CS 591: Data Systems Architectures

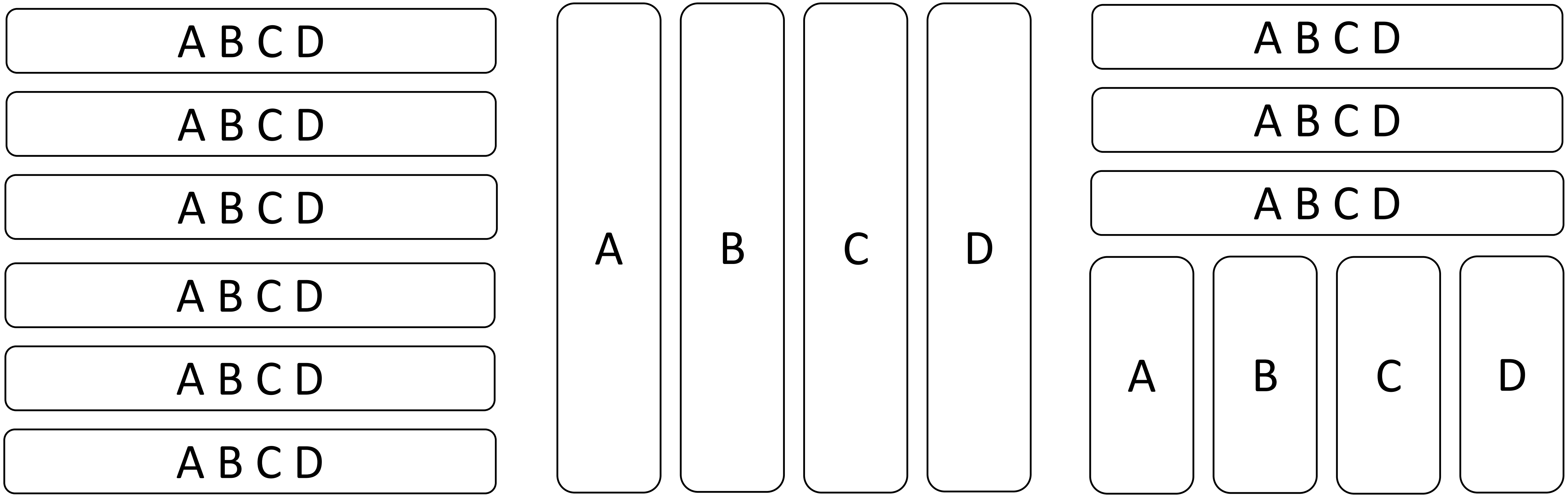
Prof. Manos Athanassoulis

mathan@bu.edu

http://manos.athanassoulis.net/classes/CS591

CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

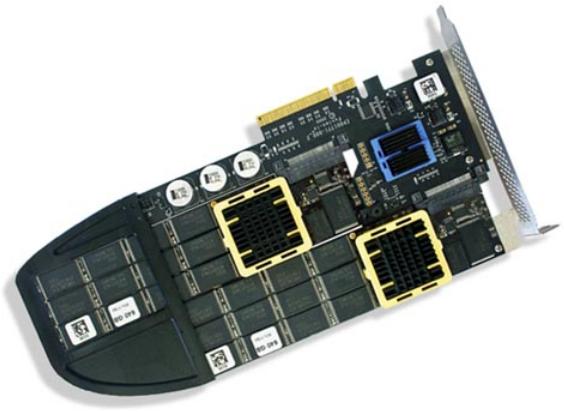


BOSTON UNIVERSITY

CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

