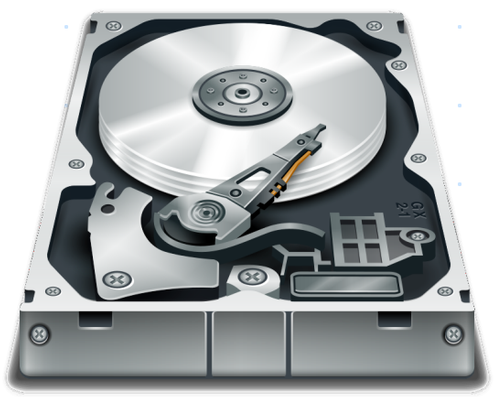
CSE 350

Advanced Data Structures

Topic 7: Serialization

Abstraction Mismatch

Permit #	Zip	State	Type
123	14228	NY	Electrical
125	14228	NY	Construction
127	10001	NY	Electrical



4K 4K 4K ...

openclipart.org

Assumptions

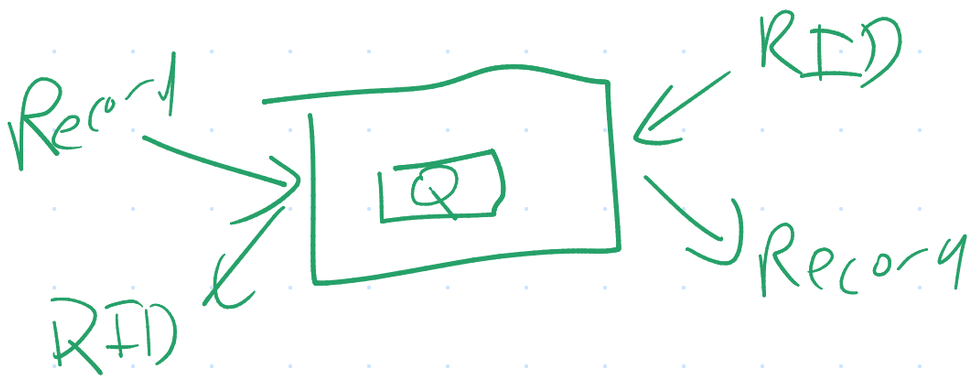
- DB is one table
- We know schema
- We don't need to store it

given schema

provide RID → Record (get)

Record → RID (put)

