

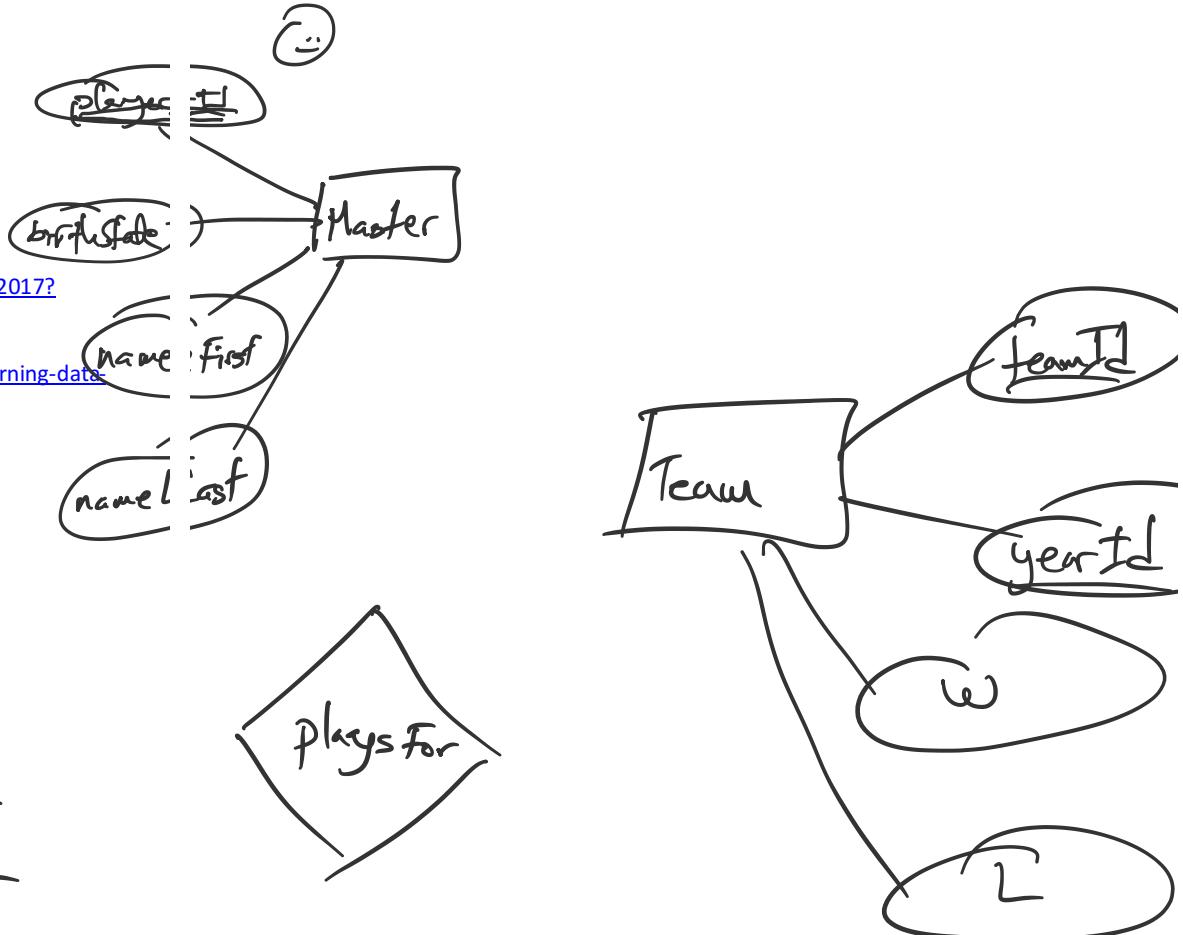
## Lesson Plan 2/13

Saturday, February 10, 2018 3:38 PM

- Admin
  - o How to knit PDFs for submission if you have errors?
  - o SQL examples from last times added to calendar
- HDSC:
  - o Kaggle Data Science Survey: <https://www.kaggle.com/surveys/2017?utm=cade>
  - o One take on it:  
<https://www.theverge.com/2017/11/1/16589246/machine-learning-data-science-dirty-data-kaggle-survey-2017>
- Finishing from last time
  - o The arrow discussion
  - o Group by and summarize
  - o Subqueries
- Operations with two tables (joins)
  - o All the joins
- SQL as a data definition language
  - o Create, Insert, Delete, Update
  - o Views
  - o NULLs
- Parting shots
  - o Optimization
  - o JSON

Entity set  $E_1$   
Entity set  $E_2$

$R: E_1 \times E_2$  e.g.  $(e_1, e_2)$



where  $e_1, e_2$

Toropese

Maste

$(e_1, e_2)$

("bryce"; "Nat")

---

Join:

Combine    a  
+ - .

$\vdash E_1$  and  $c_2 \in E_2$

八〇

at "player only plays for one  
team"



i. e., appears once in R

$b^{(r)}$ ) a boyce appears once in R



library from two tables

— — — 8 — — —      — — — 1 — — —

Table 1 join

— What to do .

