

CS 4604: Introduction to Database Management Systems

B. Aditya Prakash

Final Review

Final Exam

- **30% of the grade**
- **No books, no notes, no laptops**
- **Allowed:**
 - **Only** 2 letter-size pages
 - You can use both sides
 - Must be **hand-written**
 - **And** a calculator (recommended)
- **Duration: 2 hours. 7:45-9:45am Dec 7**
Location: regular classroom

Syllabus

- **Comprehensive exam**
 - But main focus towards and emphasis on post-midterm stuff (= starting from lecture 10)
 - Will cover all material in all lectures
 - **EXCEPT (i.e. things NOT in exam)**
 1. NoSQL/MapReduce
(No PHP too of course)



Office Hours this week

- All posted on Piazza.

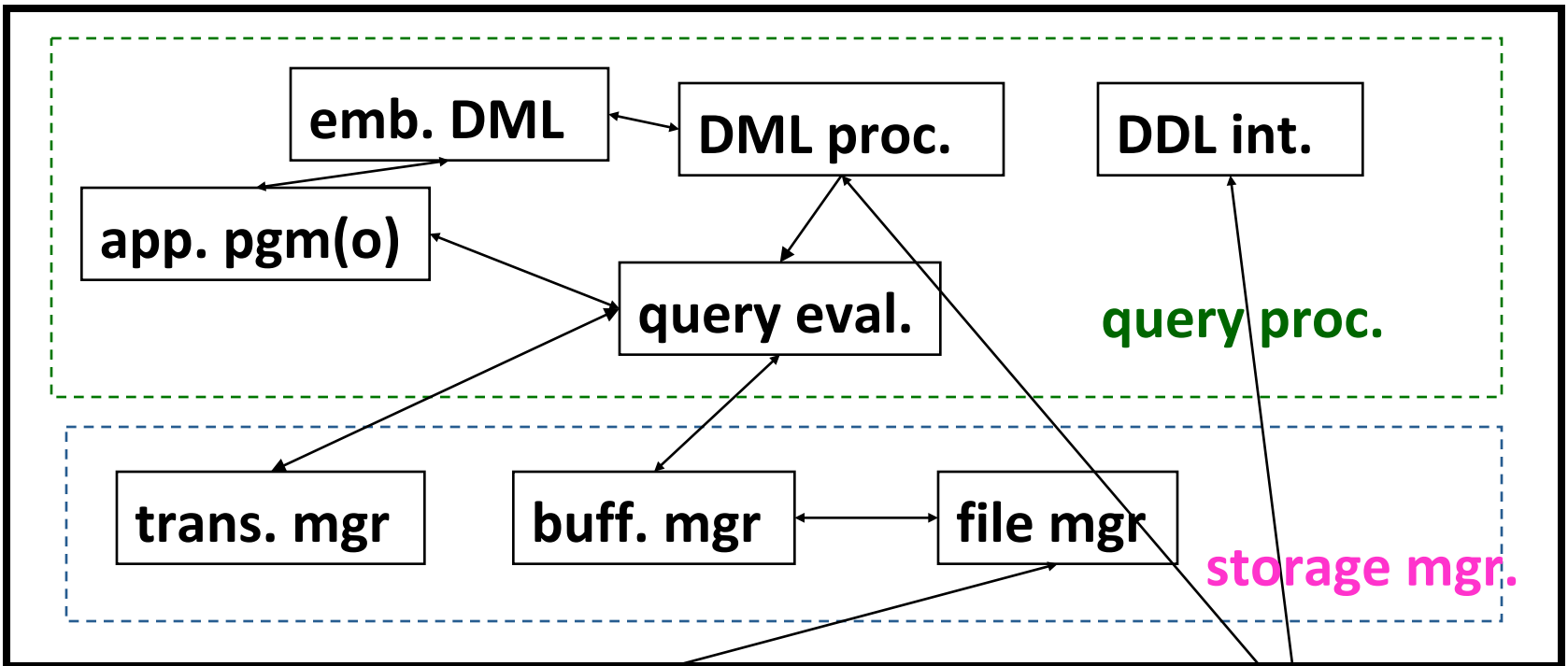
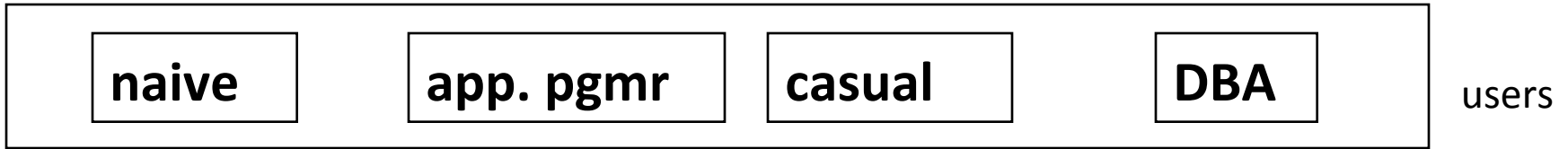
Reminder

- Project Demos on Dec 5 in Torg 3160F.

