

Chapter 5 : Extended RA and Datalog

Extension to Relational Algebra :

Extension :

① 扩展 Set RA 到 Bag RA (Relation as Bag)

⇒ 包上的关系操作

② 扩展传统 RA 的 Operators.

⇒ 扩展操作符

Datalog (Database logic) :

Basic Concepts

Comparison between Relational Algebra and Datalog

Lecture 8 Extended Relational Algebra and Datalog

Why we need to extend RA:

① 商用数据库系统中，Relations 为 Bag 而非 Set。

同一个元组允许重复多次

② 基于包的关系，使得 并 · 投影 操作无需去重步骤，实现效率高
↓
包 元

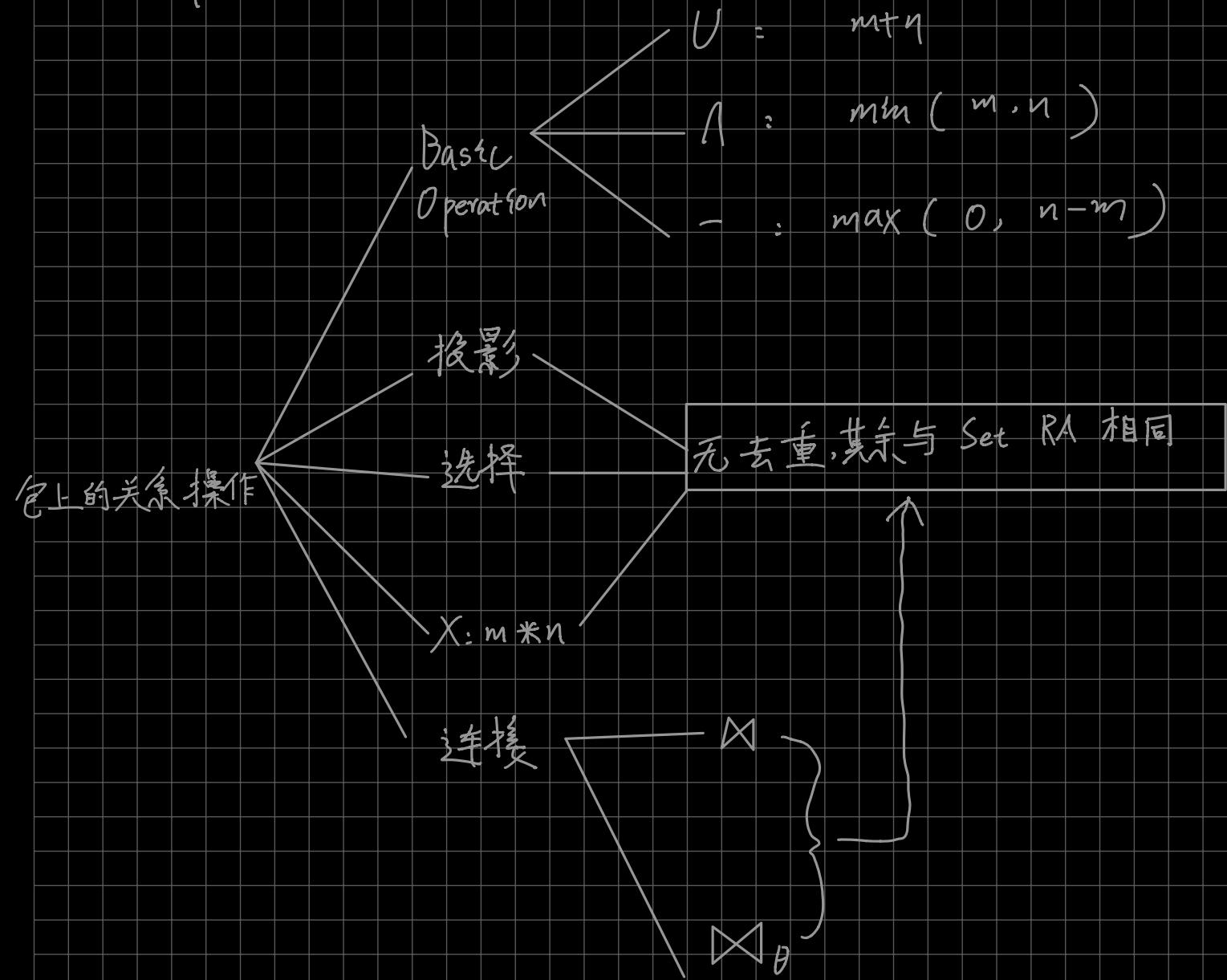
③ 某些结果（如求 属性平均值·众数）只有在包的情形下才能看到

因此需将基于 Set 理论的 RA $\xrightarrow{\text{extend}}$ 基于 Bag 的 RA

Why we need to extend RA operators?

为适合现代查询语言（SQL）的能力，需要一些传统RA中不具备的 Operators。

Assumption: R, S 为 bag, tuple t 于 R 中出现 n 次, S 中出现 m 次



△ 包上的代数定律与集合上代数定律 略有不同.

E.g. 集合差对于并的分配律: $(R \cup S) - T = (R - T) \cup (S - T)$ 于包上并不成立.

消除重复 $\Rightarrow \Sigma$

聚集 $\Rightarrow \text{SUM}, \text{AVG}, \text{MIN}, \text{MAX}, \text{COUNT}$

分组 $\Rightarrow \pi_L$: ① 若有分组，聚集操作于分组进行
② 分组属性 · 聚集属性 concept
③ Σ 为 L 的特殊小情况.