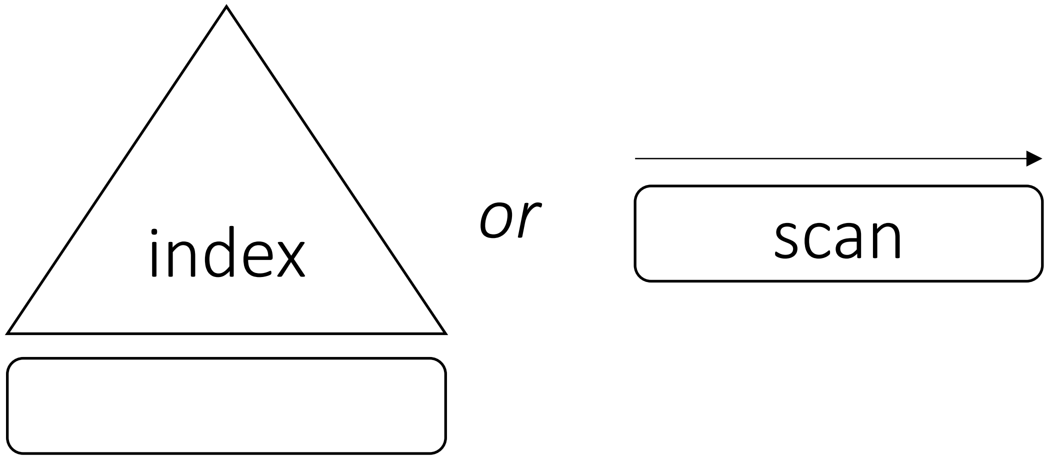


CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

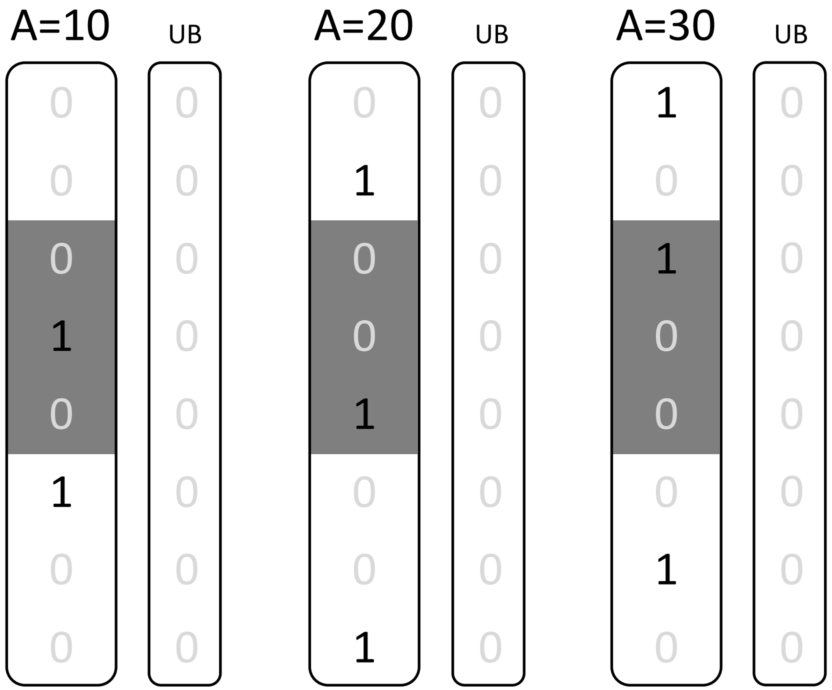


CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit



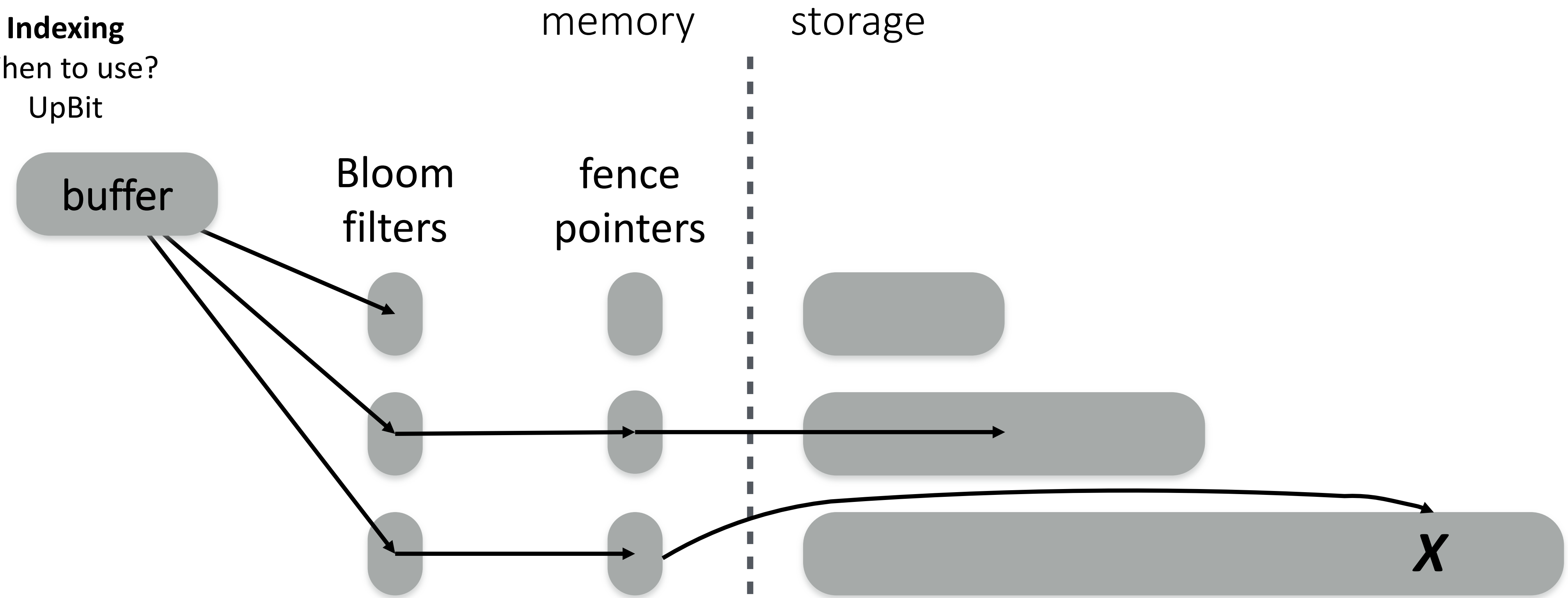
CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based



