SimpleDB Overview

CSE 444 – Section 1

What you should NOT do

- Modify given classes
  - Removing, renaming, relocating to other packages
- Modify given methods
  - Changing parameters or return types
- Use third-party libraries
  - Except the ones under lib/ directory
  - You can do everything using regular Java libraries

What you CAN do

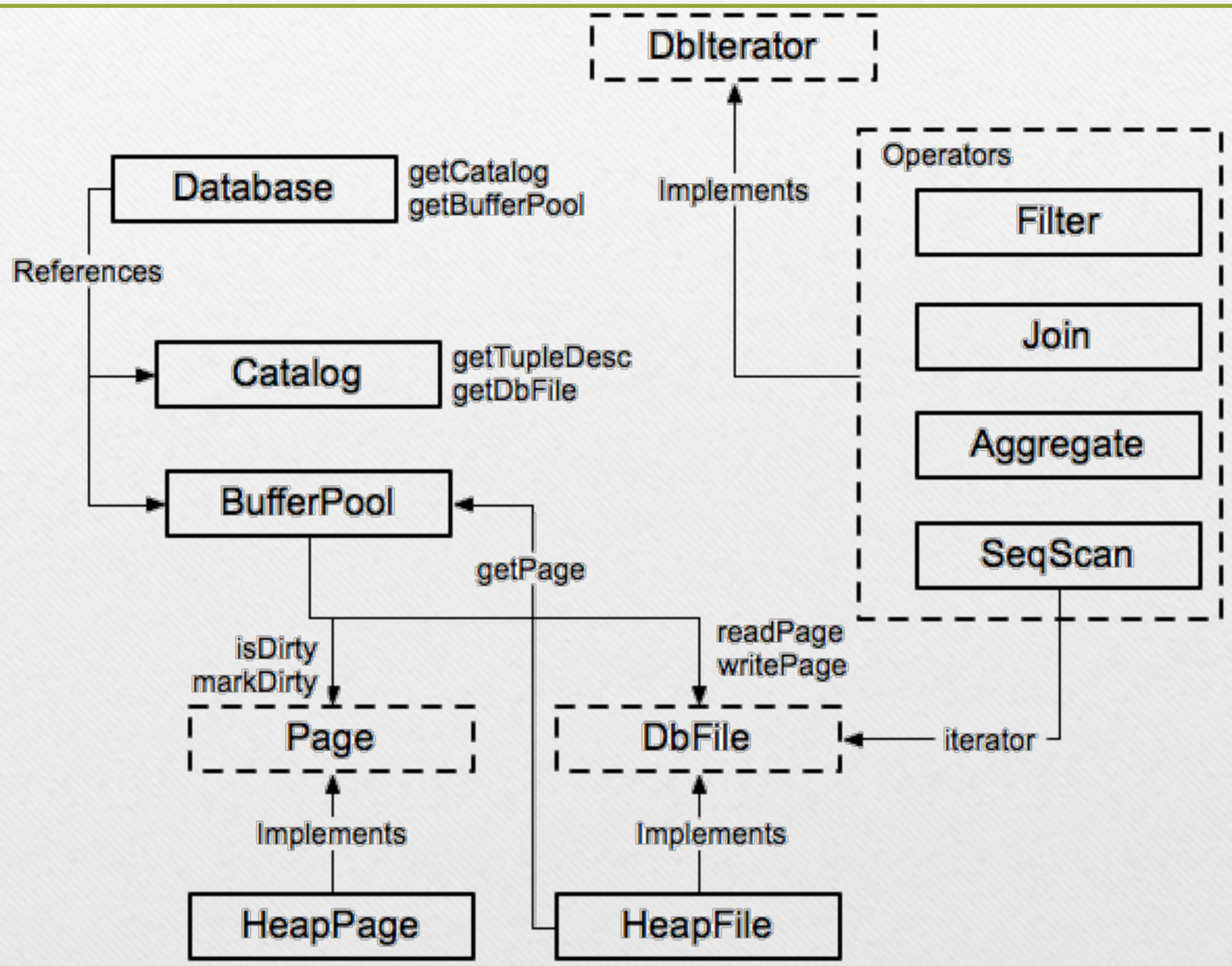
- Add new classes/interfaces/methods/packages
  - Watch out for name conflicts with future labs!
  - Safer choice: use inner classes or new packages
- Re-implement provided methods
  - Just don't destroy correctness or specification!
- Find bugs!
  - Free candy bar (or a salad) for reporting new bugs

