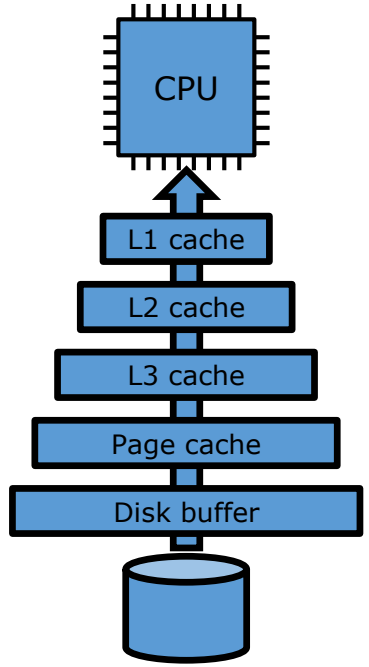
Cost of accessing data (approximations)

- Sequential reads
  - Option to maximize the effective read ratio
    - Depends on DB design
  - Enables pre-fetching

  Cost = seek+rotation+n*transfer
- Random Access
  - Requires indexing structures
  - Ignores data locality

  $Cost_{single\ cylinder\ files}$ = seek+n*(rotation+transfer)

  $Cost_{multi\text{-}cylinder\ files}$ = n*(seek+rotation+transfer)



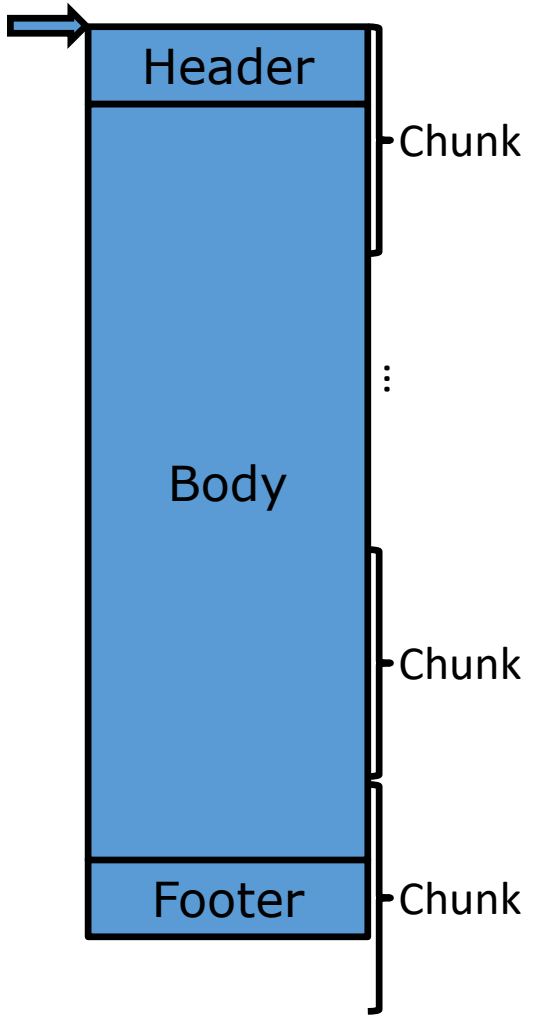
seek	~12ms
rotation	~3ms
transfer (8KB)	~0.03ms

General formulas for size estimation

$$Size(Layout) = Size(Header_{Layout}) + Size(Body_{Layout}) + Size(FooterLayout)$$

$$UsedChunks(Layout) = \frac{Size(Layout)}{Size(chunk)}$$

$$Seeks(Layout) = \lceil UsedChunks(Layout) \rceil$$



"Layout" can be either "Horizontal", "Vertical" or "Hybrid"