扩展投影 $\Sigma \Rightarrow \pi_L : \pi_{\underbrace{x+z}_{\rightarrow \text{表重命名}} \rightarrow y, \underbrace{x-z \rightarrow q}_{(R)}}$

排序 $\Rightarrow \tau$

外连接 $\Rightarrow \bowtie$

自然外连接

左

右

左&右

右外连接

Steps: ① 进行相应的 Set 连接操作

② 加入悬游元组 (Dangling)

③ 用 null(符号) $\rightarrow \perp$ 补齐无值属性

Datalog \Rightarrow 逻辑查询语言

I. Comparison With Relational Algebra

① Datalog 能表示 RA 可表示的所有查询语言

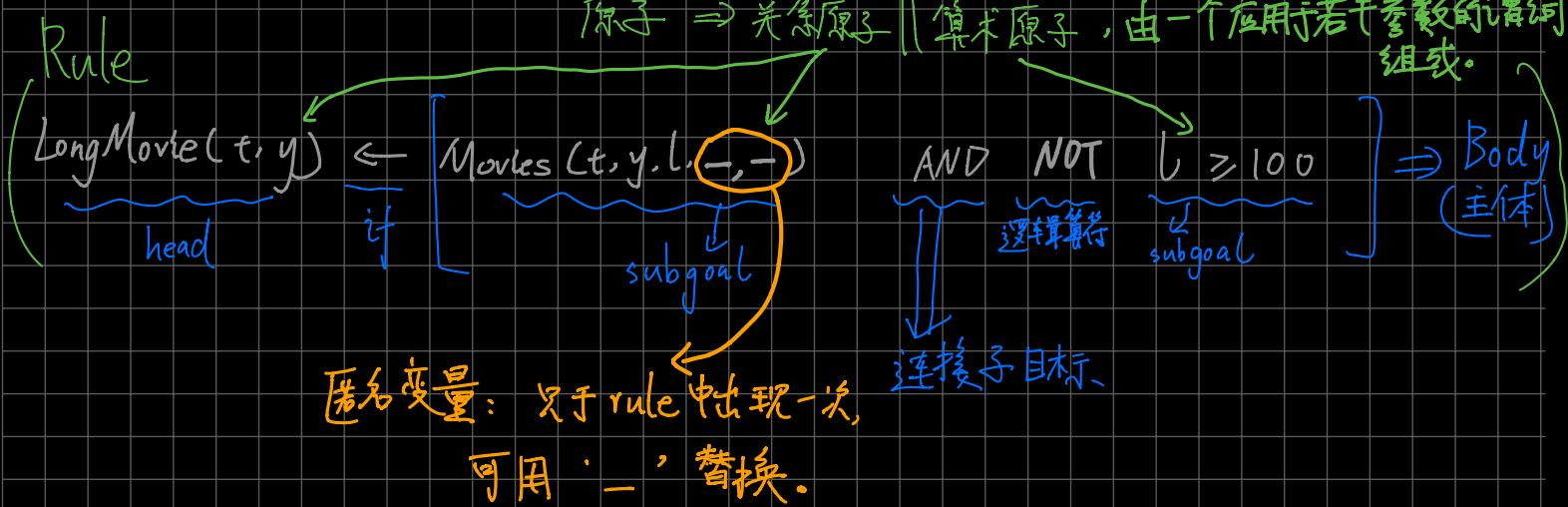
RA 能表示所有单个 Datalog 的规则.

② Datalog 能表示递归而 RA不行。

③ 扩展 RA 的一些操作 (e.g., 分组, 聚集)

不能用 Datalog 表达。 Datalog 也不能支持包操作。

II. Basic Concepts of Datalog



△ Datalog 中的 Query 由一个或多个 Rule 组合。若 Rule's head

① 只有一个关系，query 查询结果即为该 Relation 的值。② 若有多个关系，query 用 Answer 指定查询结果名，其余关系均为辅助定义。

△ 安全规则 (Safe Rule)：规则中每个变量均出现于主体的一些非否定关系子目标中。

⇒ 保证了 EDB 关系有限 → IDB 关系有限



Extended DataBase

Internal DataBase

EDB 谓词对应已存储的关系。
必须出现于 Rule's head.

IDB 谓词对应规则定义的关系，可不出现于 Rule's head

III. 用 Datalog Rules 表示 RA 操作

$\text{Answer}(a_1, a_2, \dots, a_n) \quad R(a_1, a_2, \dots, a_n) \quad S(a_1, a_2, \dots, a_n)$

$$\begin{array}{l} \text{Bool Operation} \\ \text{RUS} \Rightarrow \begin{cases} A \leftarrow R \\ A \leftarrow S \end{cases} \\ \text{RNS} \Rightarrow A \leftarrow R \text{ AND } S \\ \text{R-S} \Rightarrow A \leftarrow R \text{ AND NOT } S \end{array}$$

投影 $\Rightarrow R(t, y, l) \leftarrow \text{Movies}(t, y, l, g, s, p)$

选择 Θ 只包含 AND 操作 $\Rightarrow A \leftarrow C \text{ AND } G \text{ AND } \dots$

Θ 包含 OR 操作 $\Rightarrow n$ 个条件的 OR 可用 n 条带有相同 head 的 Rules

积 $\Rightarrow P(a, b, c, x, y, z) \leftarrow R(a, b, c) \text{ AND } S(x, y, z)$

自然连接 \Rightarrow 与 积 同样

连接

Θ 连接 \Rightarrow 先做积，后面用 AND 附上 条件

△ 复杂选择条件 \Rightarrow 将逻辑表达式 整理成 析取范式。

再按上述转换。



多重操作：观察关系代数表达式树，为树的每个内部结点创建一个 IDB 谓词。最终再合并中间 Rules 为一条复合 Rule。