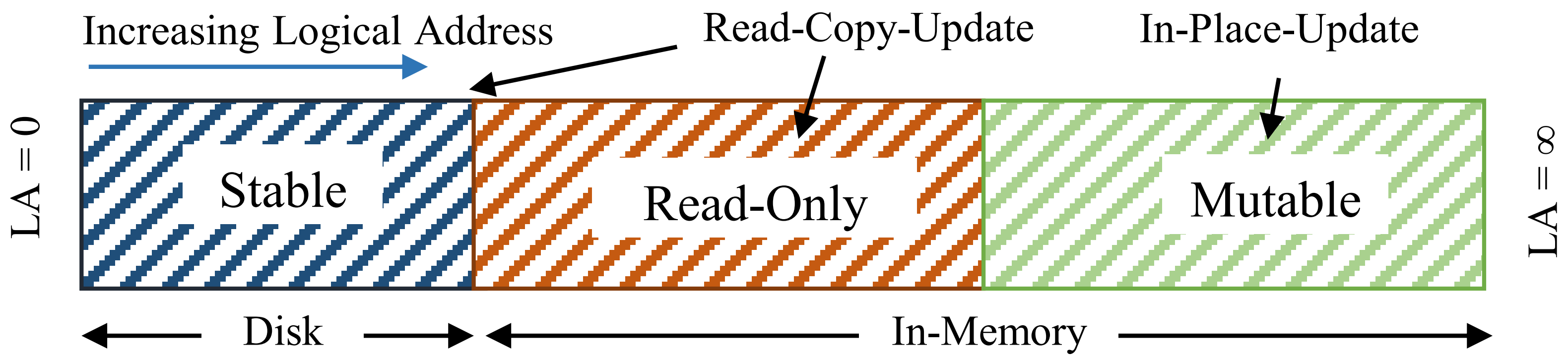
CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based



CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based

Indexing
Data Skipping
Adaptive Indexing

	year	grade	course
t_1	2012	A	DB
t_2	2011	A	AI
t_3	2011	B	OS
t_4	2013	C	DB

	year	grade	course
t_2	2011	A	AI
t_3	2011	B	OS
t_1	2012	A	DB
t_4	2013	C	DB

	grade		year	course
t_1	A	t_2	2011	AI
t_2	A	t_3	2011	OS
t_3	B	t_1	2012	DB
t_4	C	t_4	2013	DB

CS591 progress bar

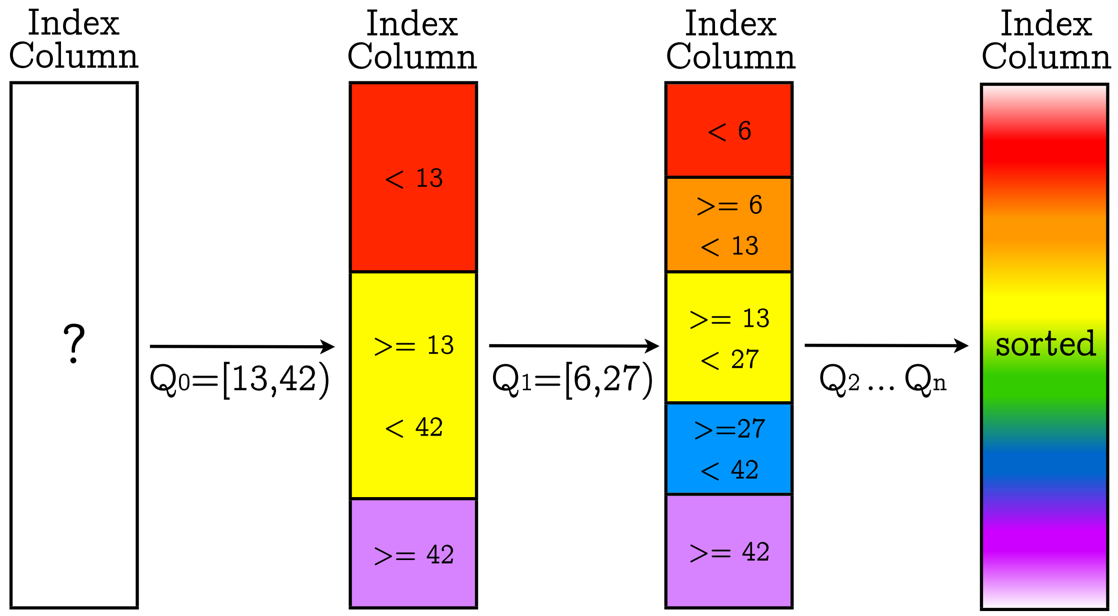
Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based