Horizontal layout size estimation

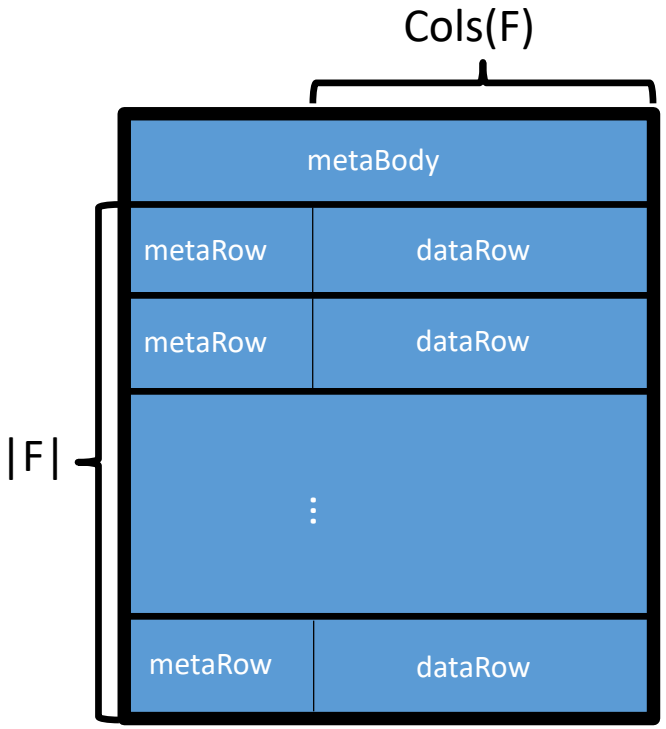


Table 1

A	B	C	D
101	201	301	401
102	202	302	402
103	203	303	403

Cols(F) = 3 → cell

metaBody	
metaRow	101,201,301,401
metaRow	102,202,302,402
metaRow	103,203,303,403

|F| = 3

What is the size of the body in Horizontal layouts?

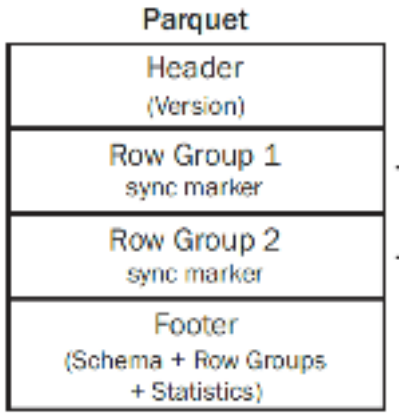
$$Size(dataRow) = Cols(F) * Size(cell)$$

$$Size(Body_{Horizontal}) = Size(metaBody) + |F| * (Size(metaRow) + Size(dataRow))$$

