

ΤΕΙ ΠΕΙΡΑΙΑ | ΤΜΗΜΑ ΗΥΣ

Θέμα:

**ΜΕΛΕΤΗ – ΣΧΕΔΙΑΣΗ ΕΦΑΡΜΟΓΗΣ ΣΕ
ΥΠΟΛΟΓΙΣΤΙΚΟ ΝΕΦΟΣ (CLOUD COMPUTING) ΜΕ
ΕΜΦΑΣΗ ΣΤΗΝ ΚΑΤΑΣΚΕΥΗ ΔΕΝΤΡΩΝ.**

Εισηγητής: Δ. Ν. Καλλέργης, MSc.

Φοιτήτρια: Κοντζοπούλου Παναγιώτα

Εισαγωγή

- Η ανάγκη για αποθήκευση και επεξεργασία δεδομένων έχει αυξηθεί κατακόρυφα τα τελευταία χρόνια.
- Το Facebook δημιουργεί 10TB δεδομένων/ημέρα.
- Εκτιμάται πως 2020 το σύνολο των δεδομένων θα είναι 35ZB.
- Την λύση για την αποτελεσματική διαχείριση των δεδομένων έρχονται να την δώσουν τα κατανεμημένα συστήματα και το υπολογιστικό νέφος (Cloud Computing).

Υπολογιστικό Νέφος

- Αποτελεί μια νέα σχετικά προσέγγιση στον χώρο των κατανεμημένων συστημάτων.
- Σχετικά με τον ορισμό:
 - Έχουν γίνει πολλές προσπάθειες οι οποίες δε μπορούν να καλύψουν όλες τις πτυχές του. Αυτό έχει σαν αποτέλεσμα την γενίκευση της έννοιας με τέτοιο τρόπο που κάθε σύστημα το οποίο επιτρέπει την ανάθεση υπολογιστικών και αποθηκευτικών υπηρεσιών «εξωτερικά» να θεωρείται υπολογιστικό νέφος.
- Ιστορικά αποτελεί εξέλιξη του υπολογιστικού πλέγματος, των υπολογιστικών δημόσιων υπηρεσιών και των γραφείων παροχής υπηρεσιών.

Χαρακτηριστικά Υπολογιστικού Νέφους

- On demand Self-Service
- Multi-tenancy
- Ελαστικότητα
- Ευρεία πρόσβαση μέσω δικτύου
- Πληρωμή ανάλογα με τη χρήση
- Κλιμακωσιμότητα
- Διαθεσιμότητα
- Εξισορρόπηση φορτίου

Μοντέλα Υπολογιστικού Νέφους (1)

- Λογισμικό ως Υπηρεσία (Software as a Service)
 - Βασίζεται στη λογική της υπενοικίασης λογισμικού από έναν πάροχο υπηρεσιών.
 - Η υπηρεσία διατίθεται μέσω διαδικτύου
 - Δημοφιλές μοντέλο λόγω ευελιξίας, ποιότητας των υπηρεσιών, σταθερότητας και ελάχιστης δυνατής συντήρησης.
 - Εφαρμογές: Google Apps, Office 365
- Πλατφόρμα ως Υπηρεσία (Platform as a Service)
 - Πλατφόρμα που διατίθεται για την ανάπτυξη λογισμικού
 - Παροχή κατάλληλων υπηρεσιών για τη δυνατότητα ανάπτυξης, δοκιμής, διάθεσης και συντήρησης εφαρμογών.
 - Λεπτομερής έλεγχος του λειτουργικού συστήματος και της πλατφόρμας
 - Εφαρμογές: Google App Engine, Heroku

Μοντέλα Υπολογιστικού Νέφους (2)

