Lab#4 Extension -1

Lab#4 is extended for another week for

i) Midterm exams
ii) Synchronizing the labs and lectures

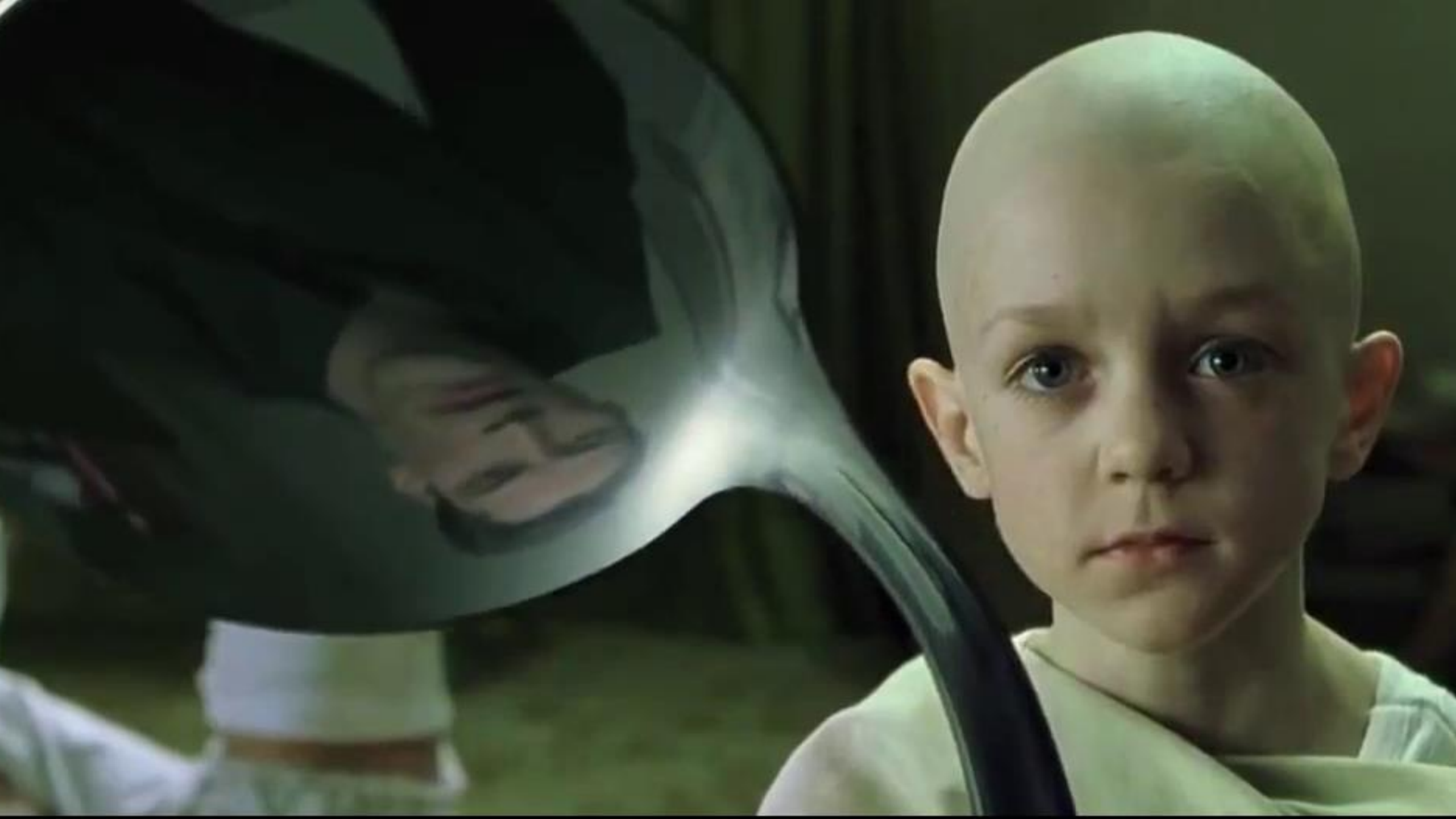
Lab#4 becomes a 3-week lab and its due date is March 13, 2019.

All other remaining labs and due dates are pushed back for one week accordingly.

