

Yandex

Binary formats 2

RCFile, Parquet

Row-based & column-based formats

Row-based & column-based formats

Table representation

X	Y	Z
x1	y1	z1
x2	y2	z2
x3	y3	z3
x4	y4	z4
x5	y5	z5

Row-based & column-based formats

Table representation

x	y	z
x1	y1	z1
x2	y2	z2
x3	y3	z3
x4	y4	z4
x5	y5	z5

Row format

x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4	z4	x5	y5	z5
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Row-based & column-based formats

Table representation

X	Y	Z
x1	y1	z1
x2	y2	z2
x3	y3	z3
x4	y4	z4
x5	y5	z5

Row format

x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4	z4	x5	y5	z5
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Columnar format

x1	x2	x3	x4	x5	y1	y2	y3	y4	y5	z1	z2	z3	z4	z5
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Row-based & column-based formats

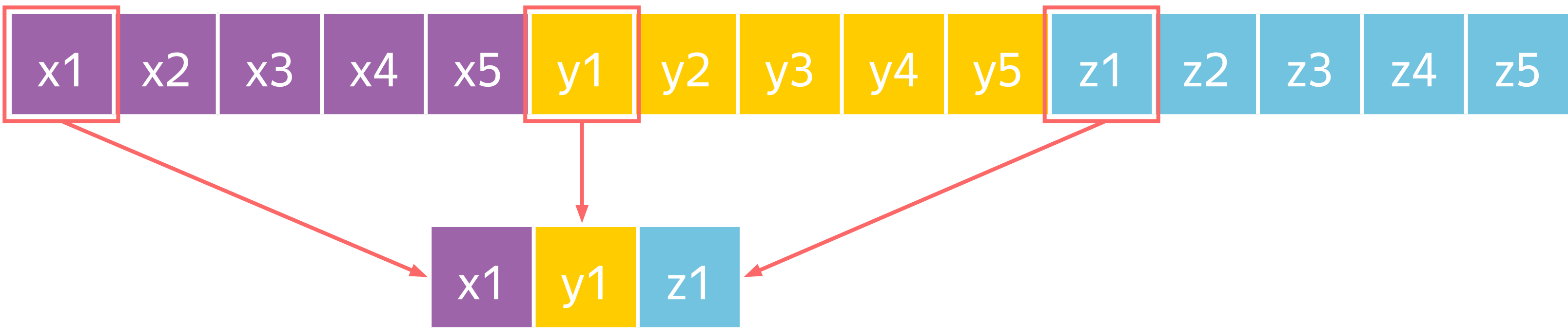
Table representation

X	Y	Z
x1	y1	z1
x2	y2	z2
x3	y3	z3
x4	y4	z4
x5	y5	z5

Row format



Columnar format

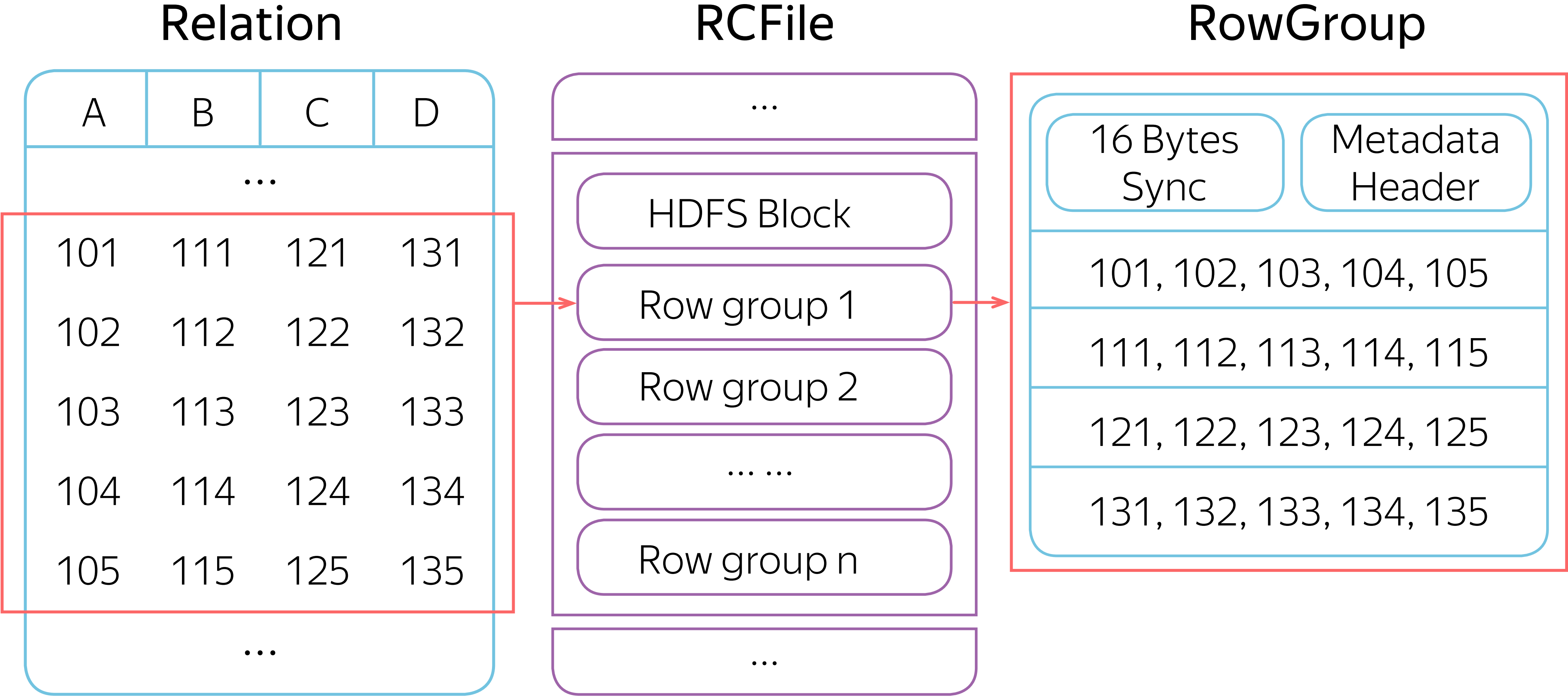


RCFile

- › First columnar format* in Hadoop()
- › Horizontal/vertical partitioning
 - › split rows into row groups
 - › transpose values within a row group

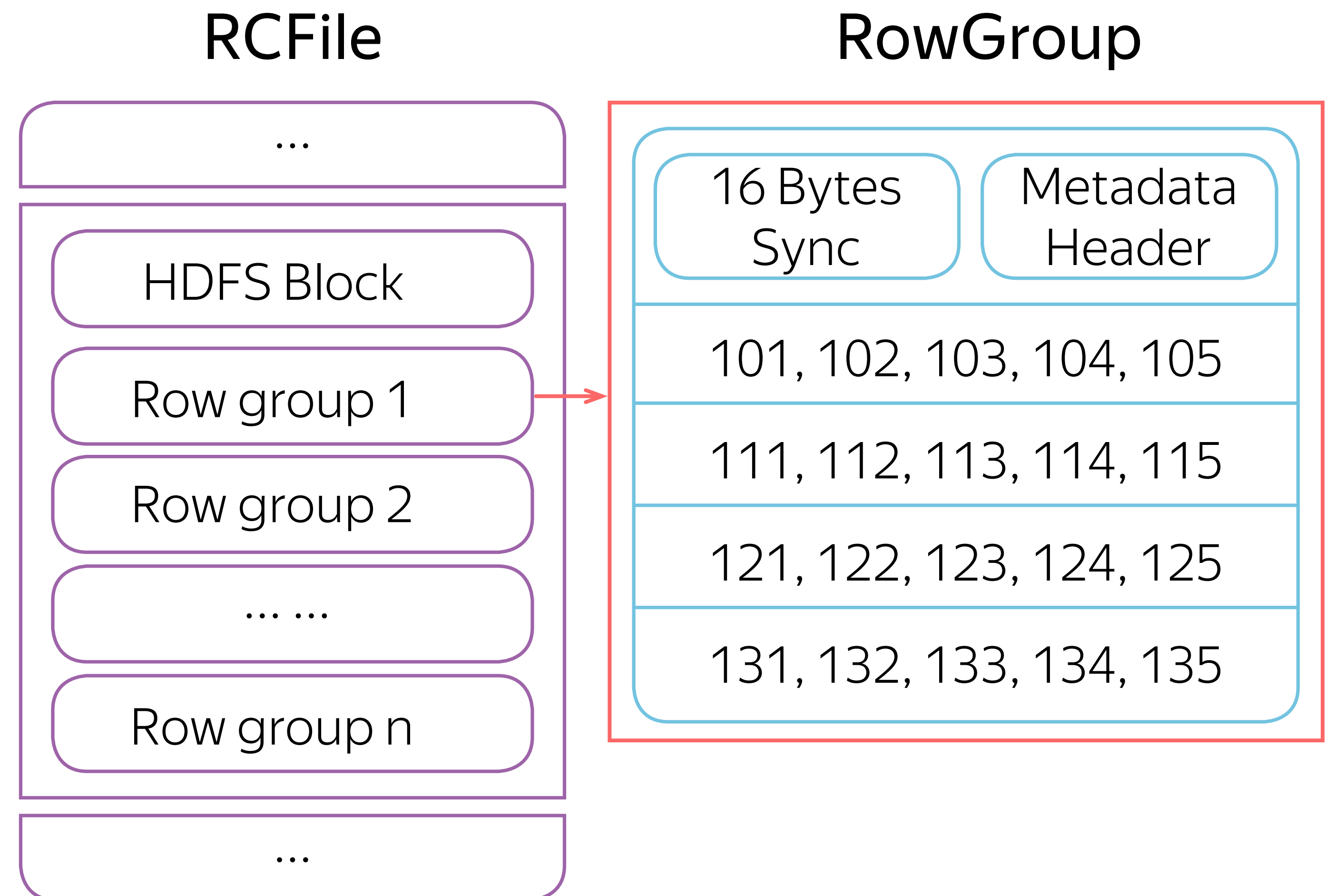
* "RCFile: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems", by He et. al; Facebook, Ohio State University; Chinese Academy of Sciences