(record store

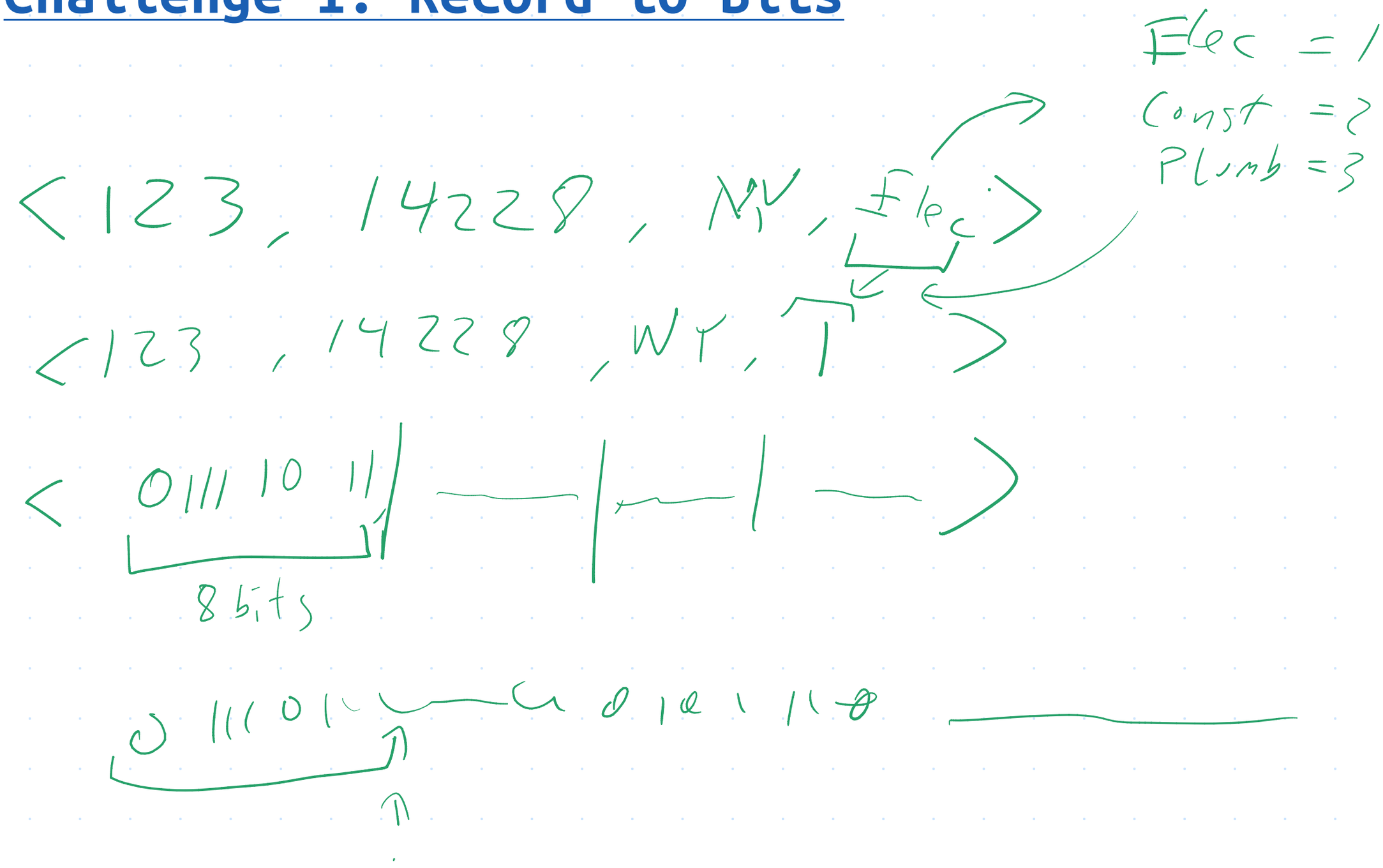


1. Record → Bits ✓

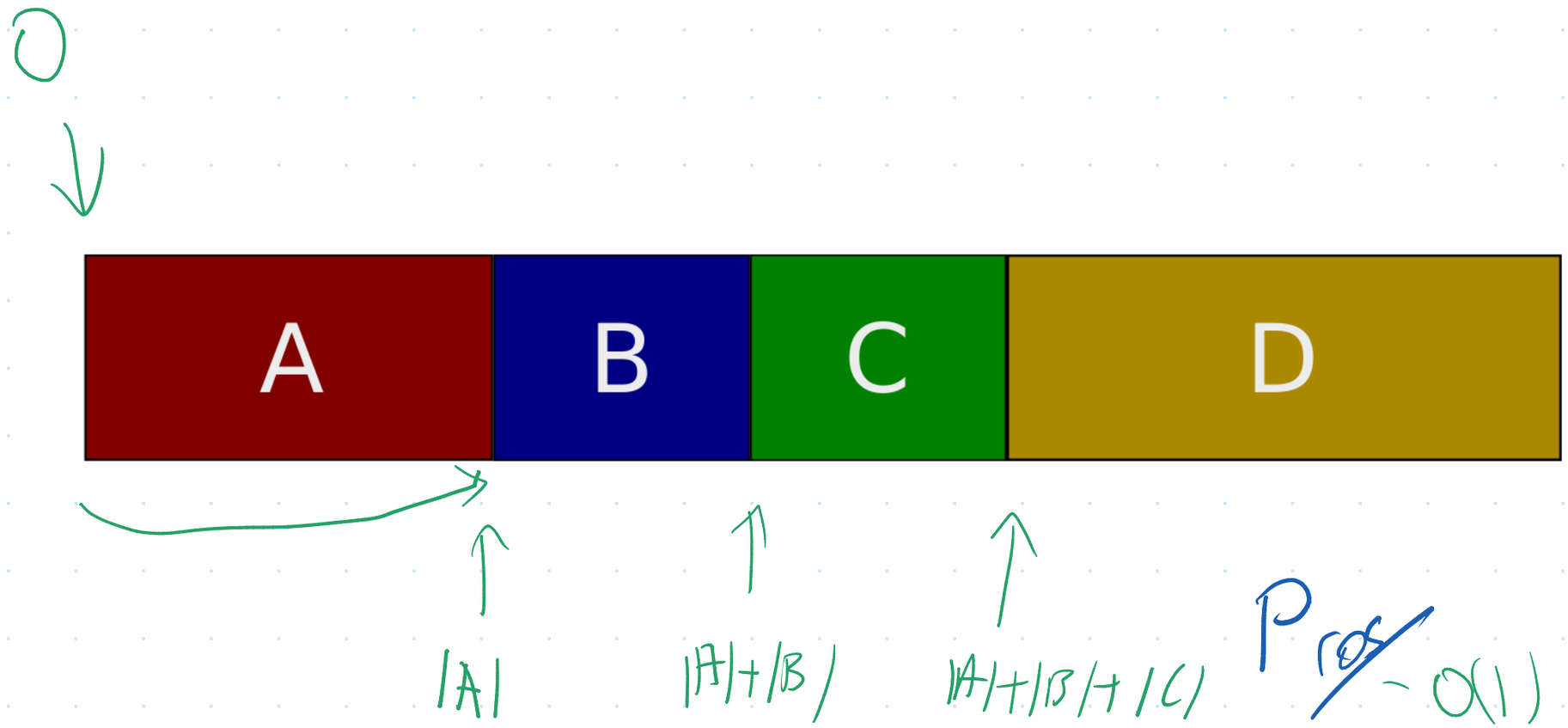
2. Bits R_1
   Bits R_2
   ⋮
   ] → Pages

