2. descriva le tecniche da utilizzare nella fase di *preprocessing* al fine di produrre un *dataset* adeguato all’addestramento del modello;
3. illustri il tema del *data drift* e del *concept drift* e le possibili strategie di mitigazione;
4. illustri le differenze tra curve ROC (*Receiver Operating Characteristic*) e curve *Precision-Recall* ed indichi quale delle due tecniche potrebbe essere adottata per valutare le prestazioni del modello, argomentando la scelta effettuata.

##### QUESITO 4B

Una catena di agenzie immobiliari intende sviluppare un modello di *machine learning* per stimare il prezzo di vendita dei nuovi immobili che entrano nel suo portafoglio. Per lo sviluppo del modello, la catena immobiliare può avvalersi della base dati storica delle transazioni in cui sono registrate alcune caratteristiche dell’immobile venduto e il relativo prezzo. In particolare, per ciascuna transazione sono riportati i seguenti attributi (con alcuni valori di esempio):

- *Prezzo\_di\_vendita\_in\_euro* (120.000,0 ; 435.000,0 ; 1.500.000,0)
- *Anno\_di\_fabbricazione* (1984 ; 2013 ; 2018 ; 1997)
- *Numero\_stanze* (3; 5; 7)
- *Metri\_quadri\_totali* (72.5 ; 90.0 ; 125.5 ; 200.0 )
- *Distanza\_dal\_centro\_km* (12.5 ; 25.0 ; 2.5,..)
- *Presenza\_Garage* (Si ; No)
- *Stato\_di\_conservazione* ( Da\_ristrutturare ; Medio; Buono ; Ottimo)

La candidata / il candidato:

1. indichi la tipologia del problema di *machine learning*, il tipo delle variabili indipendenti e le possibili metriche di valutazione di modelli;
2. proponga un modello di *machine learning* applicabile al caso in oggetto motivando la scelta, descrivendone sinteticamente i principi di funzionamento ed eventuali trasformazioni richieste per i dati di *input*;
3. descriva cosa si intende per importanza di una *feature* per un modello addestrato e illustri sinteticamente una metrica relativa;
4. ipotizzando che i campi *Stato\_di\_conservazione* e *Distanza\_dal\_centro\_km* presentino una quota di dati mancanti non superiore al 15% del totale delle osservazioni, proponga una possibile strategia di imputazione da utilizzare durante l'addestramento e il *test* del modello in modo tale da prevenire il fenomeno del *data leakage*.

### Prova in lingua inglese

Some say that the COVID-19 crisis and increased remote working/learning will lead to a change in the residential preferences for urban or non-urban living. Discuss.