Setting up SimpleDB

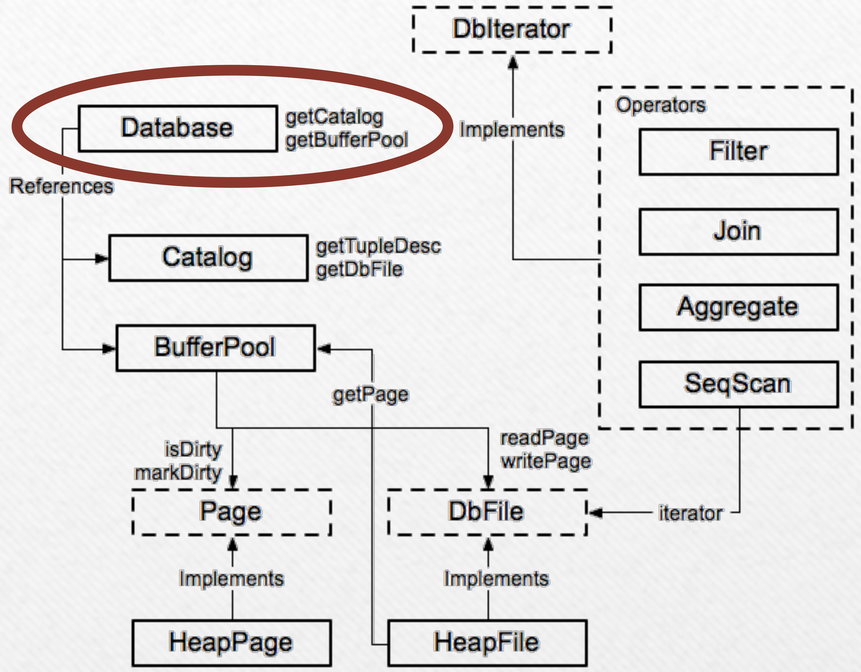
Any questions or concerns? Demo?

JUnit testing. Everybody familiar?