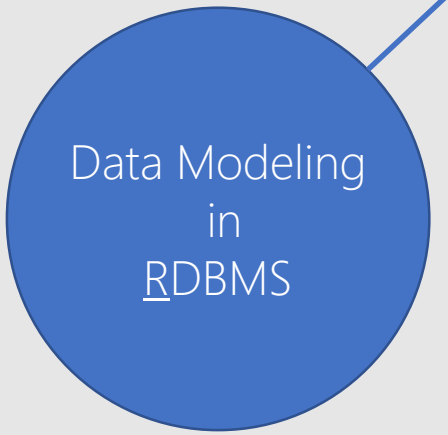
Make-up Class 0

For university closure on Feb. 12, 2019
Within {Feb. 26, 27, Mar. 6, 7, 8}, the only availability is:

Date	Thursday, March 7, 2019
Time	9:00-12:00
Location	VIC608



Today



Real World Entity

Conceptual Level | Entity-Relationship Model (E/R)

| Logical Level | Relational Model

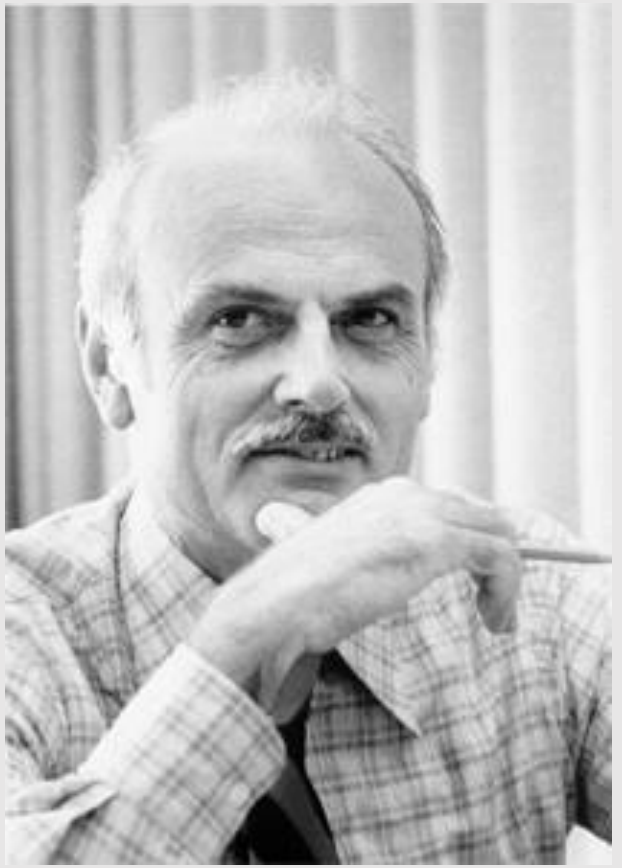
| Physical Level | SQL

Computable Entity

Welcome | Relational Algebra | Unary Op | Binary Op

Relational

Edgar Frank "Ted" Codd, IBM, 1969, 1970



Information Retrieval

A Relational Model of Data for Large Shared Data Banks

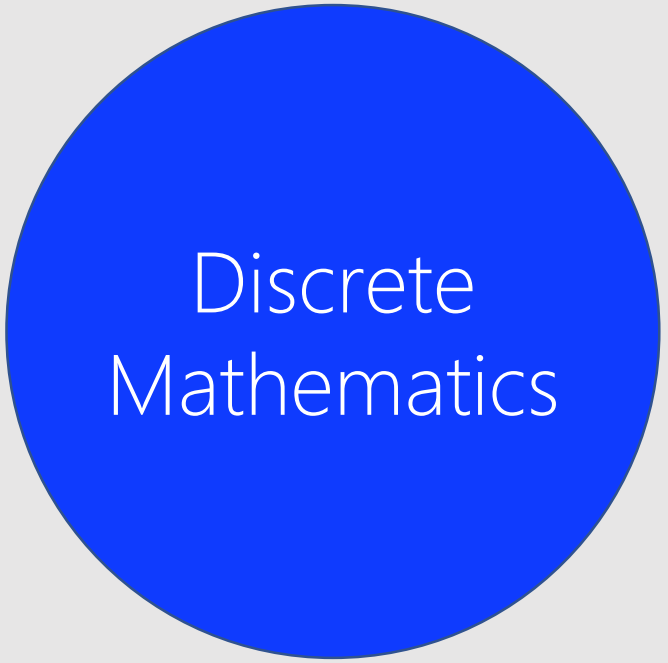
E. F. Codd
IBM Research Laboratory, San Jose, California

Relational Model × Algebra

Given a set, defining operations on elements of the set!

