



SELEZIONE PUBBLICA, PER TITOLI ED ESAMI, PER IL RECLUTAMENTO DI N. 1 UNITÀ DI TECNOLOGO DI SECONDO LIVELLO CATEGORIA D - POSIZIONE ECONOMICA D3, CON RAPPORTO DI LAVORO SUBORDINATO A TEMPO DETERMINATO, DELLA DURATA DI 24 MESI, PRESSO IL DIPARTIMENTO DI ECONOMIA, MANAGEMENT E METODI QUANTITATIVI, PER LO SVOLGIMENTO DI ATTIVITÀ DI SUPPORTO TECNICO E AMMINISTRATIVO COME FIGURA DI *DATA MANAGER*, DEDICATA ALLA PREPARAZIONE E GESTIONE DEI DATASET FACILITANDO IL REPERIMENTO E L'ANALISI DI BIG DATA, NELL'AMBITO DEL PROGETTO DIPARTIMENTO DI ECCELLENZA - BANDITA CON DETERMINA N. 8818/2019 DEL 27/05/2019 E PUBBLICATA SUL SITO INTERNET D'ATENEIO IN DATA 27/05/2019 - CODICE 20025

TRACCE PROVA SCRITTA A CONTENUTO TEORICO-PRATICO

La Commissione giudicatrice della selezione, nominata con determina n. 10127/2019 del 14/06/2019 e così composta:

PROF. RENTOCCHINI FRANCESCO - PRESIDENTE

PROF.SSA LEORATO SAMANTHA - COMPONENTE

DOTT. TARASCONI GIANLUCA - COMPONENTE

SIG.RA SCOLARI LAURA - SEGRETARIO

comunica le tracce relative alla prova scritta a contenuto teorico-pratico:

TEMA n. 1

Database relazionali

La compagnia UNTER è specializzata nel campo del trasporto persone.

Per gestire la sua attività UNTER usa un database relazionale in cui vengono immagazzinati i dati relativi ai guidatori (età, indirizzo, contatti, tipo patente) e il parco macchine (modello, targa, data acquisto, date manutenzioni, km effettuati).

Il database gestisce anche i turni dei guidatori (giornaliero, mattino e notturno), l'assegnazione dei guidatori alle macchine e quanto incassato da ogni corsa.

DOMANDA 1) disegnare il diagramma entità relazioni del relazionale sopra descritto.

DOMANDA 2) per vincoli di sicurezza i guidatori non possono lavorare più di 8 ore al giorno e solo cinque giorni alla settimana, mentre le macchine devono essere sottoposte a revisione ogni 10.000 Km. Produrre due scripts SQL/R/Stata che estraggano alla data odierna eventuali guidatori che violino i vincoli descritti ed autovetture da mandare in revisione.

PYTHON

Definire la differenza tra i tipi dato lista e tupla. Fornire un esempio di tupla ed un possibile esempio di codice per estrarne gli ultimi due elementi.

Altri tipi di DB:

Descrivete le principali differenze tra un database di tipo relazionale e un cluster di tipo HDFS, ponendo enfasi su vantaggi e svantaggi dell'uno e dell'altro tipo.

TEMA n. 2

Database relazionali

La compagnia FLUGANSA è specializzata nel campo del trasporto aereo passeggeri.

FR LA SA GL



UNIVERSITÀ DEGLI STUDI DI MILANO

Per gestire la sua attività FLUGANSA usa un database relazionale in cui vengono immagazzinati i dati relativi ai piloti (età, indirizzo, contatti, data conseguimento brevetto di volo), il parco velivoli (modello, data acquisto, date manutenzioni, km effettuati).

Il database gestisce anche i turni dei piloti, l'assegnazione dei piloti agli aerei e quanto incassato da ogni volo a/r.

DOMANDA 1) disegnare il diagramma entità relazioni del relazionale sopra descritto.

DOMANDA 2) per vincoli di sicurezza i piloti non possono lavorare più di 60 ore alla settimana e 290 ore al mese, mentre gli aerei devono essere sottoposti a controllo ogni 600 ore volo. Produrre due scripts SQL/R/Stata che estraggano alla data odierna eventuali piloti che violino i vincoli descritti e aerei da mandare al controllo per la manutenzione.

