# Πρότυπο αναφοράς άσκησης

Συστήματα Διαχείρισης Δεδομένων Μεγάλου Όγκου

Εργαστηριακή Άσκηση 2023/24

|  |  |  |
| --- | --- | --- |
| **Ονομα** | **Επώνυμο** | **ΑΜ** |
|  |  |  |
|  |  |  |

Βεβαιώνω ότι είμαι συγγραφέας της παρούσας εργασίας και ότι έχω αναφέρει ή παραπέμψει σε αυτήν, ρητά και συγκεκριμένα, όλες τις πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, προτάσεων ή λέξεων, είτε αυτές μεταφέρονται επακριβώς (στο πρωτότυπο ή μεταφρασμένες) είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για το συγκεκριμένο μάθημα/σεμινάριο/πρόγραμμα σπουδών.

Έχω ενημερωθεί ότι σύμφωνα με τον εσωτερικό κανονισμό λειτουργίας του Πανεπιστημίου Πατρών άρθρο 50§6, τυχόν προσπάθεια αντιγραφής ή εν γένει φαλκίδευσης της εξεταστικής και εκπαιδευτικής διαδικασίας από οιονδήποτε εξεταζόμενο, πέραν του μηδενισμού, συνιστά βαρύ πειθαρχικό παράπτωμα.

|  |  |
| --- | --- |
| Υπογραφή  \_\_\_ / \_\_\_ / 2024 | Υπογραφή  \_\_\_ / \_\_\_ / 2024 |

**Συνημμένα αρχεία κώδικα**

Μαζί με την παρούσα αναφορά υποβάλλουμε τα παρακάτω αρχεία κώδικα

|  |  |  |
| --- | --- | --- |
| Αρχείο | Αφορά το ερώτημα | Περιγραφή/Σχόλιο |
| *Erotima1.py* | *1* | *Περιέχει όλα τα ερωτήματα για το ερ. 1* |
|  |  |  |
|  |  |  |

# Τεχνικά χαρακτηριστικά περιβάλλοντος λειτουργίας

*[Τεχνικά χαρακτηριστικά φυσικού Η/Υ που χρησιμοποιήθηκε για την εργασία, αν χρησιμοποιήθηκε hosted υπηρεσία μπορείτε απλά να αναφέρετε αυτό αντί για τον πίνακα]*

|  |  |
| --- | --- |
| **Χαρακτηριστικό** | **Τιμή** |
| CPU model | *Intel i5-10400F* |
| CPU clock speed | *2.9GHz* |
| Physical CPU cores | *6* |
| Logical CPU cores | *12* |
| RAM | *16* |
| Secondary Storage Type | *~~HDD~~/SSD* |

# Ερώτημα 1: Παραγωγή δεδομένων

*[περιγράψτε τη δημιουργία δεδομένων εξομοίωσης και την λογική του script παραγωγής δεδομένων ως kafka producer. Δώστε screenshots από την επιτυχή εγκατάσταση του Kafka, και τη δοκιμή της λειτουργίας του μοντέλου producer-consumer]*

# Ερώτημα 2: Κατανάλωση και επεξεργασία με Spark

*[περιγράψτε τη λειτουργία του script κατανάλωσης και επεξεργασίας δεδομένων από το Spark, με κατάλληλα screenshots που δείχνουν την ορθή λειτουργία της διαδικασίας]*

# Ερώτημα 3: Αποθήκευση σε MongoDB

*[παραθέστε: 1. Το σχεδιασμό της βάσης και τα statements δημιουργίας των συλλογών, 2. Τη λειτουργία του script αποθήκευσης των ωμών δεδομένων και των αποτελεσμάτων επεξεργασίας από το Spark στη MongoDB με screenshots των αποτελεσμάτων, 3. Την εκτέλεση του script επερωτημάτων προς τη MongoDB μαζί με τα queries που σχεδιάσατε, και τα αποτελέσματα της εκτέλεσης αυτών με κατάλληλα screenshots].*

# Σχολιασμός αποτελεσμάτων

*[Συνοψίστε τα αποτελέσματα της εμπειρίας σας από το project.]*

# Βιβλιογραφία

*[πηγές που χρησιμοποιήσατε για την εργασία]*