- Υποδομή ως Υπηρεσία (Infrastructure as a Service)
 - Χρησιμοποιείται κυρίως για υπολογισμούς.
 - Υπενοικίαση υπολογιστικών και δικτυακών υποδομών ανάλογα με τις ανάγκες.
 - Παροχή εικονικών μηχανών που βρίσκονται στο «σύννεφο» και ελέγχονται μέσω ενός API.
 - Κατηγοριοποίηση των υπηρεσιών:
 - Υπολογιστικοί πόροι: Εικονικά μηχανήματα (Amazon Web Services, Microsoft Azure, OpenStack, Eucalyptus)
 - Αποθηκευτική δυνατότητα: Data-Storage as a Service (Windows Live SkyDrive, Ubuntu One, DropBox)
 - Δικτυακή επικοινωνία: Communication as a Service (Cisco Hosted Unified Communications Services)

Μοντέλα Ανάπτυξης Υπολογιστικού Νέφους

- Δημόσιο Υπολογιστικό σύννεφο (Public Cloud)
- Ιδιωτικό Υπολογιστικό σύννεφο (Private Cloud)
- Κοινοτικό Υπολογιστικό σύννεφο (Community Cloud)
- Υβριδικό Υπολογιστικό σύννεφο (Hybrid Cloud)

Apache Hadoop

- Το Hadoop είναι ένα open source framework για την εκτέλεση εφαρμογών σε δίκτυα υπολογιστών χωρίς την ανάγκη χρήσης εξειδικευμένου hardware.
- Είναι γραμμένο σε Java.
- Επιτρέπει τη δημιουργία και την εκτέλεση εφαρμογών επεξεργασίας μεγάλου όγκου δεδομένων.



Ιστορική Αναδρομή

- Δημιουργήθηκε από τον Doug Cutting.
- Αναπτύχθηκε στα πλαίσια της διαδικτυακής μηχανής Apache Nutch το 2002.
- Το 2003 υλοποιήθηκε το NDFS και το MapReduce για το Nutch.
- Το 2006 το NDFS και το MapReduce αποτέλεσαν ξεχωριστό έργο που ονομάστηκε Hadoop.
- Τον Φεβρουάριο του 2008 εγκαθίσταται από τη Yahoo!

