

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Κατανεμημένα Συστήματα

**Ενότητα # 10: Κατανεμημένα συστήματα αποθήκευσης**

**Διδάσκων: Γεώργιος Ξυλωμένος**

**Τμήμα: Πληροφορικής**



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Οικονομικό Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Οι εικόνες προέρχονται από το βιβλίο «Κατανεμημένα Συστήματα με Java», Ι. Κάβουρας, Ι. Μήλης, Γ. Ξυλωμένος, Α. Ρουκουνάκη, 3<sup>η</sup> έκδοση, 2011, Εκδόσεις Κλειδάριθμος.



# Σκοποί ενότητας

- Εξοικείωση με τα βασικά συστατικά της κατανεμημένης αποθήκευσης (ονόματα, αναγνωριστικά, εξυπηρετητές, προσωρινή αποθήκευση, σημασιολογία και συνέπεια).
- Κατανόηση των προσεγγίσεων πελάτη-εξυπηρετητή στην κατανεμημένη αποθήκευση μέσω των συστημάτων NFS και AFS.
- Εισαγωγή στην κατανεμημένη αποθήκευση μέσω ομοτίμων στα συστήματα PAST και CFS.

# Περιεχόμενα ενότητας (1 από 2)

- Διεπαφές αρχείων και ευρετηρίων
- Ονόματα και αναγνωριστικά αρχείων
- Εξυπηρετητές αρχείων και ευρετηρίων
- Προσωρινή αποθήκευση αρχείων
- Σημασιολογία καταμερισμού αρχείων
- Συνέπεια συστήματος αρχείων

# Περιεχόμενα ενότητας (2 από 2)

- Το σύστημα NFS
- Το σύστημα AFS
- Το σύστημα PAST
- Το σύστημα CFS

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Διεπαφές αρχείων και ευρετηρίων

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Αρχεία και ευρετήρια (1 από 6)

- Υπηρεσίες συστήματος αρχείων
  - Αποθήκευση και προσπέλαση λογικών αρχείων
  - Διαχείριση αρχείων και ιδιοτήτων τους
- Οργάνωση αρχείων σε ευρετήρια
  - Τα ευρετήρια είναι αρχεία ειδικού τύπου
  - Αντιστοίχιση ονομάτων σε αναγνωριστικά
  - Ένα ευρετήριο μπορεί να περιέχει άλλα



# Αρχεία και ευρετήρια (2 από 6)

- Κατανεμημένη υπηρεσία αρχείων
  - Προσπέλαση σε απομακρυσμένα αρχεία
  - Ίδια διεπαφή με αυτή των τοπικών αρχείων
  - Οι χρήστες βλέπουν τα αρχεία τους παντού
  - Χρησιμοποιείται ως βάση για άλλες υπηρεσίες
    - Κατανεμημένες υπηρεσίες ονομασίας και ευρετηρίων

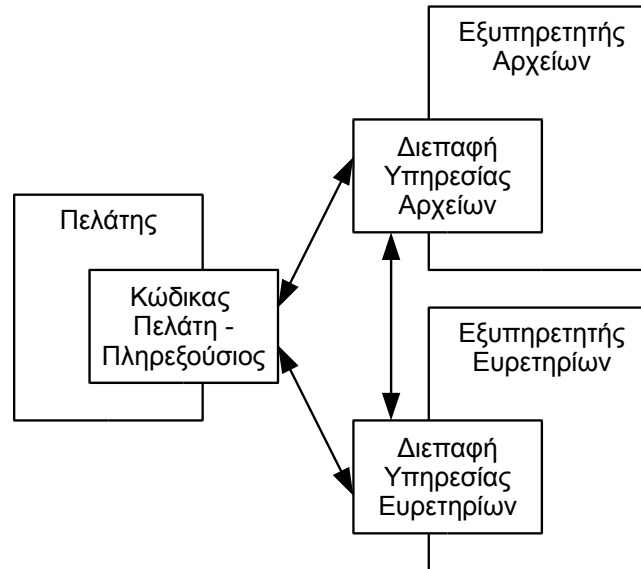
# Αρχεία και ευρετήρια (3 από 6)

- Μοντέλο πελάτη-εξυπηρετητή
  - Περιορισμένο και σταθερό σύνολο εξυπηρετητών
    - Είτε ο πελάτης τους γνωρίζει όλους
    - Είτε οι εξυπηρετητές γνωρίζονται μεταξύ τους
- Μοντέλο ομοτίμων
  - Απεικόνιση αιτημάτων σε εξυπηρετητές
    - Με κατανεμημένους πίνακες κατακερματισμού
    - Μεγάλο και ασταθές σύνολο εξυπηρετητών

# Αρχεία και ευρετήρια (4 από 6)

- Λειτουργίες συστήματος αρχείων
  - Άνοιγμα, κλείσιμο, δημιουργία και διαγραφή
  - Ανάγνωση και εγγραφή στοιχείων, αναζήτηση
  - Ανάγνωση και εγγραφή ιδιοτήτων
- Υλοποίηση σε κατανεμημένο σύστημα
  - Διάκριση υπηρεσίας αρχείων και ευρετηρίων
    - Χρήση διαφορετικών εξυπηρετητών

# Αρχεία και ευρετήρια (5 από 6)



- Συνδυασμός υπηρεσιών αρχείων και ευρετηρίων
  - Κάθε νέο αρχείο λαμβάνει ένα αναγνωριστικό
  - Καταχωρείται στο ευρετήριο μαζί με το όνομα
  - Όλες οι λειτουργίες γίνονται με το αναγνωριστικό

# Αρχεία και ευρετήρια (6 από 6)

- Υλοποίηση υπηρεσίας ευρετηρίων
  - Τα ευρετήρια είναι αρχεία με ειδική δομή
  - Η υπηρεσία ευρετηρίων είναι και αυτή πελάτης
- Λειτουργίες κώδικα διεπαφής
  - Απόκρυψη απομακρυσμένων κλήσεων
  - Παροχή κατάλληλης διεπαφής για κάθε πελάτη
    - Επικοινωνία Windows και UNIX

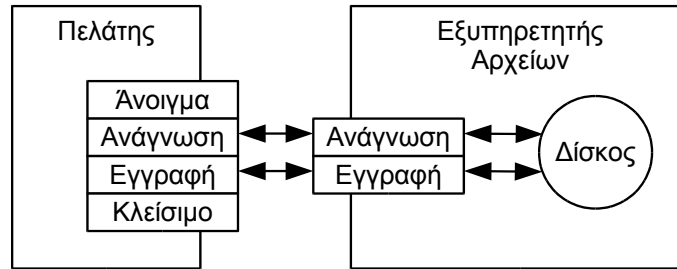
# Διεπαφή αρχείων (1 από 4)

- Αναγνωριστικά αρχείων
  - Αντιστοιχίζονται στο όνομα μέσω ευρετηρίων
  - Εξαρτώνται από το σύστημα
    - Στο UNIX είναι οι αριθμοί των i-node
- Μορφή κλήσεων ανάγνωσης και εγγραφής
  - Στο UNIX κάθε κλήση συνεχίζει από προηγούμενη
    - Αποθηκεύεται η τρέχουσα θέση του αρχείου
    - Για να αλλάξει κάνουμε αναζήτηση

# Διεπαφή αρχείων (2 από 4)

- Μορφή κλήσεων ανάγνωσης και εγγραφής
  - Στο NFS κάθε αίτηση είναι αυτόνομη
    - Περιέχει τη θέση ανάγνωσης/εγγραφής
  - Επιτρέπεται επανάληψη εντολών
    - Δεν αλλάζει το τελικό αποτέλεσμα
  - Ο εξυπηρετητής δεν διατηρεί κατάσταση
    - Δεν χρειάζεται κλήσεις ανοίγματος / κλεισίματος
    - Τρέχουσα θέση στον κώδικα διεπαφής

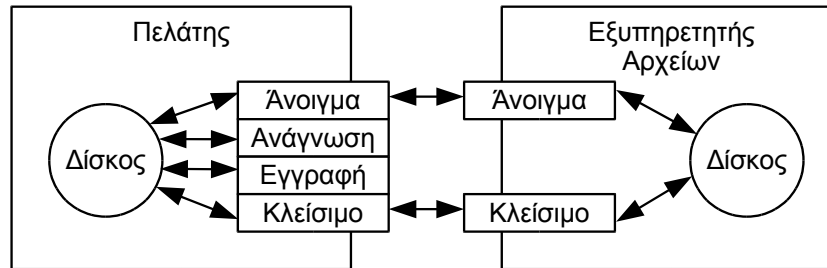
# Διεπαφή αρχείων (3 από 4)



- Μοντέλο απομακρυσμένης πρόσβασης
  - Υλοποιείται από το NFS
  - Ανάγνωση / εγγραφή στον εξυπηρετητή
    - Απαιτούν επικοινωνία μέσω δικτύου
  - Κάθε αίτηση είναι αυτόνομη
  - Δεν υπάρχει άνοιγμα και κλείσιμο