3. Finding Pages in a Table

Challenge 1: Record to Bits



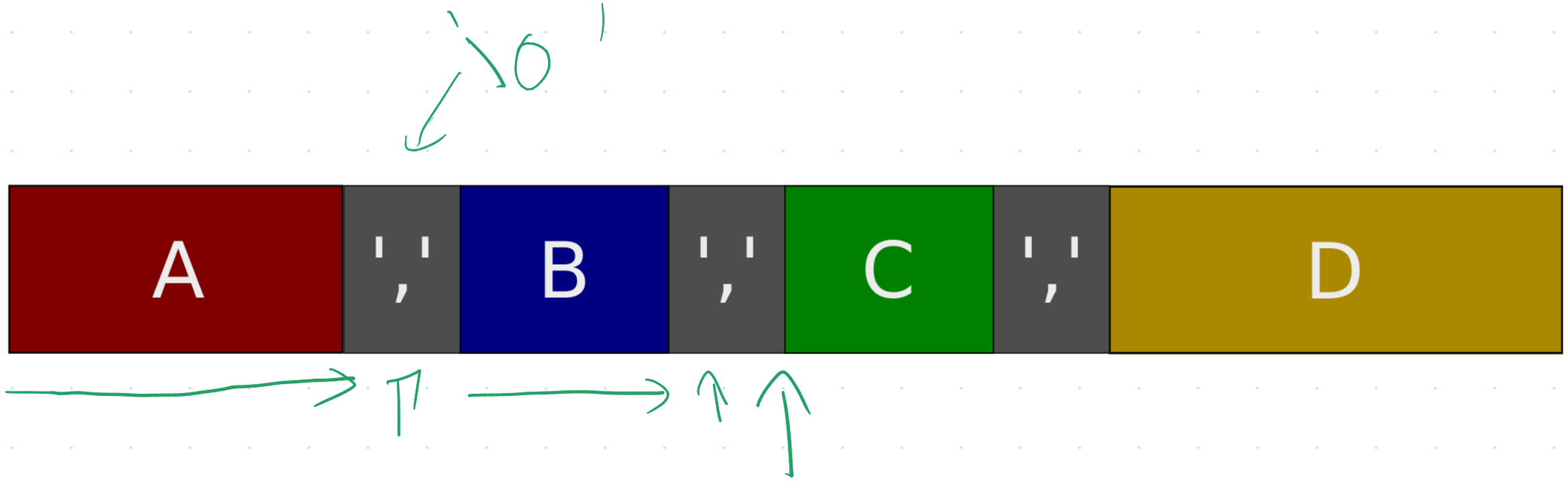
Variation 1: Fixed Layout



Pros
- O(1) field access
- limited space overhead

Cons
- Wasted space or not useful at all for variable sized fields

Variation 2: Delimiters

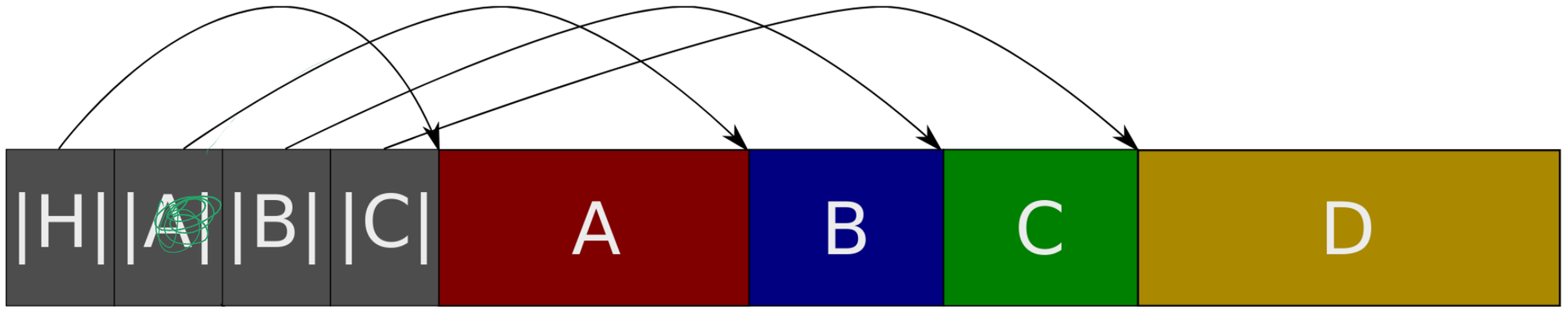


Pros/ - Variable length fields

Cons/ - symbol can't appear in data

- O(|record|) CPU to access a field

Variation 3: Headers



16 → ← 16 → ← 16 → ← 16 → ← 30 bytes →

bits
= 2 bytes

8 3.8

Pros
- Variable Size Fields
- O(1) access a field

Cons
- slightly slower than fixed
- A few extra bytes per field

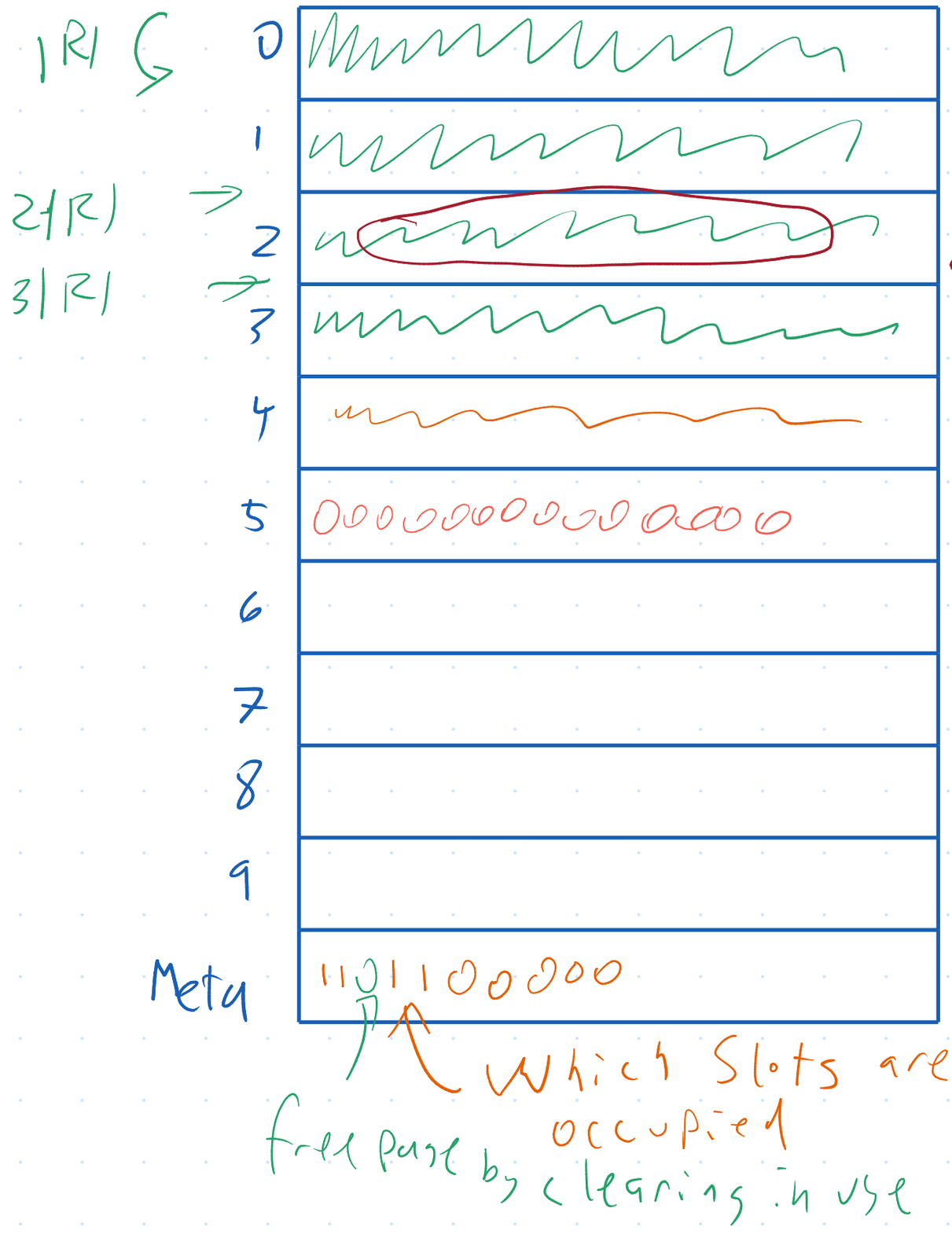
Challenge 2: Records to Pages

All layers are the same
↳ sort of

Tricks: Same

Identity: RID instead of field
↳ Array, not struct
↙ ↘
variable length, fixed length
fixed type, variable types