Given Z={integers}={..., -2, -1, 0, 1, 2, ...}
Operators & Operands:
- Unary: -(2), 2!
- Binary: 2+3, 2*3, 2^3

Closure:
- Result is also an element of the set



Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations

Query in Natural Language → Query in Math Formula

Relational Model × Algebra

Who made 'Pulp Fiction'?

$$\pi_{\substack{\text{Director.FirstName}\\ \text{Director.LastName}}} \left(\sigma_{\substack{\text{Movie.Title='PulpFiction'} \text{ AND}\\ \text{Movie.Id=MovieDirctor.MovieId} \text{ AND}\\ \text{Director.Id=MovieDirctor.DirectorId}}} (\text{Movie} \times \text{MovieDirector} \times \text{Director})\right)$$

Relational Model × Algebra

Who acted in 'Pulp Fiction'?

$$\pi_{\substack{\text{Actor.FirstName}\\ \text{Actor.LastName}}} \left(\sigma_{\substack{\text{Movie.Title='PulpFiction' AND}\\ \text{Movie.Id=StarIn.MovieId AND}\\ \text{Actor.Id=StarIn.ActorId}}} (\text{Movie} \times \text{StarIn} \times \text{Director})\right)$$

Relational Model × Algebra 8

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations

Unary Operation	Binary Operation (Set Theory)
π(R) : Project	R1 $\cup$ R2: Union
σ(R) : Select	R1 $\cap$ R2: Intersection
ρ(R) : Rename	R1 \ R2 : Set Difference
	R1 × R2: Cartesian Product

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

ρ(R) : Rename

Binary Operation (Set Theory)

R1 $\cup$ R2: Union

R1 $\cap$ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

Algebra × Project (π)

π, pi, is used to select a subset of attributes (columns) from a relation

$$A = \pi_{<\text{attribute list}>}(R)$$

Vertical Filtering

R is a relation

<attribute list> subset of attributes of R

A is a relation including all tuples in R with only attributes in list

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

What are directors' name?

$$A = \pi_{\text{FirstName, LastName}}(\text{Director})$$

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

What are directors' name?

$$A = \pi_{\text{FirstName, LastName}}(\text{Director})$$

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

How many movies each director made?

$$A = \pi_{\text{FirstName, LastName, MovieCount}}(\text{Director})$$

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

How many movies each director made?

$$A = \pi_{\text{FirstName, LastName, MovieCount}}(\text{Director})$$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

ρ(R) : Rename

Binary Operation (Set Theory)

R1 $\cup$ R2: Union

R1 $\cap$ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

Algebra × Selection (σ)

σ, sigma, is used to select a subset of tuples from a relation based on a condition (θ) over relation's attributes.

$$A = \sigma_{\theta}(R)$$

R is a relation

θ is a Boolean expression on the attributes of R

A is a relation including tuples that make θ true

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which director was born in US?

$$A = \sigma_{\text{PlaceOfBirth}=\text{'USA'}}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which director was born in US?

$$A = \sigma_{\text{PlaceOfBirth}='\text{USA}'}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which director make more than 20 movies?

$$A = \sigma_{\text{MovieCount} > 20}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which director make more than 20 movies?

$$A = \sigma_{\text{MovieCount} > 20}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which director has same first and last names?

$$A = \sigma_{\text{FirstName} = \text{LastName}} (\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which director has same first and last names?