# Διεπαφή αρχείων (4 από 4)

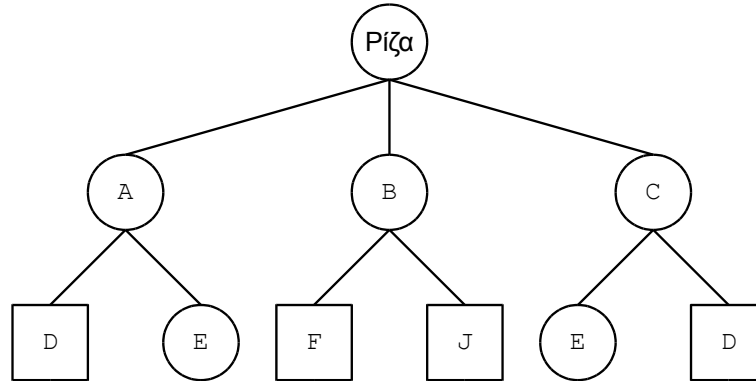


- Μοντέλο ανεβάσματος - κατεβάσματος
  - Παρέχονται μόνο εντολές για ολόκληρα αρχεία
  - Κατέβασμα αρχείου στο άνοιγμα
  - Ανέβασμα αρχείου στο κλείσιμο
  - Ανάγνωση/εγγραφή υλοποιούνται τοπικά
    - Χρήση δικτύου μόνο στην αρχή και στο τέλος

# Διεπαφή ευρετηρίων (1 από 3)

- Προσθήκες στην υπηρεσία αρχείων
  - Μετάφραση ονομάτων σε αναγνωριστικά
  - Οργάνωση αρχείων σε ευρετήρια
  - Τα στοιχεία αποθηκεύονται σε αρχεία
- Λειτουργίες υπηρεσίας ευρετηρίων
  - Καταχώρηση νέου ονόματος - αναγνωριστικού
  - Διαγραφή ονόματος - αναγνωριστικού
    - Χωριστή προσθήκη και διαγραφή του αρχείου
  - Αναζήτηση αναγνωριστικού αρχείου

# Διεπαφή ευρετηρίων (2 από 3)



- Αναζήτηση αρχείων
  - Ξεκινάμε από κάποιο γνωστό σημείο
    - Το αναγνωριστικό του ευρετηρίου της ρίζας
  - Διαβάζουμε τα ευρετήρια με τη σειρά
  - Ανάλυση ονομάτων επαναληπτικά ή αναδρομικά

# Διεπαφή ευρετηρίων (3 από 3)

- Παραγωγή αντιγράφων αρχείων
  - Αντιστοίχιση πολλών αναγνωριστικών σε όνομα
  - Επιλογή κατάλληλου από σύστημα ή πελάτη
- Κώδικας διεπαφής πελάτη
  - Εμφανίζει υπηρεσίες παρόμοιες με τις τοπικές
  - Μπορεί να υποστηρίζει συνδέσμους
    - Απλοί: πολλές εγγραφές του αναγνωριστικού
    - Συμβολικοί: ειδικά αρχεία, διερμηνεία στον πελάτη

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Όνόματα και αναγνωριστικά αρχείων

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα  
συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Όνόματα / αναγνωριστικά (1 από 4)

- Ομάδα αρχείων (ή σύστημα αρχείων)
  - Ενιαίο σύνολο αρχείων και ευρετηρίων
    - Παράδειγμα: διαμέριση ενός δίσκου
  - Όνόματα / αναγνωριστικά ως προς την ομάδα
- Εξυπηρετητής + διαδρομή
  - Μπορεί να φαίνεται ως `/dias/home/users`
    - Διαδρομή `/home/users` του εξυπηρετητή `/dias`
  - Διαφάνεια ονομασίας: παρόμοιο με τοπικό όνομα
  - Διαφάνεια τοποθεσίας: δε φαίνεται που είναι ο `dias`

# Ονόματα / αναγνωριστικά (2 από 4)

- Εγκατάσταση ομάδων αρχείων
  - Ο εξυπηρετητής εξάγει το /home/users
  - Ο πελάτης το στο ευρετήριο /remote
  - Η διαδρομή γίνεται /remote/home/users
  - Επιτρέπει τη μετακίνηση των ομάδων αρχείων
  - Κάθε μηχανή μπορεί να την εγκαθιστά αλλού
- Επιθυμητή: ανεξαρτησία τοποθεσίας
  - Ανεξαρτησία από θέση αρχείου και θέση χρήστη
  - Λειτουργία εφαρμογών παντού στο σύστημα

# Όνόματα / αναγνωριστικά (3 από 4)

- Αναγνωριστικό και όνομα αρχείου
  - Οι χρήστες χρησιμοποιούν το όνομα
  - Το σύστημα χρησιμοποιεί το αναγνωριστικό
- Αναγνωριστικά με βάση τη μηχανή
  - i-κόμβος: αναγνωριστικό μέσα σε ομάδα αρχείων
  - Ομάδα αρχείων: αναγνωριστικό μέσα σε μηχανή
  - Μηχανή: αναγνωριστικό μέσα στο σύστημα
  - Δεν επιτρέπει μετακίνηση των ομάδων αρχείων



# Ονόματα / αναγνωριστικά (4 από 4)

- Αναγνωριστικά με βάση την ομάδα αρχείων
  - Κάθε ομάδα έχει μοναδικό αναγνωριστικό
  - Δίνεται ο i-κόμβος μέσα στην ομάδα αρχείων
  - Η ομάδα αρχείων μπορεί να μετακινηθεί εύκολα
  - Χρειάζεται υπηρεσία εντοπισμού των ομάδων
- Αναγνωριστικά σε ομότιμα συστήματα
  - Επίπεδα αναγνωριστικά (ακολουθίες bit)
  - Προαιρετικά χωριστή υπηρεσία ευρετηρίων

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Εξυπηρετητές αρχείων και ευρετηρίων

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα  
συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο

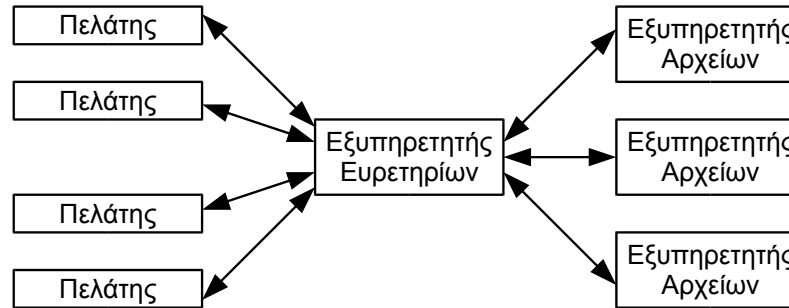


ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

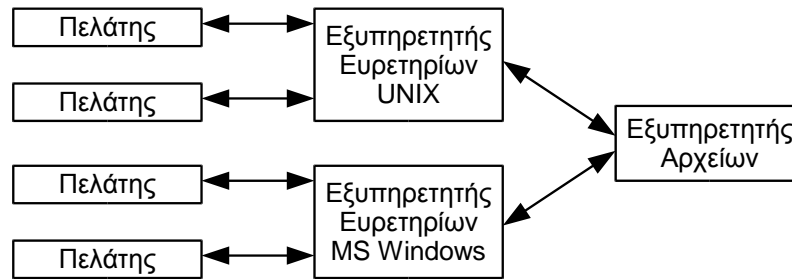


# Εξυπηρετητές (1 από 4)



- Διάκριση υπηρεσίας αρχείων και ευρετηρίων
  - Αρχικά μετάφραση ονόματος από ευρετήριο
  - Στη συνέχεια προσπέλαση με αναγνωριστικό
  - Ίδια διεργασία: λιγότερα μηνύματα
  - Χωριστές διεργασίες: μεγαλύτερη ευελιξία
    - Κρυφή μνήμη για αποφυγή μηνυμάτων

# Εξυπηρετητές (2 από 4)



- Ένα ευρετήριο, πολλοί εξυπηρετητές
  - Κατανομή φόρτου διαφανώς προς του πελάτες
- Ένας εξυπηρετητής, πολλά ευρετήρια
  - Διαφορετικές όψεις ευρετηρίου
  - Κατάλληλες για διαφορετικούς πελάτες

# Εξυπηρετητές (3 από 4)

- Χωρίς κατάσταση: δεν υπάρχουν ανοιχτά αρχεία
  - Κάθε μήνυμα περιέχει όλες τις πληροφορίες
  - Ο εξυπηρετητής δεν τηρεί στοιχεία για τον πελάτη
  - Ανοχή στα σφάλματα γιατί δε χάνεται κατάσταση
- Με κατάσταση: υπάρχουν ανοιχτά αρχεία
  - Χρήση σύντομου αναγνωριστικού σε κάθε κλήση
  - Μικρότερα μηνύματα ανάγνωσης και εγγραφής
  - Διευκολύνει προσωρινή αποθήκευση και κλείδωμα