Overview of SimpleDB

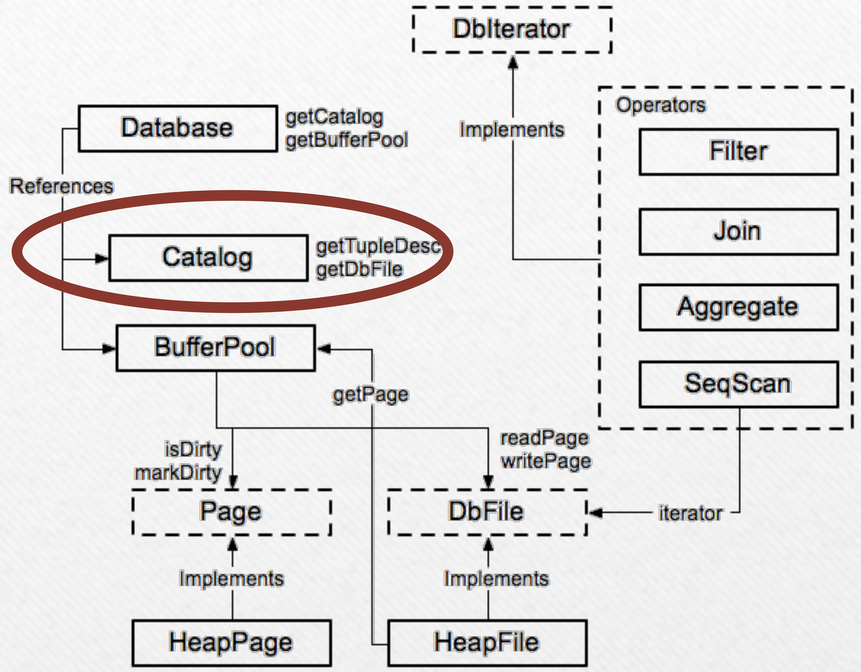


Database



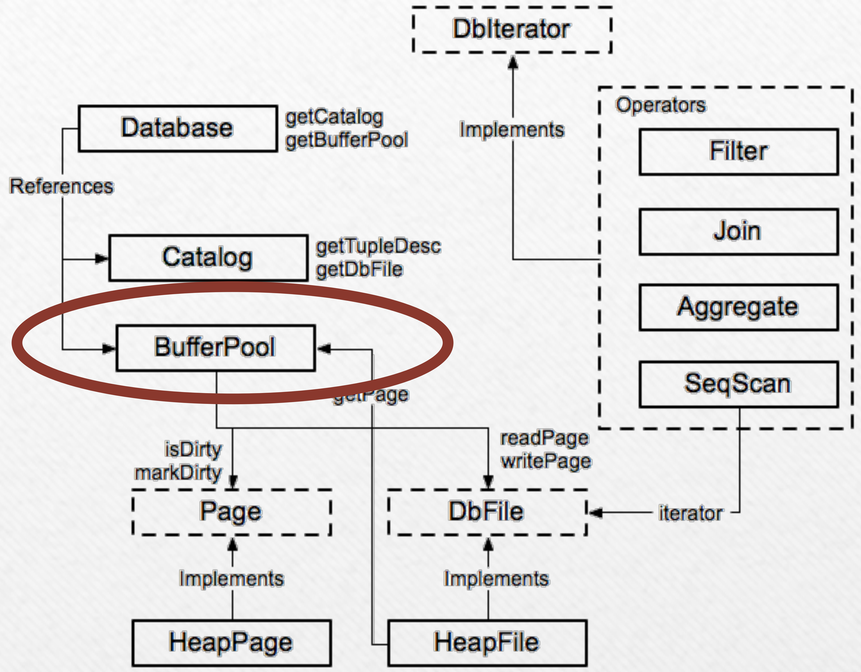
- A single database
  - One schema
  - List of tables
- References to major components
  - Global instance of Catalog
  - Global instance of BufferPool

Catalog

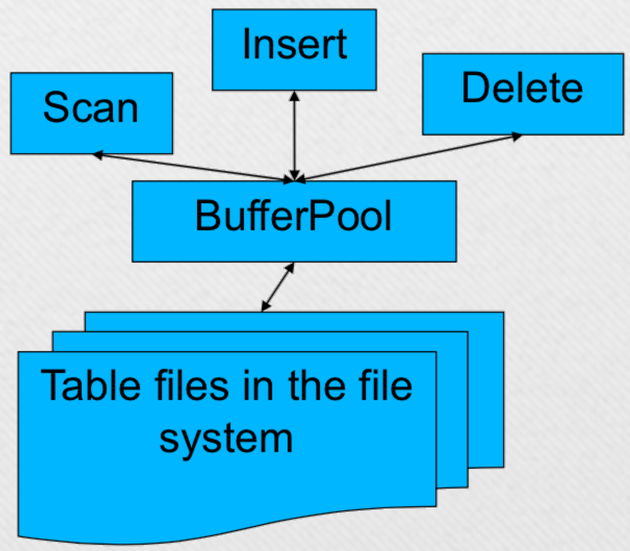


- Stores meta information about tables in the database
  - void addTable(DbFile d, TupleDesc d)
  - DbFile getTable(int tableid)
  - TupleDesc getTupleDesc(int tableid)
- NOT persisted to disk
  - Catalog info is reloaded every time SimpleDB starts up

