UNIVERSITY *of* WASHINGTON

TCSS 445 A: Au–17

AGENDA

- COURSE INTRODUCTION / SYLLABUS
- OVERVIEW OF DBMS
- RELATIONAL ALGEBRA

WRITE THE FOLLOWING ATTRIBUTES

1. PREFERRED NAME
2. YOUR MAJOR
3. BRIEFLY DESCRIBE (< 20 WORDS) YOUR DATABASE EXPERIENCE
4. BRIEFLY DESCRIBE WHAT YOU MOST WANT TO GAIN/LEARN FROM THIS CLASS
5. A FUN FACT ABOUT YOURSELF
   – (THAT YOU DON'T MIND OTHERS KNOWING)

WHAT TO EXPECT THIS QUARTER

You'll find what you need on canvas...

Autumn 2017

- Home
- Announcements
- Assignments
- Discussions
- Grades
- People
- Pages
- Files
- Syllabus
- Outcomes
- Quizzes
- Modules
- Conferences

TCSS 445 A Au 17: Database Systems Design

Edit

Welcome to TCSS 445 A for Autumn Quarter, 2017. Below is a the tentative schedule for this course. Links to materials, slides and notes for each week will be added to this page each week. This page will also serve to communicate what is expected of students prior to each class, as well as due dates for homework assignments. Check back frequently!

Please notify me if you have any questions or observe any errors or inaccuracies. For detailed information regarding the design and structure of this course, please refer to the Syllabus.

Course Schedule

WEEK # Date of Lecture	**TOPICS** Discussed During Class	**READINGS** Due Before Class	**GRADED ACTIVITIES** Due On Date Shown	**LINKS** Slides and Content
WEEK 1 Sep 30	Course Overview Overview of DBMS Intro to Relational Algebra	None	None	Syllabus Week 1 Slides.pdf

INTRODUCTION TO DATA

- What is/are data?
  - Why the confusion on the plural?
    - "Data" is the Latin plural for datum.
    - Academia = Plural, Industry = Singular
  - How is *data* defined today
    - "Data is a set of discrete, objective facts about events. In an organizational context, data is more usefully described as structured records of transactions."
      - Davenport T., Prusak L. (1998). Working Knowledge. Harvard Business School Press: Boston, MA.

INTRODUCTION TO DATA

> What is/are data?

INTRODUCTION TO DATA

- What is/are data?
  - Why the confusion on the plural?
    - "Data" is the Latin plural for datum.
    - Academia = Plural, Industry = Singular

INTRODUCTION TO DATA

- What is/are data?
  - Why the confusion on the plural?
    - "Data" is the Latin plural for datum.
    - Academia = Plural, Industry = Singular

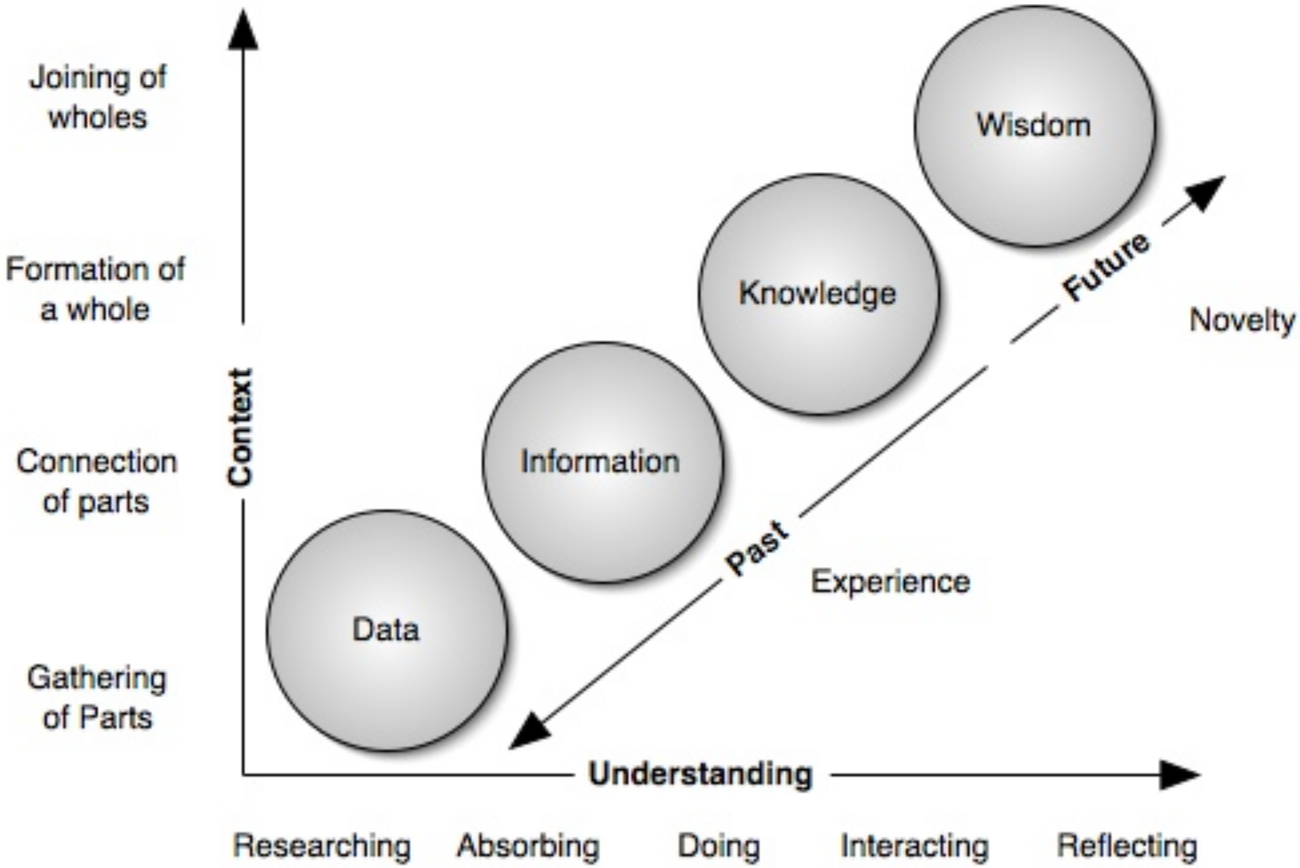
  - How is *data* defined today
    - "Data is a set of discrete, objective facts about events. In an organizational context, data is more usefully described as structured records of transactions."
      - Davenport T., Prusak L. (1998). Working Knowledge. Harvard Business School Press: Boston, MA.

CONVERSATION POINT

> How does data differ from information?

> How do they relate?

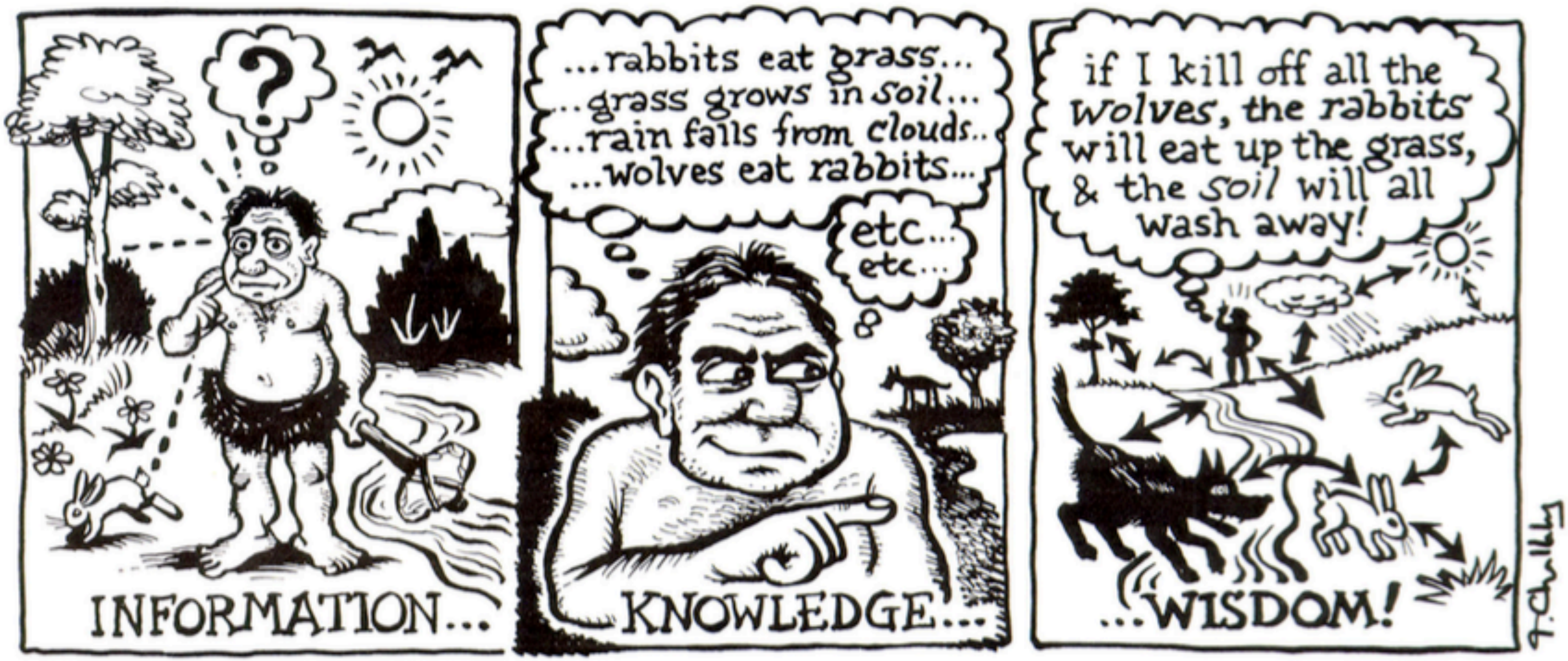
THE CONTINUUM OF KNOWLEDGE



Cleveland H. "Information as a Resource", *The Futurist,* December 1982 p. 34-39

THE CONTINUUM OF KNOWLEDGE

TOM CHALKLEY



The information-knowledge-wisdom hierarchy. The caveman (left) has lots of information (facts and ideas); he selects and organizes useful information into knowledge (center), but he does not achieve wisdom until he has integrated his knowledge into a whole that is more useful than the sum of its parts.