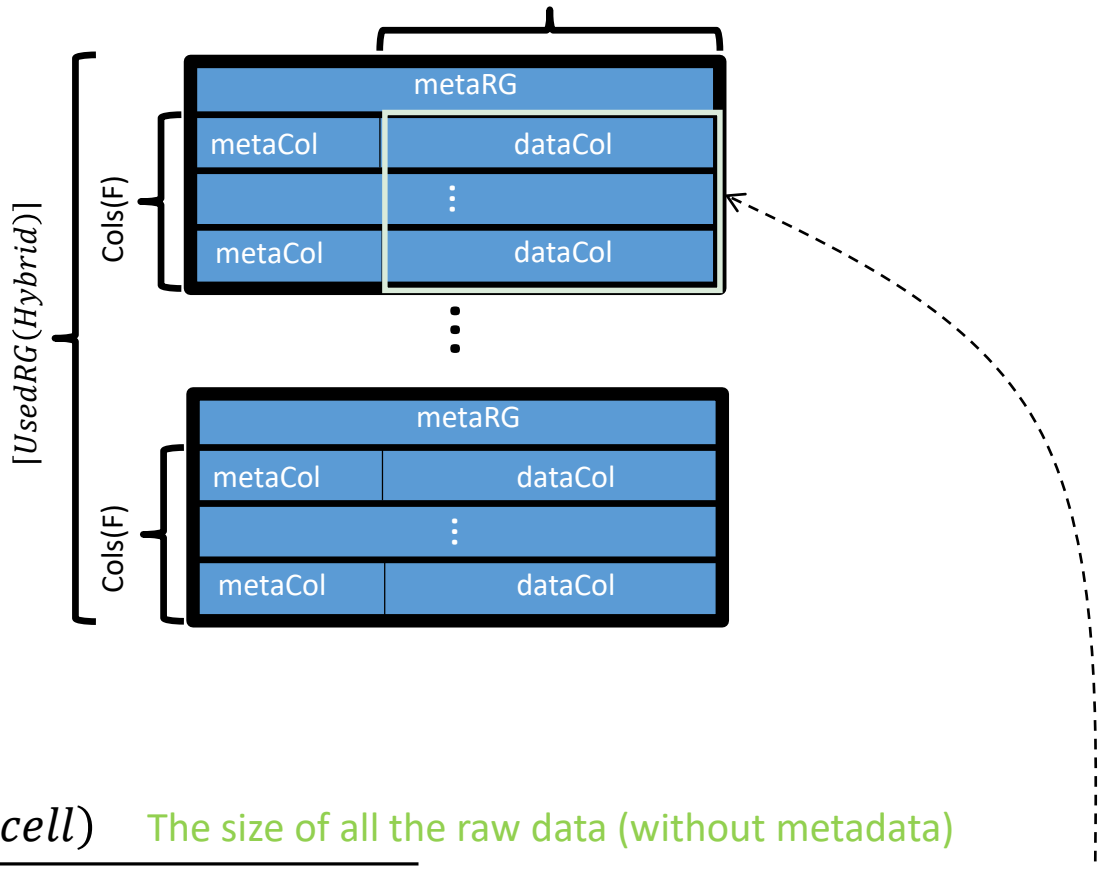
"F" is the content of the file

|F| is the cardinality (number of rows)

Hybrid layout size estimation



$$UsedRows(RG) = \frac{|F|}{UsedRG(Hybrid)}$$



What is the size of the body in Vertical layouts?

$$UsedRG(Hybrid) = \frac{Cols(F) * |F| * Size(cell)}{(Size(RG) - Size(metaRG) - Cols(F) * Size(metaCol))}$$

The size of all the raw data (without metadata)

The size of an RG without metadata

$$Size(Body_{Hybrid}) = \lceil UsedRG(Hybrid) \rceil * (Size(metaRG) + Cols(F) * Size(metaCol)) + Cols(F) * |F| * Size(cell)$$

Multiply each RG with the metadata it stores

Plus the size of the raw data

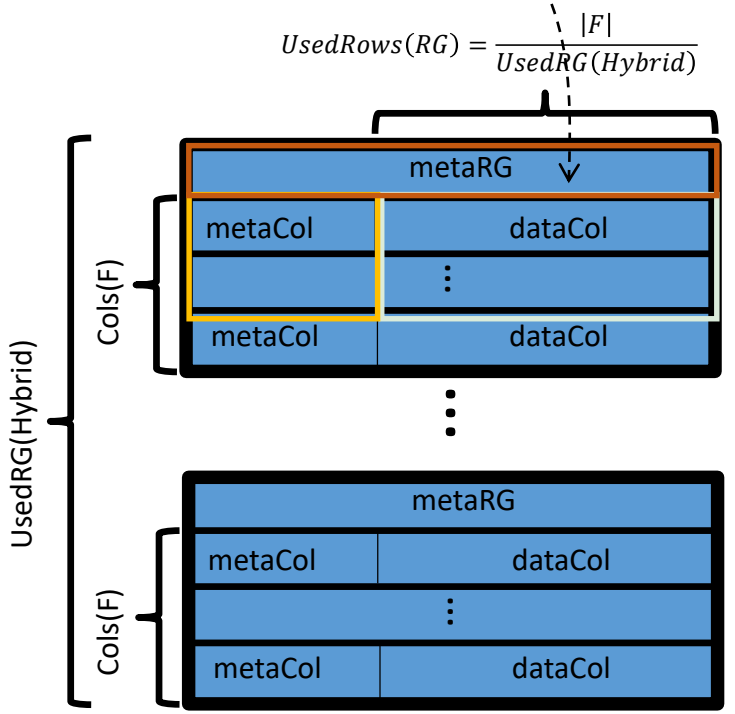
Cost of projection in hybrid layouts

$$Size(projCols) = Proj(F) * UsedRows(RG) * Size(cell)$$

$$Size(Project_{Hybrid}) = Size(Header_{Hybrid}) + Size(Footer_{Hybrid}) + \lceil UsedRG(Hybrid) \rceil * (Size(metaRG) + proj(F) * Size(metaCol)) + UsedRG(Hybrid) * Size(projCols)$$

$$Cost(Project_{Hybrid}) = \text{UsedChunks}(\text{Project}_{\text{Hybrid}}) * W_{\text{ReadTransfer}} + \text{Seeks}(\text{Hybrid}) * (1 - W_{\text{ReadTransfer}})$$