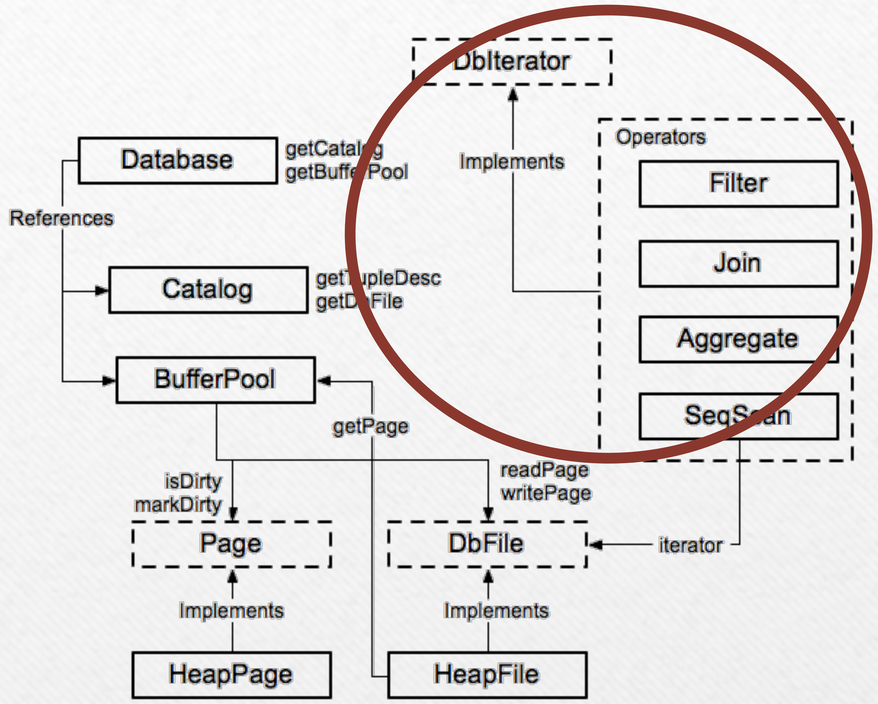
BufferPool



- The ONLY bridge between data-processing operators and actual data files
- Data files are never accessed directly
- Later labs:
  - Locking for transactions
  - Flushing pages for recovery