# Εξυπηρετητές (4 από 4)

- Έλεγχος προσπέλασης
  - Με λίστες ελέγχου προσπέλασης ή ικανοτήτων
  - Έλεγχος ταυτότητας και έλεγχος προσπέλασης
  - Απαιτεί μηχανισμούς ελέγχου μέσω δικτύου
    - Κατανεμημένη πιστοποίηση ταυτότητας
  - Χωρίς κατάσταση: συνεχείς έλεγχοι
    - Εναλλακτικά: έκδοση κωδικοποιημένης ικανότητας
    - Χρήση ικανότητας σε επόμενες προσπελάσεις

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Προσωρινή αποθήκευση αρχείων

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο

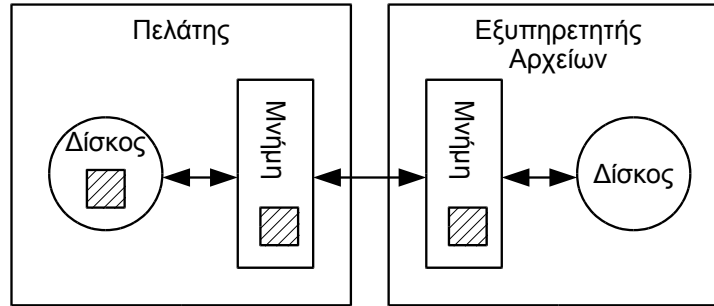


ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Προσωρινή αποθήκευση (1 από 5)



- Πιθανές θέσεις αποθήκευσης αρχείων
  - Μόνιμα: Δίσκος εξυπηρετητή
  - Προσωρινά (κρυφή μνήμη)
    - Μνήμη εξυπηρετητή
    - Μνήμη πελάτη
    - Δίσκος πελάτη



# Προσωρινή αποθήκευση (2 από 5)

- Μνήμη εξυπηρετητή
  - Δεν αποφεύγει την κίνηση στο δίκτυο
  - Γενικά διαφανής προς τον πελάτη
  - Αποθήκευση τμημάτων
  - Αποθήκευση ολόκληρων αρχείων
    - Επιτρέπει ακολουθιακή εγγραφή στο δίσκο
  - Χρήση πολύ σύνθετων πολιτικών (όπως LRU)
    - Οι εξυπηρετητές είναι πολύ ισχυρές μηχανές

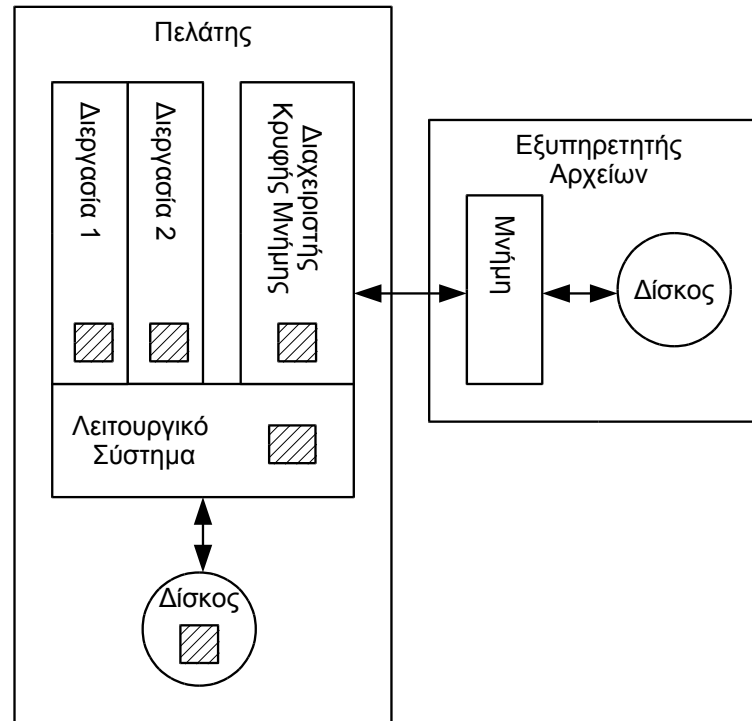
# Προσωρινή αποθήκευση (3 από 5)

- Προσωρινή αποθήκευση στον πελάτη
  - Αποφεύγει την ανταλλαγή μηνυμάτων μέσω δικτύου
  - Μνήμη πελάτη: για αποθήκευση τμημάτων
    - Συνηθίζεται στο μοντέλο απομακρυσμένης πρόσβασης
  - Δίσκος πελάτη: για αποθήκευση αρχείων
    - Συνηθίζεται στο μοντέλο ανεβάσματος – κατεβάσματος
  - Δημιουργεί όμως προβλήματα συνέπειας
    - Αν πολλοί πελάτες έχουν αποθηκευμένο το ίδιο αρχείο;
    - Πώς ενημερώνουν για τις αλλαγές τους άλλους πελάτες;

# Προσωρινή αποθήκευση (4 από 5)

- Πού είναι η κρυφή μνήμη του πελάτη;
  - Στο δίσκο: αρχεία διαθέσιμα σε όλους
  - Στη μνήμη της διεργασίας
    - Γρήγορη υλοποίηση κλήσεων αρχείων
    - Αδυναμία καταμερισμού με άλλες διεργασίες
  - Στον πυρήνα
    - Μεγάλο κόστος προσπέλασης
    - Δυναμική διαχείριση διαθέσιμης μνήμης

# Προσωρινή αποθήκευση (5 από 5)



- Πού είναι η κρυφή μνήμη του πελάτη;
  - Διαχειριστής κρυφής μνήμης
    - Ενδιάμεση λύση, αν έχουμε κοινή μνήμη

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**

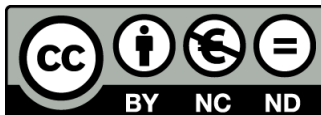


**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Σηματολογία καταμερισμού αρχείων

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα  
συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Σημασιολογία (1 από 4)

- Κλασική σημασιολογία
  - Παρέχεται στα συστήματα ενός επεξεργαστή
  - Η ανάγνωση επιστρέφει στοιχεία τελευταίας εγγραφής
  - Απαιτεί καθολική διάταξη όλων των γεγονότων
- Κατανεμημένη κλασική σημασιολογία
  - Χρήση ενός μόνο εξυπηρετητή αρχείων
  - Κάθε πελάτης αριθμεί τις πράξεις που στέλνει
  - Ο εξυπηρετητής εκτελεί τις πράξεις με τη σειρά

# Σημασιολογία (2 από 4)

- Προβλήματα κλασικής σημασιολογίας
  - Πολύ χαμηλή επίδοση
    - Απαγορεύεται η προσωρινή αποθήκευση
  - Δεν κλιμακώνεται
    - Δεν μπορούμε να κατανείμουμε το φόρτο
  - Μοναδικός εξυπηρετητής
    - Μοναδικό σημείο αποτυχίας

# Σηματολογία (3 από 4)

- Σηματολογία συνόδου
  - Αλλαγές στα ανοικτά αρχεία μόνο τοπικά
  - Ενημέρωση εξυπηρετητή στο κλείσιμο αρχείου
  - Διάταξη πράξεων σε επίπεδο συνόδου
    - Ανάγνωση: δίνει στοιχεία τελευταίας συνόδου
    - Της τελευταίας που ολοκληρώθηκε!
  - Επαρκής όταν τα αρχεία καταμερίζονται σπάνια



# Σημασιολογία (4 από 4)

- Σημασιολογία ατομικών συναλλαγών
  - Ομάδες πράξεων περικλείονται σε συναλλαγές
  - Οι πράξεις εκτελούνται όλες με τη σειρά
  - Αλλιώς, δεν εκτελείται καμία
  - Ατομική και ακολουθιακή εκτέλεση συναλλαγών
  - Πολύ υψηλό κόστος για γενική χρήση

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Συνέπεια συστήματος αρχείων

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο

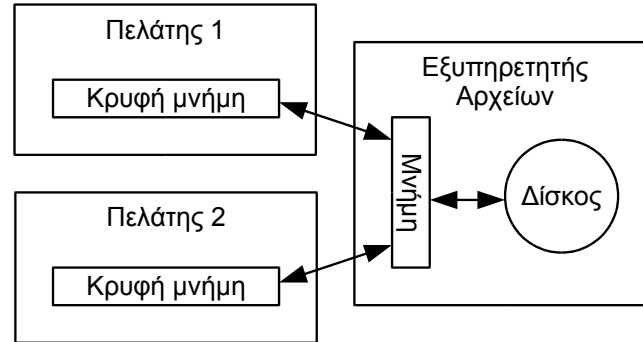


ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Συνέπεια (1 από 4)



