IVASS
ISTITUTO PER LA VIGILANZA SULLE ASSICURAZIONI

CONCORSO PER L'ASSUNZIONE DI 4 LAUREATI NELLE DISCIPLINE DELL'ICT (INFORMATION AND COMMUNICATION TECHNOLOGY) CON CONOSCENZE NELLA SICUREZZA INFORMATICA O NELLE TECNOLOGIE DEI SISTEMI INFORMATICI DISTRIBUITI E A REGISTRO DISTRIBUITO E DELL'INTELLIGENZA ARTIFICIALE

(Bando pubblicato sulla Gazzetta Ufficiale della Repubblica Italiana – 4ª serie speciale "Concorsi ed Esami" n. 86 del 28 ottobre 2022)

PROVA SCRITTA DEL 27 APRILE 2023

TESTO N. 2

Il candidato scelga una coppia di materie (A e B oppure C e D) e svolga due quesiti, uno per ciascuna materia della coppia prescelta. La traccia di lingua inglese è invece unica.

A. *ARCHITETTURE DEI SISTEMI DISTRIBUITI (un quesito a scelta)*

Quesito A.1

Un *pattern* architetturale fornisce una descrizione strutturata volta a supportare la progettazione dell'architettura di un sistema *software*. Nella progettazione di sistemi distribuiti rileva la scelta della modalità di comunicazione remota tra i diversi componenti del sistema. La candidata/il candidato:

- descriva i *pattern* di comunicazione remota applicabili in un sistema distribuito, con particolare riferimento al ruolo dello strato di *middleware* nell'interazione tra i componenti delle applicazioni;
- illustri, nel caso dell'utilizzo di servizi *web*, e, prendendo a riferimento uno specifico protocollo (ad es. REST, SOAP), come avviene la comunicazione tra i componenti delle applicazioni e come sia possibile garantire la confidenzialità dei messaggi scambiati.

Si consideri il caso di un'azienda di medie dimensioni che utilizza un'applicazione *web* di CRM (*Customer Relationship Management*) sviluppata *in house* e installata *on premise*. L'applicazione gestisce l'autenticazione degli utenti interni mediante *Kerberos* e quella dei clienti mediante *Form Based Authentication* su canale sicuro. Nell'ottica di far evolvere il servizio, l'azienda decide di valutare possibili soluzioni alternative basate su servizi di CRM *in cloud* pubblico di tipo SaaS (*Software as a Service*). La candidata/il candidato:

- descriva quali nuove modalità di *identity, credential and access management* dovrebbero essere utilizzate per garantire la sicurezza della soluzione sulla rete *internet*.

Quesito A.2

Nelle moderne organizzazioni si sono affermate, nell'ambito del processo di sviluppo *software*, soluzioni tecnologiche e organizzative per la gestione del ciclo di vita delle applicazioni. La candidata/il candidato:

- descriva le principali funzionalità di un servizio di *repository* del codice, focalizzandosi sulle operazioni di caricamento (*push*) e scaricamento (*pull*) del codice, archiviazione delle modifiche, ramificazione e unione delle diverse versioni e in particolare sui conflitti che possono verificarsi durante queste operazioni e sui metodi per risolverli;
- illustri i principi su cui si basa un approccio di *Continuous Integration e Continuous Delivery* e i relativi vantaggi, descrivendo quali ambienti sono necessari per organizzare i rilasci del codice in maniera efficace; discuta inoltre le problematiche connesse alla predisposizione e all'esecuzione dei test sul *software* sviluppato.

Si consideri un *team* che si occupa della gestione di sistemi utilizzati da applicazioni *web* basate su microservizi e quindi in grado di scalare agilmente su più *server*, orizzontalmente, in funzione del carico applicativo. La candidata/il candidato:

- descriva le soluzioni che possono essere utilizzate per automatizzare le operazioni di installazione, configurazione e gestione dei moduli applicativi e dei sistemi elaborativi che li ospitano (solitamente indicate come soluzioni di tipo IaC, *Infrastructure as Code*).

