

- Που μπορεί να χρησιμοποιηθεί η μέθοδος MapReduce;
- Στο Hadoop, τα αρκετά μεγάλα αρχεία συνήθως σε τι μεγέθους τμήματα χωρίζονται και σε πόσους διαφορετικούς κόμβους αποθηκεύονται;
- Ποια είναι η χρήση του namenode στο Hadoop;
- Τι είναι το PIG;
- Τι συμβαίνει για κάθε project που πρέπει να εκτελεστεί σε ένα πρόγραμμα Hadoop:
- Τι γνωρίζετε σχετικά με τα δεδομένα εξόδου ενός έργου Απεικόνισης και τα δεδομένα εισόδου ενός Συγχώνευσης;
- Τι είναι ο Combiner;
- Τι συμβαίνει σε περίπτωση σφάλματος εκτέλεσης ενός έργου;
- Σε ποιο κόμβο εκτελείται συνήθως ένα έργο Απεικόνισης ή Συγχώνευσης;
- Ποιο είναι το μοντέλο λειτουργίας του HDFS;
- Ποια είναι η μικρότερη μονάδα δεδομένων που μπορεί να γίνει ανάγνωση ή εγγραφή στο HDFS;
- Ποιο είναι το default μέγεθος ενός block;
- Ποια είναι τα κύρια χαρακτηριστικά του Apache Spark?
- Ποια είναι τα μέρη του οικοσυστήματος του Apache Spark;
- Ποια γλώσσα προγραμματισμού δεν υποστηρίζεται από το Apache Spark
- Ποιος είναι ο πιο απλός τύπος βάσης δεδομένων NoSQL;
- Ποιο σύστημα χρησιμοποιεί το Apache Spark για να αναλύσει δεδομένα σε πραγματικό χρόνο;
- Ποιο τύπο δεδομένων αξιοποιεί πλήρως η βιβλιοθήκη Mlib του Apache Spark: Που βρίσκεται το μεγαλύτερο cluster Hadoop;
- Πως ονομάζεται η ανάλυση δεδομένων με σκοπό την ανέρευση κρυφών μοτίβων ή συσχετίσεων;
- Πόσα είναι τα στάδια επεξεργασίας ενός προγράμματος MapReduce?