Cleveland H. "Information as a Resource", *The Futurist,* December 1982 p. 34-39

DIKW IN THE CONTEXT BUSINESS INTELLIGENCE

> TRANSACTIONS, DATA CAPTURE (DATA)

> REPORTING: METRICS & KPIs (INFORMATION)
*KPI = Key Performance Indicator

> INSIGHTS / DECISION SUPPORT (KNOWLEDGE)

> MODELS / FORECASTS (WISDOM)

WHAT IS A DATABASE?

> A collection of logically related data

Image Source: Wikipedia.org

WHAT IS IN DATABASE?

- DATA
  - Obviously! But what kind of data?

WHAT IS IN DATABASE?

- DATA
  - Obviously! But what kind of data?
- STRUCTURED
  - Numbers, Text, Dates

WHAT IS IN DATABASE?

> DATA
  - Obviously! But what kind of data?

> STRUCTURED
  - Numbers, Text, Dates

> UNSTRUCTURED (Semi-Structured or Poly-Structured)
  - Images, Videos, Documents

WHAT IS IN DATABASE?

> DATA
  – Obviously! But what kind of data?

> STRUCTURED
  – Numbers, Text, Dates

> UNSTRUCTURED (Semi-Structured or Poly-Structured)
  – Images, Videos, Documents

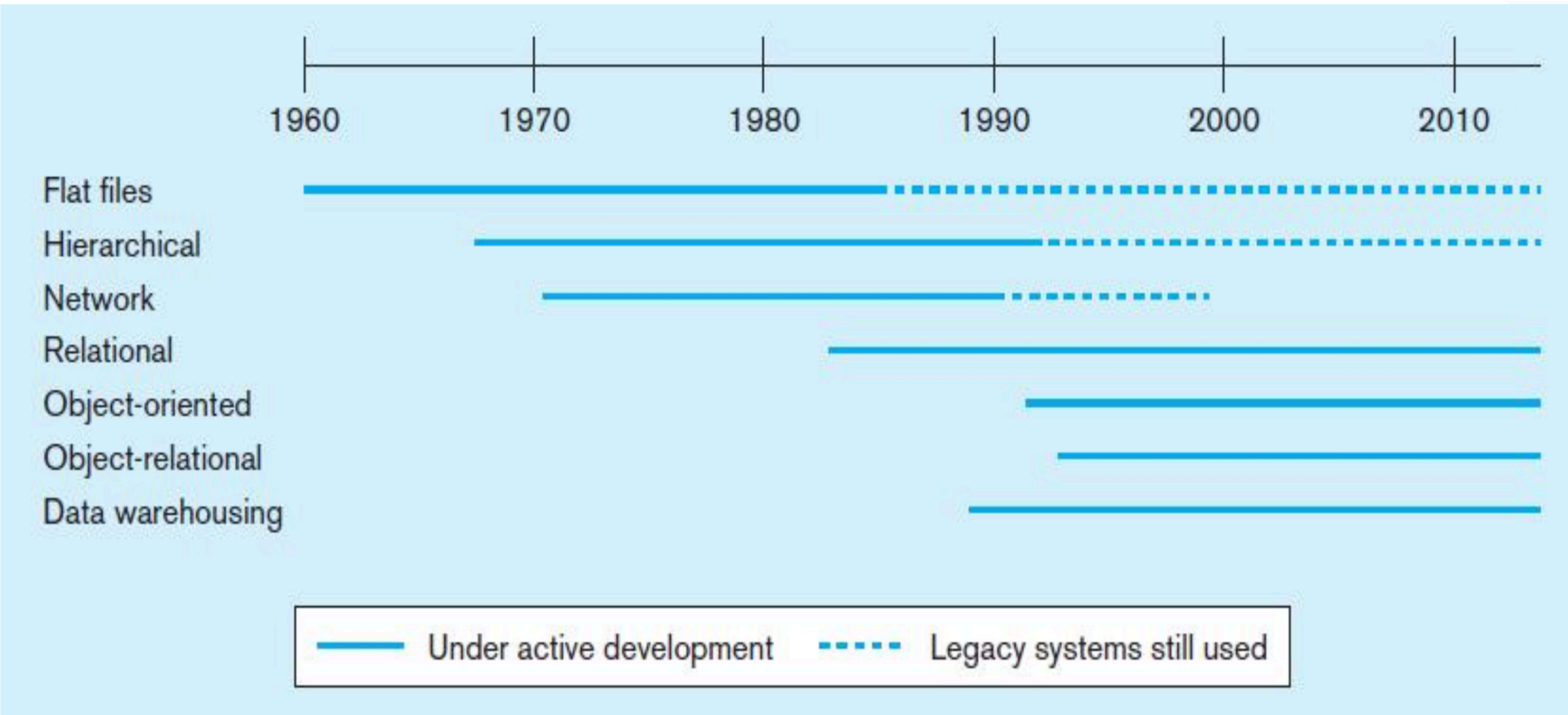
> METADATA
  – DATA ABOUT THE DATA

WHAT IS A DATABASE MANAGEMENT SYSTEM (DBMS)?

> A SYSTEM DESIGNED TO CREATE, MAINTAIN, AND EXPOSE INFORMATION (OR NOT)

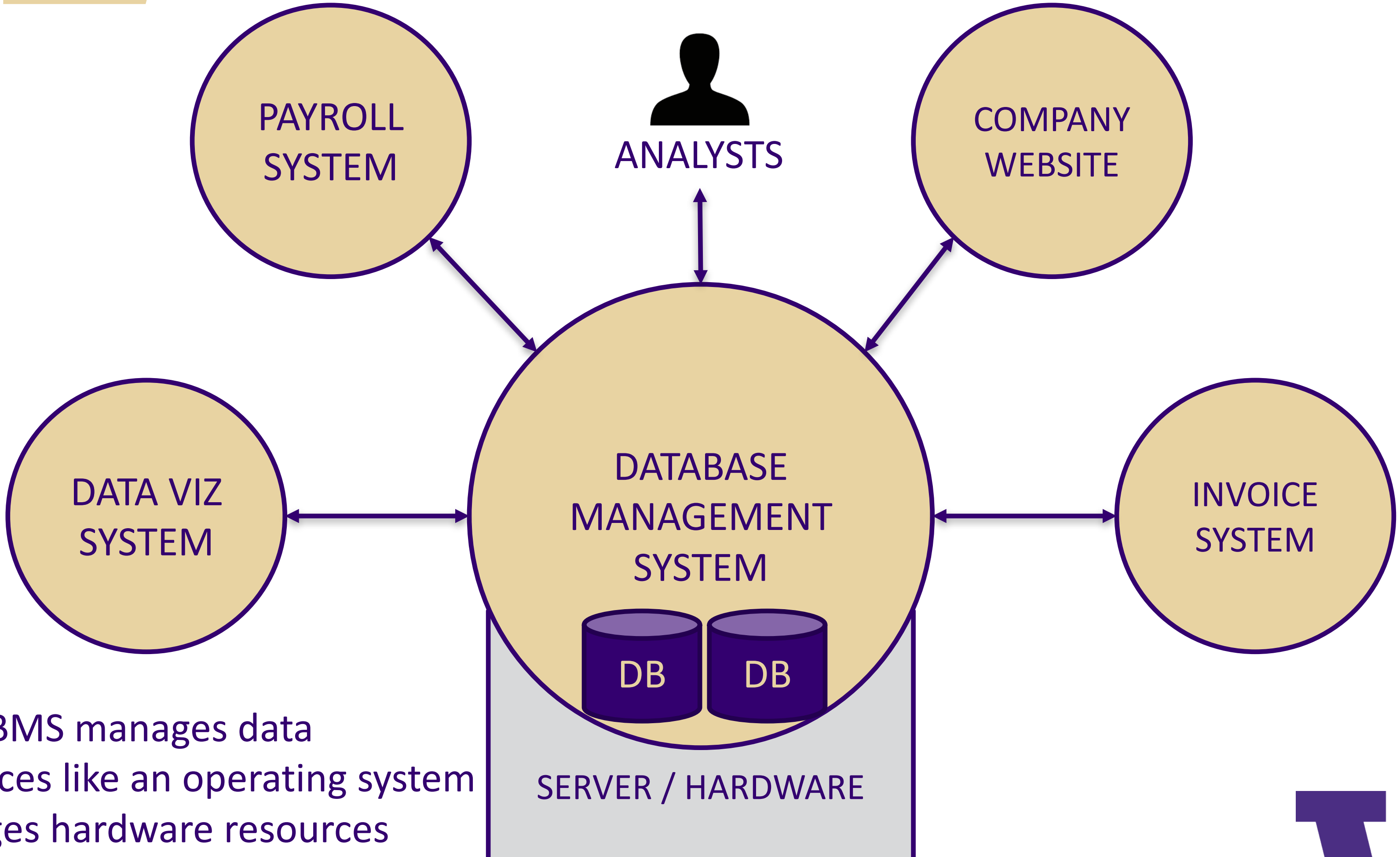
Image Source: Wikipedia.org

THE HISTORY OF DIGITAL DATABASES



Source: TCSS445 - Menaka Abraham

A TYPICAL DBMS CONTEXT DIAGRAM



The DBMS manages data resources like an operating system manages hardware resources