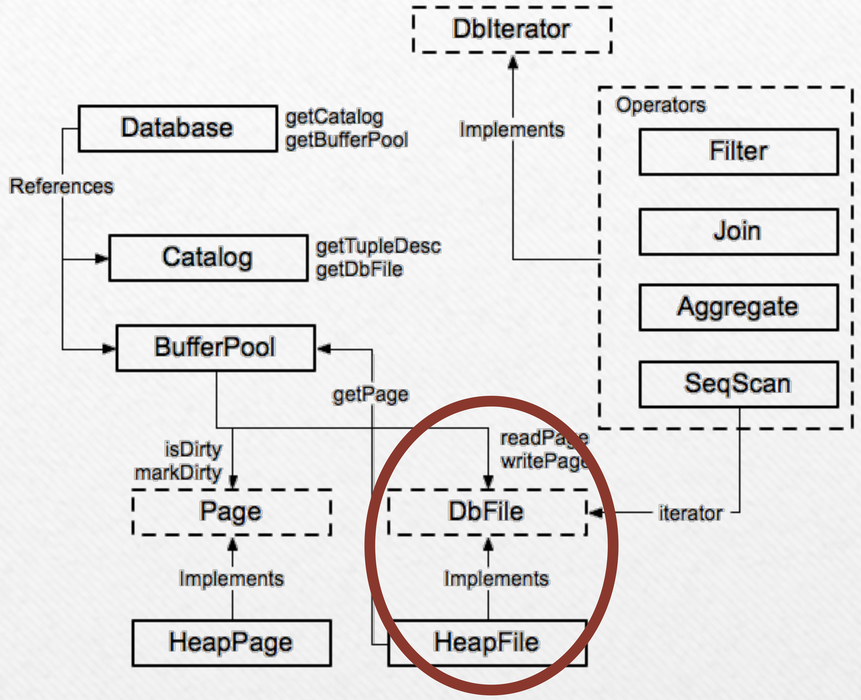


DbIterator



- Ancestor class for all operators
  - Join, Project, SeqScan, etc…
- Each operator has methods:
  - open(), close(), getTupleDesc(), hasNext(), next(), rewind()
- Iterator model: chain iterators together

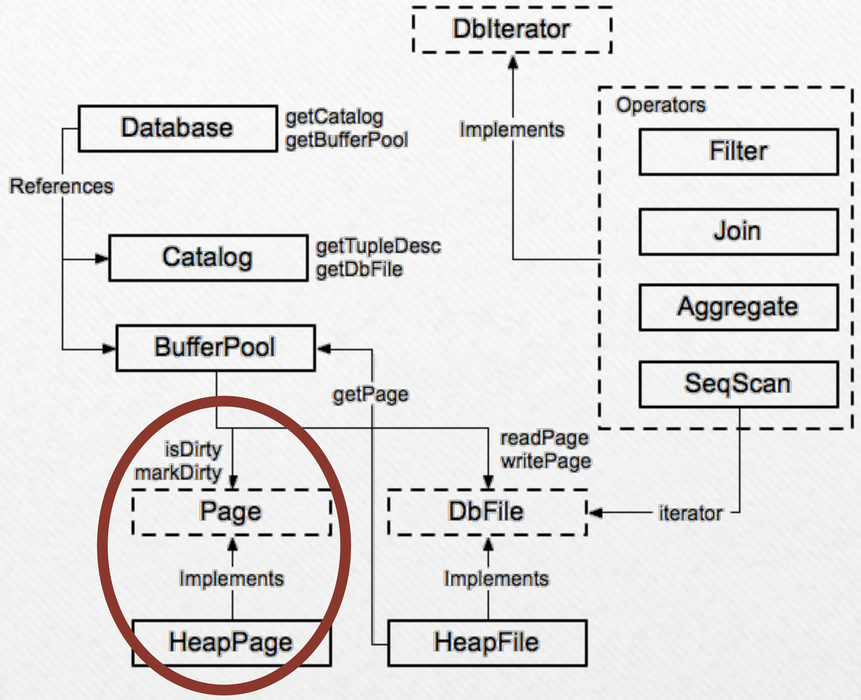
HeapFile



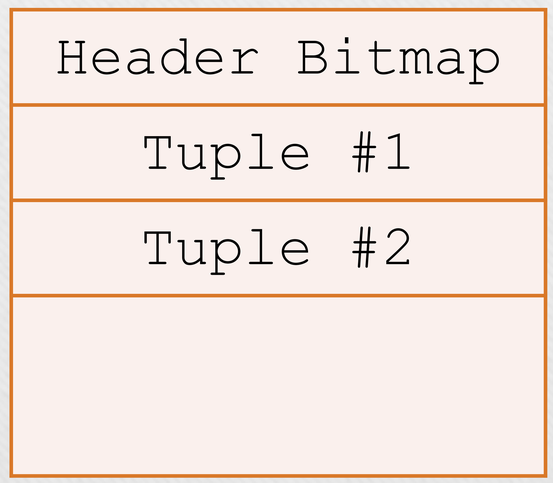
- Main class that organizes the physical storage of tables
- Collection of HeapPages on disk
  - One HeapFile for each table
  - Fixed-size pages means efficient lookup of pages

HeapPage #1
HeapPage #2
HeapPage #3

HeapPage



- A chunk of data that can reside in the BufferPool
- Format: Header + Tuples
  - # of 1 bits in Bitmap = # of active tuples on page
- Fixed size: BufferPool.PAGE_SIZE



Questions?