- Προβλήματα ασυνέπειας
  - Προσωρινή αποθήκευση: ασυνέπεια
    - Κάθε πελάτης έχει διαφορετική εικόνα
    - Οι πράξεις έχουν απρόβλεπτα αποτελέσματα
  - Πιθανές ασυνέπειες και στα ευρετήρια

# Συνέπεια (2 από 4)

- Προσωρινή αποθήκευση με άμεση εγγραφή
  - Κάθε τροποποίηση στέλνεται στον εξυπηρετητή
    - Είτε ενημερώνονται και όλοι οι πελάτες
    - Είτε ο πελάτης ρωτάει τον εξυπηρετητή
  - Μεγάλη κυκλοφορία στο δίκτυο
  - Αποφεύγεται με περιοδική ενημέρωση
    - Λιγότερα μηνύματα σε πολλαπλές εγγραφές
    - Εισάγει (μικρότερη) πιθανότητα ασυνέπειας

# Συνέπεια (3 από 4)

- Προσωρινή αποθήκευση με κλείδωμα
  - Χρήση συγκεντρωτικού αλγορίθμου
  - Ο εξυπηρετητής γνωρίζει ποιος διαβάζει ή γράφει
    - Μπορούμε να έχουμε πολλούς αναγνώστες
    - Μπορούμε να έχουμε ένα μόνο συγγραφέα
  - Κίνδυνος υποσιτισμού των συγγραφέων
    - Οι αναγνώστες μπορεί να είναι συνεχώς πολλοί
    - Οι συγγραφείς πρέπει κάποτε να έχουν προτεραιότητα

# Συνέπεια (4 από 4)

- Αποφυγή υποσιτισμού συγγραφέων
  - Αποστολή μηνυμάτων σε αναγνώστες
    - Απαγόρευση τοπικής αποθήκευσης
    - Απαιτεί χειρισμό αυτόκλητων μηνυμάτων
- Προσωρινή αποθήκευση συνόδου
  - Αλλαγές στέλνονται στο τέλος της συνόδου
  - Ουσιαστικά επιτρέπουμε την ασυνέπεια
  - Λογική προσέγγιση με λίγα κοινά αρχεία

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Το σύστημα NFS

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Υλοποίηση NFS (1 από 5)

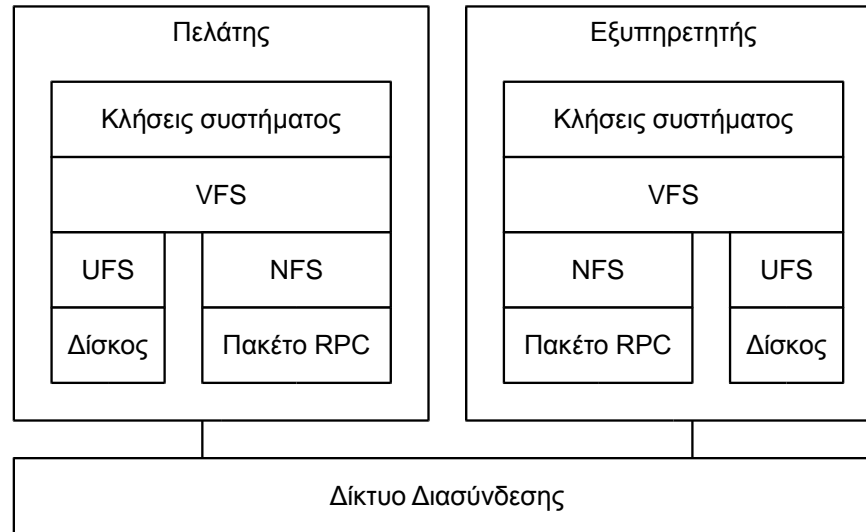
- NFS: Network File System
  - Δημιουργήθηκε από τη Sun για το UNIX
  - Η μηχανή μπορεί να είναι πελάτης και εξυπηρετητής
- Πρωτόκολλο NFS
  - Απομακρυσμένες κλήσεις διαδικασιών (RPC)
  - Απεικονιστής θυρών RPC σε γνωστή διεύθυνση
  - Χρήση UDP σε LAN ή TCP σε WAN
  - Μηνύματα ίσου μεγέθους με ομάδων αρχείων



# Υλοποίηση NFS (2 από 5)

- Χειριστήρια αρχείων NFS
  - Αναγνωριστικά των αρχείων σε εξυπηρετητή
    - Αδιαφανή στον πελάτη
  - Περιέχουν πληροφορίες εντοπισμού
  - UNIX: i-node, ομάδα αρχείων, αριθμός γενιάς
    - Η ομάδα αρχείων και το i-node ορίζουν το αρχείο
    - Ο αριθμός γενιάς κάνει τα αναγνωριστικά μοναδικά
    - Επιτρέπει επαναχρησιμοποίηση i-node/ομάδας

# Υλοποίηση NFS (3 από 5)



- Εικονικό σύστημα αρχείων (VFS)
  - Χρήση ίδιων κλήσεων για απομακρυσμένα αρχεία
  - Τα ευρετήρια εγκαθίστανται στον τοπικό χώρο
    - Κάθε ευρετήριο / αρχείο περιγράφεται από v-node
    - Κλήση κατάλληλου κώδικα ανάλογα v-node

# Υλοποίηση NFS (4 από 5)

- Χρήση RPC για επικοινωνία
  - Όταν το v-node αντιστοιχεί σε NFS
  - Η κλήση μεταφέρεται στον εξυπηρετητή
- Υλοποίηση πελάτη NFS στον πυρήνα
  - Ενσωμάτωση μέσω του VFS
  - Τοπικός καταμερισμός αρχείων
    - Ενιαίος χώρος προσωρινής αποθήκευσης
  - Πιστοποίηση ταυτότητας χρήστη από πυρήνα

# Υλοποίηση NFS (5 από 5)

- Υλοποίηση εξυπηρετητή NFS
  - Ο εξυπηρετητής δεν γνωρίζει τον πελάτη
    - Δεν έχει ελεγχθεί από τον δικό του πυρήνα
  - Η πιστοποίηση απαιτεί πρόσθετο μηχανισμό
    - Το ίδιο το NFS δεν παρέχει κάποιες ευκολίες
  - Πρέπει να πιστοποιεί κάθε φορά την ταυτότητα
    - Αφού δεν διατηρεί κατάσταση για τους πελάτες
  - Μπορεί να επιστρέφεται ικανότητα στον πελάτη
    - Χρησιμοποιείται για να επιταχύνει επόμενες αιτήσεις

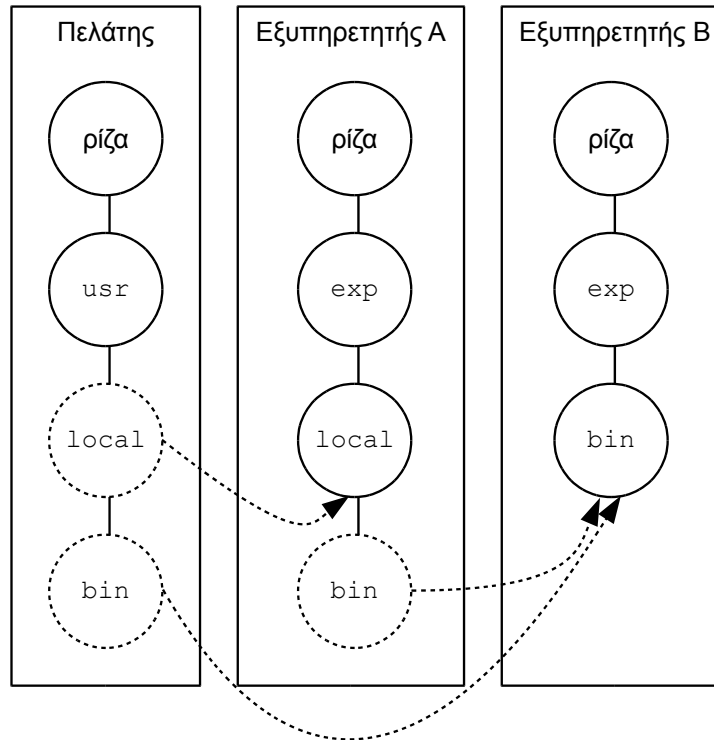
# Διεπαφή NFS (1 από 5)

- Διεπαφή NFS version 3 (χωρίς κατάσταση)
  - Κοινή διεπαφή αρχείων και ευρετηρίων
    - Δημιουργία/διαγραφή αρχείων με μία κλήση
  - Επιτρέπεται τμηματική ανάγνωση ευρετηρίων
    - Ο πελάτης παρακολουθεί τη θέση ανάγνωσης
- Υπηρεσία εγκατάστασης
  - Χωριστή διεργασία στον εξυπηρετητή
  - Εξάγονται ευρετήρια, όχι ομάδες αρχείων
    - Αρχείο διάρθρωσης με ευρετήρια και δικαιώματα