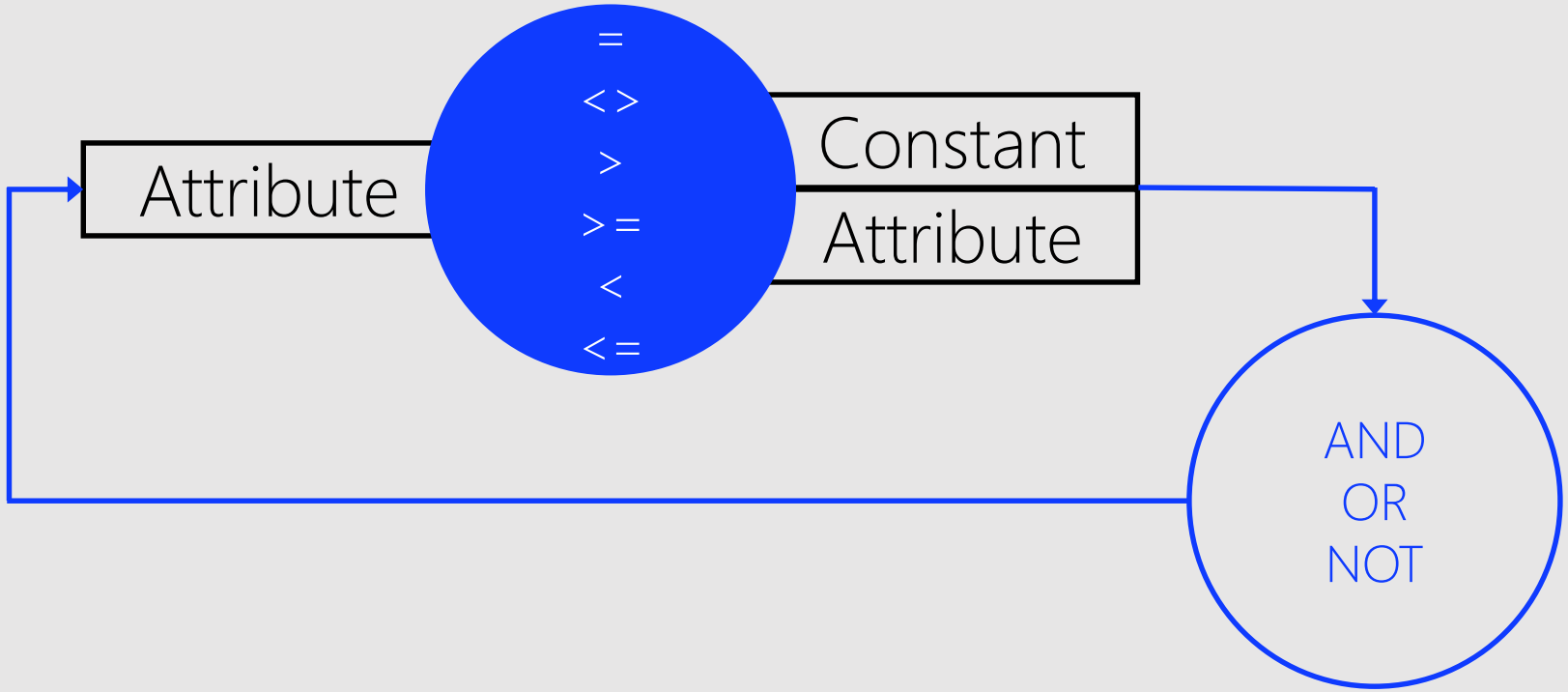
$$A = \sigma_{\text{FirstName} = \text{LastName}} (\text{Director})$$

$$= \emptyset$$

Algebra × Selection (σ)

θ can be made up of number of Boolean clauses

$A = \sigma_{\theta} (R)$



Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies?

$$A = \sigma_{\text{PlaceOfBirth='USA' AND MovieCount > 20}}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies or is not American?

$A1 = \sigma_{\text{PlaceOfBirth='USA' AND MovieCount > 20 OR PlaceOfBirth<>'USA'}}$ (Director)

$A2 = \sigma_{\text{PlaceOfBirth='USA' AND (MovieCount > 20 OR PlaceOfBirth<>'USA')}}$ (Director)

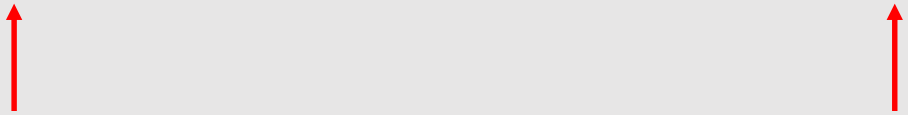
$A3 = \sigma_{\text{(PlaceOfBirth='USA' AND MovieCount > 20) OR (PlaceOfBirth<>'USA')}}$ (Director)

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies or is not American?