PYTHON

Dato il seguente codice Python spiegare quale sarà il possibile output se, per esempio line= "00014 UNIV HELSINKI, DEPT FOREST ECOL, HELSINKI, FINLAND"

```
rx_dict = {'City': re.compile(r'(?P<City>(\w+)), \w+$'),
           'Country': re.compile(r', (?P<Country>(\w+))$'),
           'Zip': re.compile(r'(?P<Zip>(\d{3,5}))' ) }
```

```
for key, rx in rx_dict.items():
    match = rx.search(line)
    if match:
        print(key, match.group(key))
```

Altri tipi di DB

In che relazione sono tra loro in Pig Latin il tipo dati BAG e il tipo TUPLA? Fate un esempio evidenziando qual è il dato BAG e quale la TUPLA.

TEMA n. 3

Database relazionali

La compagnia TIR è specializzata nel campo del trasporto di merci.

Per gestire la sua attività TIR usa un database relazionale in cui vengono immagazzinati i dati relativi ai piloti (età, indirizzo, contatti, data di scadenza patente) e il parco autovetture (modello, targa, data acquisto, date manutenzioni, km effettuati).

Il database gestisce anche i turni dei piloti, l'assegnazione dei piloti alle autovetture e quanto incassato da ogni spedizione.

DOMANDA 1) disegnare il diagramma entità relazioni del relazionale sopra descritto.

DOMANDA 2) per vincoli di sicurezza i guidatori non possono lavorare più di 10 ore al giorno e solo quattro giorni alla settimana, mentre gli automezzi devono essere sottoposti a revisione ogni 25.000 Km. Produrre due scripts SQL/R/Stata che estraggano alla data odierna eventuali guidatori che violino i vincoli descritti ed autovetture da mandare in revisione.

PYTHON

Dato un PANDAS dataframe contenente una lista di pubblicazioni scientifiche con i seguenti campi:

	UID	addr_num	organization	suborganization	full_address
0	WOS:000071806400001	1	UNIVERSITA' DI MODENA E REGGIO EMILIA	NaN	UNIV MODENA, MILAN, ITALY

FR AS 52



UNIVERSITÀ DEGLI STUDI DI MILANO

1	WOS:00007180640000 1	2	UNIV STATALE MILANO	NaN	UNIV STATALE DI MILANO, MILANO, ITALY
---	-------------------------	---	---------------------	-----	---

[...]

Aggiungete un campo STATALE_BIN che venga popolato tramite una funzione lambda con i valori 1, se l'affiliazione è UNIV STATALE MILANO, 0 altrimenti.

Altri tipi di DB

Se avessi un file esempio.csv nella directory /temp con i seguenti campi, in formato comma separated:

patent title	publication_number	owner_id	publication_date
HYDRAULIC ACTUATOR WITH MECHANICAL PISTON POSITION FEEDBACK	WO2018136981(A2)	US140689340	26/07/2018
SENSORLESS MANIFOLD ASSEMBLY WITH PRESSURE-BASED REVERSING FLUID CIRCUIT	WO2018136980(A1)	US140689340	26/07/2018
WIRELESS CHARGING SYSTEM WITH MULTI-COIL SCANNING	WO2018136940(A1)	NL60166878 US930584541	26/07/2018
EQUIPMENT AND METHOD FOR THREE-DIMENSIONAL RADIANCE AND GAS SPECIES FIELD ESTIMATION IN AN OPEN COMBUSTION ENVIRONMENT	WO2018136911(A1)	US222640650	26/07/2018
RANSOMWARE RESILIENT CLOUD SERVICES	WO2018136862(A1)	US911144442	26/07/2018
EXPERT-AUGMENTED MACHINE LEARNING FOR CONDITION MONITORING	WO2018136841(A1)	US222640650	26/07/2018
SEMI-PERSISTENT SCHEDULING FOR AUTONOMOUS TRANSMISSION ACTIVATION AND RELEASE	WO2018136815(A1)	US941672743	26/07/2018
SECONDARY TIMING ADVANCE GROUPS WITH ONLY LICENSE ASSISTED ACCESS SECONDARY CELLS	WO2018136770(A1)	US953685934	26/07/2018

Scrivere le istruzioni per importare in ambiente Hive i dati usando l'istruzione SerDe

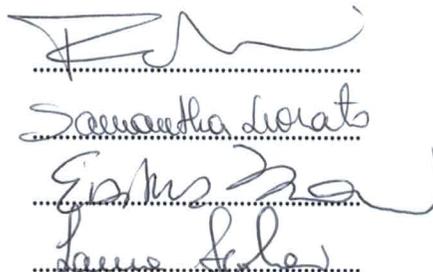
LA COMMISSIONE

PROF. RENTOCCHINI FRANCESCO - PRESIDENTE

PROF.SSA LEORATO SAMANTHA - COMPONENTE

DOTT. TARASCONI GIANLUCA - COMPONENTE

SIG.RA SCOLARI LAURA - SEGRETARIO


.....
.....
.....
.....