# Διεπαφή NFS (2 από 5)

- Εγκατάσταση στον πελάτη
  - Κλήση RPC προς τον εξυπηρετητή
    - Επιστρέφει το χειριστήριο του ευρετηρίου
    - Δημιουργία δομής με στοιχεία επικοινωνίας
- Αναζήτηση ονομάτων
  - Η διαδρομή μπορεί να διασχίζει εξυπηρετητής
  - Μόνο επαναληπτική ανάλυση ονομάτων
  - Προσωρινή αποθήκευση χειριστηρίων

# Διεπαφή NFS (3 από 5)



- Επανεξαγωγή ευρετηρίων
  - Τα ευρετήρια αυτά δεν επανεξάγονται αυτόματα
  - Πρέπει να τα εγκαταστήσει ρητά ο πελάτης

# Διεπαφή NFS (4 από 5)

- Εγκατάσταση ευρετηρίων
  - Εξαναγκασμένη: ο πελάτης περιμένει για πάντα
  - Προαιρετική: επιστρέφει σφάλμα αν αποτύχει
- Ενιαία εικόνα ευρετηρίων
  - Κάθε πελάτης εγκαθιστά όπου θέλει τα ευρετήρια
    - Μπορεί να έχει διαφορετική εικόνα των αρχείων
  - Συνήθως τυποποιημένη μορφή καταλόγων
    - Είναι μία σύμβαση σε κάθε οργανισμό



# Διεπαφή NFS (5 από 5)

- Αυτόματη εγκατάσταση ευρετηρίων
  - Διεργασία εγκατάστασης στον πελάτη
  - Αρχικά λειτουργεί σαν τοπικός εξυπηρετητής
  - Πρώτη προσπέλαση: εγκατάσταση ευρετηρίου
    - Σύνδεση σημείου εγκατάστασης με συμβολικό σύνδεσμο
    - Οι επόμενες προσπελάσεις παρακάμπτουν τη διεργασία
- Πολλοί εξυπηρετητές ανά ομάδα
  - Επιλέγεται όποιος απαντήσει πρώτος

# Αποθήκευση NFS (1 από 3)

- Προσωρινή αποθήκευση στον εξυπηρετητή
  - Δεν δημιουργεί προβλήματα διαφάνειας
  - Ο εξυπηρετητής γράφει περιοδικά τις αλλαγές
    - Κάθε 30 sec στο UNIX
    - Αποφυγή απώλειας σε περίπτωση αποτυχίας
  - Προαιρετικά πιο γρήγορη εγγραφή
    - Άμεση εγγραφή σε κάθε εντολή
    - Ομαδική εγγραφή με χωριστή εντολή

# Αποθήκευση NFS (2 από 3)

- Προσωρινή αποθήκευση στον πελάτη
  - Τμήματα αρχείων ή ευρετηρίων
    - Τα χειριστήρια δεν δημιουργούν πρόβλημα
  - Δύο χρονοσφραγίδες για κάθε ομάδα
    - Χρόνος τελευταίου ελέγχου με εξυπηρετητή
    - Χρόνος τελευταίας τροποποίησης σε εξυπηρετητή
  - Περιοδικός έλεγχος εγκυρότητας
    - Το διάστημα ελέγχου είναι ρυθμιζόμενο
    - Ελέγχεται ο χρόνος τροποποίησης του αρχείου

# Αποθήκευση NFS (3 από 3)

- Προσωρινή αποθήκευση στον πελάτη
  - Έλεγχος όλων των ομάδων αρχείου σε αλλαγή
    - Πρέπει να ακυρωθούν όλες
  - Επιστροφή χρόνου τροποποίησης σε κάθε κλήση
    - Γλυτώνουμε χωριστά μηνύματα ελέγχου
- Αποστολή αλλαγών στον εξυπηρετητή
  - Όταν κλείνει το αρχείο
  - Όταν το ζητάει ρητά ο πελάτης
  - Ασύγχρονα μέσω πρόσθετης διεργασίας

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**

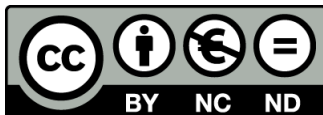


**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Το σύστημα AFS

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



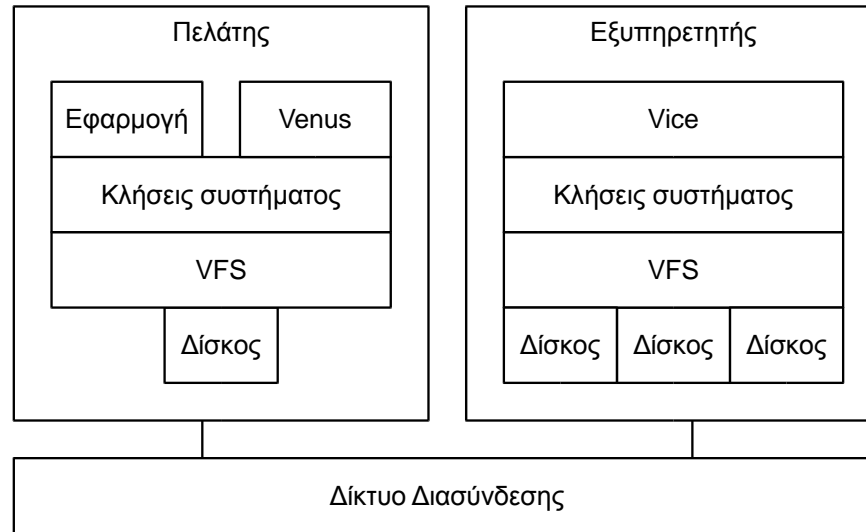
# Υλοποίηση AFS (1 από 6)

- AFS: Andrew File System
  - Δημιουργήθηκε στο CMU
  - Απευθύνεται σε πολύ μεγάλα συστήματα
    - Πολλοί χρήστες και εξυπηρετητές
  - Χρησιμοποιεί το πρωτόκολλο του NFS
  - Όλα τα αρχεία χρησιμοποιούν χειριστήρια NFS
    - Ίδια όψη τοπικών / απομακρυσμένων αρχείων
  - Μοντέλο ανεβάσματος / κατεβάσματος

# Υλοποίηση AFS (2 από 6)

- Προσωρινή αποθήκευση ολόκληρων αρχείων
  - Τα αρχεία αποθηκεύονται στο δίσκο του πελάτη
    - Ο πελάτης σταδιακά κατεβάζει όλα τα αρχεία που θέλει
  - Όλες οι πράξεις εκτελούνται τοπικά
    - Μικρή κίνηση στο δίκτυο κατά τη λειτουργία
    - Το αρχείο επιστρέφεται όταν κλείσει
  - Βελτιστοποίηση για βιβλιοθήκες και αρχεία χρηστών
    - Είτε δεν αλλάζουν είτε δεν καταμερίζονται
    - Δυνατότητα χρήσης ομάδων 64 KB σε μεγάλα αρχεία

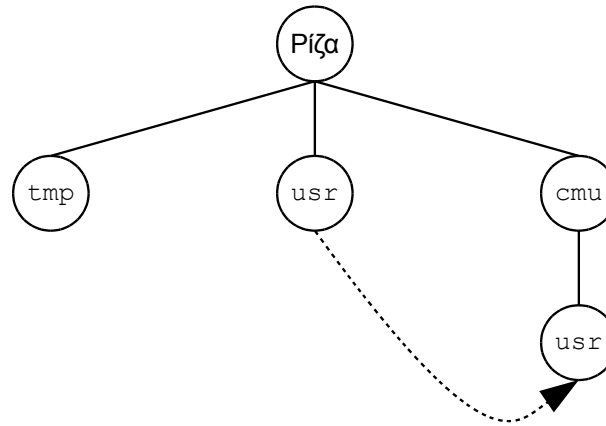
# Υλοποίηση AFS (3 από 6)



- Τροποποιήσεις στον πελάτη (διεργασία Venus)
  - Οι κλήσεις για άνοιγμα και κλείσιμο περνάνε στη Venus
  - Αν ένα αρχείο δεν υπάρχει, προσκομίζεται τοπικά
    - Αποθηκεύεται σε χωριστή διαμέριση του δίσκου
  - Το λειτουργικό σύστημα ανοίγει το τοπικό αντίγραφο
  - Ενημέρωση Venus με το κλείσιμο του αρχείου