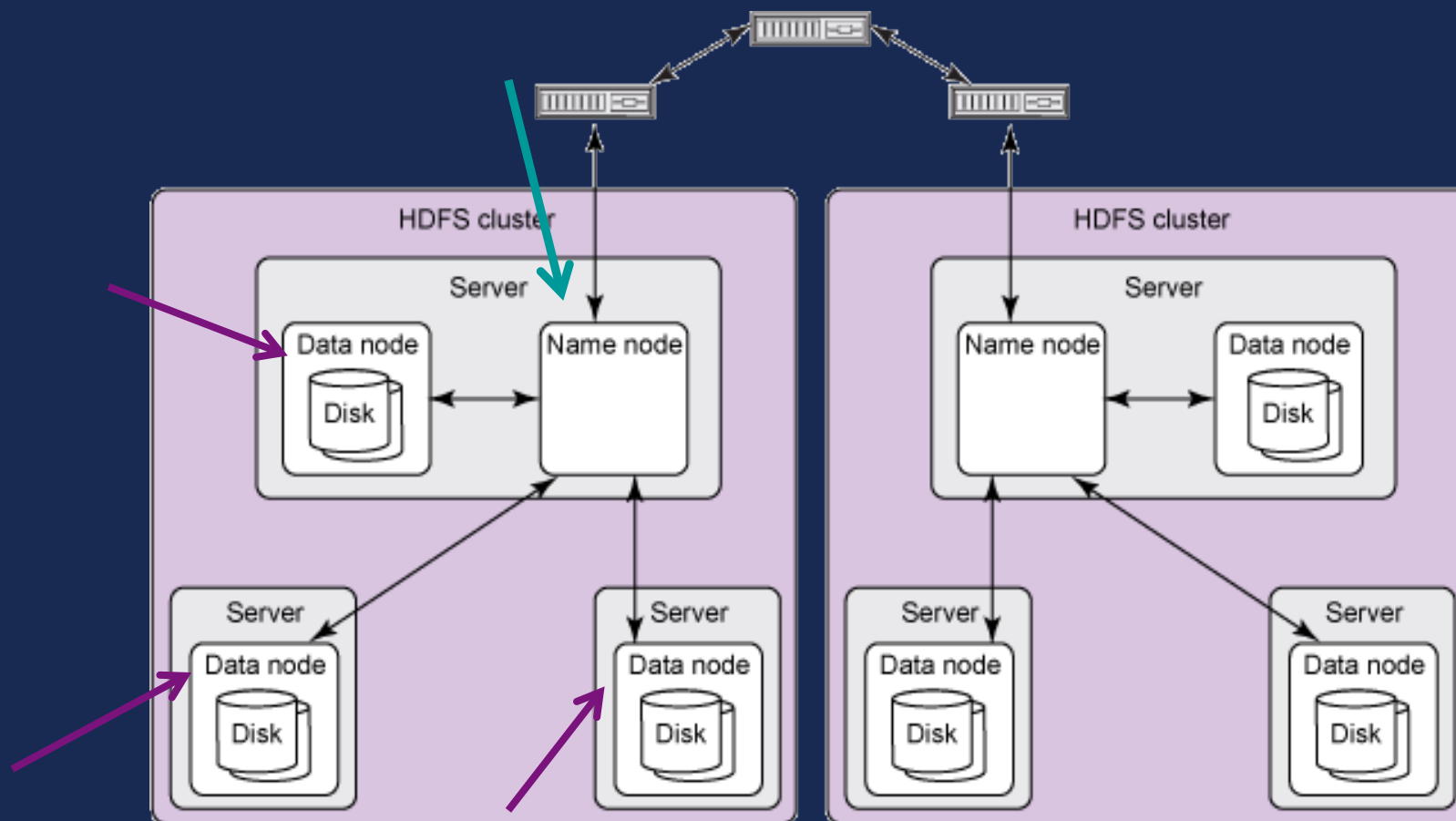
HDFS

- Είναι το κατανεμημένο σύστημα αρχείων που χρησιμοποιεί το Hadoop.
- Χαρακτηριστικά:
 - Υψηλή ανοχή σε σφάλματα
 - Streaming πρόσβαση στα δεδομένα
 - Μεγάλος όγκος δεδομένων εισόδου
 - Μοντέλο WORM (Write-Once-Read-Many)
 - Εκτέλεση επεξεργασίας κοντά στα δεδομένα

Αρχιτεκτονική του HDFS

- Το HDFS ακολουθεί την αρχιτεκτονική Master/Slave.
- Ένα Cluster αποτελείται από:
 - Έναν NameNode, που παίζει τον ρόλο του master.
 - Διαχειρίζεται το namespace του συστήματος αρχείων.
 - Ρυθμίζει την πρόσβαση των χρηστών στα αρχεία του HDFS.
 - Διατηρεί πληροφορίες σχετικά με τη χαρτογράφηση των block στους DataNodes.
 - Τους DataNodes, που παίζουν τον ρόλο των slave.
 - Συνήθως τρέχει ένας σε κάθε κόμβο.
 - Εξυπηρετούν αιτήσεις ανάγνωσης και εγγραφής .
 - Δημιουργούν, διαγράφουν και αντιγράφουν block σύμφωνα με τις οδηγίες του NameNode.

Αρχιτεκτονική του HDFS



MapReduce

- Framework σχεδιασμένο για κατανεμημένους υπολογισμούς μεγάλου όγκου δεδομένων.
- Καταμερισμός εργασιών σε μεγάλο αριθμό μηχανημάτων.
- Προγράμματα σε MapReduce μπορούν να γραφούν σε διάφορες γλώσσες προγραμματισμού όπως Java, C++, Python.
- Αποτελείται από δύο βασικές λειτουργίες, την Map και την Reduce.
- Το MapReduce χρησιμοποιείται κυρίως για:
 - Log Processing
 - Web Index Building
 - Data mining
 - Machine Learning

Λειτουργία του MapReduce

- Συνάρτηση map
 - Παίρνει ως είσοδο ζευγάρια δεδομένων τύπου (κλειδί, τιμή).
 - Τα αποτελέσματα ομαδοποιούνται με βάση το κλειδί και δίνονται ως είσοδος στη συνάρτηση reduce.
- Συνάρτηση reduce
 - Παίρνει ως είσοδο ζευγάρια δεδομένων τύπου (κλειδί, λίστα τιμών).
 - Τα αποτελέσματα της συνάρτησης ομαδοποιούνται και αυτό είναι το τελικό ζητούμενο.