B. *SICUREZZA INFORMATICA (un quesito a scelta)*

Quesito B.1

Il panorama delle minacce *cyber* evolve rapidamente, con attaccanti sempre più motivati, dotati di elevate competenze tecniche e ingenti risorse economiche, capaci di condurre attacchi pianificati e prolungati nel tempo. In tale contesto, un'adeguata mitigazione del rischio *cyber* presuppone la capacità delle organizzazioni di rilevare tempestivamente e di gestire efficacemente gli attacchi. La candidata/il candidato:

- illustri le fasi di un processo per la gestione degli incidenti di sicurezza, riferendosi eventualmente a *framework* esistenti, discutendone obiettivi, aspetti organizzativi e strumenti a supporto;
- descriva come la conoscenza delle tattiche, delle tecniche e delle procedure utilizzate dagli attaccanti possa migliorare le capacità di prevenzione, rilevazione e risposta agli attacchi *cyber*, facendo riferimento ad almeno una *knowledge base* esistente;
- illustri le finalità e le modalità di raccolta delle evidenze digitali da un singolo sistema fisico (ad es. *server* o *laptop*), indicando le misure tecniche e organizzative da adottare per la loro corretta conservazione.

Quesito B.2

L'utilizzo degli algoritmi e dei protocolli di crittografia consente di proteggere lo scambio di informazioni in rete. La candidata/il candidato:

- descriva le differenze tra la crittografia simmetrica e la crittografia asimmetrica; illustri inoltre come queste tecniche di cifratura possano essere impiegate per garantire la confidenzialità, l'integrità e la non ripudiabilità delle informazioni;
- descriva quali approcci possono essere utilizzati per attaccare i sistemi di cifratura a chiave simmetrica e asimmetrica e indichi possibili contromisure.

Si consideri un'azienda che abbia deciso di utilizzare servizi in *cloud* pubblico per la memorizzazione, in forma cifrata, dei propri dati. La candidata/il candidato:

- illustri le possibili soluzioni per la gestione delle chiavi di cifratura e decifratura in tale contesto, indicando i principali vantaggi e svantaggi.

C. *PROGRAMMAZIONE, ALGORITMI, STRUTTURE E MODELLI DATI (un quesito a scelta)*

Quesito C.1

Lo sviluppo di sistemi informativi complessi spesso prevede di integrare dati provenienti da sorgenti esterne. In questo contesto, capita di dover gestire dati di scarsa qualità e accompagnati da documentazione carente. Per integrare i dati di queste sorgenti nel nostro sistema è quindi necessario definire vincoli che permettano di verificarne la qualità, ricostruire l'informazione rappresentata dai dati, affrontare le eterogeneità di rappresentazione dell'informazione. Si supponga di dover importare nel

proprio sistema la base di dati (relazionale) mostrata in Figura 1 (è riportato un insieme rappresentativo dei dati).

CLIENTI

PIVA	Nome	Cognome	Email
ES 234567890	Jose R.	Blanco	j.blanco@live.com
IT-1234**G**6**Z**89	Paolo	Rossi	paolo.rossi@gmail.com
ES.345678901	Diego	Fuerte	diego.fuerte@abc.es

FASCICOLI

Codice	Data_apertura	PIVA	Data_chiusura
ABC	23/05/2021	ES 234567890	02/03/2022
BCD	12/07/2022	ES.345678901	null
CDE	15/06/2021	IT-1234**G**6**Z**89	24/02/2022
DEF	12/07/2022	IT-1234**G**6**Z**89	null

ABILITAZIONI

Codice	Nome
a11	Contabilità
b22	Finanza

DOCUMENTI

Numero	Codice	Descrizione	Data	Responsabile
1	ABC	Bla bla	02/08/2021	X1Y1
2	ABC	Ble ble	07/11/2021	X1Y1
1	BCD	Bli bli	06/09/2022	Z1Z1
1	CDE	Blo blo	27/02/2022	X1Y1

D_A

Matricola	Codice
X1Y1	a11
X1Y1	b22
Z1Z1	a11

DIPENDENTI

Matricola	Nome	Cognome	Telefono
X1Y1	Luigi	La Pastina	333000222
Z1Z1	Maria	De Santis	Null

Figura 1 - Base di dati da importare

La candidata/il candidato:

- analizzando le tabelle della base di dati in Figura 1, definisca i vincoli di chiave, i vincoli di integrità referenziale (*foreign-key*), i vincoli sui valori nulli, ed eventuali altre tipologie di vincoli che ritiene utili al fine di valutare la consistenza dei dati;
- definisca uno schema *Entity-Relationship* (ER) che descriva la base di dati in Figura 1 e da cui lo schema relazionale fornito possa essere derivato nella fase di progettazione logica.

I dati della tabella CLIENTI nella base di dati in Figura 1 dovranno essere integrati con i dati di una tabella ANAGRAFICA già presente nel nostro sistema, per la quale riportiamo in Figura 2 alcune ennuple.