VARIOUS DATA MODELS

- HIERARCHICAL
- NETWORK
- **RELATIONAL**
- OBJECT-ORIENTED
- (MULTI) DIMENSIONAL MODELS
  - STAR SCHEMA
  - SNOWFLAKE
  - HYBRID

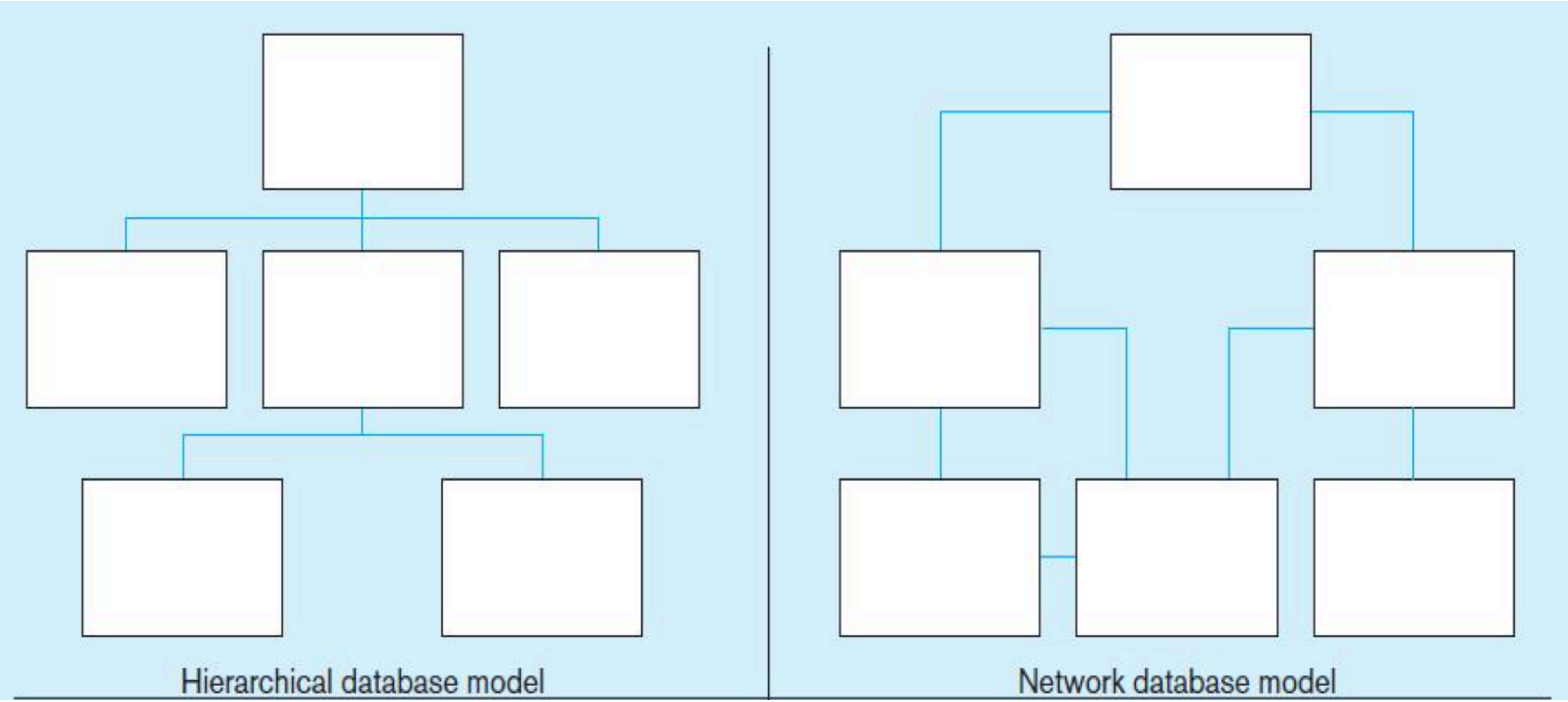
HIERARCHICAL & NETWORK

- Hierarchical
  - *Records* connected through *Links*
  - Links connects exactly two records
  - Tree Like Structure
  - Supports only "*Many to 1*" relationships.

- *Network*
  - *Records* connected through *Links*
  - Links connect exactly two records
  - A record can have multiple "parents"
  - More like a DAG

Source: TCSS445 - Menaka Abraham

HIERARCHICAL & NETWORK



Source: TCSS445 - Menaka Abraham

RELATIONAL

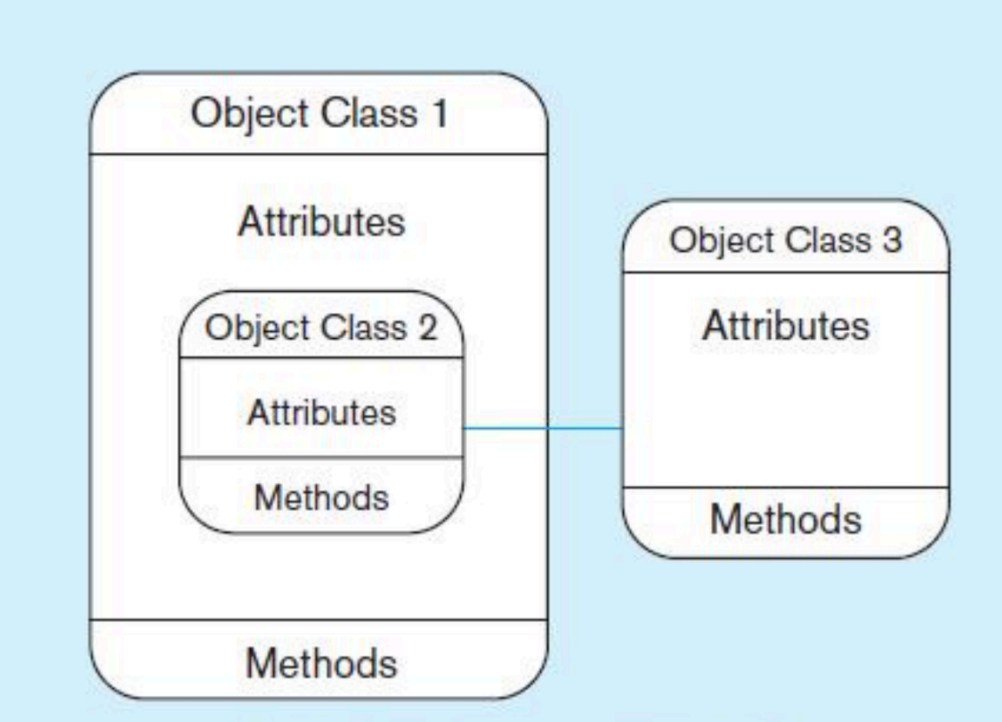
- "Tables" are known as "Relations"
- "Relations" have Keys and "Attributes"
- Records in Relations are known as "Tuples"

```
RELATION X (PRIMARY KEY, ATTRIBUTES...)

RELATION Y (PRIMARY KEY, FOREIGN KEY, ATTRIBUTES...)
```

OBJECT–ORIENTED

- "Tables" implemented as classes with attributes.
- Supports Inheritance
- Attributes can "key" on an object in another class.
- "Is a" relationships and "Has a" relationships



Object-oriented database model

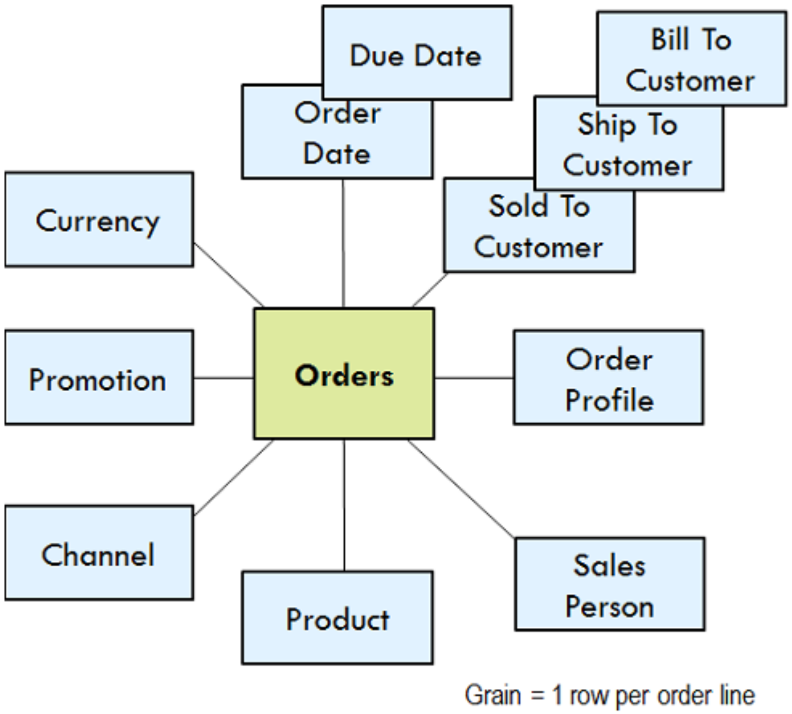
Image Source: TCSS445 - Menaka Abraham

DIMENSIONAL MODELS & CUBES

> Tables split into "FACTS" and "DIMENSIONS"
  – FACTS = Measurements (Verbs)
  – DIMENSIONS = Details / Attributes (Adjectives)