$A2 = \sigma_{\text{PlaceOfBirth='USA' AND (MovieCount > 20 OR PlaceOfBirth<>'USA')}}$ (Director)



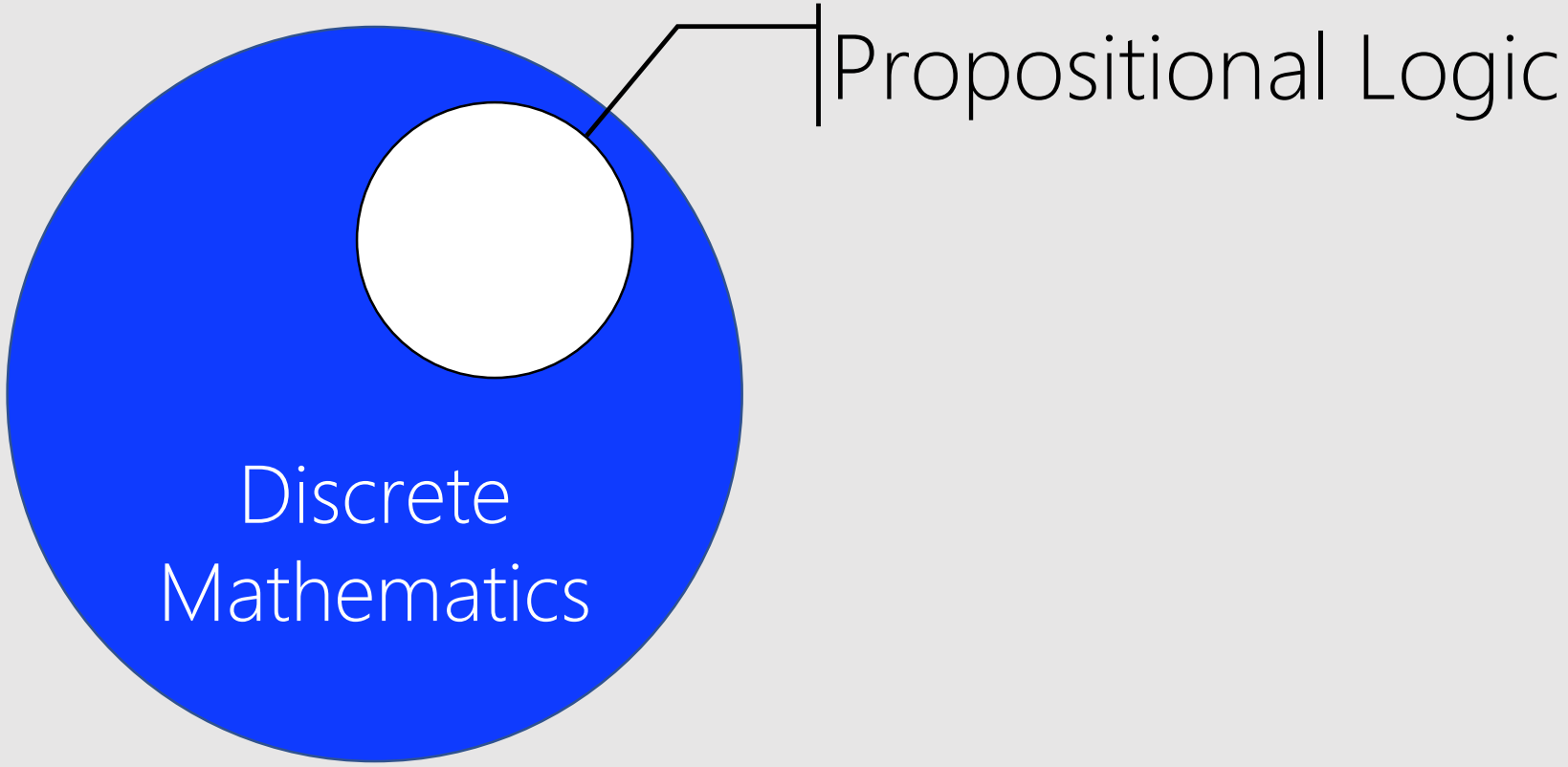
Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