Υλοποίηση Εφαρμογής

Υλοποίηση του προγράμματος

- Η υλοποίηση του προγράμματος έγινε σε Java.
- Για τη δημιουργία δέντρων επιθεμάτων χρησιμοποιήθηκε ο αλγόριθμος του Ukkonen.
- Έγινε παραλληλοποίηση του αλγορίθμου με τη βοήθεια του MapReduce.
- Σκοπός του προγράμματος είναι η εύρεση της μεγαλύτερης παλίνδρομης υπό-συμβολοσειράς.

Αλγόριθμος του Ukkonen

- Παρουσιάστηκε το 1995 από τον Esko Ukkonen.
- Είναι ένας online αλγόριθμος κατασκευής δέντρων επιθεμάτων.
- Διαβάζει την είσοδο από αριστερά προς τα δεξιά.
- Για μια συμβολοσειρά n χαρακτήρων λειτουργεί σε n φάσεις, προσθέτοντας κάθε φορά τον επόμενο χαρακτήρα.
- Η χρονική και χωρική πολυπλοκότητα είναι $O(n)$

Εφαρμογή

- Για την υλοποίηση της εφαρμογής και την πραγματοποίηση των μετρήσεων δημιουργήθηκε η ανάγκη να στηθεί το Hadoop σε ένα δίκτυο υπολογιστών.
- Το Hadoop έχει σχεδιαστεί ώστε να τρέχει σε οποιοδήποτε hardware, ενώ οι κόμβοι που συνιστούν το δίκτυο υπολογιστών μπορούν να έχουν διαφορετικά χαρακτηριστικά.
- Στην παρούσα εφαρμογή επιλέχθηκε να γίνει η εγκατάσταση στη πλατφόρμα Okeanos Public Cloud.

Okeanos Public Cloud

- Υλοποιήθηκε από το Εθνικό Δίκτυο Έρευνας και Τεχνολογίας.
- Προσφέρει δωρεάν υπηρεσίες υπολογιστικού νέφους στην ακαδημαϊκή κοινότητα.
- Παρέχει αποθηκευτικό χώρο για τη φιλοξενία αρχείων και εικονικά μηχανήματα με τη δυνατότητα δημιουργίας διαφορετικών δικτύων μεταξύ αυτών.
- Η διαχείριση γίνεται μέσω ενός διαδικτυακού περιβάλλοντος.