> Two main schema layouts:
  – Snowflake
  – Star
  – (Hybrid)*

DIMENSIONAL MODELS



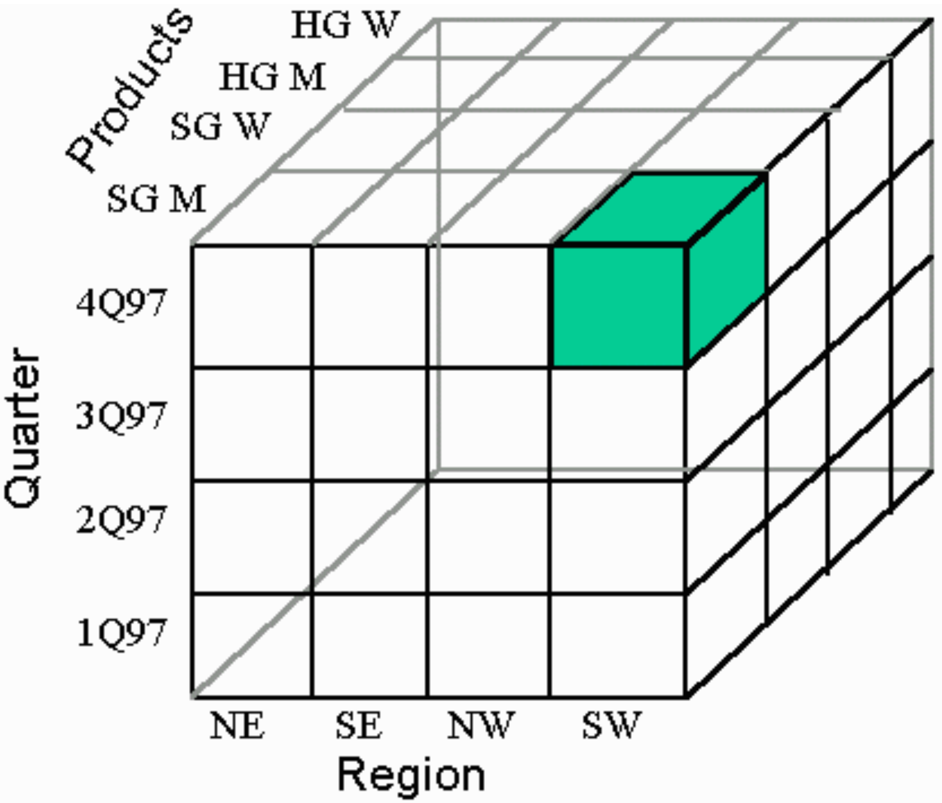


Image Source: KimballGroup.Com (left) holowczak.com (right)

UNIVERSITY of WASHINGTON

THE DIGITAL DBMS

PRIMARY DBMS FUNCTIONS

> QUERY PROCESSING

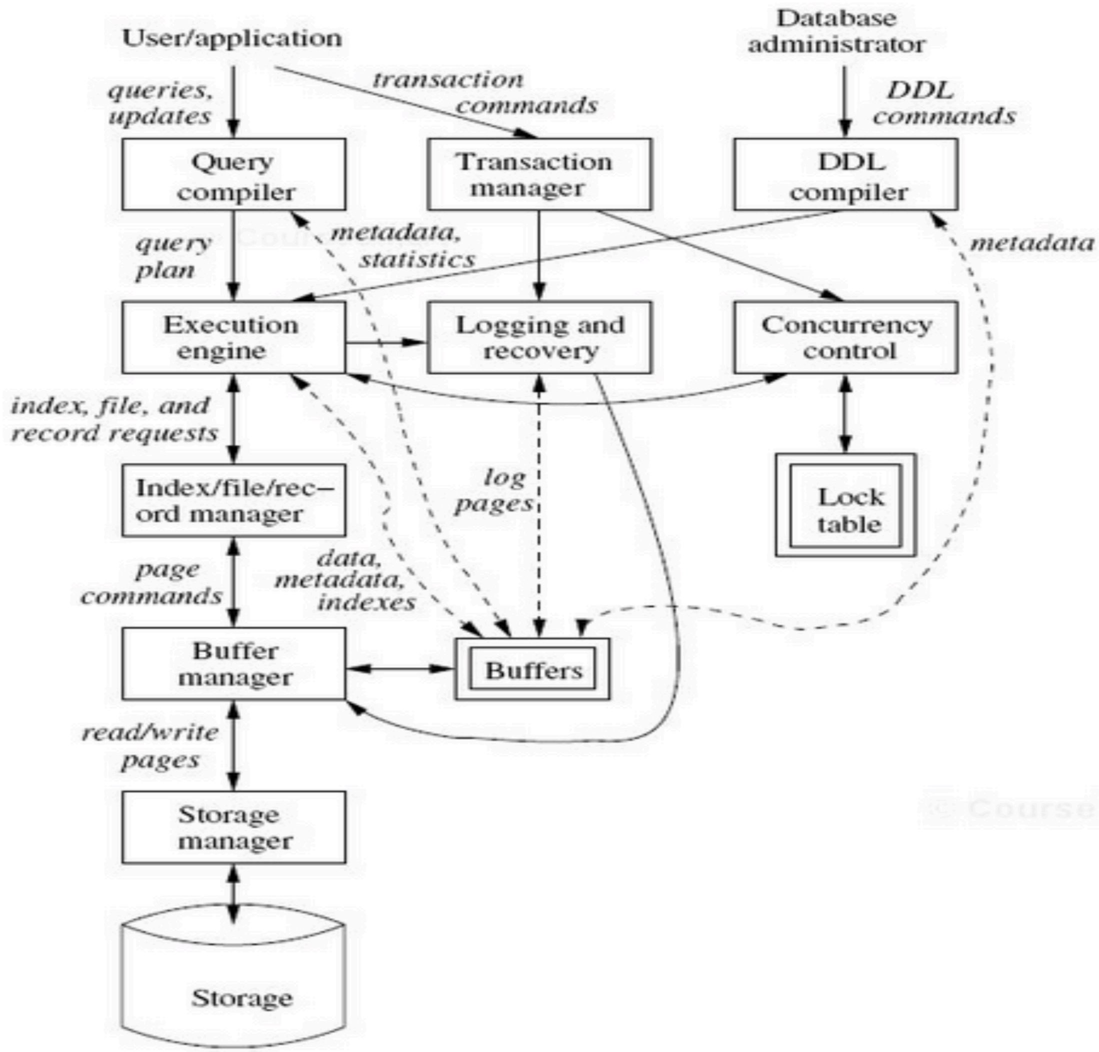
> DATA MANAGEMENT
> (Storage, Memory, Buffer Mgmt)

> TRANSACTION PROCESSING

REQUIREMENTS: THE ACID TEST

- **Atomicity** – Group actions into atomic units commonly known as ‘transactions’
- **Consistent** – Keys & Constraints are met, deterministic, proper state
- **Isolation** – ”Don’t tread on me…” Concurrent actions should not unpredictably impact one another
- **Durability** – Never fail, but be safe when you do!

DBMS CONTROL & DATA FLOW



Solid Lines	Dashed Lines
Data + Control	Data Only

Image Source: Course text book

SQL SERVER ENGINE
HIGH LEVEL ARCHITECTURE

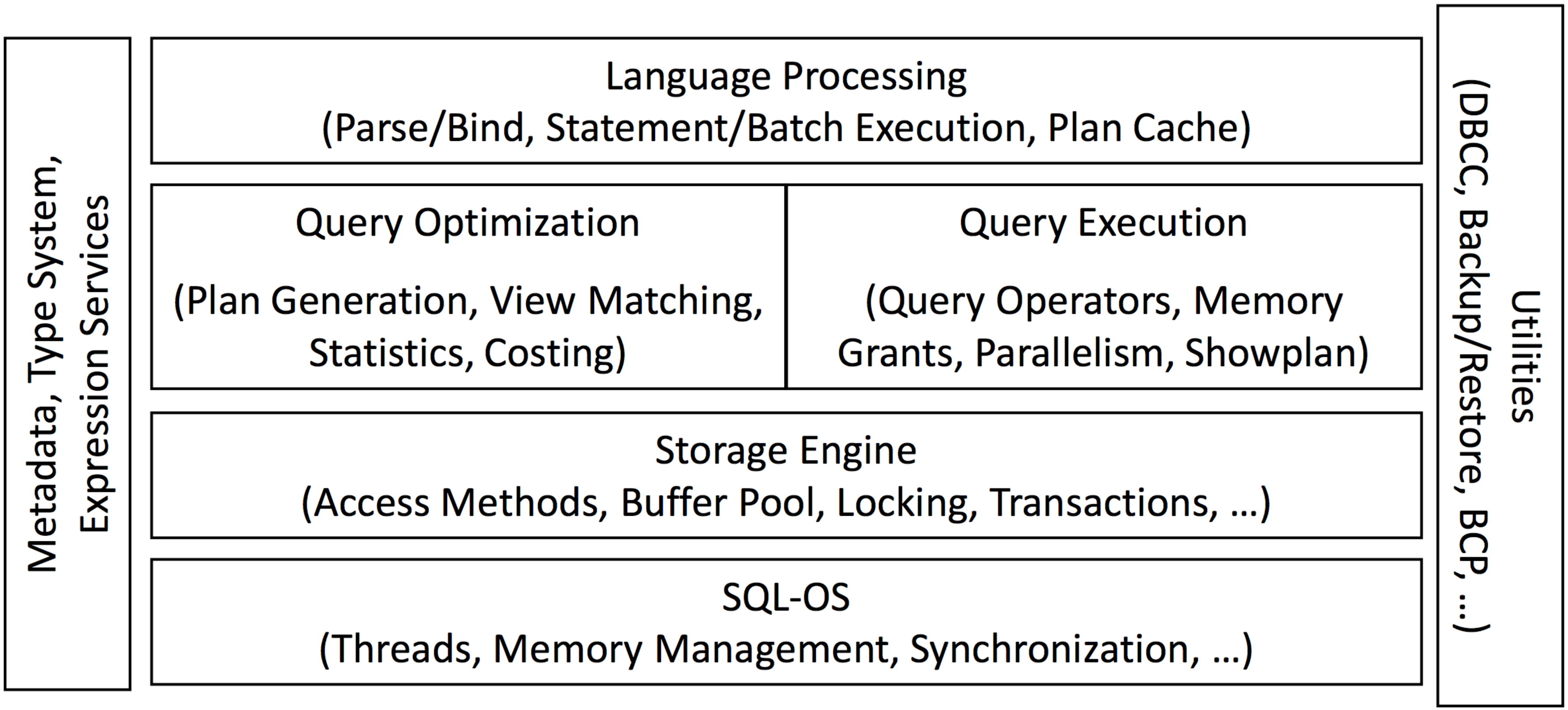


Image Source: Craig Freedman's SQL Server Blog

UNIVERSITY of WASHINGTON

RELATIONAL ALGEBRA

RELATIONAL STRUCTURE (AGAIN)

