

Τεχνικές Ανάκαμψης

Επιθυμητές Ιδιότητες μιας Δοσοληφίας

Ιδιότητες Δοσοληφιών

- **Atomicity (ατομικότητα)** - είτε όλες οι πράξεις είτε καμία
- **Consistency (συνέπεια)** - διατήρηση συνέπειας της ΒΔ
- **Isolation (απομόνωση)** - δεν αποκαλύπτει ενδιάμεσα αποτελέσματα
- **Durability (μονιμότητα ή διάρκεια)** - μετά την επικύρωση μιας δοσοληφίας οι αλλαγές δεν είναι δυνατόν να χαθούν

Επιθυμητές Ιδιότητες μιας Δοσοληφίας

- **Atomicity (ατομικότητα)** → **ΤΕΧΝΙΚΕΣ ΑΝΑΚΑΜΨΕΙΣ**
- **Consistency (συνέπεια)** → ΥΠΕΥΘΥΝΟΤΗΤΑ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΙΣΤΗ
- **Isolation (απομόνωση)** → ΕΛΕΓΧΟΣ ΣΥΝΔΡΟΜΙΚΟΤΗΤΑΣ
- **Durability (μονιμότητα ή διάρκεια)** → **ΤΕΧΝΙΚΕΣ ΑΝΑΚΑΜΨΕΙΣ**

Ανάκαμψη από Αποτυχίες

Είδη Αποτυχιών

Δυο κατηγορίες: καταστροφή ή όχι της μόνιμης αποθήκευσης (δίσκου)

Παραδείγματα αποτυχιών ...

Παραδείγματα Αποτυχιών

Είδη Αποτυχιών

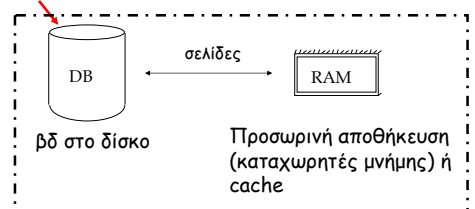
1. Αποτυχία του υπολογιστή (system crash)
2. Σφάλμα δοσοληφίας ή συστήματος (π.χ., προγραμματιστικό λάθος)
3. Εφαρμογή ελέγχου συνδρομικότητας
4. Πρόβλημα δίσκου
5. Φυσικά προβλήματα και καταστροφές

Στόχος της ανάκαμψης: επαναφορά στην πιο πρόσφατη συνεπή κατάσταση ακριβώς πριν τη στιγμή της αποτυχίας

Έννοιες Ανάκαμψης

Ημερολόγιο Συστήματος (system log)
Πληροφορίες για κάθε δοσοληφία

το μοντέλο του συστήματος



Ημερολόγιο Συστήματος (Log)

- Για να είναι δυνατή η ανάκαμψη από αποτυχίες, καταχωρούνται πληροφορίες για τις πράξεις των δοσοληψιών
- Αποθηκεύονται στο δίσκο
- Τύποι πληροφορίας: έναρξη δοσοληψίας
εγγραφή στοιχείου (παλιά, νέα τιμή)
ανάγνωση στοιχείου
επικύρωση/ακύρωση

Ανάκαμψη από Αποτυχίες

Περιπτώσεις αποτυχίας

1. Καταστροφική αποτυχία (δίσκου)
Backup + (backupted) log
2. Μη καταστροφική αποτυχία
Χρήση μόνο του disk log

Επιθυμητές Ιδιότητες μιας Δοσοληψίας

Ιδιότητες Δοσοληψιών

- **Atomicity (ατομικότητα)** - είτε όλες οι πράξεις είτε καμία
- **Durability (μονιμότητα ή διάρκεια)** - μετά την επικύρωση μιας δοσοληψίας οι αλλαγές δεν είναι δυνατόν να χαθούν

Ένοιες Ανάκαμψης

⌘ Ατομικότητα:

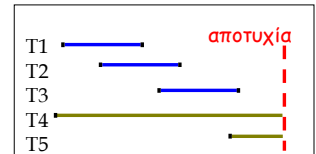
☑ Οι δοσοληψίες μπορεί να αναιρεθούν ("Rollback").

⌘ Διάρκεια:

☑ Τι θα συμβεί αν το ΣΔΒΔ καταρρεύσει?