ANAGRAFICA

SIGLA_NAZIONE	NUM_PIVA	NOME	COGNOME
IT	123456789	Paolo	Rossi
IT	543210987	Sergio	Navi
AT	U78901234	Hans	Muller
ES	234567890	José Roberto	Blanco

Figura 2 - Tabella ANAGRAFICA già presente nel nostro sistema

La candidata/il candidato:

- descriva una strategia (espressa mediante pseudocodice o diagramma di flusso) per individuare possibili duplicati tra la tabella CLIENTI della Figura 1 e la tabella ANAGRAFICA della Figura 2, considerando i problemi di eterogeneità di rappresentazione e di qualità dei dati del *database* da importare. A titolo di esempio, si osservi che la prima ennupla della tabella CLIENTI e l'ultima ennupla della tabella ANAGRAFICA fanno riferimento alla stessa persona (Josè Roberto Blanco). Tuttavia, per la presenza di errori nei dati e per le diverse forme di rappresentazione dell'informazione, i dati nelle due tabelle non coincidono. Si tenga in considerazione che le tabelle CLIENTI e ANAGRAFICA potrebbero contenere decine di migliaia di ennuple.

Quesito C.2

In un'applicazione *data-intensive* solitamente la persistenza dei dati è affidata ad un sistema di gestione di basi di dati (DBMS). La candidata/il candidato:

- illustri il concetto di transazione nel contesto di un DBMS, spiegando, anche attraverso esempi, le cosiddette proprietà ACID ed i livelli di isolamento in SQL.

Recentemente si sono affermati DBMS cosiddetti NoSQL, che non adottano un modello relazionale dei dati. La candidata/il candidato:

- illustri almeno tre tipologie di sistemi NoSQL evidenziando per ciascuno gli scenari di utilizzo più appropriati.

Una applicazione *web* è naturalmente esposta ad attacchi *cyber*. La candidata/il candidato:

- descriva, anche attraverso esempi concreti, la vulnerabilità SQL-*injection* e illustri le contromisure da adottare in fase di progettazione e sviluppo del codice.

D. *DISTRIBUTED LEDGER TECHNOLOGY* E INTELLIGENZA ARTIFICIALE (un quesito a scelta)

Quesito D.1

Le *Distributed Ledger Technologies* (DLT) si basano su "registri distribuiti", archivi di informazioni conservati su nodi connessi tra loro e costantemente sincronizzati e verificati mediante l'utilizzo di meccanismi di consenso. La *blockchain* è una particolare DLT nella quale i dati sono organizzati in blocchi concatenati in maniera crittograficamente sicura. Una caratteristica importante delle *blockchain* è quella della programmabilità, cioè la capacità di eseguire specifiche transazioni secondo logiche predeterminate (*smart contract*), implementando la *business logic* direttamente all'interno del protocollo. La candidata/il candidato:

- illustri le caratteristiche principali delle *blockchain Bitcoin* ed *Ethereum*, facendo riferimento, in particolare, agli obiettivi perseguiti e al modello di stato adottato;
- con riferimento agli aspetti di programmabilità, descriva le caratteristiche e gli ambiti di utilizzo dei linguaggi di *scripting* (es. *Bitcoin*) e dei linguaggi di alto livello (es. *Ethereum*).

Inoltre, considerata la *blockchain Ethereum*, la candidata/il candidato:

- illustri il concetto di *gas* in relazione ai costi di *deployment* ed esecuzione di *smart contract*, la problematica del *out-of-gas-condition* e i possibili rimedi;
- descriva la finalità e la modalità di funzionamento degli oracoli e il loro potenziale impatto sulle *performance* della *blockchain*.

Quesito D.2

Con *machine learning* si intende un'area dell'intelligenza artificiale costituita da algoritmi che apprendono sulla base dell'osservazione empirica dei dati mediante un processo di generalizzazione. Facendo riferimento ai processi di apprendimento, si distinguono approcci supervisionati, non supervisionati e di rinforzo (*reinforcement learning*). La candidata/il candidato:

- descriva le tecniche di apprendimento non supervisionato e in tale ambito illustri cosa si intende per problemi di clusterizzazione;
- descriva il funzionamento dell'algoritmo *K-means* specificando i passi di calcolo che conducono all'individuazione dei centroidi e all'assegnazione degli oggetti a ciascun *cluster* evidenziando punti di forza e di debolezza dell'algoritmo;
- illustri due esempi di metriche di qualità dei *cluster* ottenuti e una metodologia per la determinazione del numero K di *cluster*;
- discuta il problema della *curse of dimensionality* fornendo almeno un esempio di misura di mitigazione.

TRACCIA DI LINGUA INGLESE

Do you think the constant growth of online shopping will sooner or later lead to the 'death' of physical shops? Why/why not?