- "Tables" are known as "Relations"
- "Relations" have Keys and "Attributes"
- Records in Relations are known as "Tuples"

```
RELATION X (PRIMARY KEY, ATTRIBUTES...)

RELATION Y (PRIMARY KEY, FOREIGN KEY, ATTRIBUTES...)
```

WHAT IS AN ALGEBRA?

- A specific language consisting of operators and <u>atomic</u> operands.
  - e.g. $(3 + x) * \pi = 1$

RELATIONAL ALGEBRA

> **Operands**

Variables (relations)

Constants (finite relations)

> **Operators**

<u>Binary</u>

$\cup$ (union)

$\cap$ (intersection)

— (difference)

× (cross product)

$\bowtie$ (natural join)

<u>Unary</u>

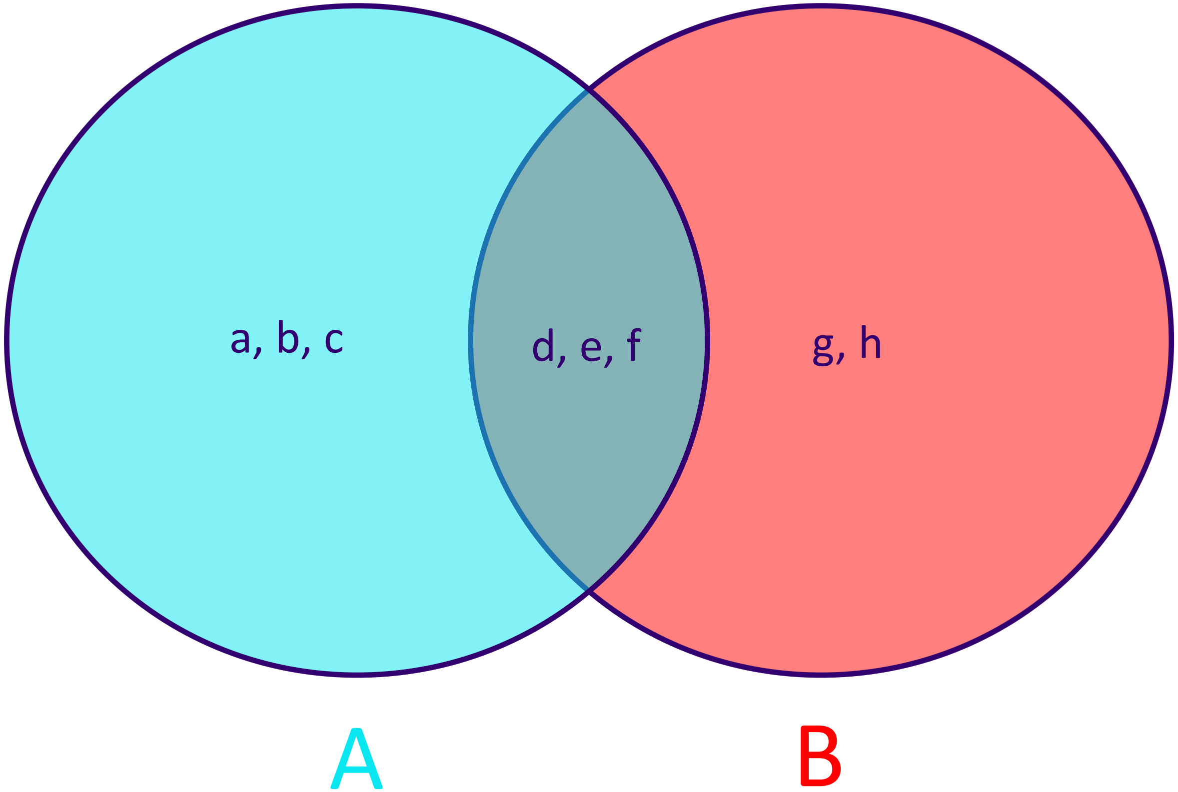
σ (selection)

π (projection)