Variation 1: Fixed



What is a RID
Index of Slot

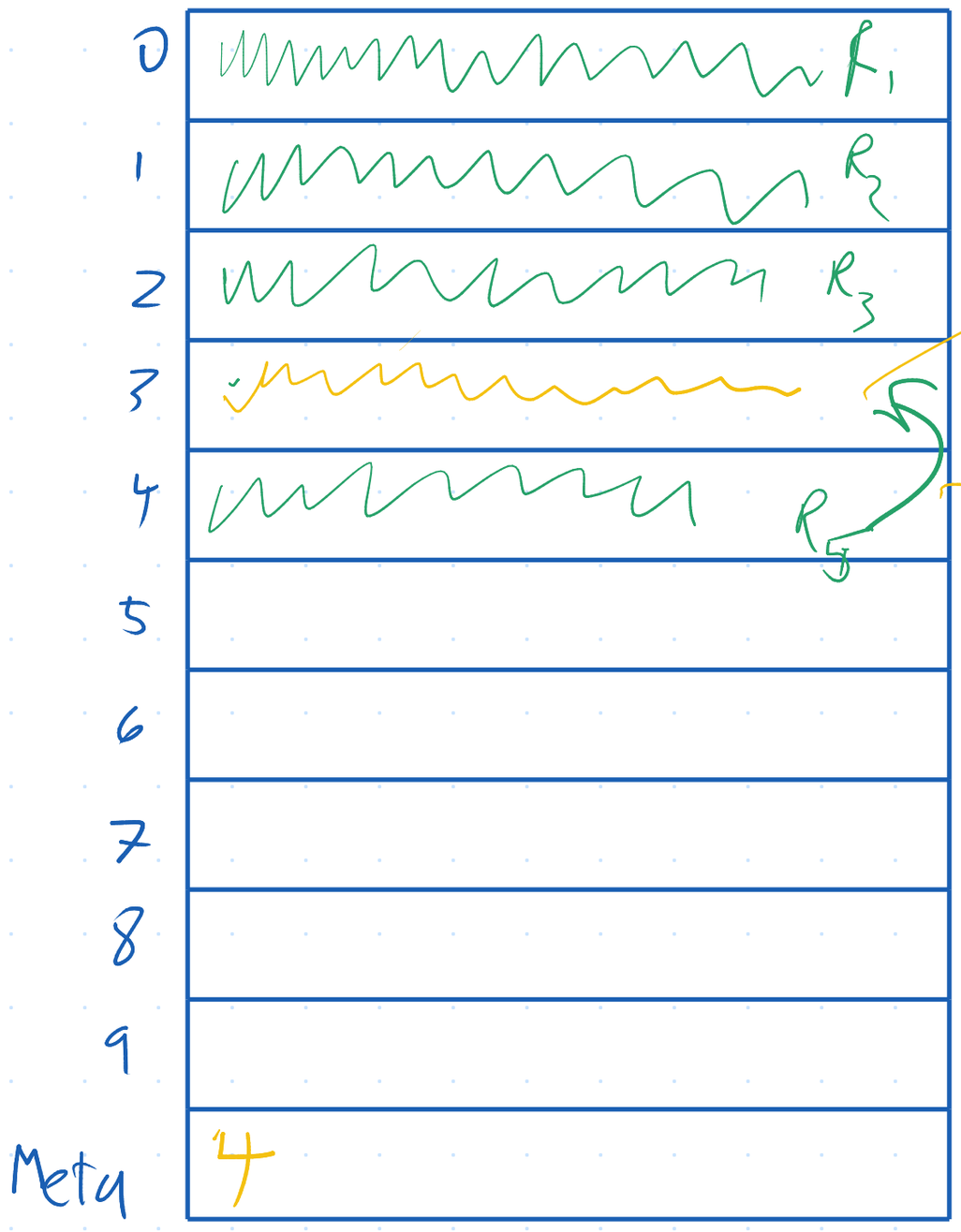
Where can I add the next record
store in use bits

Pro/ -somewhat fast allocation
-fast retrieve RID

Cons/ -stuck w/ fixed size Records

Variation 1b: Fixed (Packed)