# Υλοποίηση AFS (4 από 6)



- Τα ευρετήρια εγκαθίστανται στο /cmu
  - Συμβολικοί σύνδεσμοι από συνήθη ευρετήρια
  - Οι χρήστες βλέπουν τα ίδια σε όλες τις μηχανές
    - Όταν ανοίγει αρχείο στο /cmu ειδοποιείται η Venus
    - Όταν κλείνει αρχείο στη διαμέριση ειδοποιείται η Venus
  - Χρήση LRU για διαχείριση της ειδικής διαμέρισης

# Υλοποίηση AFS (5 από 6)

- Τροποποιήσεις εξυπηρετητή (διεργασία Vice)
  - Διαχείριση όλων των αρχείων με χειριστήρια
- Ομαδοποίηση αρχείων σε τόμους
  - Αναγνωριστικά αρχείων ως προς τους τόμους
  - Αρχείο αντιστοίχισης τόμων με εξυπηρετητές
  - Υποστήριξη παραγωγής αντιγράφων τόμων
    - Η ενημέρωση γίνεται στον πρωτεύοντα εξυπηρετητή
    - Στη συνέχεια ενημερώνονται όλα τα αντίγραφα

# Υλοποίηση AFS (6 από 6)

- Αναγνωριστικά αρχείων στο AFS
  - Αριθμός τόμου
  - Αριθμός i-κόμβου στον τόμο
  - Αριθμός σειράς (ανάλογος με αριθμό γενιάς NFS)
  - Ο εξυπηρετητής δέχεται κλήσεις με αναγνωριστικά
    - Δεν δέχεται χωριστές κλήσεις ευρετηρίων
    - Τα ευρετήρια διαβάζοντας σαν απλά αρχεία από τη Vice
    - Η Venus τα παρουσιάζει σαν ευρετήρια στον πελάτη

# Αποθήκευση AFS (1 από 2)

- Προσωρινή αποθήκευση στο AFS
  - Τα περισσότερα αρχεία δεν καταμερίζονται
  - Οι πελάτες τα αποθηκεύουν πολύ καιρό
  - Δεν πρέπει να ρωτάνε συνεχώς
- Υποσχέσεις επανάκλησης από εξυπηρετητή
  - Ο εξυπηρετητής γνωρίζει που έδωσε το αρχείο
    - Άρα διατηρεί κατάσταση για τους πελάτες
  - Σε περίπτωση αλλαγών ενημερώνει τους πελάτες
    - Ακυρώνει την υπόσχεση επανάκλησης

# Αποθήκευση AFS (2 από 2)

- Υποσχέσεις επανάκλησης από εξυπηρετητή
  - Άνοιγμα: ελέγχουμε την υπόσχεση επανάκλησης
    - Αν είναι άκυρη, φέρνουμε ξανά το αρχείο
  - Αποτυχία: ελέγχουμε ξανά εγκυρότητα αρχείων
    - Μπορεί να έχουμε χάσει κάποιες επανακλήσεις
  - Περιοδικός έλεγχος εγκυρότητας από πελάτη
- Γενικά παρέχεται σημασιολογία συνόδου
  - Εκτός αν οι διεργασίες είναι στον ίδιο πελάτη

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**

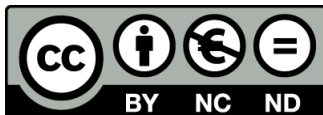


**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Το σύστημα PAST

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Εισαγωγή στο PAST (1 από 5)

- PAST: αποθήκευση αρχείων μέσω Pastry
  - Αντιστοίχιση κάθε αρχείου σε  $k$  κόμβους
    - Αριθμητικά πλησιέστεροι στο κλειδί του αρχείου
  - Βαθμός αναπαραγωγής  $k$ 
    - Αύξηση αξιοπιστίας και αποδοτικότητας
  - Δεν υποστηρίζει πλήρες σύστημα αρχείων
    - Απλά αποθηκεύει τα αρχεία αυτόνομα

# Εισαγωγή στο PAST (2 από 5)

- Διεπαφή προσανατολισμένη σε αρχεία
  - Βασίζεται σε σύστημα δημόσιας κρυπτογραφίας
    - Θεωρούμε ότι γνωρίζουμε τα δημόσια κλειδιά
  - `key=insert(name,k,file)`
    - Αποθήκευση `file` με όνομα `name` σε `k` κόμβους
    - Το `key` παράγεται με κατακερματισμό των στοιχείων
    - Όνομα, δημόσιο κλειδί και τυχαία επιλεγμένη τιμή



# Εισαγωγή στο PAST (3 από 5)

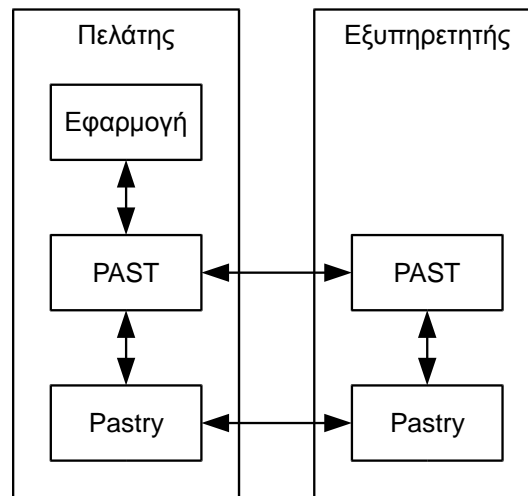
- Διεπαφή προσανατολισμένη σε αρχεία
  - `file=lookup(key)`
    - Ανάκτηση ενός αντιγράφου του αρχείου με κλειδί `key`
    - Συνήθως από τον τοπολογικά πλησιέστερο κόμβο
  - `reclaim(key)`
    - Απελευθέρωση του χώρου του αρχείου με κλειδί `key`
    - Δεν εγγυάται ότι θα διαγραφούν όλα τα αντίγραφα

# Εισαγωγή στο PAST (4 από 5)

- Ιδιαιτερότητες της διεπαφής του PAST
  - Δεν επιτρέπει την τροποποίηση των αρχείων
  - Αν εισάγουμε ξανά το αρχείο θα αλλάξει κλειδί
    - Θα χρησιμοποιηθεί διαφορετική τυχαία τιμή
    - Τα αρχεία είναι αμετάβλητα (immutable)
  - Μόνο ο ιδιοκτήτης γνωρίζει το κλειδί του αρχείου
    - Αυτός έχει την ευθύνη και για τη διανομή του
  - Δεν γίνεται έλεγχος πρόσβασης στα lookups
    - Έλεγχος μόνο στη διαγραφή των αρχείων

# Εισαγωγή στο PAST (5 από 5)

- Μοντέλο ανεβάσματος-κατεβάσματος
  - Παρόμοιο με το AFS αλλά χωρίς σύστημα αρχείων
  - Η εφαρμογή βλέπει τη διεπαφή του PAST
  - Η εφαρμογή δεν τρέχει σε όλους τους κόμβους



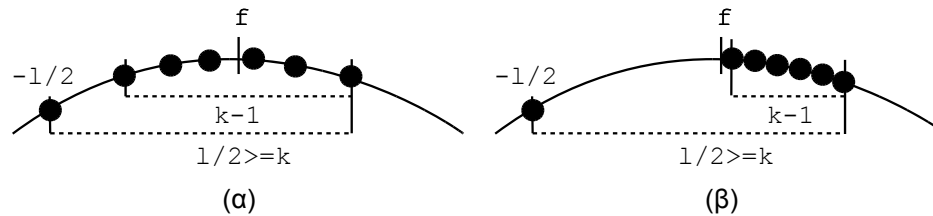
# Εισαγωγή αρχείων (1 από 3)

- Εισαγωγή αρχείων στο PAST
  - Υπολογισμός κλειδιού με κατακερματισμό
    - Όνομα αρχείου, δημόσιο κλειδί, τυχαία τιμή
  - Δημιουργία πιστοποιητικού αρχείου
    - Κλειδί, hash περιεχομένου, τιμή  $k$ , τυχαία τιμή, ώρα
    - Δεν μεταβάλλεται αφού το αρχείο δεν αλλάζει
    - Υπογράφεται ψηφιακά από τον ιδιοκτήτη
    - Αποθηκεύεται μαζί με το αρχείο

# Εισαγωγή αρχείων (2 από 3)

- Εισαγωγή αρχείων στο PAST
  - Αρχείο και πιστοποιητικό δρομολογούνται
  - Όταν φτάσουν σε έναν από τους  $k$  κόμβους
    - Ελέγχεται η εγκυρότητα αρχείου και πιστοποιητικού
    - Το αρχείο αποθηκεύεται
    - Το αρχείο στέλνεται στους  $k-1$
    - Κάθε παραλήπτης επιστρέφει απόδειξη αποθήκευσης
  - Ο πελάτης θα λάβει  $k$  έγκυρες αποδείξεις

# Εισαγωγή αρχείων (3 από 3)



- Εντοπισμός των  $k$  πλησιέστερων κόμβων
  - Τα μηνύματα δρομολογούνται προς το κλειδί  $f$
  - Κάθε κόμβος διατηρεί ένα σύνολο  $l$  φύλλων
    - Οι  $l/2$  μικρότεροι και  $l/2$  μεγαλύτεροι κόμβοι
    - Απαιτούμε  $l/2 \geq k$
  - Έστω ότι το μήνυμα φτάνει στον ακραίο κόμβο
  - Οι  $k-1$  άλλοι κόμβοι βρίσκονται στο σύνολο φύλλων
  - Άρα ο πρώτος γνωρίζει πάντα τους άλλους  $k-1$

# Διαγραφή αρχείων

- Διαγραφή αρχείων στο PAST
  - Δημιουργία πιστοποιητικού ανάκτησης
    - Υπογράφεται ψηφιακά από τον ιδιοκτήτη
  - Η αίτηση και το πιστοποιητικό δρομολογούνται
  - Όταν φτάσουν σε έναν από τους  $k$  κόμβους
    - Ελέγχεται η εγκυρότητα του πιστοποιητικού
    - Το αρχείο διαγράφεται
    - Η αίτηση δρομολογείται στους  $k-1$
    - Κάθε παραλήπτης επιστρέφει μια απόδειξη διαγραφής
  - Ο πελάτης θα λάβει  $k$  έγκυρες αποδείξεις (ίσως!)