SET OPERATORS

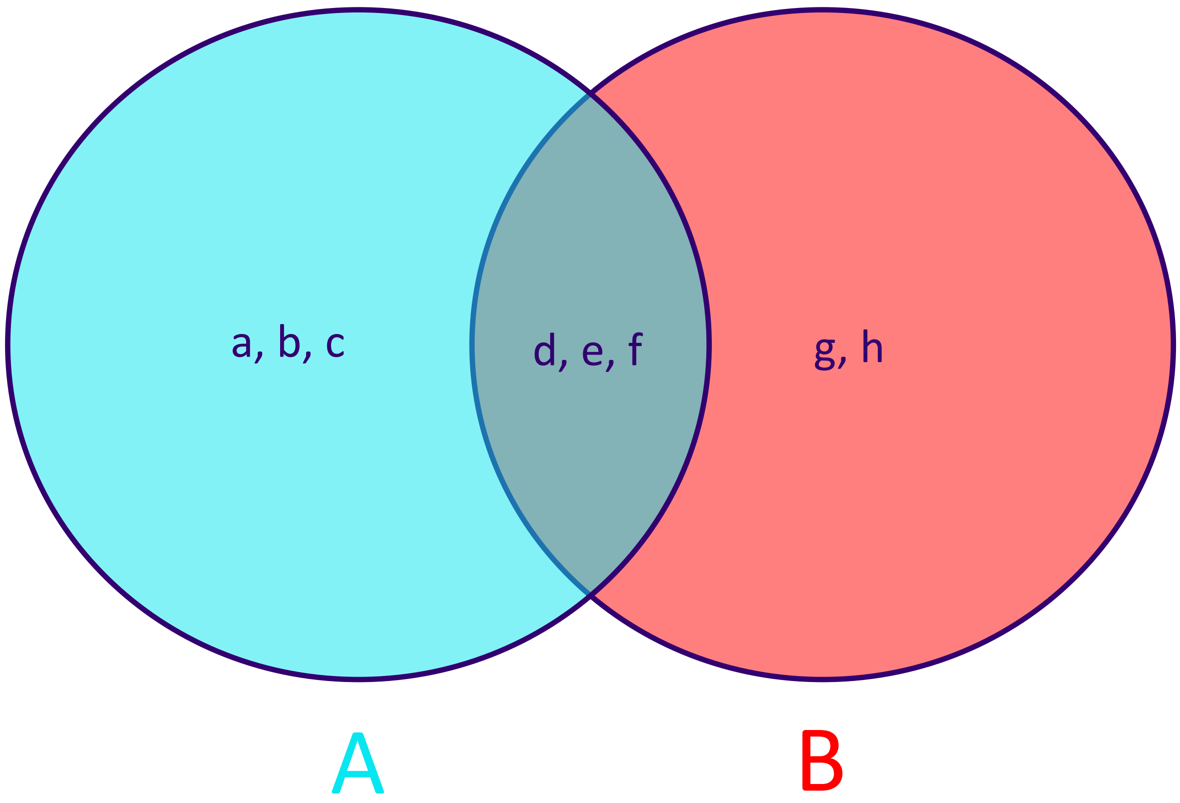
Let set A = {a, b, c, d, e, f}

Let set B = {d, e, f, g, h}



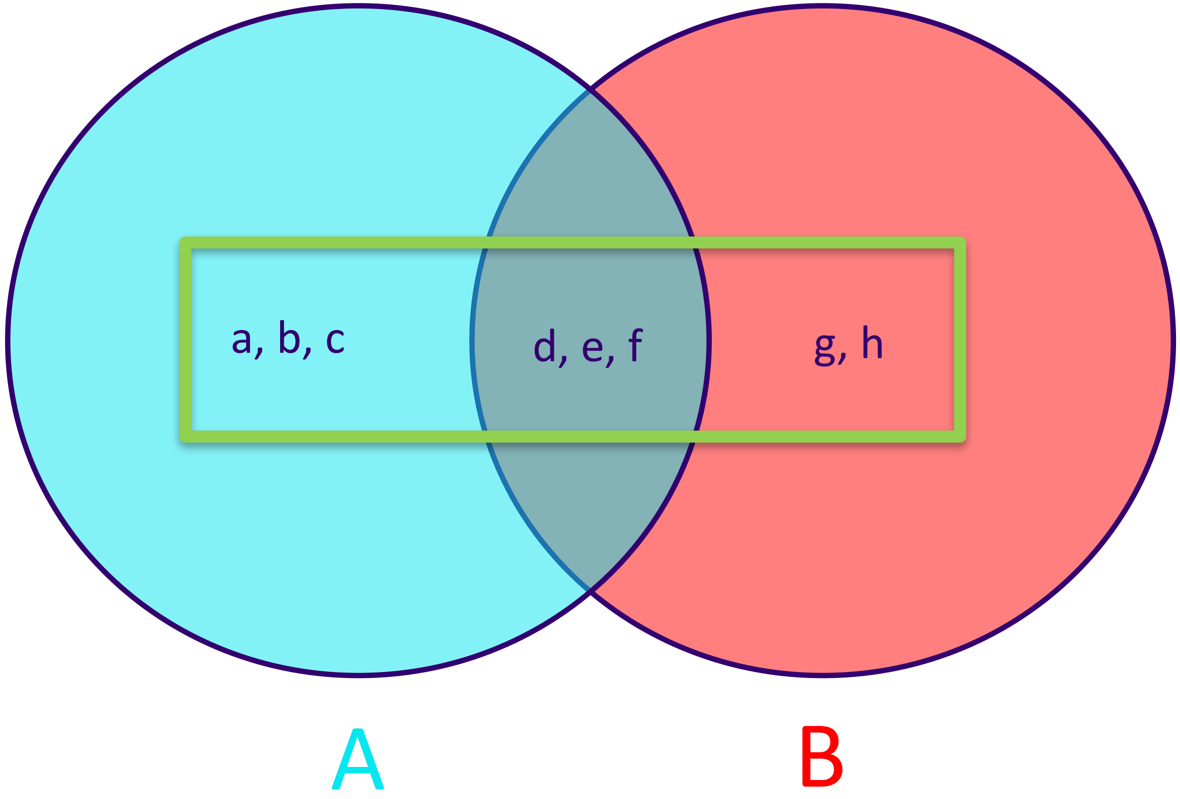
SET OPERATORS: $A \cup B$

$A \cup B = ???$



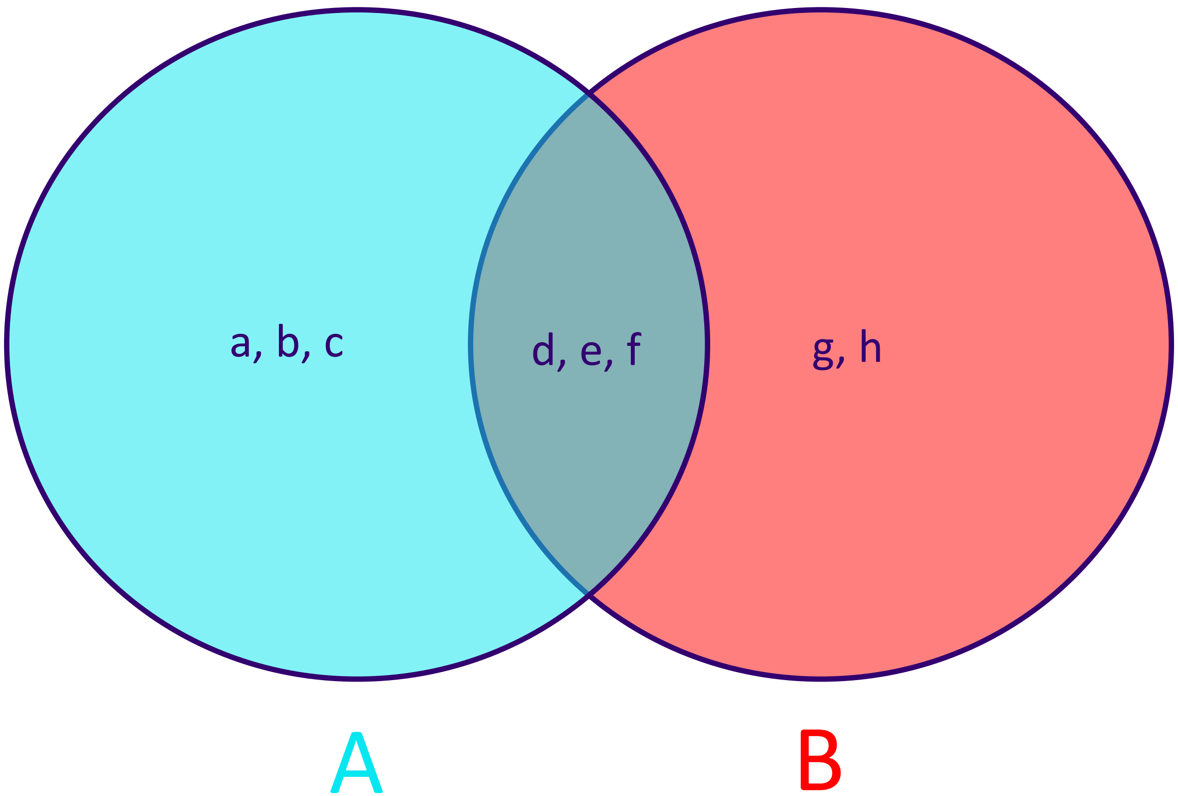
SET OPERATORS: $A \cup B$