"inner-join" : only  
"natural join" j

Left

R<sub>j</sub><sup>ht</sup>

Oster Siau

case a new table

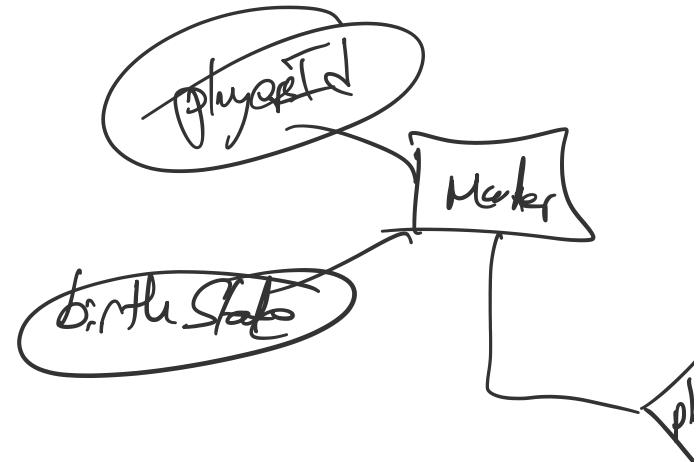
- Table 2 join condition  
about non-matching entities

include entities that satisfy  
or condition

Join : keep all entities  
in left table

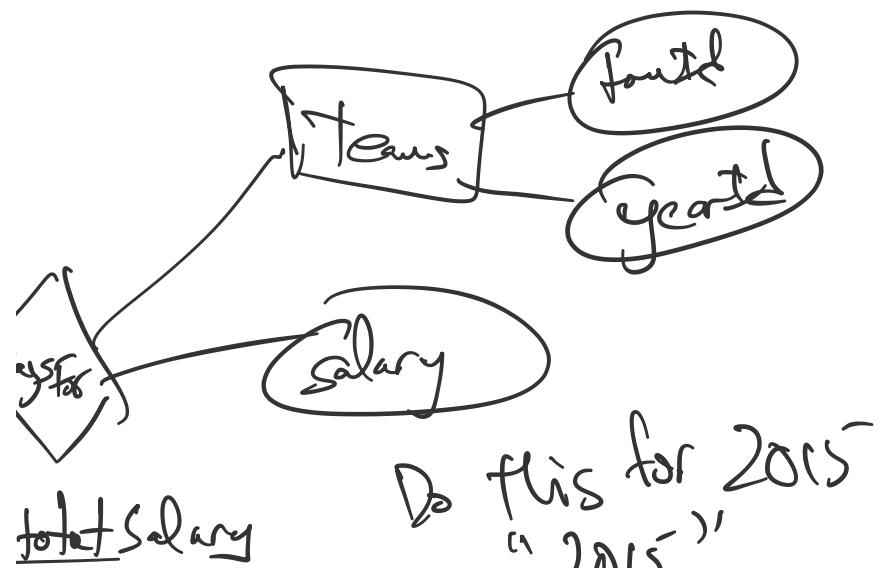
Join : give for right table

: keep all entities



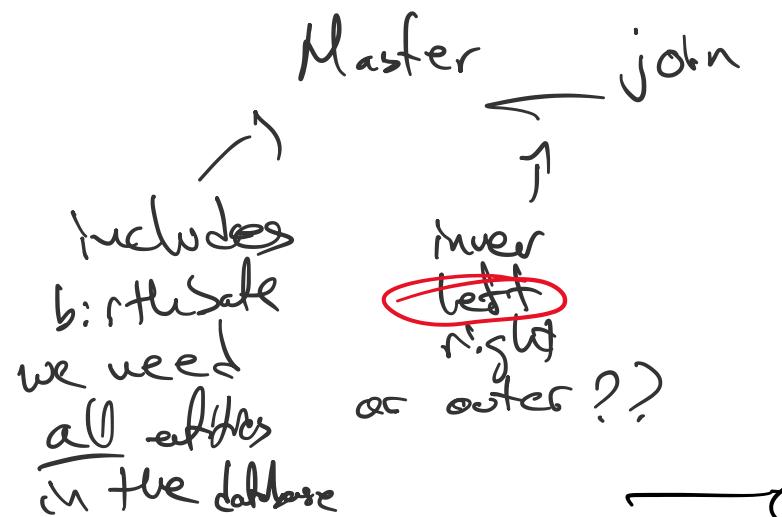
What kind of Join  
database is

—



to list all states in  
even if no players

Playing in :