If UsedRows=dataCol and Proj(F) = 2



What about selection?

We have to access the entire RG, even if only some data are actually requested by the user(e.g., there is a filter).

So, we need to compute how many RGs we need to access based on the rows that are defined by the predicate!

"Proj(F)" is the number of projected columns

Selection in hybrid layouts

$$P(RGSelected) = 1 - (1 - SF)^{UsedRows(RG)}$$

$$Size(RowsSelected) = \left\lceil \frac{SF * |F|}{UsedRows(RG)} \right\rceil \left(Size(metaRG) + Cols(F) * Size(metaCol)\right)$$

The size of the metadata depending on SF

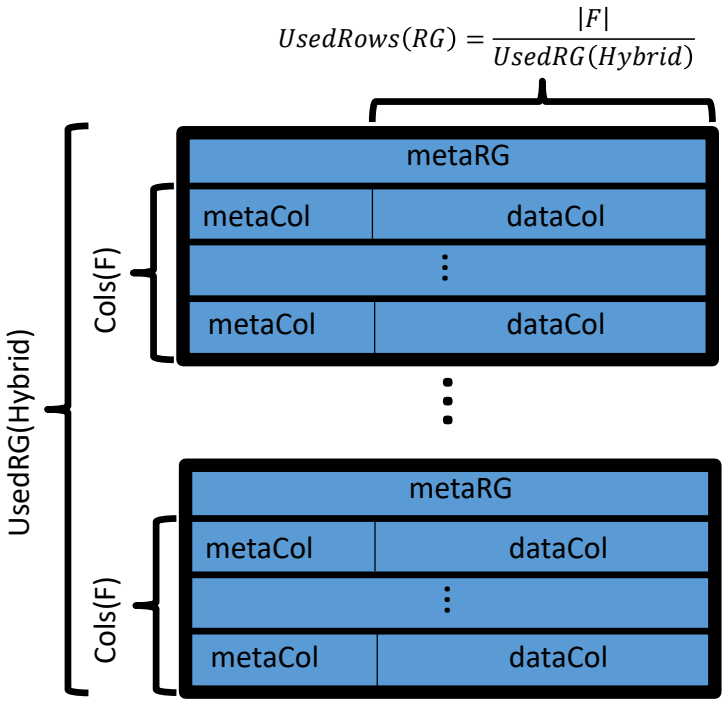
$$+SF * |F| * Cols(F) * Size(cell)$$

The size of the raw data multiplied by SF

$$UsedRG(Select_{Hybrid}) = \begin{cases} if\ unsorted:\ P(RGSelected) * UsedRG(Hybrid) \\ if\ sorted: \left\lceil \frac{Size(RowsSelected)}{Size(RG)} \right\rceil \end{cases}$$

$$Size(Select_{Hybrid}) = Size(Header_{Hybrid}) + Size(Footer_{Hybrid}) + UsedRG(Select_{Hybrid}) * Size(RG)$$

$$Cost(Select_{Hybrid}) = \text{UsedChunks}(\text{Select}_{\text{Hybrid}}) * \text{W}_{\text{ReadTransfer}} + \text{Seeks}(\text{SelectHybrid}) * (1 - \text{WReadTransfer})$$