❖ Επιθυμητή συμπεριφορά μετά από μια αποτυχία:

- T1, T2 & T3 πρέπει να έχουν διάρκεια (redo-επανάληψη)
- T4 & T5 πρέπει να αναιρεθούν (undo-αναίρεση)



Ένοιες Ανάκαμψης

⌘ Ατομικότητα:

☑ όχι επικυρωμένες -- undo - αναίρεση

⌘ Διάρκεια:

☑ επικυρωμένες -- redo -- επανάληψη

Ένοιες Ανάκαμψης

Στην πιο εύκολη περίπτωση

- όλες οι ενημερώσεις μιας δοσοληψίας γράφονται στον τοπικό χώρο εργασίας της δοσοληψίας
- η βδ ενημερώνεται *μόνον* αφού μια δοσοληψία φτάσει στο σημείο επικύρωσής της

Στην πραγματικότητα --Steal (flush) ::

- κάποιες ενημερώσεις μιας δοσοληψίας γράφονται στον δίσκο (βδ) πριν η δοσοληψία επικυρωθεί
- Γιατί; βελτίωση throughput
- Πρόβλημα:
Ατομικότητα (ανάρρηση (undo) στην περίπτωση που η δοσοληψία αποτύχει)

Στην πραγματικότητα -- No Force ::

- κάποιες ενημερώσεις μιας δοσοληψίας δε γράφονται στον δίσκο (βδ) ακόμα και αφού η δοσοληψία επικυρωθεί
- Γιατί; βελτίωση χρόνου απόκρισης
- Πρόβλημα: Διάρκεια (επανάληψη (redo) στην περίπτωση που το σύστημα αποτύχει)

	No Steal	Steal
Force	Εύκολο	
No Force		Επιθυμητό

Κάποια ορολογία:

όλες οι ενημερώσεις μιας δοσοληψίας γράφονται στον τοπικό χώρο εργασίας της δοσοληψίας -- η βδ ενημερώνεται *μόνον* αφού μια δοσοληψία φτάσει στο σημείο επικύρωσης της

ετεροχρονισμός των ενημερώσεων -- deferred updates

- χωρίς ανάρρηση/ με επανάληψη (no undo/redo)

κάποιες ενημερώσεις μιας δοσοληψίας γράφονται στη βδ πριν μια δοσοληψία φτάσει στο σημείο επικύρωσης της

άμεση ενημέρωση -- immediate updates

- με ανάρρηση/ με επανάληψη (undo/no redo)

- Κατάλογος σελίδων που είναι στη μνήμη (cache)
- Δυαδικό ψηφίο τροποποίησης (dirty bit)

- Καταχώρηση τύπου ΕΠΑΝΑΛΗΨΗ (REDO)
Περιέχει την νέα τιμή (AFter Image)
- Καταχώρηση τύπου ΑΝΑΙΡΕΣΗ (UNDO)
Περιέχει την παλιά τιμή (BeFore Image)

Το πρωτόκολλο προεγγραφής ημερολογίου (Write Ahead Log)

- (1) Η καταχώρηση στο log για μια εγγραφή γράφεται στο δίσκο **πριν** οι αντίστοιχες σελίδες να γραφούν στο δίσκο
- (2) Πριν την επικύρωση (commit) μιας δοσοληψίας όλες οι καταχωρήσεις του log που την αφορούν γράφονται στο δίσκο

Το (1) δίνει ατομικότητα

Το (2) δίνει διάρκεια

Θα δούμε μια συγκεκριμένη υλοποίηση (ARIES)

Σειριακή εγγραφή στο log

Για κάθε εγγραφή στοιχείου, στο log μια καταχώρηση:

$\langle \text{XID, pageID, offset, length, old data, new data} \rangle$

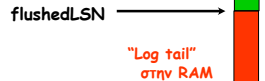


☒ Κάθε καταχώρηση στο log έχει ένα μοναδικό αριθμό **Log Sequence Number (LSN)**.

☒ LSNs πάντα αυξάνουν.