Same basic idea



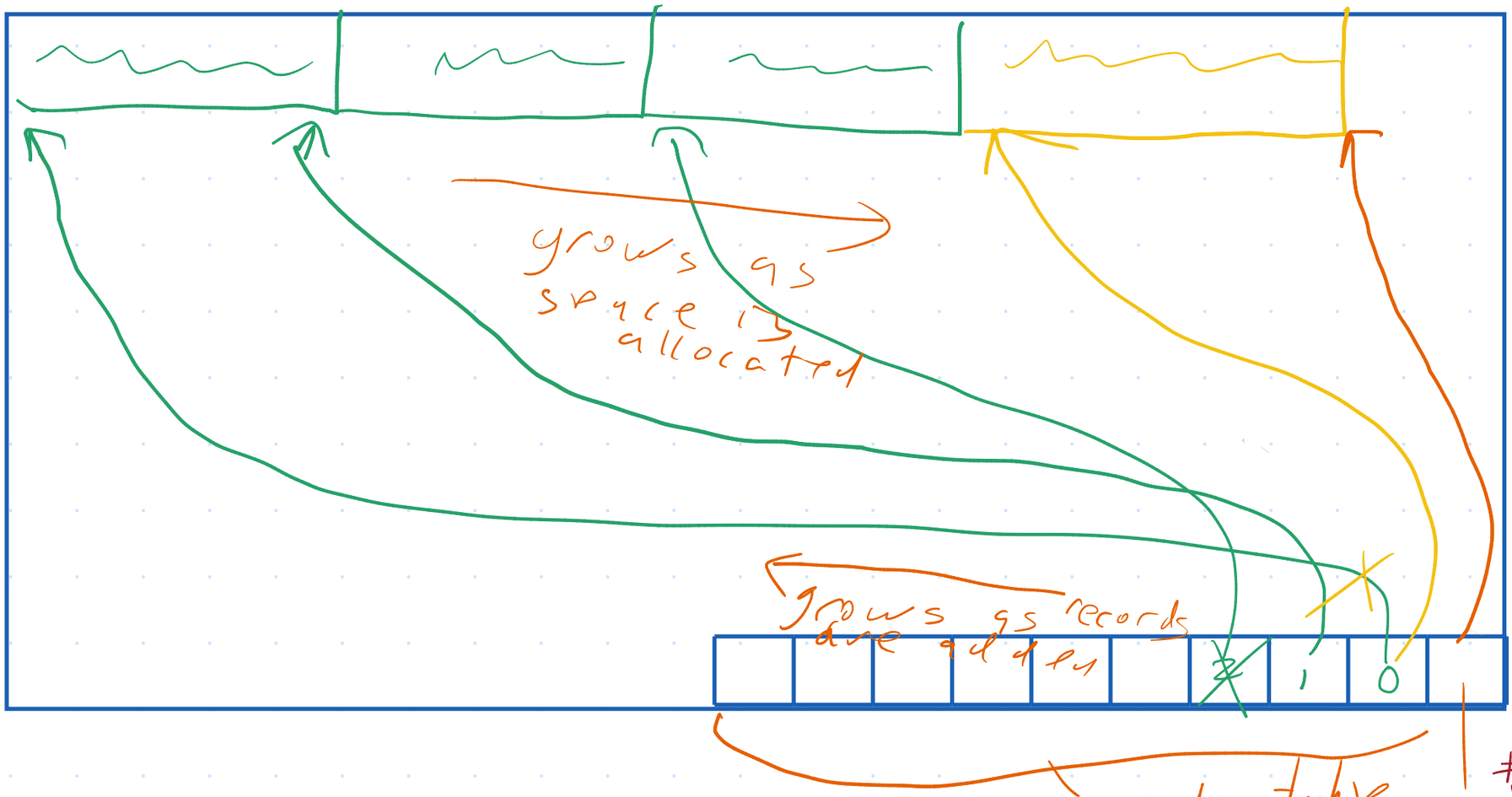
Pros/ -Records together
-Faster allocatio