OVERVIEW

What you learnt in the course

- Weeks 1–4: Query/ Manipulation Languages and Data Modeling
 - Relational Algebra
 - Data definition
 - Programming with SQL
 - Entity-Relationship (E/R) approach
 - Specifying Constraints
 - Good E/R design
- Weeks 5–8: Indexes, Processing and Optimization
 - Storing
 - Hashing/Sorting
 - Query Optimization
 - NoSQL and Hadoop
- Week 9-10: Relational Design
 - Functional Dependencies
 - Normalization to avoid redundancy
- Week 11-12: Concurrency Control
 - Transactions
 - Logging and Recovery
- Week 13–14: Students' choice
 - Practice Problems
 - XML
 - Data mining and warehousing



SQL/RA

- Make sure you know all the operators for SQL and RA
 - Select, From, Where, Group-by, Having, Order-by
 - Set-semantics/Bag-semantics
- The base for DB

ER

- You should already have enough practice!

FDs

- Definitions of FDs, closures (Attributes vs FDs), cover, normal forms, decompositions etc. etc.
 - Pay attention to multiple ways of defining the same thing!
 - E.g. ‘Key’: multiple ways of defining and understanding
- Various procedures to compute the above

Indexing and Hashing

- Know your basic structure, and definitions
- Less emphasis (as we have covered this in the midterm)

Query Processing

- Estimating costs
 - What are you estimating? = #disk accesses
 - How to estimate?
 - sorting
 - Different types of joins (NLJ, Block-NLJ, SMJ, HJ)
 - Don't just memorize the formulae, understand how they are derived, the 'best-case' 'worst-case' scenarios

Query Optimization

- Algebraic manipulation
- Selectivity estimation
 - Many cases
 - How to use selectivities to get the output size

Transactions

- ACID
- Problems with concurrency and Serializability concept
- Conflict-Serializability, how to detect
- 2PL, when, why, what, how, limitations
- Strict 2PL, when, why, what, how, limitations
- Know your venn diagrams!
- Deadlocks, how to detect and avoid them
- Dependency graph vs Waits-for graphs