☒ Το log γράφεται στο δίσκο. Το σύστημα κρατά τη μεταβλητή **flushedLSN**: το μέγιστο LSN που γράφτηκε στο δίσκο μέχρι τώρα

Καταχωρήσεις του log ήδη flushed στο δίσκο



Αν το σύστημα αποτύχει, μένουμε με το «πράσινο» κομμάτι

☒ Κάθε σελίδα δεδομένων περιέχει έναν **pageLSN**: το πιο πρόσφατο LSN για εγγραφή σε αυτή τη σελίδα

☒ **WAL**: Πριν γραφτεί μια σελίδα, **pageLSN ≤ flushedLSN**

Καταχωρήσεις του log ήδη flushed στο δίσκο



WAL(1) Η καταχώρηση στο log για μια εγγραφή γράφεται στο δίσκο πριν οι αντίστοιχες σελίδες να γραφούν στο δίσκο. Ακόμα και αν η δοσοληψία αναρριθεί μπορούμε να χρησιμοποιήσουμε το log για να γίνει undo η εγγραφή

Πεδία LogRecord:

Προηγούμενο LSN για τη δοσοληψία

prevLSN
XID
type

Μόνο για εγγραφές

pageID
length
offset
before-image
after-image

Τύποι καταχωρήσεων

- ☒ Update
- ☒ Commit
- ☒ Abort
- ☒ End
- ☒ Compensation Log Records (CLRs)
- ☒ για πράξεις UNDO

⌘ Πίνακας Δοσοληψιών:

- ☒ Μια εγγραφή για κάθε ενεργή δοσοληψία
- ☒ Πεδία εγγραφής:
 - XID**,
 - status**, (running/committed/aborted),
 - lastLSN** (τελευταίο LSN που αφορά τη δοσοληψία).

XID	status	lastLSN
-----	--------	---------

⌘ Πίνακας Τροποποιημένων Σελίδων:

- ☒ Μια εγγραφή για κάθε τροποποιημένη σελίδα στη μνήμη.
- ☒ Πεδία εγγραφής
 - recLSN** -- LSN της καταχώρησης του log της πρώτης αλλαγής στη σελίδα

PID	recLSN
-----	--------



LogRecords

prevLSN
XID
type
pageID
length
offset
before-image
after-image



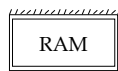
Σελίδες Δεδομένων
σε κάθε σελίδα
pageLSN



Πίνακας Δοσοληψιών
lastLSN
status

Πίνακας Τροποποιημένων Σελίδων
recLSN

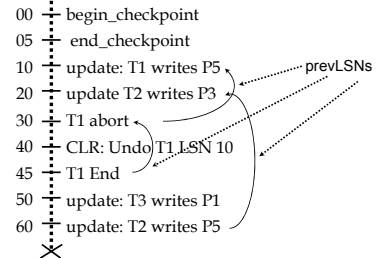
flushedLSN



Πίνακας Δοσοληψιών
lastLSN
status

Πίνακας Τροποποιημένων Σελίδων
recLSN
flushedLSN

LSN LOG



- Μια ακολουθία από reads & writes, που τελειώνουν με commit ή abort.
 - Υποθέτουμε ότι οι εγγραφές στο δίσκο είναι ατομικές
- Αυστηρό 2PL ⇒ αυστηρά χρονοπρογράμματα
- STEAL, NO-FORCE, με προεγγραφή log (Write-Ahead Logging)

1. Γράψε μια καταχώρηση τύπου **commit** στο log
2. Γράψε στο δίσκο όλες τις καταχωρήσεις του log έως το lastLSN της δοσοληψίας
flushedLSN ≥ lastLSN (WAL)
3. Τέλος του Commit
4. Γράψε μια καταχώρηση τύπου **end** στο log

• Αρχικά, ας θεωρήσουμε ότι δε συμβαίνουν αποτυχίες

ΣΤΟΧΟΣ: UNDO (αναίρεση) των εγγραφών (updates)

«Παίζουμε» το log της δοσοληψίας ανάποδα

1. Γράφουμε στο log μια καταχώρηση τύπου **abort**
2. Παίρνουμε το **lastLSN** της δοσοληψίας από τον Πίνακα Δοσοληψιών
3. Ακολουθούμε την αλυσίδα των καταχωρήσεων στο log βάσει του πεδίου **prevLSN**

• Για να το εκτελέσουμε την αναίρεση πρέπει να έχουμε *κλειδί* στα δεδομένα

• Πριν να γράψουμε την προηγούμενη τιμή σε μια σελίδα, γράφουμε στο log μια καταχώρηση τύπου **CLR**