Cons/ -Array index no longe
stable

Variation 2: Delimited

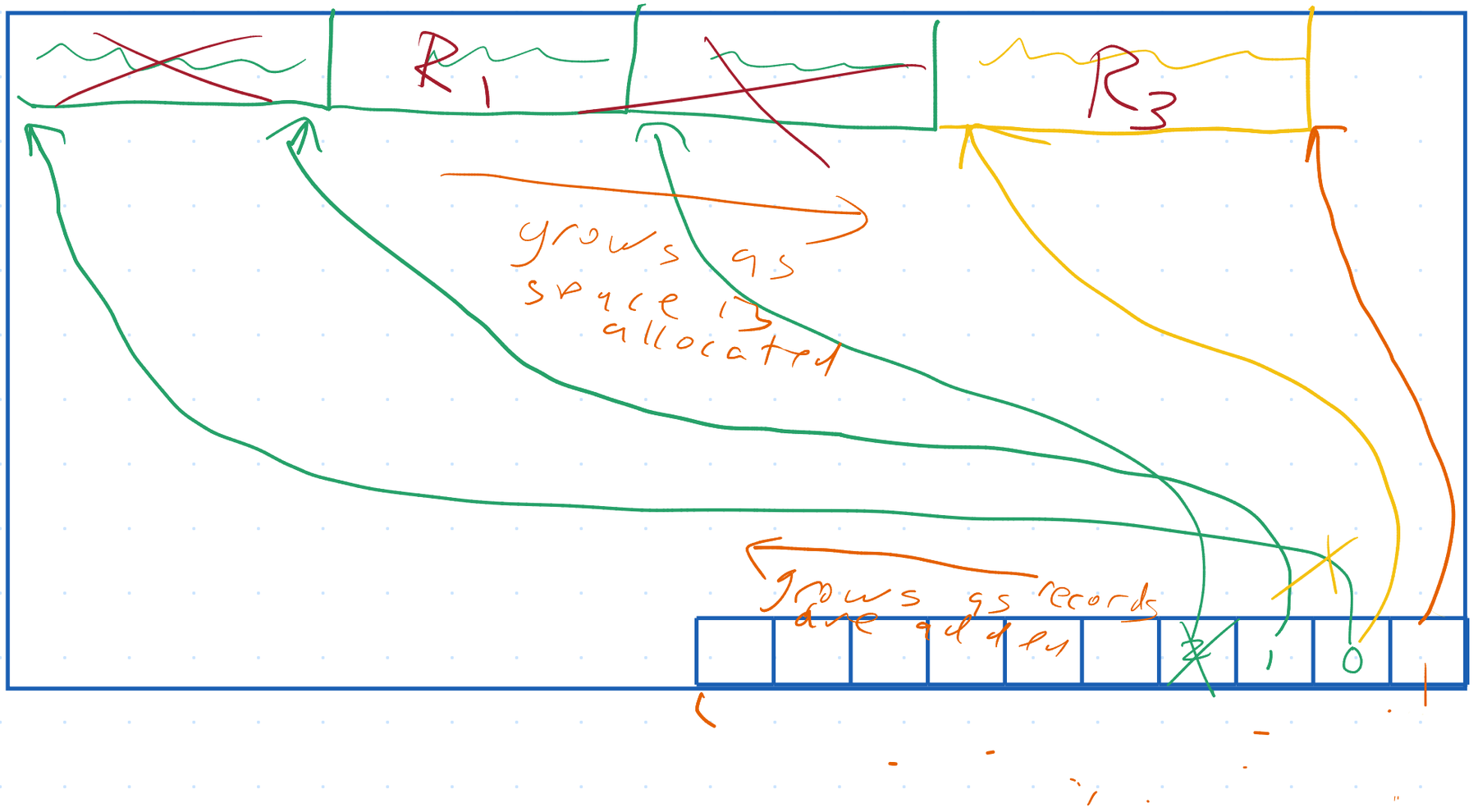
Don't

Variation 3: Footers

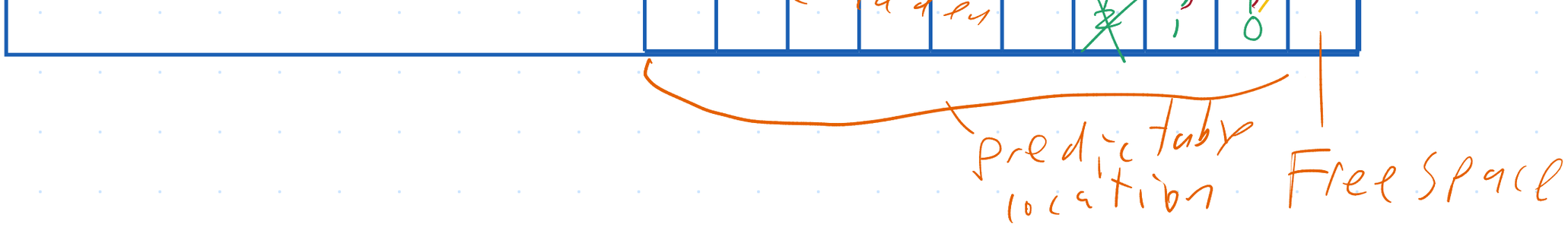


Pros/ - Variable size records
- O(1) record access

Cons/ - Needs periodic compaction