Indexing
Data Skipping
Adaptive Indexing



BOSTON UNIVERSITY

CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

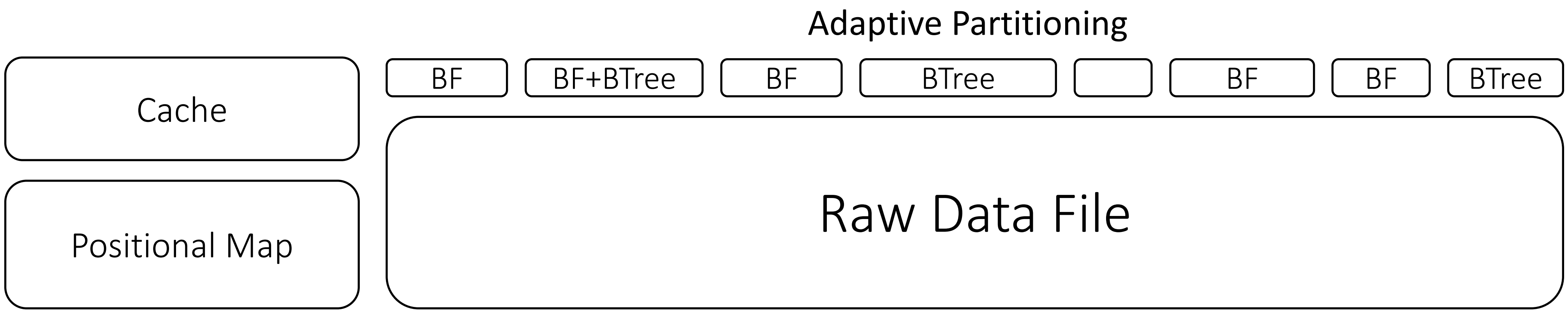
New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based

Indexing
Data Skipping
Adaptive Indexing

Scientific Data Management
In-situ Query Processing



BOSTON UNIVERSITY

CS591 progress bar

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based

Indexing
Data Skipping
Adaptive Indexing

Scientific Data Management
In-situ Query Processing
Today: Array Data

Today: Array Data Storage Manager

Up to now: ***uni-dimensional*** data (integers, real, string)

Array Data: ***multi-dimensional*** data

why is this a challenge?

No unique order (cannot sort!)

How to store?

Concepts: multi-dimensional arrays, storage manager, tiles, thread-safe, dense vs. sparse arrays, global cell order, fragments, dense vs. sparse fragments, consolidation

BOSTON UNIVERSITY

CS591 progress bar

New Paradigms

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based

Indexing
Data Skipping
Adaptive Indexing

Scientific Data Management
In-situ Query Processing
Today: Array Data

CS591 progress bar

New Paradigms

Storage Layouts
Rows vs Cols vs Hybrid

New Hardware
Flash Storage
Multi-core

Indexing
When to use?
UpBit

NoSQL Engines
LSM-Trees
Hash-based

Indexing
Data Skipping
Adaptive Indexing

Scientific Data Management
In-situ Query Processing
Today: Array Data

Distributed DB
Database Systems
at Global Scale

MapReduce
Computing at Scale

Systems for ML
ML building blocks

ML for Systems
Automatic Data
System Design

Learned Indexes
Learn Data Distributions
for Indexing

Data Calculator
Synthesize Indexes

BOSTON UNIVERSITY

Do not forget: *reviews*

You can skip up to 3 reviews

18 classes: 5 long + 10 short + 3 skipped

new rule: you can do extra long reviews, 1 long counts as 3 short

Normally for full marks: 5 long + 10 short

or 6 long + 7 short

or 7 long + 4 short

or 8 long + 1 short

Do not forget: *project*

Do not leave your project work for last minute!

Until *Tuesday April 16th* every group in OH to discuss progress

April 30 and May 2 project presentations:

problem + approach + results + open questions

Project presentations will also be peer-evaluated

BOSTON UNIVERSITY