Η καταχώρηση **CLR** έχει ένα πεδίο **undonextLSN**: επόμενο LSN προς αναίρεση

Μια καταχώρηση τύπου **CLR** δε χρειάζεται να ανααιρεθεί (αλλά μπορεί να χρειαστεί να επαναληφθεί)

• Στο τέλος, γράφουμε μια καταχώρηση τύπου **end**

Περιοδικά, το ΣΔΒΔ δημιουργεί ένα **σημείο ελέγχου (checkpoint)**, με σκοπό τη μείωση του χρόνου που χρειάζεται για ανάκαμψη στην περίπτωση αποτυχίας

• Γράφει τον Πίνακα Δοσοληψιών και τον Πίνακα Τροποποιημένων Σελίδων στο log

⌘ Γράφει στο log:

για εγγραφή **begin_checkpoint**: που δηλώνει πότε άρχισε το chkpt.

για εγγραφή **end_checkpoint**: που δηλώνει πότε τέλειωσε το chkpt

▪ Ασφές (fuzzy)

▪ Οι δοσοληψίες εξακολουθούν να εκτελούνται. Οι πίνακες είναι ακριβείς ως προς το **begin_checkpoint**

▪ Δεν γράφονται τροποποιημένες σελίδες στο δίσκο

Το LSN της εγγραφής του chkpt γράφεται σε ασφαλή χώρο (**master record**)



LogRecords

prevLSN
XID
type
pageID
length
offset
before-image
after-image



Σελίδες Δεδομένων
σε κάθε σελίδα
pageLSN

master record

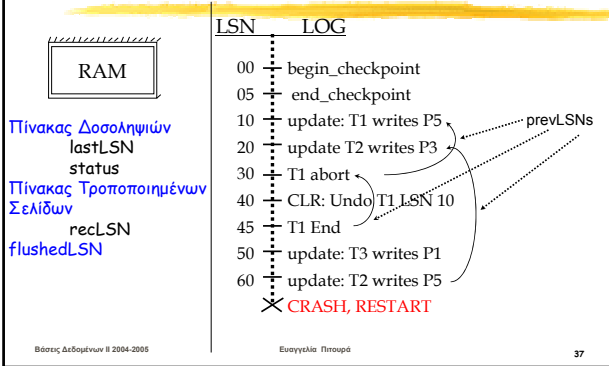


Πίνακας Δοσοληψιών
lastLSN
status

Πίνακας Τροποποιημένων Σελίδων
recLSN

flushedLSN

ARIES: Παράδειγμα



ARIES: Ανάκαμψη από Αποτυχία

Τρεις Φάσεις

ΑΝΑΛΥΣΗ -- ανακατασκευή των πινάκων του συστήματος: ποιες σελίδες είναι τροποποιημένες, ποια είναι η κατάσταση των δοσοληψιών

ΕΠΑΝΑΛΗΨΗ (REDO) -- επανάληψη όλων των δοσοληψιών

ΑΚΥΡΩΣΗ (UNDO) -- αφαίρεση των δοσοληψιών που δεν ήταν επικυρωμένες

ARIES: Η Φάση της Ανάλυσης

ΑΝΑΛΥΣΗ -- ανακατασκευή των πινάκων του συστήματος (ΠΙΝΑΚΑΣ ΔΟΣΟΛΗΨΙΩΝ & ΠΙΝΑΚΑΣ ΤΡΟΠΟΠΟΙΗΜΕΝΩΝ ΣΕΛΙΔΩΝ) χρησιμοποιώντας το τελευταίο ολοκληρωμένο checkpoint:

ποια είναι η κατάσταση των δοσοληψιών από τότε

ποιες σελίδες είναι τροποποιημένες από τότε

ARIES: Η Φάση της Ανάλυσης

- Αρχή από το τελευταίο checkpoint
- Εύρεση του με χρήση του master record
- Βρίσκουμε το αντίστοιχο begin_checkpoint

ARIES: Η Φάση της Ανάλυσης

Διαβάζουμε το log forward ξεκινώντας από το **begin_checkpoint**. Αν βρούμε καταχώρηση τύπου:

end: σβήνουμε τη δοσοληψία από τον Πίνακα Δοσοληψιών

άλλο τύπο: εισαγωγή δοσοληψίας (αν δεν υπάρχει ήδη)

αλλαγή **lastLSN = LSN**, και αλλαγή κατάστασης αν commit (αλλιώς πρέπει να αναιρεθεί)

update (εγγραφή):

Αν η σελίδα P δεν είναι στον Πίνακα Τροποποιημένων Σελίδων

εισαγωγή του P στον πίνακα με **recLSN = LSN**.