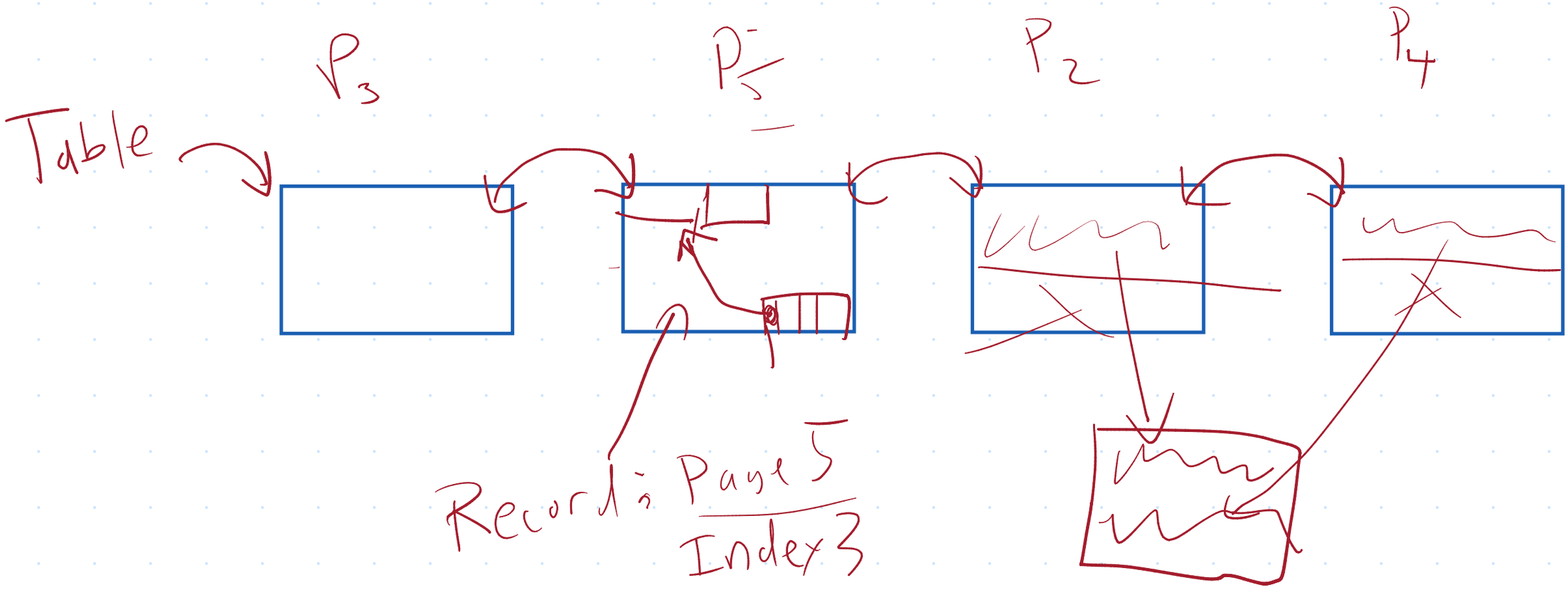

R1
R3
grows as space is allocated
grows as records are added
1
0


R1
R3
grows as records

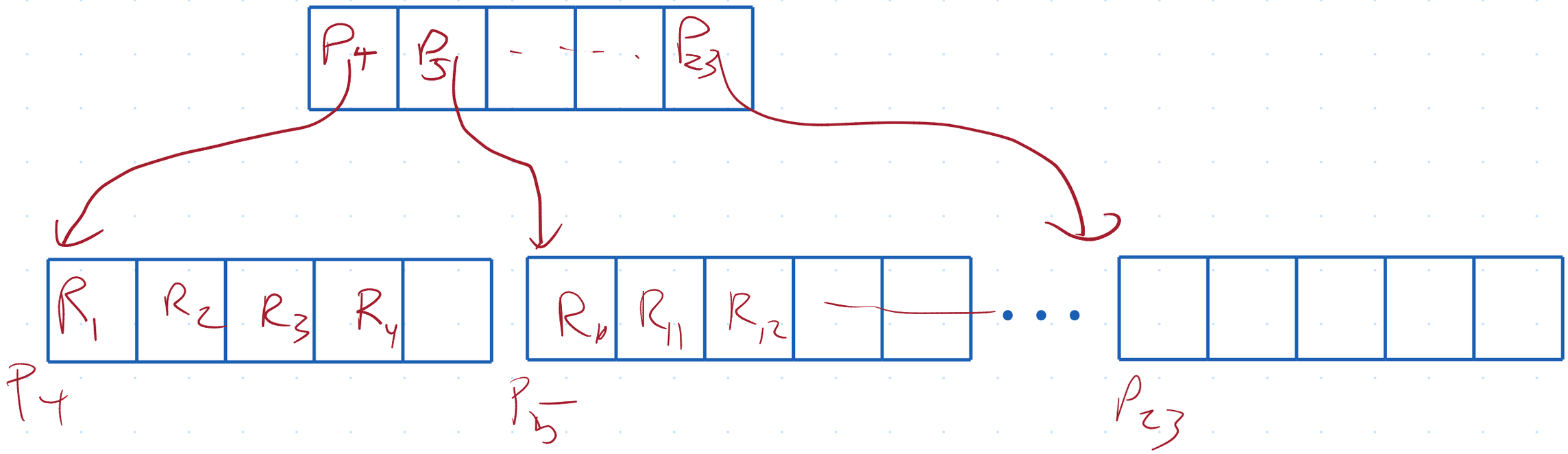

predictable location
Free Space

