

Ma. 1 - Verde 5

COLLOQUIO- LISTA DOMANDE 1

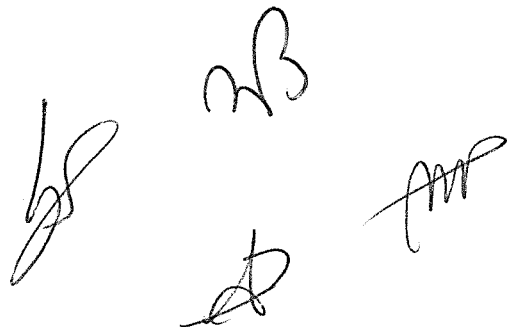
A)

Si supponga di voler memorizzare circa 100 TB di dati da rendere disponibili ad un gruppo di ricercatori dislocati in diverse zone del pianeta. Il candidato descriva, secondo le sue conoscenze, vantaggi e svantaggi di un sistema di memorizzazione tradizionale (ad esempio NAS file server) rispetto ad un sistema di tipo cloud storage (ad. es. Google Drive, Amazon Web Services, Dropbox, ...) con particolare riferimento a:

- 1) facilità di accesso ai dati
- 2) problemi di privacy
- 3) costi di acquisizione / manutenzione
- 4) resilienza dei dati (intesa come disponibilità del dato rispetto ad un guasto)

B)

Il professor Carlo Goschei lavora nell'ambito della patologia digitale e uno degli obiettivi della sua ricerca è quello di definire algoritmi per rilevare la presenza di cellule tumorali da immagini istologiche digitalizzate. Utilizzando metodologie di Deep Learning deve lavorare con centinaia o migliaia di immagini ad alta risoluzione. Si descrivano i componenti principali e il dimensionamento di un server per gestire i dati del Prof. Goschei e per eseguire i suoi frequenti esperimenti.



Handwritten signatures and initials, including '2B' and 'AM'.

COLLOQUIO– LISTA DOMANDE 2

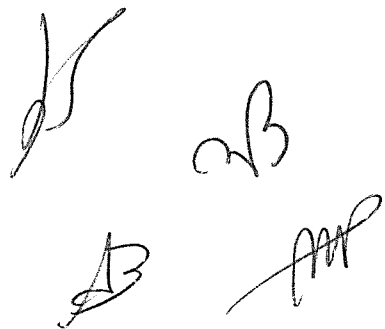
A)

Il prof. Tagliacavalli dirige un laboratorio in cui uno strumento diagnostico acquisisce quotidianamente 200 GB di dati. I dati vengono memorizzati in un PC la cui affidabilità non è sufficiente a far dormire tranquillo il professore. Si descriva un sistema di memorizzazione dei dati "gerarchico" che permetta, in momenti successivi, di:

- 1) acquisire velocemente i dati
- 2) scaricare velocemente i dati acquisiti su uno storage molto capacitivo
- 3) archiviare a lungo termine i dati del punto 2

B)

La professoressa Doris Codata si occupa dell'analisi testuale dei rapporti medici dai quali estrae entità e relazioni rilevanti atte alla creazione di grafi di conoscenza che rappresentino il contenuto dei rapporti medici. La Prof. Codata ha la necessità di gestire, accedere, elaborare e ricercare i rapporti medici e i grafi creati a partire dai rapporti. Quali sono i passi principali che seguireste per arrivare alla definizione e implementazione di una o più basi di dati utili alla Prof. Codata?.



Handwritten signatures and initials, including a stylized signature, the initials 'AB', and another signature.

COLLOQUIO– LISTA DOMANDE 3

A)

Il candidato descriva, secondo le sue conoscenze, un possibile sistema di backup per i due casi seguenti:

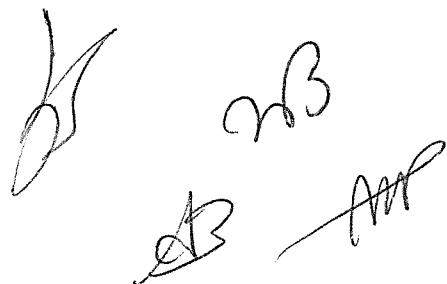
1. 500 GB di dati
2. 500 TB di dati

Si considerino, tra i possibili parametri:

- la retention (quanto indietro nel tempo si vuole poter andare)
- i tempi per effettuare backup / restore

B)

Il professor Gennaro Doschei si occupa dell'analisi testuale dei rapporti medici con l'obiettivo di estrarre entità e relazioni rilevanti alla creazione di grafi di conoscenza che rappresentino in modo "machine-readable" contenuto dei rapporti medici. Doschei lavora con milioni di documenti testuali di piccole-medie dimensioni che deve continuamente leggere, elaborare, modificare e salvare in un nuovo formato, tutto in modo automatico. Si descrivano i componenti principali e il dimensionamento di un server per gestire i dati e gli esperimenti del Prof. Doschei.

Handwritten signatures and initials in black ink, including a stylized signature, the letters 'NB', and another signature.

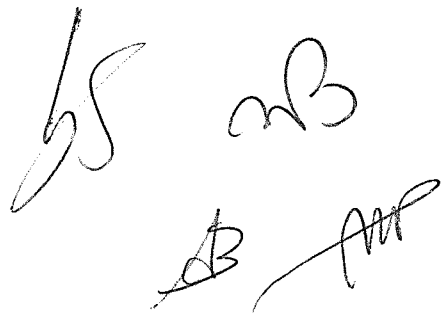
COLLOQUIO– LISTA DOMANDE 4

A)

Si vuole acquisire un sistema di storage per memorizzare i dati personali (servire le home degli utenti) di un gruppo eterogeneo di utenti: client Linux, Windows, Mac (?). Il candidato descriva, secondo le sue conoscenze, quali sono le problematiche da affrontare relativamente all' accesso ai dati da parte di sistemi operativi diversi.

B)

La professoressa Giovanna Maidati deve gestire una grande quantità di dati riguardanti geni, malattie, proteine, molecole e cellule. In particolare, la sua ricerca si basa sull'individuare le interazioni o i "pathways" tra le molecole di una cellula o tra le proteine nel corpo umano. Un esempio di query frequente che la Prof. Maidati dovrà eseguire è "trovare tutte le interazioni comuni (percorsi comuni) delle proteine HBA e HBB" oppure "trovare il percorso più breve che collega e porta a interagire due specifiche molecole". Il candidato spieghi quale DBMS utilizzerebbe per gestire i dati della Prof. Maidati e quale linguaggio di querying è il più indicato per rispondere alle query tipiche della professoressa.

Handwritten signatures and initials in black ink, including a large stylized signature, the initials 'MB', and other scribbled marks.