Logging and Recovery: Big Picture



LogRecords

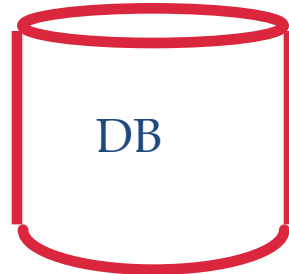
prevLSN
XID
type

update
CLR

pageID
length
offset
before-image
after-image

CLR

undoNextLSN



Data pages
each with a
pageLSN

master record
LSN of most
recent checkpoint

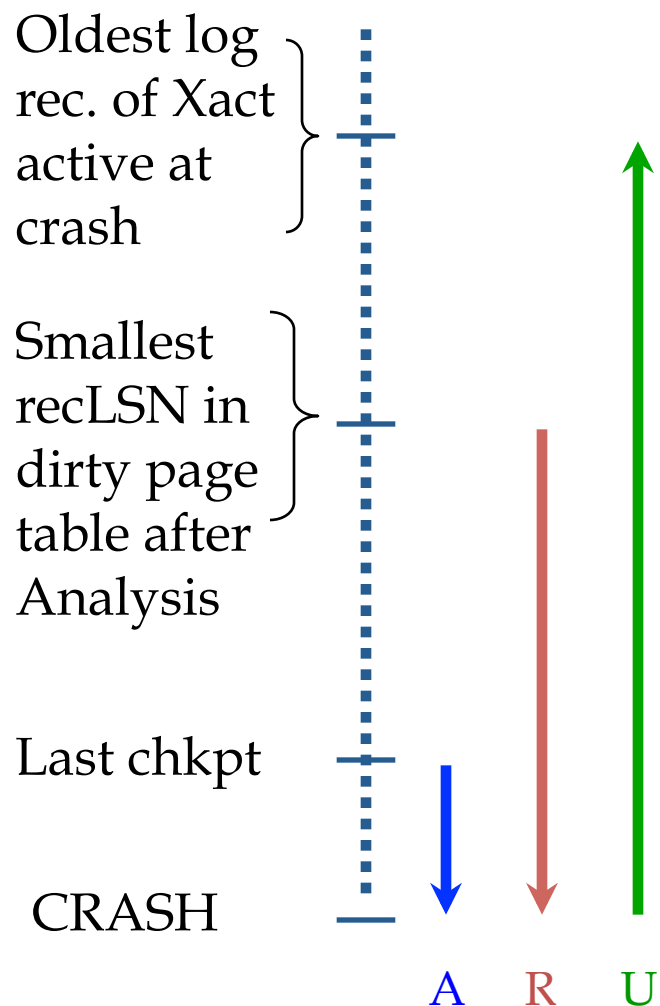


Xact Table
lastLSN
status

Dirty Page Table
recLSN

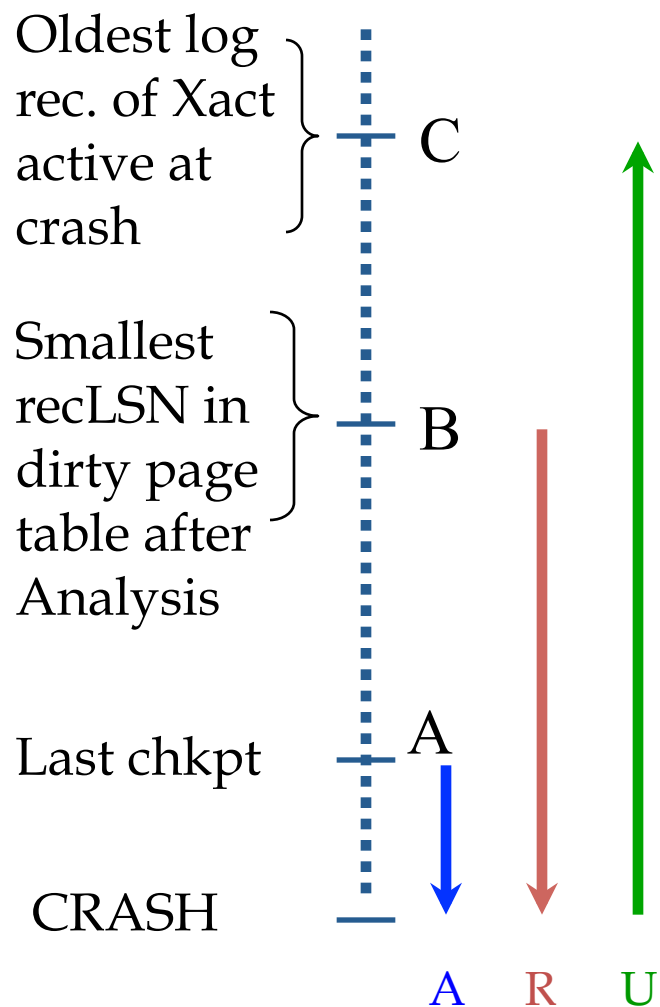
flushedLSN

Crash Recovery: Big Picture



- Start from a **checkpoint** (found via **master** record).
- Three phases.
 - **Analysis** - Figure out which Xacts committed since checkpoint, which failed.
 - **REDO** all actions (repeat history)
 - **UNDO** effects of failed Xacts.

Crash Recovery: Big Picture



- Notice: relative ordering of A, B, C may vary!

Logging and Recovery

- Make sure you know **exactly** how recovery takes place, and what is logged
 - Practice, practice
 - Check out problems in lectures, practice problems and hws
 - Be comfortable with small conceptual questions (see practice problems)

Tips

- Know your definitions!
 - Different ways of defining same thing e.g. keys
- Go through the slides
 - Checking the textbook if you are unclear
- **Go through HWs, Handouts, Exams, and Practice problems**
 - Textbook also has good problems! Even numbered problems have solutions on-line
 - Take advantage of our office hours
- Make use of your 2 allowed written notes!
- Bring a calculator

More

- Read the whole question carefully before answering
- Raise your hand if you need any clarification

Data Management

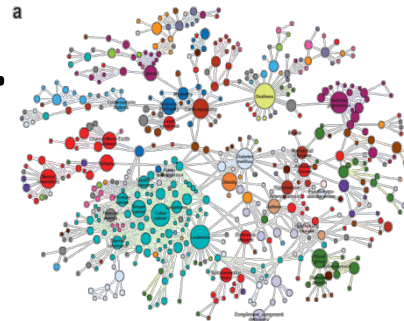
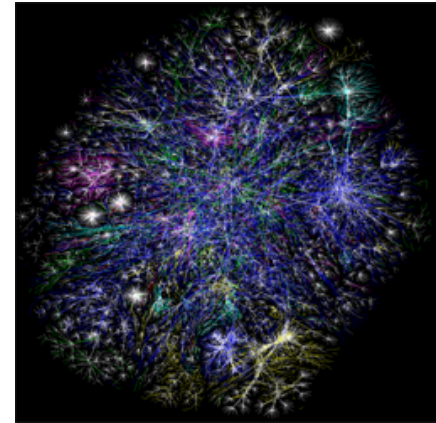
- Is a really exciting field ('BIG-Data')
- High commercial *and* academic research interest

Lots more stuff we did not cover

- Storage Manager
 - File organization
- More details about query processing
 - Fine-tuning Join algorithms
- Other powerful query languages
 - Datalog etc.
- More sophisticated locking, concurrency control
 - E.g. Hierarchical locking, time-stamped CC
- Spatial Data Management
- Distributed Databases
- More advanced data mining
- More details on NoSQL/Map Reduce etc.
-

Course Plug: CS 5614: (Big) Data Management Systems

- Spring 2019: Graduate level course
- Project, research papers
 - Will cover the state-of-the-art
- Would be exciting and fun!
- Good way to get exposed to the state-of-the-art in large data management and mining, network analysis, graph databases etc.



Good Luck!

- Especially for those of you will graduate!
- Feel free to keep in touch 😊