# Ανάκτηση αρχείων

- Ανάκτηση αρχείων στο PAST
  - Το μήνυμα αποστέλλεται μέσω του Pastry
  - Ο πρώτος από τους  $k$  κόμβους απαντά
  - Επιστρέφεται και το πιστοποιητικό αρχείου
    - Ο πελάτης επιβεβαιώνει την εγκυρότητα του αρχείου
- Χρήση πιστοποιητικών
  - Το PAST δημιουργεί πιστοποιητικά
    - Κατά την εισαγωγή και διαγραφή αρχείων
    - Στις άλλες λειτουργίες απλά τα επιβεβαιώνει



# Αποθήκευση σε πολλά σημεία

- Κάθε αρχείο αποθηκεύεται σε  $k$  κόμβους
  - Οι κόμβοι είναι διάσπαρτοι στο σύστημα
  - Μικρή πιθανότητα ταυτόχρονης αποτυχίας
  - Κάποιος μπορεί να είναι κοντά στον πελάτη
- Προσωρινή αποθήκευση αρχείων
  - Κάθε κόμβος αξιοποιεί τον ελεύθερο χώρο του
  - Αποθηκεύει προσωρινά διερχόμενα αρχεία
  - Διαχείριση χώρου με κάποια πολιτική
  - Προτεραιότητα έχουν τα αντίγραφα αρχείων

# Κατανομή φόρτου

- Βασίζεται στην τυχαία κατανομή αρχείων
  - Αυτό δεν οδηγεί σε ομοιόμορφη κατανομή
- Ανομοιόμορφες δυνατότητες κόμβων
  - Ορισμένοι κόμβοι έχουν πολύ διαθέσιμο χώρο
    - Οι κόμβοι διασπώνται σε εικονικούς κόμβους
    - Μοιράζονται το φυσικό χώρο αποθήκευσης
- Ανομοιόμορφη κατανομή αρχείων
  - Αρχεία με πάρα πολύ μεγάλο μέγεθος
    - Κάποιοι κόμβοι μπορεί να φορτωθούν περισσότερο

# Εκτροπή αντιγράφων (1 από 4)

- Εκτροπή αντιγράφων
  - Έστω ότι ο A λαμβάνει μία αίτηση εισαγωγής
  - Αν δεν μπορεί να την ικανοποιήσει;
    - Επιλέγει ένα από τα φύλλα του, τον B
      - Δεν πρέπει να είναι στους  $k$  πλησιέστερους
      - Δεν πρέπει να έχει δεχτεί ήδη άλλο αντίγραφο
    - Στέλνει το αρχείο στον B
    - Αποθηκεύει έναν δείκτη προς το αντίγραφο
    - Εκδίδει την απόδειξη αποθήκευσης

# Εκτροπή αντιγράφων (2 από 4)

- Εκτροπή αντιγράφων
  - Περιοδικά οι δείκτες στέλνονται και αλλού
    - Για κάθε δείκτη εντοπίζεται ο  $k+1$  πλησιέστερος
      - Έστω ότι είναι ο  $C$
    - Ο δείκτης εκτροπής στέλνεται και στον  $C$
    - Αν αποτύχει ο  $A$ , ο  $C$  παίρνει αυτόματα τη θέση του
    - Αν αποτύχει ο  $C$ , ο  $A$  εντοπίζει κάποιον άλλο κόμβο

# Εκτροπή αντιγράφων (3 από 4)

- Κριτήρια εκτροπής αντιγράφων
  - Κάθε αίτηση εισαγωγής εξετάζεται χωριστά
  - Εκτρέπονται αρχεία που θέλουν πολύ χώρο
    - Πάνω από κάποιο ποσοστό του διαθέσιμου χώρου
  - Δεν γίνεται από υποφορτωμένους κόμβους
  - Εκτρέπονται κυρίως μεγάλα αρχεία

# Εκτροπή αντιγράφων (4 από 4)

- Αν δεν δεχτεί κανένα φύλλο το αρχείο;
  - Μπορεί όλα τα φύλλα να είναι φορτωμένα
    - Χρήση παρόμοιας πολιτικής με αυτή της εκτροπής
  - Αναγκαστικά γίνεται εκτροπή αρχείου
    - Επιστροφή σφάλματος στον πελάτη
  - Ο πελάτης δημιουργεί νέο κλειδί για το αρχείο
    - Απλά αλλάζει την τυχαία τιμή
  - Το αρχείο αποθηκεύεται με το νέο κλειδί
    - Το παλιό κλειδί διαγράφεται

# Διαχείριση αλλαγών (1 από 3)

- Αποτυχία, αποχώρηση, προσχώρηση κόμβων
  - Αλλάζει το σύνολο φύλλων ορισμένων κόμβων
  - Αν αλλάξουν οι  $k$  πλησιέστεροι ενός αρχείου;
  - Δημιουργία νέων αντιγράφων του αρχείου
    - Πρέπει να έχουμε  $k$  στους σωστούς κόμβους
    - Περιλαμβάνονται οι κόμβοι με εκτροπή αντιγράφων
    - Και αυτοί ανήκουν στο σύνολο φύλλων ενός εκ των  $k$

# Διαχείριση αλλαγών (2 από 3)

- Αποτυχία ή αποχώρηση
  - Ένα κόμβος μπορεί να χρειάζεται αντίγραφα
  - Οι  $k-1$  θα βρίσκονται στο σύνολο φύλλων του
    - Αρκεί να ρωτήσει τα φύλλα του ποια αρχεία έχουν
    - Μετά αποφασίζει ποια αρχεία του αντιστοιχούν
    - Τέλος ζητάει τα αρχεία από αυτούς τους κόμβους



# Διαχείριση αλλαγών (3 από 3)

- Προσχώρηση κόμβων
  - Το σύστημα περιέχει ήδη  $k$  αντίγραφα
  - Αρκεί ο νέος κόμβος να προσθέσει δείκτες
    - Αρχικά εντοπίζει τον κόμβο που έφυγε από τους  $k$
    - Στη συνέχεια αποθηκεύει δείκτες προς αυτόν
  - Σταδιακά τα αντίγραφα έρχονται στον κόμβο
    - Τα αντίγραφα μπορεί να απομακρυνθούν από τους  $k$
    - Αν έχουμε αποτυχίες θα μείνουν λιγότερα από  $k$

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**

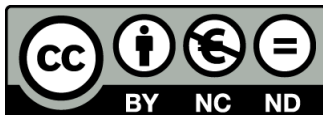


**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Το σύστημα CFS

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

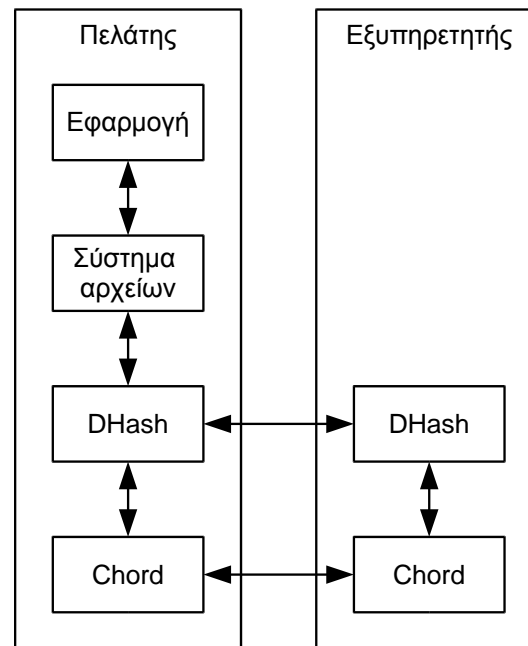
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# Εισαγωγή στο CFS (1 από 3)