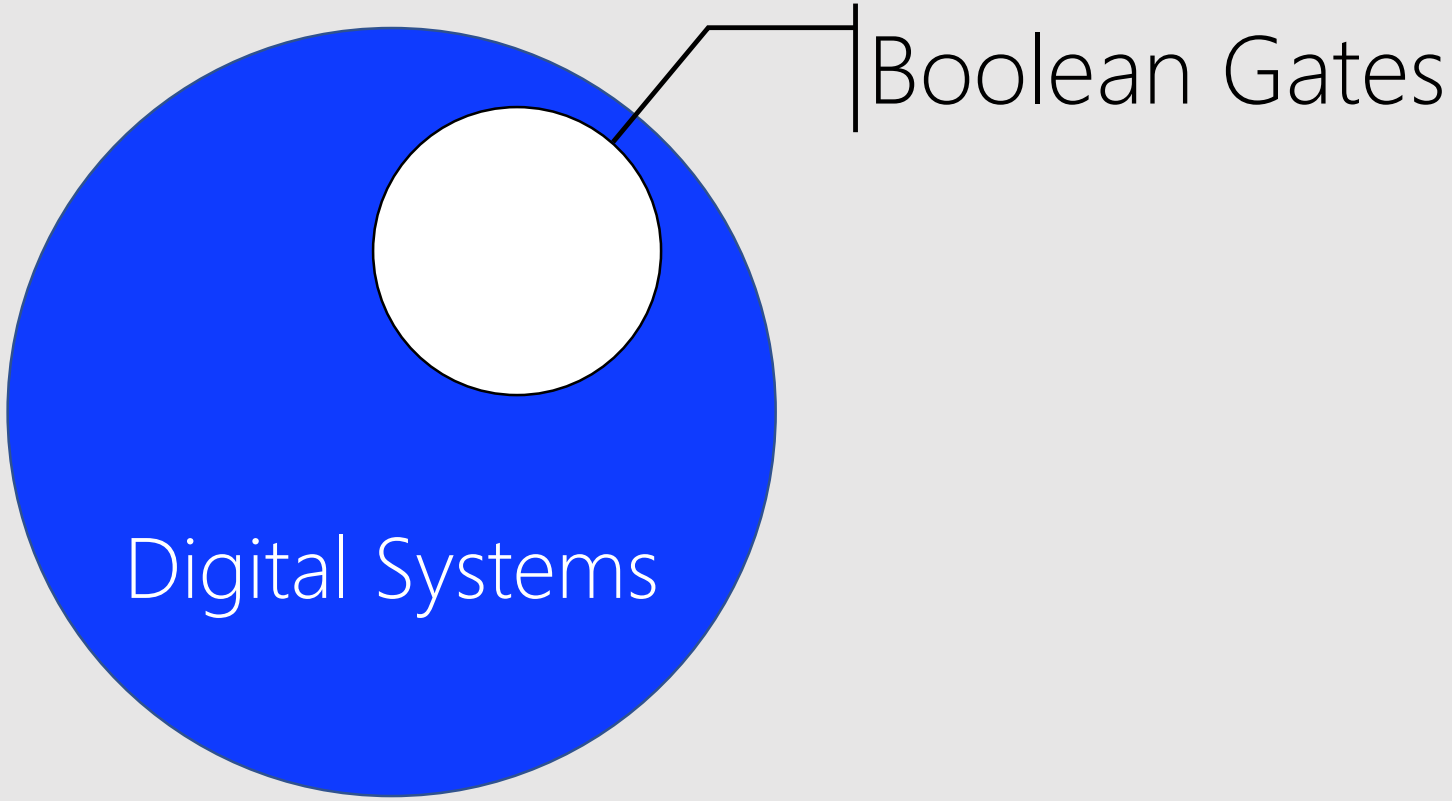
Which American director made more than 20 movies or is not American?

$$A3 = \sigma_{(PlaceOfBirth='USA' \text{ AND } MovieCount > 20) \text{ OR } (PlaceOfBirth<>'USA')}(Director)$$

Algebra × Selection (σ)



Algebra × Selection (σ)



Algebra × Selection (σ)

Operation Precedence

() > § > † > ‡

Commutative Law

A § B = B § A *(we say A commutes with B under §)*

Associative Law

A § (B § C) = (A § B) § C = A § B § C

Distributive Law

A § (B † C) = (A § B) † (A § C)