"F" is the content of the file

Client caching

Cash miss

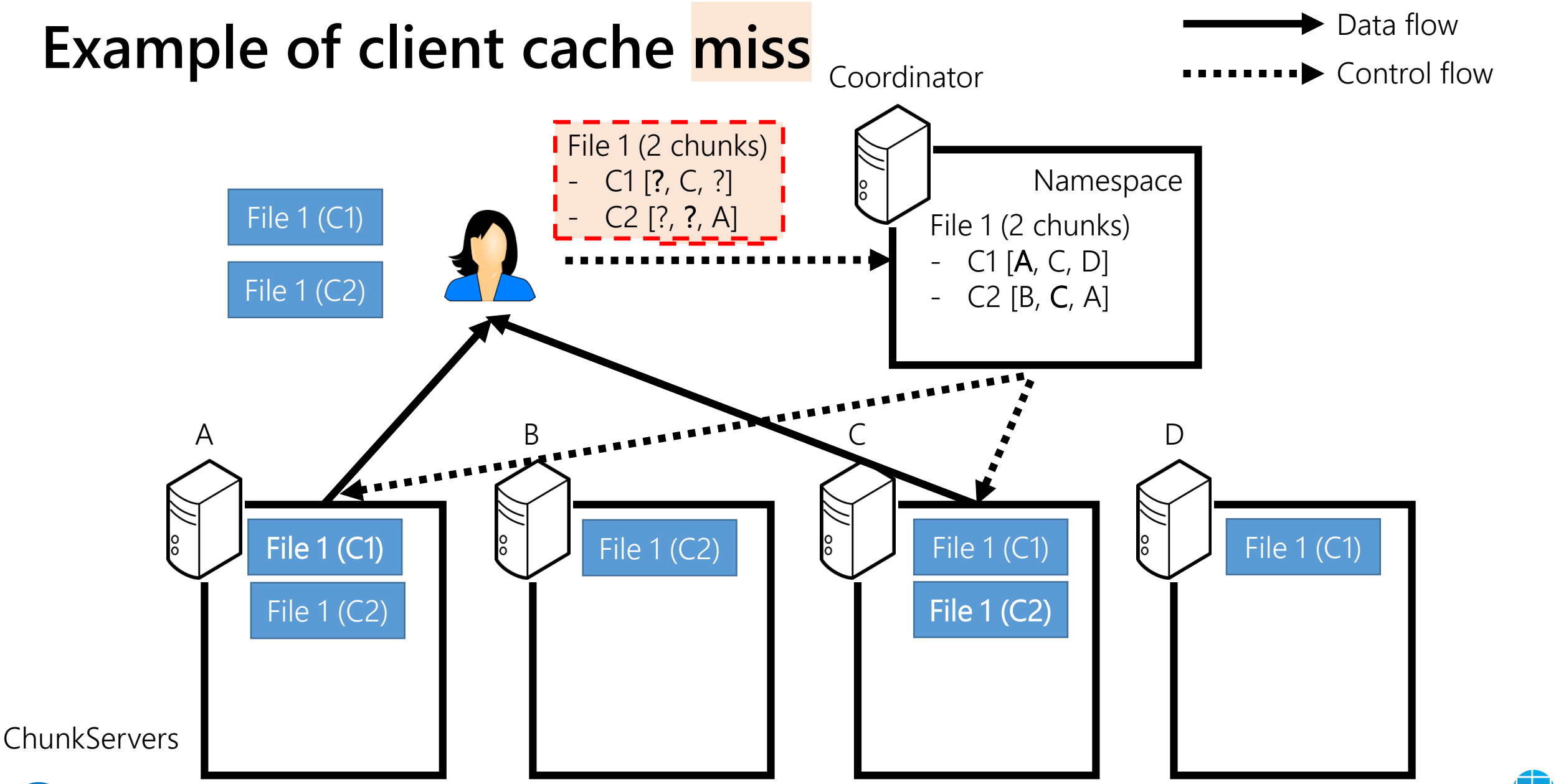
1. The client sends a READ command to the coordinator
2. The coordinator requests chunkservers to send the chunks to the client
   - Ranked according to the closeness in the network
3. The list of locations is cached in the client
   - Not a complete view of all chunks

Cash hit

1. The client reads the cache and requests the chunkservers to send the chunks

Avoid coordinator bottleneck
+
One communication step is saved

Example of client cache miss



Example of client cache hit