Εγκατάσταση του Hadoop

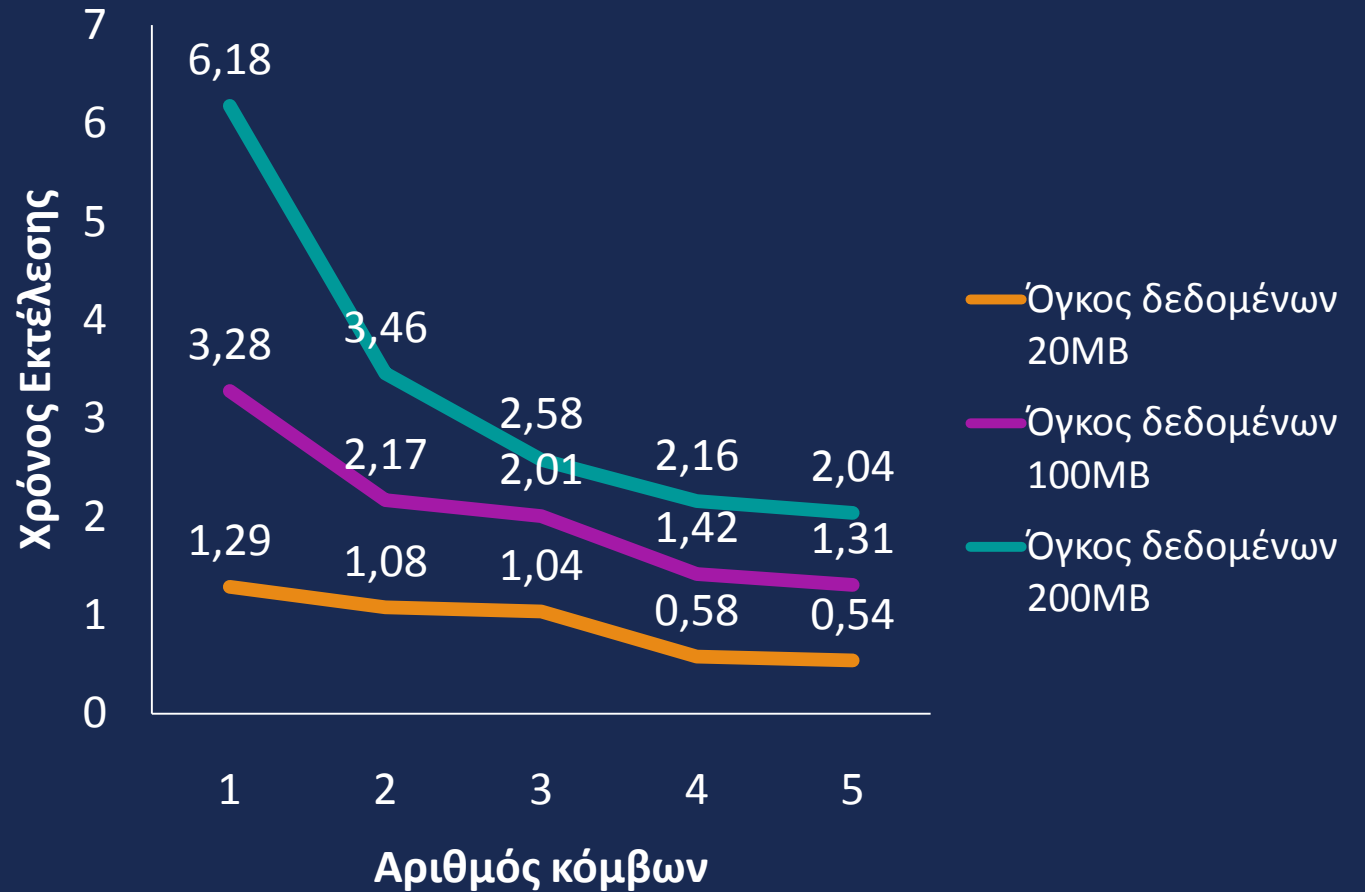
- Έγινε εγκατάσταση του Hadoop 0.20.4.
- Σε ένα δίκτυο υπολογιστών αποτελούμενο από 6 εικονικά μηχανήματα.
- Όλα τα εικονικά μηχανήματα έχουν τα ίδια χαρακτηριστικά τα οποία είναι:
 - Διπύρηνος επεξεργαστής
 - 2GB Μνήμη RAM
 - 20GB Σκληρό δίσκο
- Όλα τα εικονικά μηχανήματα τρέχουν Ubuntu 12.04.

Πειραματικό στάδιο

- Πραγματοποιήθηκαν ενδεικτικές μετρήσεις με διαφορετικού όγκου πακέτα δεδομένων εισόδου.
- Τα πακέτα ήταν της τάξης:
 - 20MB
 - 100MB
 - 200MB
- Η επεξεργασία έγινε σε ένα δίκτυο υπολογιστών αποτελούμενο, αρχικά, από 2 μηχανήματα(1 master – 1 slave) ενώ το πλήθος των slave αυξάνεται κατά 1.

Πειραματικό στάδιο

- Στο διπλανό διάγραμμα απεικονίζεται ο χρόνος εκτέλεσης του προγράμματος σύμφωνα με τον αριθμό των κόμβων για τα τρία πακέτα δεδομένων.



Τέλος

Σας ευχαριστώ για τον χρόνο σας!