- CFS: Αποθήκευση αρχείων μέσω Chord
  - Αντιστοίχιση ομάδας δεδομένων σε  $k$  κόμβους
  - Βαθμός αναπαραγωγής  $k$
  - Υποστηρίζει πλήρες σύστημα αρχείων



# Εισαγωγή στο CFS (2 από 3)

- Δομή του CFS
  - Το Chord απεικονίζει κλειδιά ομάδων σε κόμβους
  - Το DHash διαχειρίζεται τις ομάδες των δεδομένων
  - Το σύστημα αρχείων συνθέτει αρχεία και ευρετήρια
- Ομοιότητες με το PAST
  - Εικονικοί κόμβοι για εξισορρόπηση φόρτου
- Διαφορές με το PAST
  - Λειτουργία σε επίπεδο ομάδων και όχι αρχείων
    - Πιο ομοιόμορφη κατανομή δεδομένων
    - Απλούστερη διαχείριση αντιγράφων

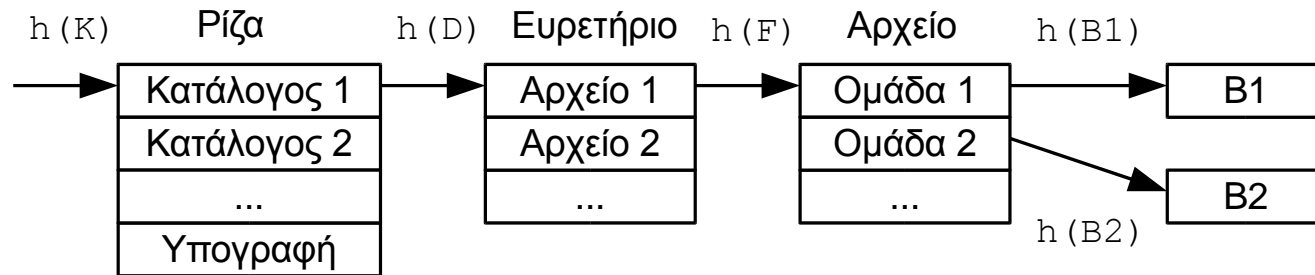
# Εισαγωγή στο CFS (3 από 3)

- Διεπαφή αρχείων του CFS
  - Παρόμοια με αυτή του UNIX
    - Υλοποιεί μοντέλο απομακρυσμένης πρόσβασης
  - Μόνο ο ιδιοκτήτης μπορεί να αλλάξει τα αρχεία
    - Για τους πελάτες τα αρχεία είναι ανάγνωσης μόνο
  - Οι ομάδες αποθηκεύονται για περιορισμένο χρόνο
    - Ο ιδιοκτήτης πρέπει να τις ανανεώνει
    - Αλλιώς οι ομάδες διαγράφονται
    - Δεν υπάρχει ρητή κλήση διαγραφής αρχείων
  - Υποθέτουμε σύστημα δημόσιας κρυπτογραφίας

# Σύστημα αρχείων (1 από 3)

- Ξεκινάμε με τις ομάδες δεδομένων B1 και B2
  - Αποθηκεύονται με κλειδιά  $h(B1)$  και  $h(B2)$
- Το αρχείο F περιέχει τα κλειδιά των ομάδων
  - Αντίστοιχο με το i-node στο UNIX
  - Αποθηκεύεται με κλειδί  $h(F)$
- Το ευρετήριο D περιέχει τα κλειδιά των αρχείων
  - Αποθηκεύεται με κλειδί  $h(D)$
- Το ευρετήριο ρίζας υπογράφεται από τον ιδιοκτήτη
  - Αποθηκεύεται με κλειδί  $h(K)$  (K είναι το δημόσιο κλειδί)
  - Ένα σύστημα αρχείων ανά δημόσιο κλειδί

# Σύστημα αρχείων (2 από 3)



- Έλεγχος δεδομένων

- Κάθε ομάδα ταιριάζει με το κλειδί της

- Η ρίζα ελέγχεται με βάση την υπογραφή

- Πρέπει να ταιριάζει με το κλειδί του ευρετηρίου

- Οι ομάδες ελέγχονται με βάση το περιεχόμενο

- Πρέπει να ταιριάζουν με τα κλειδιά τους

# Σύστημα αρχείων (3 από 3)

- Τροποποίηση αρχείων
  - Κάθε αλλαγή σε ομάδα αλλάζει το κλειδί της
    - Προέρχεται από κατακερματισμό της ομάδας
  - Τελικά αυτό απαιτεί αλλαγή της ρίζας
    - Το νέο κλειδί αλλάζει την ομάδα που καταχωρείται
  - Μόνο ο ιδιοκτήτης αλλάζει τα δεδομένα
    - Καταχωρεί νέα ρίζα και την υπογράφει ψηφιακά
    - Πρέπει να χρησιμοποιήσει το ίδιο δημόσιο κλειδί



# Διεπαφή του CFS (1 από 2)

- Η διαχείριση ομάδων γίνεται από το DHash
  - Το Chord χρησιμοποιείται για τις αναζητήσεις
  - Το DHash αξιοποιείται από το σύστημα αρχείων
- Διεπαφή του DHash
  - `put_h(block)`
    - Υπολογίζει το κλειδί κατακερματίζοντας το `block`
    - Στέλνει την ομάδα για αποθήκευση σε  $k$  κόμβους

# Διεπαφή του CFS (2 από 2)

- Διεπαφή του DHash
  - put\_s(block, pubkey)
    - Επιβεβαιώνει ότι η block είναι έγκυρη
      - Πρέπει να είναι ψηφιακά υπογεγραμμένη με pubkey
      - Χρησιμοποιείται μόνο για το ριζικό ευρετήριο
    - Υπολογίζει κλειδί με κατακερματισμό του pubkey
    - Στέλνει την ομάδα για αποθήκευση σε k κόμβους
  - block=get(key)
    - Ανακτά την ομάδα που προσδιορίζεται από το key

# Αποθήκευση ομάδων (1 από 2)

- Αποθήκευση ομάδων σε κόμβους
  - Το DHash εντοπίζει το διάδοχο της ομάδας
  - Ο κόμβος αποθηκεύει την ομάδα σε  $k$  διαδόχους
  - Οι διάδοχοι βρίσκονται στη λίστα διαδόχων
    - Η λίστα διαδόχων είναι μεγαλύτερη του  $k$
  - Αν αποτύχει αναλαμβάνει ο νέος διάδοχος
    - Έχει ήδη αποθηκευμένη την ομάδα

# Αποθήκευση ομάδων (2 από 2)

- Τυχαία κατανομή των αντιγράφων
  - Οι  $k$  κόμβοι είναι διάσπαρτοι στο δίκτυο
    - Πολύ απίθανο να αποτύχουν ταυτόχρονα
  - Στις ανακτήσεις ο διάδοχος επιστρέφει τους  $k$
  - Ο πελάτης επιλέγει τον τοπολογικά πλησιέστερο
    - Αρκετά πιθανό κάποιος από τους  $k$  να είναι κοντά

# Προσωρινή αποθήκευση

- Προσωρινή αποθήκευση ομάδων
  - Όταν γίνεται αναζήτηση οι ενδιάμεσοι κόμβοι ερωτώνται
    - Αν περιέχουν την ομάδα, την επιστρέφουν άμεσα
    - Αλλιώς θα τη λάβουν από τον πελάτη όταν την ανακτήσει
  - Δεν απαιτείται έλεγχος εγκυρότητας των ομάδων
    - Το κλειδί τους πιστοποιεί αν είναι έγκυρες
    - Αν αλλάξουν, αλλάζει και το κλειδί τους
  - Εξαιρείται το ευρετήριο ρίζας
    - Το κλειδί του δεν αλλάζει όταν τροποποιείται
    - Πρέπει να εξετάζεται πόσο πρόσφατο είναι

# Ανανέωση αρχείων

- Περιοδική ανανέωση αρχείων
  - Στο CFS οι ομάδες δεν διαγράφονται ρητά
    - Απλά διαγράφονται όταν παλιώσουν πολύ
  - Το σύστημα αρχείων διασχίζεται περιοδικά
    - Ξεκινάμε από τη ρίζα και φτάνουμε στα φύλλα
    - Κάθε ομάδα που εντοπίζεται αποθηκεύεται ξανά
    - Οι ομάδες που έχουν διαγραφεί δεν ανανεώνονται

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

# Τέλος Ενότητας # 10

**Μάθημα:** Κατανεμημένα Συστήματα, **Ενότητα # 10:** Κατανεμημένα συστήματα αποθήκευσης

**Διδάσκων:** Γιώργος Ξυλωμένος, **Τμήμα:** Πληροφορικής



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