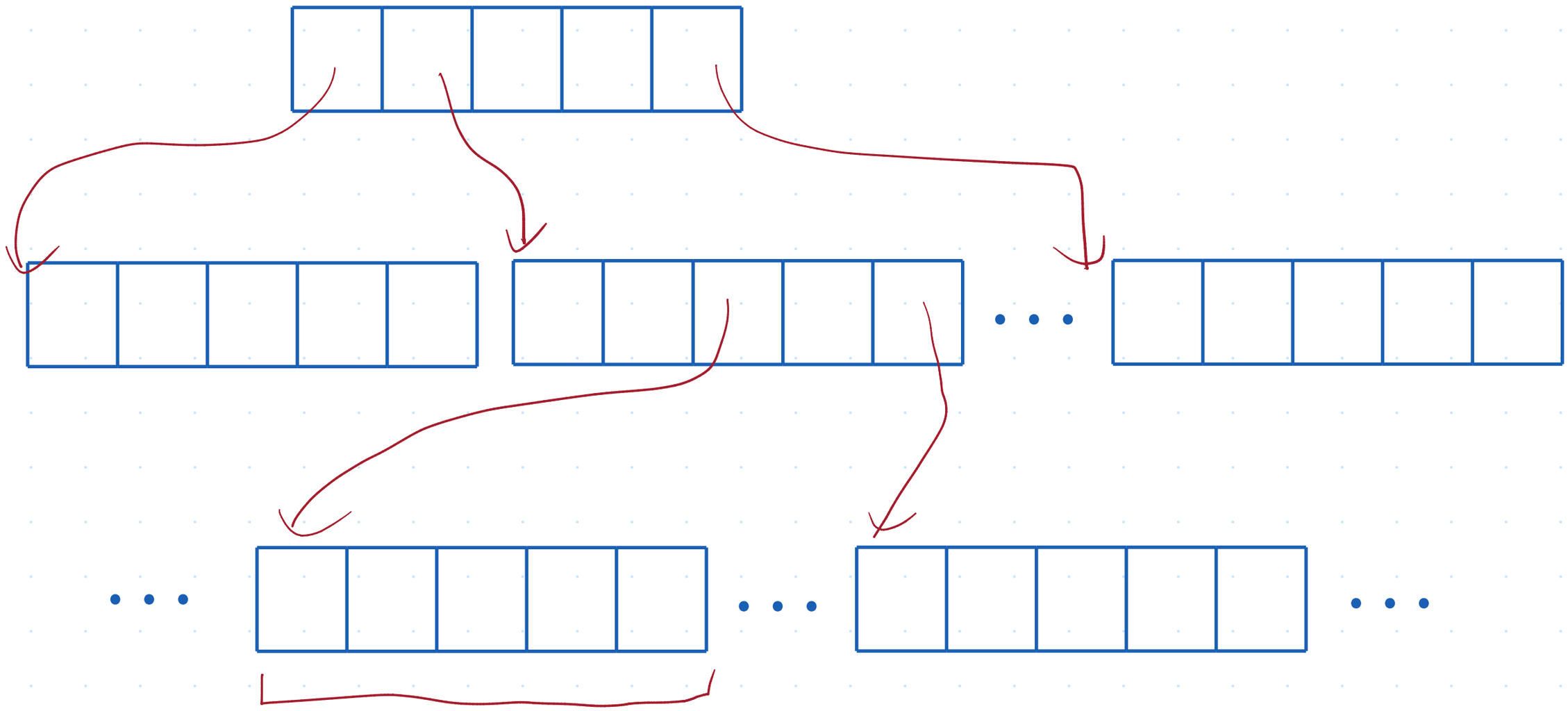
Challenge 3: Page Assignment

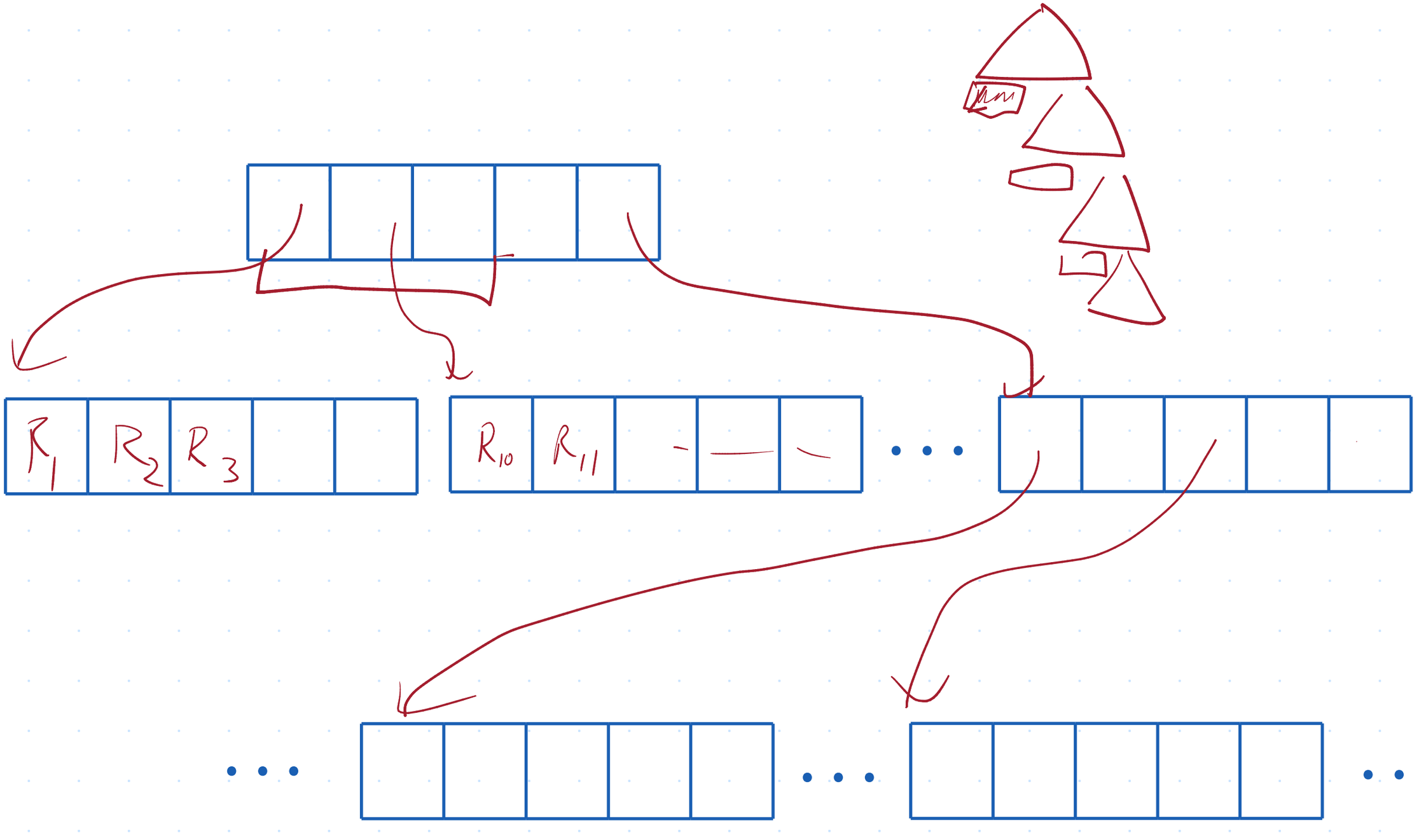
Variation 1: Linked List



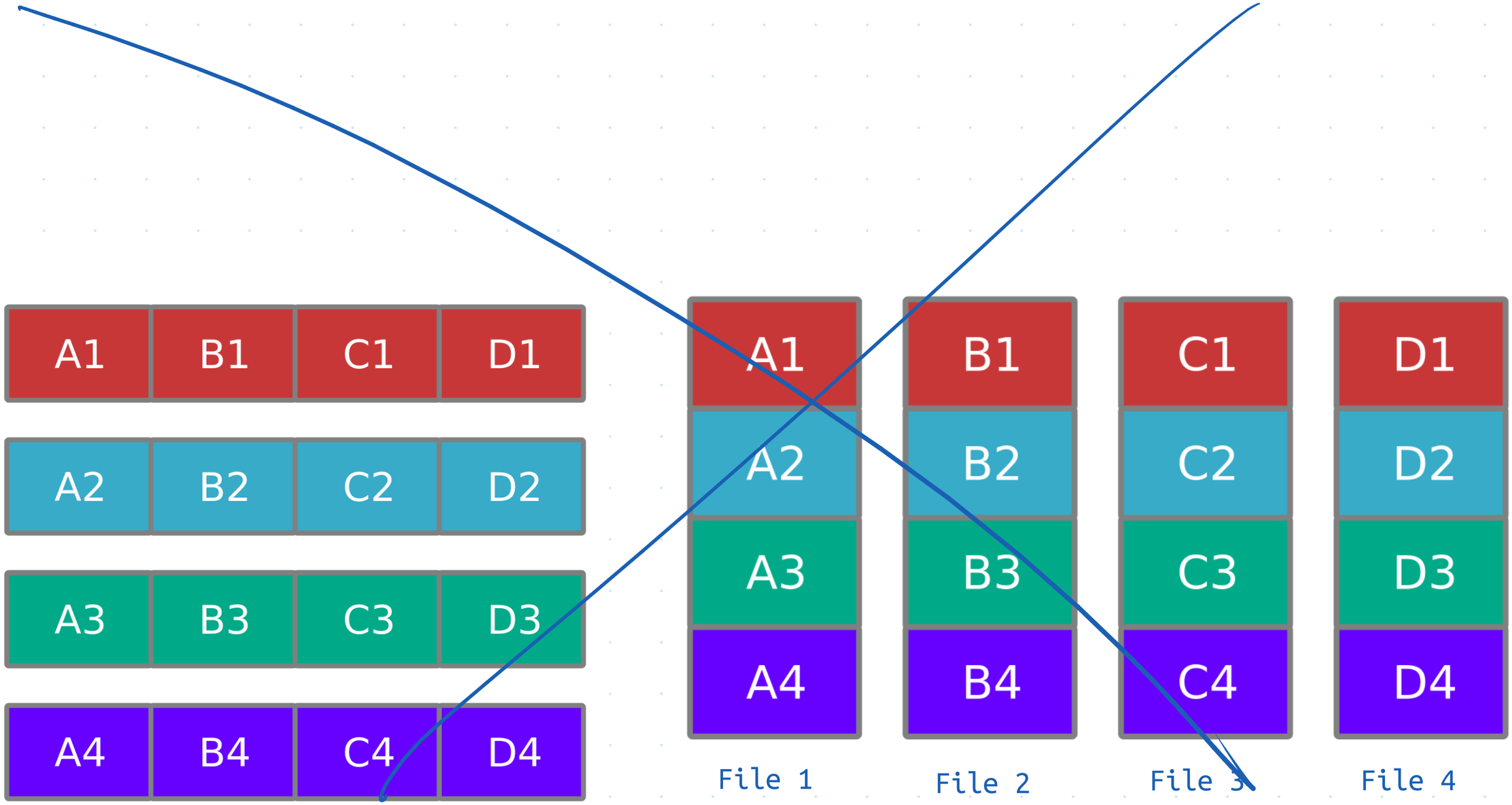
Variation 2: Inode






R1
R2
R3
R10
R11

Columnar Data



File 1

Skipped for now