$A \cup B = \{a, b, c, d, e, f, g, h\}$



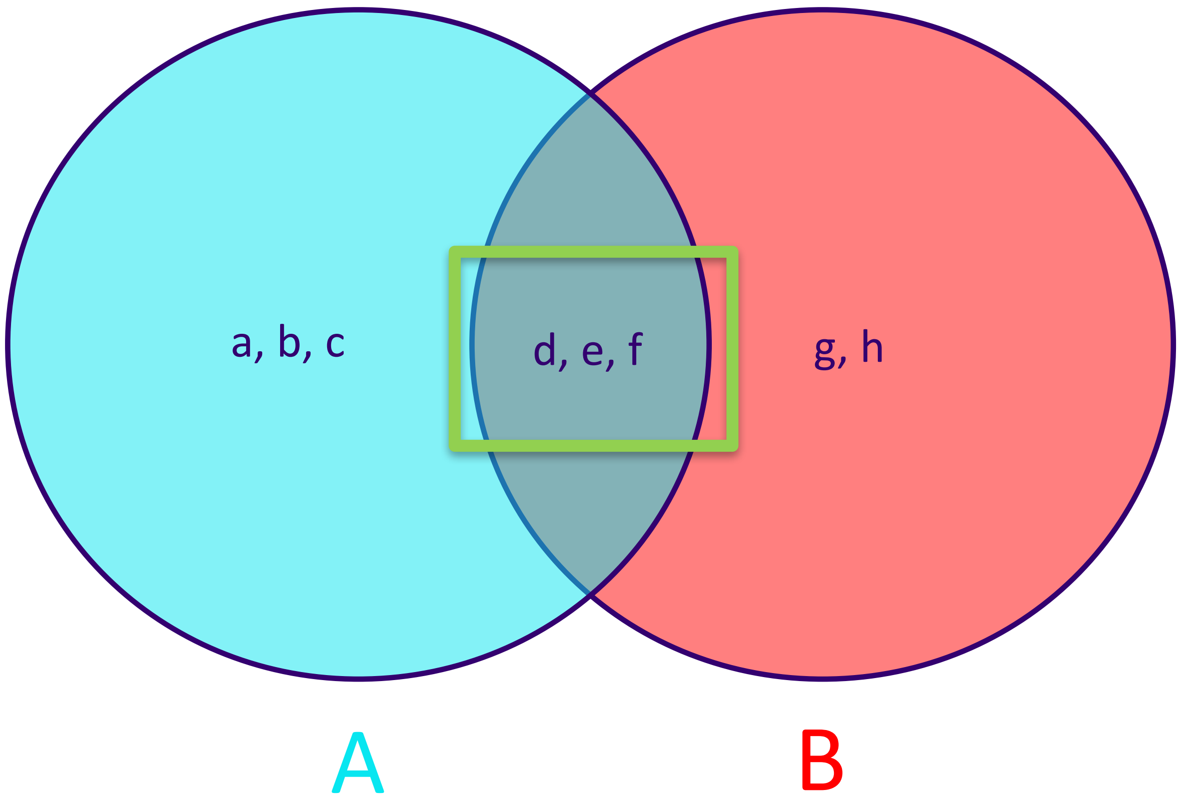
SET OPERATORS: A ∩ B

$A \cap B = ???$



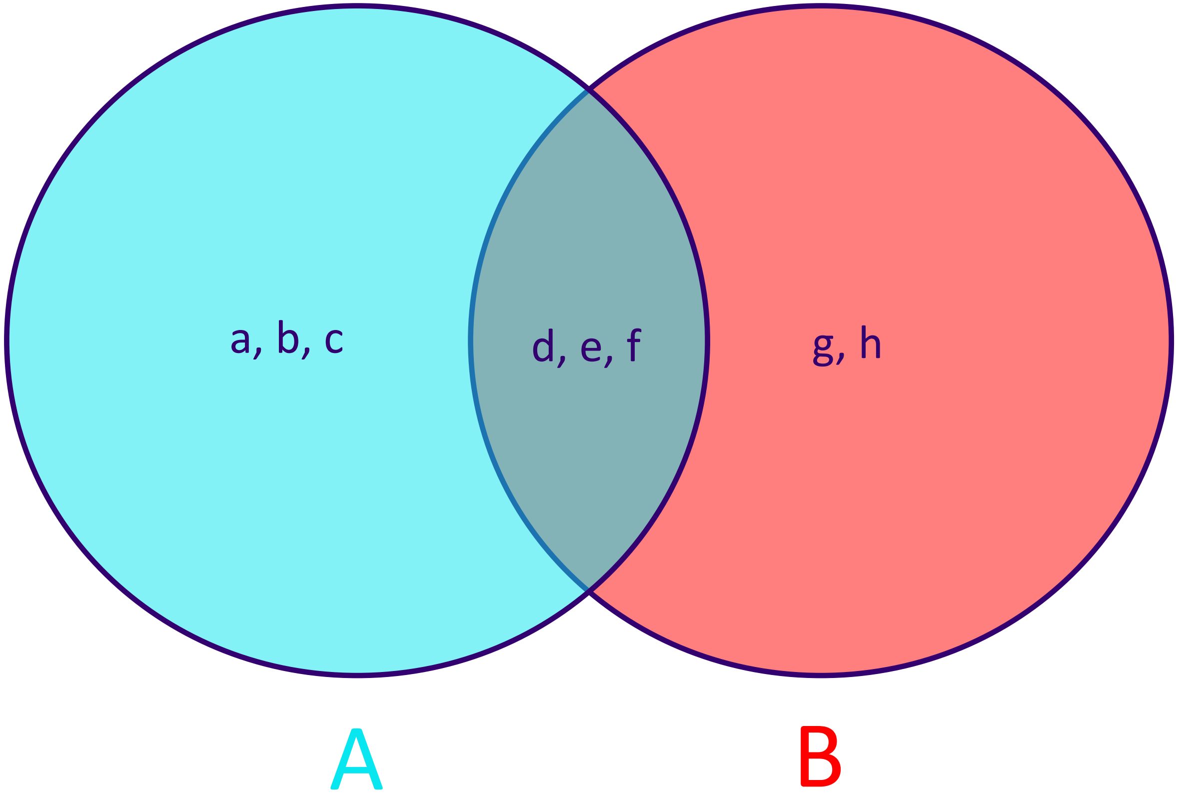
SET OPERATORS: $A \cap B$

$A \cap B = \{d, e, f\}$



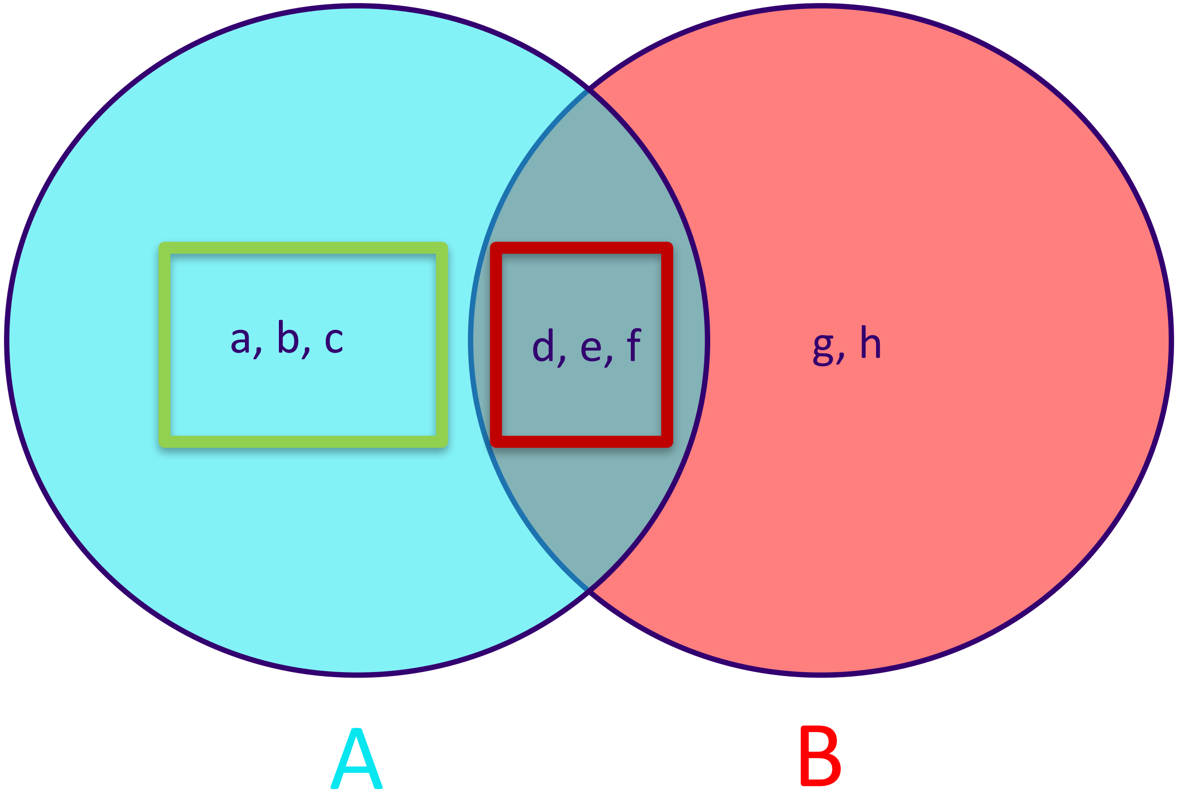
SET OPERATORS: A – B

A - B = ???