ARIES: Η Φάση της Ανάλυσης

Αποτέλεσμα οι πίνακες όπως ήταν τη στιγμή που γράψαμε το τελευταίο log στο δίσκο (τελευταία επικύρωση δοσοληψίας)

Πίνακας Δοσοληψιών: όλες οι δοσοληψίες που ήταν ενεργές τη στιγμή της αποτυχίας

Πίνακας Τροποποιημένων Σελίδων: όλες οι σελίδες που έχουν τροποποιηθεί (κάποιες μπορεί να έχουν γραφτεί ήδη στο δίσκο, μα δεν έχουμε αυτή τη πληροφορία στο log)

Τεχνικές Ανάκαμψης από Σφάλματα

Άσκηση 8

Φάση της Ανάλυσης

00 begin_checkpoint	T1 abort 70	P5 20
10 end_checkpoint	T3 running 60	P3 60
20 update: T1 writes P5	Πίνακας Δοσοληψιών	Πίνακας Τροποποιημένων Σελίδων
40 T2 Commit		
50 T2 end		
60 update: T3 writes P3		
70 T1 abort		

ARIES: Ανάκαμψη από Αποτυχία

Τρεις Φάσεις

ΑΝΑΛΥΣΗ --

ΕΠΑΝΑΛΗΨΗ (REDO) -- επανάληψη όλων των δοσοληψιών

ΑΚΥΡΩΣΗ (UNDO) -- αναιρέση των δοσοληψιών που δεν ήταν επικυρωμένες

ARIES: Η Φάση της Επανάληψης

- Επαναλαμβάνουμε το log ώστε να φτάσουμε στην κατάσταση τη στιγμή της αποτυχίας
- Επαναλαμβάνουμε τις εγγραφές ακόμα και των δοσοληψιών που ακυρώθηκαν (aborted)

Προχωράμε στο log forward ξεκινώντας από το μικρότερο $recLSN$ στον Πίνακα Τροποποιημένων Σελίδων

Αρχαιότερη ενημέρωση που ίσως να μην έχει γραφτεί στο δίσκο

ARIES: Η Φάση της Επανάληψης

Για κάθε καταχώρηση τύπου **update** ή **CLR**, επανέλαβε την πράξη εκτός αν:

- η σελίδα δεν είναι στον Πίνακα Τροποποιημένων Σελίδων
- η σελίδα είναι στον Πίνακα Τροποποιημένων Σελίδων αλλά $recLSN > LSN$
- $pageLSN$ (στο δίσκο) $\geq LSN$
- Σημείωση: δε γνωρίζουμε ποιες σελίδες έχουν γραφτεί στο δίσκο

ARIES: Η Φάση της Επανάληψης

Για να επαναλάβεις μια πράξη:

- επανεκτέλεση
- θέσε $pageLSN = LSN$
- δε χρειάζεται log

Στο τέλος end για όσες τύπου C

Τεχνικές Ανάκαμψης από Σφάλματα

Άσκηση 8

Φάση της Ανάλυσης

00 begin_checkpoint	T1 abort 70	P5 20
10 end_checkpoint	T3 running 60	P3 60
20 update: T1 writes P5	Πίνακας Δοσοληψιών	Πίνακας Τροποποιημένων Σελίδων
40 T2 Commit	Φάση της Επανάληψης (REDO)	
50 T2 end	Ξεκινάμε από το 20, γιατί:	
60 update: T3 writes P3		
70 T1 abort		

Τρεις Φάσεις

ΑΝΑΛΥΣΗ --

ΕΠΙΤΑΝΑΛΗΨΗ (REDO) -- επανάληψη όλων των δοσοληψιών

ΑΚΥΡΩΣΗ (UNDO) -- αναίρεση των δοσοληψιών που δεν ήταν επικυρωμένες

Κατασκευή λίστας

ToUndo = {lastLSN της T | T δοσοληψία προς αναίρεση}

δοσοληψίες προς αναίρεση: δοσοληψίες ενεργές κατά την αποτυχία

Repeat:

Επέλεξε το μεγαλύτερο LSN στην λίστα ToUndo.
 If το LSN είναι τύπου CLR και undonexLTSN == NULL γράψε μια καταχώρηση τύπου end για τη δοσοληψία.
 If το LSN είναι τύπου CLR και undonexLTSN != NULL Πρόσθεσε το undonexLTSN στην λίστα ToUndo
 Else if το LSN είναι τύπου update.
 Undo το update,
 Γράψε μια καταχώρηση τύπου CLR,
 Πρόσθεσε το prevLSN στην λίστα ToUndo.

Until ToUndo is empty.

Άσκηση 8

Φάση της Ανάλυσης

00 begin_checkpoint	T1 abort 70	P5 20
10 end_checkpoint	T3 running 60	P3 60

20 update: T1 writes P5 Πίνακας Δοσοληψιών Πίνακας Τροποποιημένων Σελίδων
 40 T2 Commit
 50 T2 end
 60 update: T3 writes P3
 70 T1 abort

Φάση της Επανάληψης (REDO)
 Ξεκινάμε από το 20, γιατί;

Φάση της Αναίρεσης (UNDO)
 ToUndo = {70, 60}

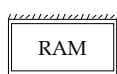
Παλιότερη καταχώρηση log δοσοληψίας ενεργής κατά την αποτυχία

Μικρότερο recLSN στον Πίνακα Τροποποιημένων Σελίδων μετά την Ανάλυση

Τελευταίο chkrpt ΑΠΟΤΥΧΙΑ



- ❖ Άρχισε από ένα σημείο ελέγχου (που βρίσκεται μέσω του master record).
- ❖ Τρεις Φάσεις.
 - Εύρεση ποιες δοσοληψίες μετά το σημείο ελέγχου επικυρώθηκαν, ποιες ακυρώθηκαν (ΑΝΑΛΥΣΗ).
 - REDO όλες τις δοσοληψίες.
 - ♦ (repeat history)
 - UNDO το αποτέλεσμα των αποτυχημένων δοσοληψιών

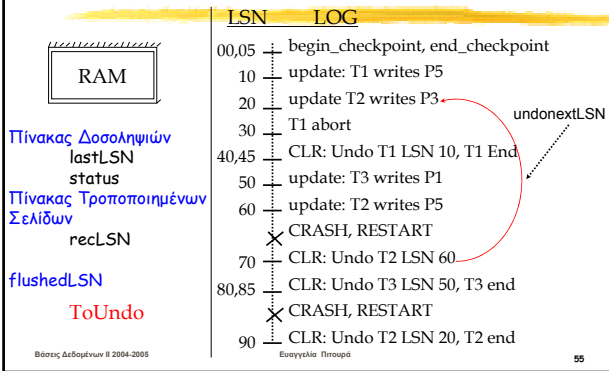


Πίνακας Δοσοληψιών
 lastLSN
 status
 Πίνακας Τροποποιημένων Σελίδων
 recLSN
 flushedLSN
 ToUndo

LSN	LOG
00	begin_checkpoint
05	end_checkpoint
10	update: T1 writes P5
20	update: T2 writes P3
30	T1 abort
40	CLR: Undo T1 LSN 10
45	T1 End
50	update: T3 writes P1
60	update: T2 writes P5

prevLSNs (pointing to LSN 10, 20, 30, 45, 50, 60)
 ✗ CRASH, RESTART

ARIES: Παράδειγμα



Τεχνικές Ανάκαμψης από Σφάλματα

Στο ARIES υποθέτουμε αυστηρό 2PL, άρα όχι διάδοση ανακλήσεων, εγγραφές

Γενικά μπορεί να χρειαστεί ανάκληση

(άρα στο log εκτός των εντολών εγγραφής και εντολής ανάγνωσης, για να ελέξουμε ποιες πρέπει να ανακληθούν)

Τεχνικές Ανάκαμψης

- Εφεδρικά αντίγραφα
- Σκιώδης Σελιδοποίηση