Algebra × Selection (σ)

Operation Precedence for Logical Operations

() > NOT > AND = OR

Commutative Law

A AND B = B AND A *(we say A commutes with B under AND)*

A OR B = B OR A *(we say A commutes with B under OR)*

Associative Law

A AND (B AND C) = (A AND B) AND C = A AND B AND C

A OR (B OR C) = (A OR B) OR C = A OR B OR C

Distributive Law

A AND (B OR C) = (A AND B) OR (A AND C)

A OR (B AND C) = (A OR B) AND (A OR C)

de Morgan´s Theorem**

NOT (A AND B) = NOT(A) OR NOT(B)

NOT (A OR B) = NOT(A) AND NOT(B)

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies?

$$A = \sigma_{\text{PlaceOfBirth='USA' AND MovieCount > 20}}(\text{Director})$$

$$= \sigma_{\text{MovieCount > 20 AND PlaceOfBirth='USA'}}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies?

$$A1 = \sigma_{PlaceOfBirth='USA'}(Director)$$

$$A = \sigma_{MovieCount > 20}(A1)$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies?

$$A = \sigma_{\text{MovieCount} > 20} \left(\sigma_{\text{PlaceOfBirth}=\text{'USA'}} (\text{Director})\right)$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Which American director made more than 20 movies?

$$A = \sigma_{\text{PlaceOfBirth='USA'}}(\sigma_{\text{MovieCount} > 20}(\text{Director}))$$

$$= \sigma_{\text{PlaceOfBirth='USA' AND MovieCount} > 20}(\text{Director})$$

$$= \sigma_{\text{MovieCount} > 20 \text{ AND PlaceOfBirth='USA'}}(\text{Director})$$

$$= \sigma_{\text{MovieCount} > 20}\ (\sigma_{\text{PlaceOfBirth='USA'}}(\text{Director}))$$

Algebra × Selection (σ)

$$\sigma_{\theta}(\sigma_{\theta'}(R))$$
$$=$$
$$\sigma_{\theta \text{ AND } \theta'}(R)$$
$$=$$
$$\sigma_{\theta' \text{AND } \theta}(R)$$
$$=$$
$$\sigma_{\theta'}(\sigma_{\theta}(R))$$

Corollary: $\sigma_{\theta}(\sigma_{\theta'}(\sigma_{\theta''\ldots}(\sigma_{\theta'''}(R)) = \sigma_{\theta \text{ AND } \theta' \text{AND } \theta''\ldots \text{ AND } \theta'''}(R)$

Algebra × Selection (σ) × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' name who make more than 20 movies?

Algebra × Selection (σ) × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	*Stanley*	*Kubrick*	*Jul. 26, 1928*	*USA*	*1*	*13*
2	*Alfred*	*Hitchcock*	*Aug. 13, 1899*	*England*	*203*	*47*
3	*Clint*	*Eastwood*	*May 31, 1930*	*USA*	*803*	*35*

Find directors' name who make more than 20 movies?

$$A1 = \sigma_{MovieCount > 20}(Director)$$

Algebra × Selection (σ) × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' name who make more than 20 movies?

$$A1 = \sigma_{MovieCount > 20}(Director)$$

$$A = \pi_{FirstName,\ LastName}(A1)$$

$$= \pi_{FirstName,\ LastName}(\sigma_{MovieCount > 20}(Director))$$

Algebra × Selection (σ) × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' name who make more than 20 movies?

✗ $A = \sigma_{MovieCount > 20} (\pi_{FirstName, LastName} (Director))$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

ρ(R) : Rename

Binary Operation (Set Theory)

R1 $\cup$ R2: Union

R1 $\cap$ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

Algebra × Rename (ρ)

ρ, rho, is used to rename a relation or its attributes or both

$$A = \rho_{<R'(a'/a,\ b'/b,\ \ldots)>}(R)$$

R is a relation
R' is the new name for R(a, b, ...)
a' is the new name for attribute a of R
b' is the new name for attribute b of R
...

Algebra × Rename (ρ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

$$A = \pi_{\text{FirstName, Name}} (\rho_{\text{ActiveDirector(Name/LastName)}} (\sigma_{\text{MovieCount} > 20}(\text{Director})))$$

We'll c its real use ☺