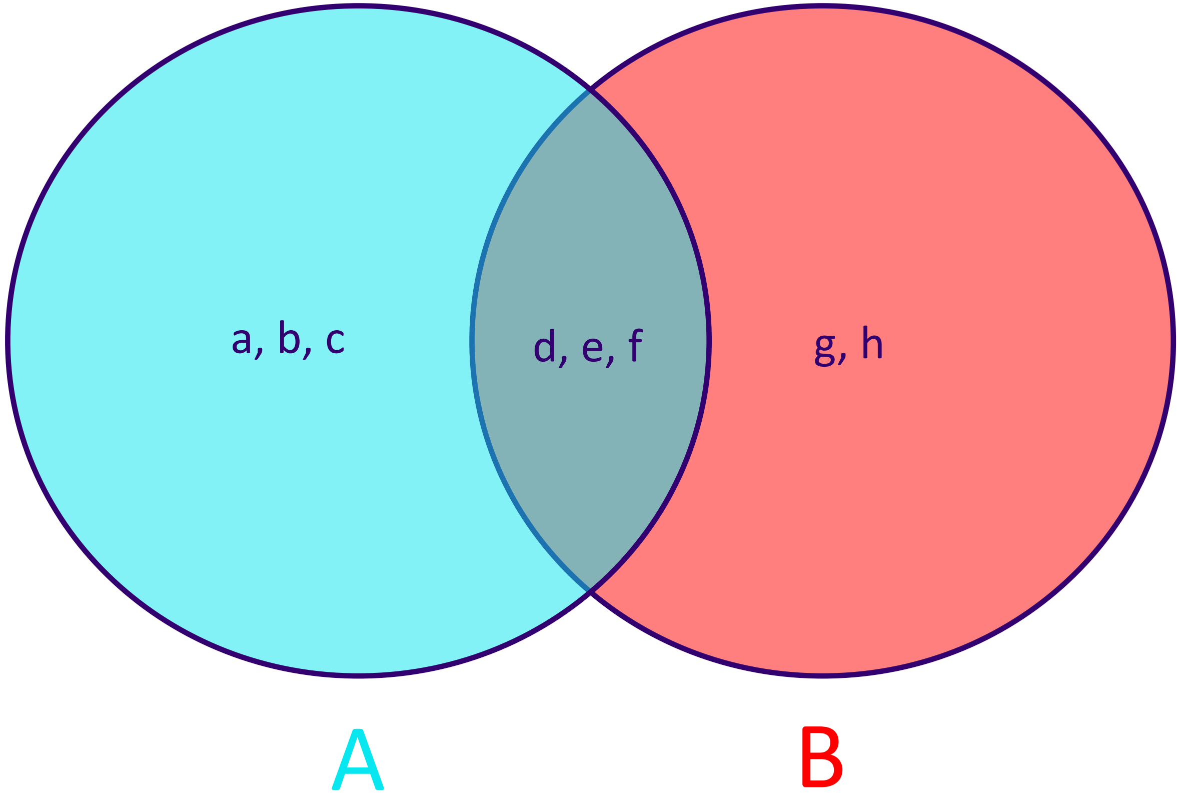
SET OPERATORS: A – B

A - B = {a, b, c}



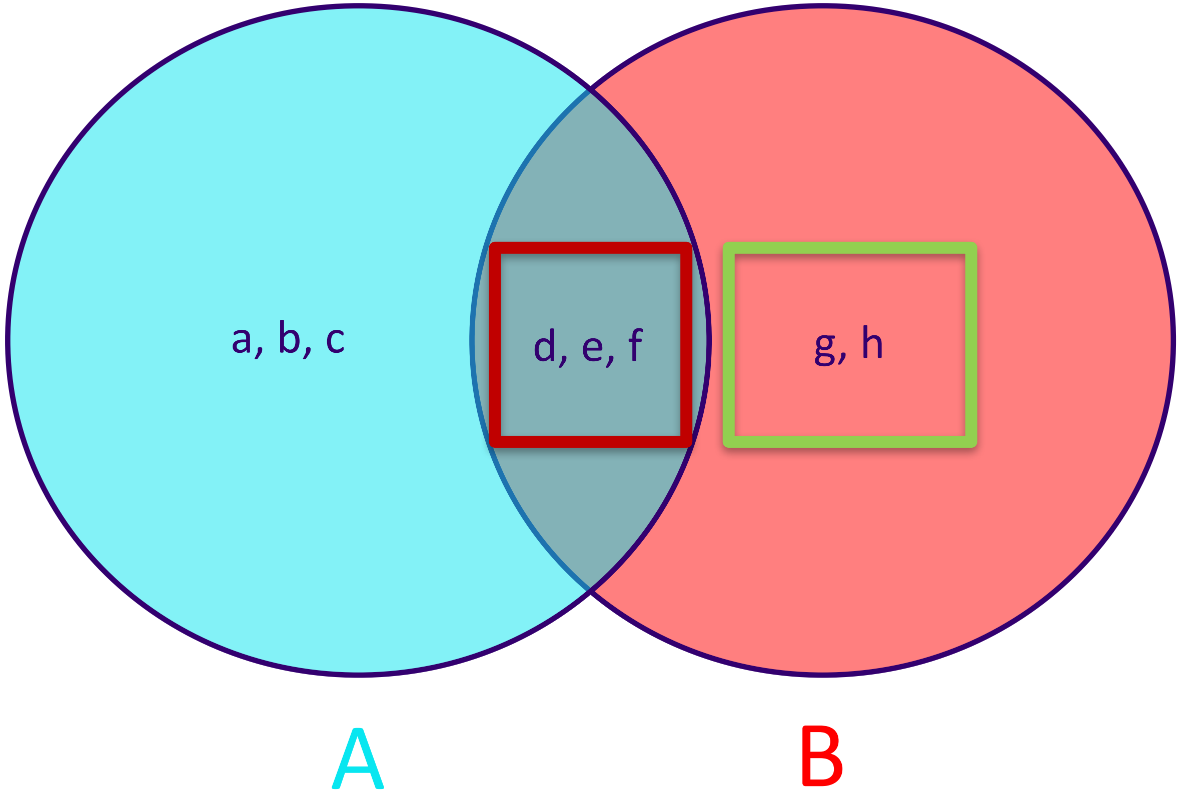
SET OPERATORS: B – A

B - A = ???



SET OPERATORS: B – A

B - A = {g, h}



OPERATING WITH RELATIONS

STUDENTS (S)

NAME	AGE
HOMER	37
MARGE	33
LISA	7
BART	10

$$S \cup F$$

FACULTY (F)

NAME	AGE
SKINNER	42
MARGE	33
CRABAPPLE	3
WILLIE	12

OPERATING WITH RELATIONS

STUDENTS (S)

NAME	AGE
HOMER	37
MARGE	33
LISA	7
BART	10

$S \cup F$

FACULTY (F)

NAME	AGE
SKINNER	42
MARGE	33
CRABAPPLE	3
WILLIE	12

Note: Attributes must match for set operations to be performed on relations.

NAME	AGE
HOMER	37
MARGE	33
LISA	7
BART	10
SKNNER	42
CRABAPPLE	3
WILLIE	12

< Notice Marge appears only once

RELATIONS EXERCISES

STUDENTS (S)

NAME	AGE
HOMER	37
MARGE	33
LISA	7
BART	10

FACULTY (F)

NAME	AGE
SKINNER	42
MARGE	33
CRABAPPLE	3
WILLIE	12

CALCULATE THE FOLLOWING

a) $F \cap S$

b) $S - F$

c) $F - S$

CROSS PRODUCT (aka Cartesian Product)

Let set A = {a, b, c}

Let set B = {d, e}

$$A \times B = \{ad, ae, bd, be, cd, ce\}$$

$$|A \times B| = |A| * |B|$$

CROSS PRODUCT (aka Cartesian Product)

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

S x CS

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
HOMER	37	TCSS360	LISA
HOMER	37	TCSS445A	LISA
LISA	7	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

PROJECTION: $\pi_{A1,A2...An}(R)$

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

S x CS

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
HOMER	37	TCSS360	LISA
HOMER	37	TCSS445A	LISA
LISA	7	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

PROJECTION: $\pi_{A1,A2\ldots An}(R)$

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

$\pi_{sname}(S)$

SNAME
HOMER
LISA

S x CS

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
HOMER	37	TCSS360	LISA
HOMER	37	TCSS445A	LISA
LISA	7	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

PROJECTION: $\pi_{A1,A2...An}(R)$

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

$\pi_{sname}(S)$

SNAME
HOMER
LISA

S x CS

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
HOMER	37	TCSS360	LISA
HOMER	37	TCSS445A	LISA
LISA	7	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

$\pi_{s.sname,\ age,\ course}(S \times CS)$

S.SNAME	AGE	COURSE
HOMER	37	TCSS445A
HOMER	37	TCSS360
HOMER	37	TCSS445A
LISA	7	TCSS445A
LISA	7	TCSS360
LISA	7	TCSS445A

SELECTION: $\sigma_c(R)$

- c = Boolean criteria on which to 'filter' relation R

SELECTION: $\sigma_c(R)$

> **c = Boolean criteria on which to 'filter' relation R**

Example:
Given a relation **Products**, select products with:
Price *less than $100 in the Category* ***Electronics***

SELECTION: $\sigma_c(R)$

> c = Boolean criteria on which to 'filter' relation R

Example:
Given a relation Products, select products with:
Price *lower than $100 in the Category* ***Electronics***

$$\sigma_{\text{price} < 100 \,\wedge\, \text{category} = \text{'electronics'}}(\text{Products})$$

SELECTION: $\sigma_c(R)$

STUDENTS (S)

SNAME	AGE
HOMER	37
MARGE	33
LISA	7
BART	10

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

$\sigma_{age>20}(S)$

SNAME	AGE
HOMER	37
MARGE	33

$\sigma_{course='TCSS445A' \wedge sname='LISA'}(CS)$

COURSE	SNAME
TCSS445A	LISA

SELECTION: $\sigma_c(R)$ [EXERCISES]

STUDENTS (S)

SNAME	AGE	IQ	HAIRCOLOR	OCCUPATION
HOMER	37	40	BROWN	SAFETY ENGINEER
MARGE	33	96	BLUE	MOTHER OF 3
LISA	7	132	YELLOW	STUDENT
BART	10	50	YELLOW	STUDENT
APU	40	108	BLACK	CLERK
LENNY	36	88	BROWN	SAFETY ENGINEER
CARL	40	79	BLACK	SAFETY ENGINEER

Select Tuples with Occupation Student with normal IQ (85-114)
Select Tuples with Occupation Student and abnormal IQ $(\infty, 85]$ $[114, \infty)$
Select Tuples with Occupation Safety Engineer who are in their 40s

SELECTION: $\sigma_c(R)$ [EXERCISES]

STUDENTS (S)

SNAME	AGE	IQ	HAIRCOLOR	OCCUPATION
HOMER	37	40	BROWN	SAFETY ENGINEER
MARGE	33	96	BLUE	MOTHER OF 3
LISA	7	132	YELLOW	STUDENT
BART	10	50	YELLOW	STUDENT
APU	40	108	BLACK	CLERK
LENNY	36	88	BROWN	SAFETY ENGINEER
CARL	40	79	BLACK	SAFETY ENGINEER

Select tuples with Occupation Student with normal IQ (85-114)

$\sigma_{\text{occupation='student'} \wedge \text{iq} > 85 \wedge \text{iq} < 114}(S)$

Select tuples with Occupation Student and abnormal IQ (∞, 85] [114, ∞)

$\sigma_{\text{occupation='student'} \wedge (\text{iq} <= 85 \vee \text{iq} >= 114)}(S)$

Select tuples with Occupation Safety Engineer who are in their 40s

$\sigma_{\text{occupation='safety engineer'} \wedge \text{age} >= 40 \wedge \text{age} < 50}(S)$

JOINS (Selection + Cross Product)

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

Say we want the ages of those registered each of the classes?

JOINS (Selection + Cross Product)

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

S x CS

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
HOMER	37	TCSS360	LISA
HOMER	37	TCSS445A	LISA
LISA	7	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

First we take the Cross Product of the relations

JOINS (Selection + Cross Product)

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

S x CS

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
HOMER	37	TCSS360	LISA
HOMER	37	TCSS445A	LISA
LISA	7	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

$\sigma_{\text{s.sname = r.sname}}$ (SxCS)

S.SNAME	AGE	COURSE	CS.SNAME
HOMER	37	TCSS445A	HOMER
LISA	7	TCSS360	LISA
LISA	7	TCSS445A	LISA

And select only the records where the names match!

NATURAL JOIN $\bowtie$

- SYNTATCIC SUGAR
  - Doesn’t add to the power of the algebraic query language
- SHORT HAND FOR JOIN
  - On matching attribute names being equal
  - Drops redundant names*
- REQUIRES RESPONSIBLE (and simple) DB SCHEMAS

NATURAL JOIN ⋈

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

NATURAL JOIN ⋈

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

S ⋈ CS

SNAME	AGE	COURSE
HOMER	37	TCSS445A
LISA	7	TCSS360
LISA	7	TCSS445A

NATURAL JOIN $\bowtie$

STUDENTS (S)

SNAME	AGE
HOMER	37
LISA	7

CLASS_SCHEDULE (CS)

COURSE	SNAME
TCSS445A	HOMER
TCSS360	LISA
TCSS445A	LISA

S $\bowtie$ CS

SNAME	AGE	COURSE
HOMER	37	TCSS445A
LISA	7	TCSS360
LISA	7	TCSS445A

$\pi_{\text{s.sname, age, course}}(\sigma_{\text{s.sname = cs.sname}}\ (\text{SxCS}))$

S.SNAME	AGE	COURSE
HOMER	37	TCSS445A
LISA	7	TCSS360
LISA	7	TCSS445A

Can be done with the longer notation that includes a projection, selection and a cross product!

FOR NEXT WEEK

- Read Sections 1.1, 1.2, 2.1, 2.2, 2.4, 5.1, and 5.2
- Review & Get Started on Homework 1 (Due 10/11)
- Check out Discussion Board on Canvas