Database      Query      E

SQL:      Declarative  
DML:      Procedural

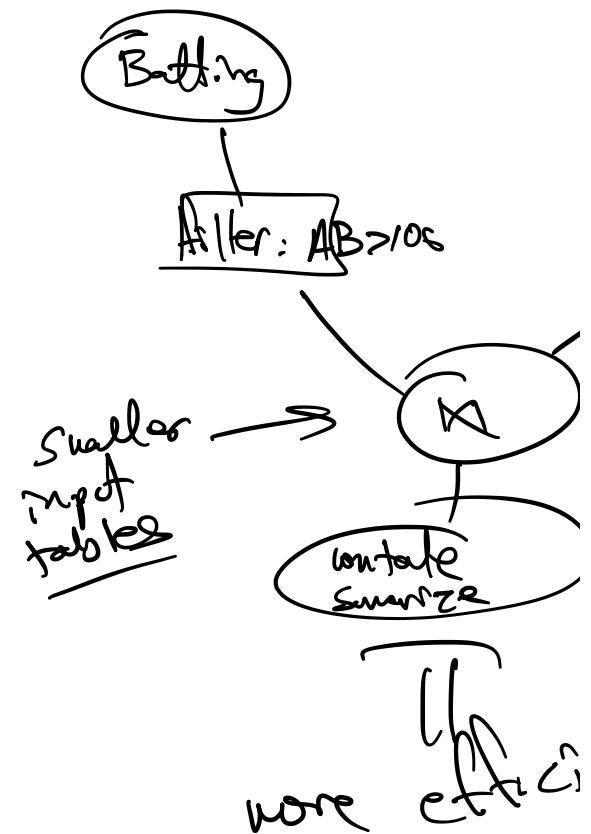
2015??

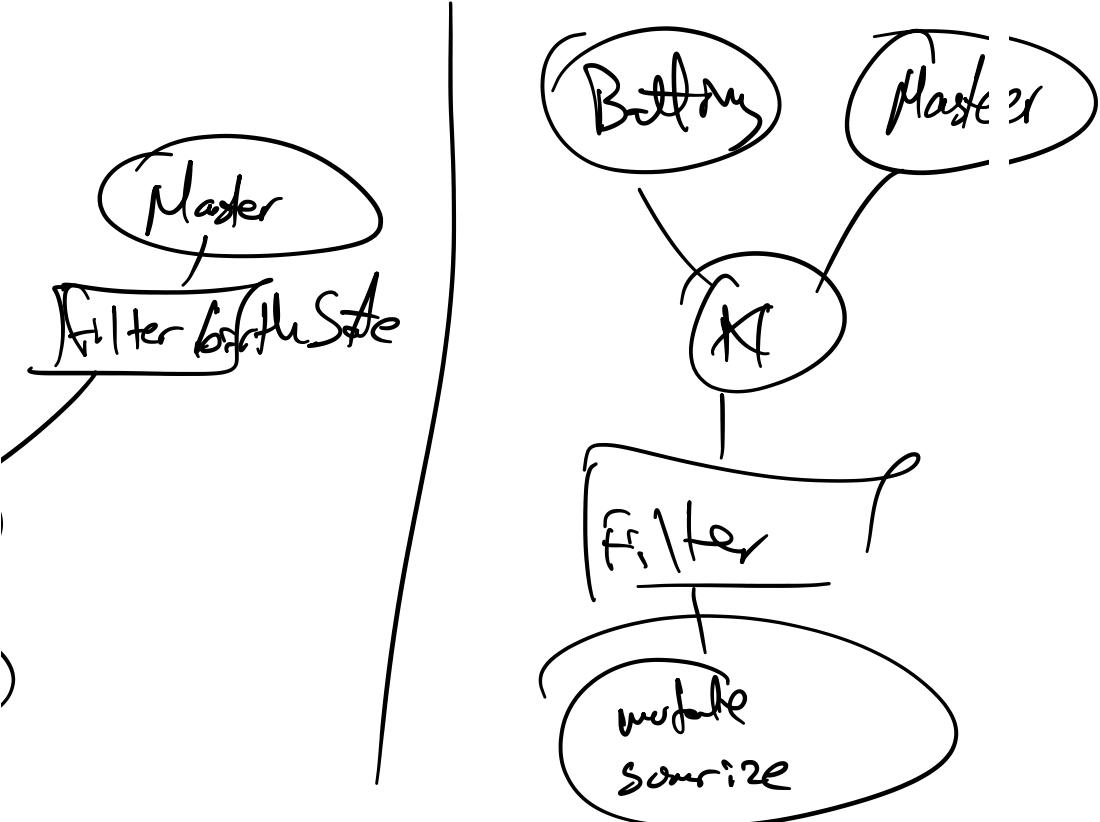
Plays for

→

Optimization

- 11 -





ent



Query

Optimizer:

find execution plan to evaluate  
in query most efficiently

- finds candidate plans
- Estimate cost each plan

"Search problem"

How to estimate cost:

→ Estimate of operation size

"How many rows satisfy  
a filter"

. . . n

- keep stats of attribute values
- Estimate of operation cost
- Cost of accessing rows  
"indexing"

SQL vs R:

- Describe Result
  - vs.
  - Describe: operations



Tidying Data

v -

.