RCFile



RCFile

- › Space efficiency
- › Speed
- › Data types
- › Splittable
- › Extensibility



RCFile

› Space efficiency

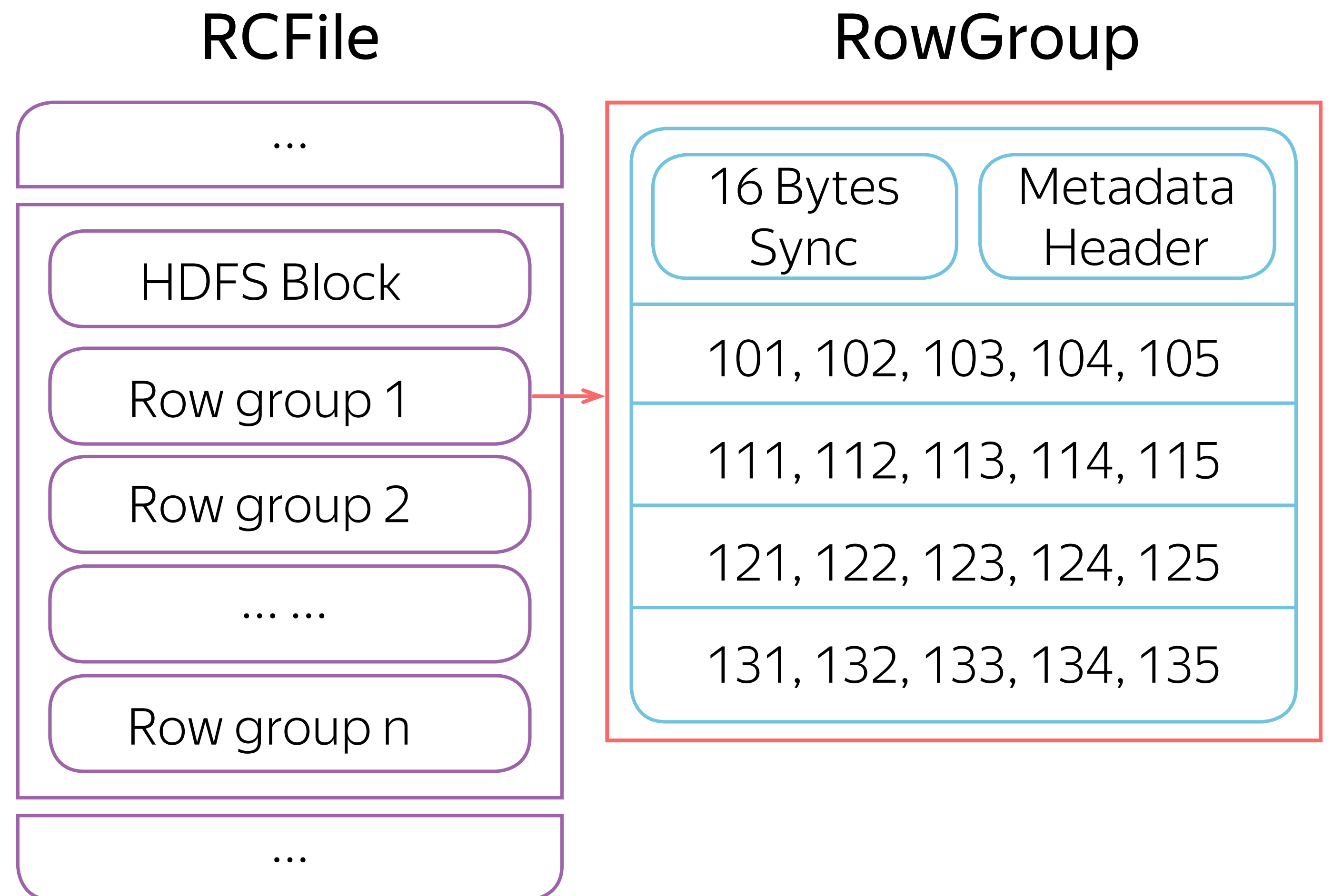
GOOD

› Speed

› Data types

› Splittable

› Extensibility



RCFile

- › Space efficiency

GOOD

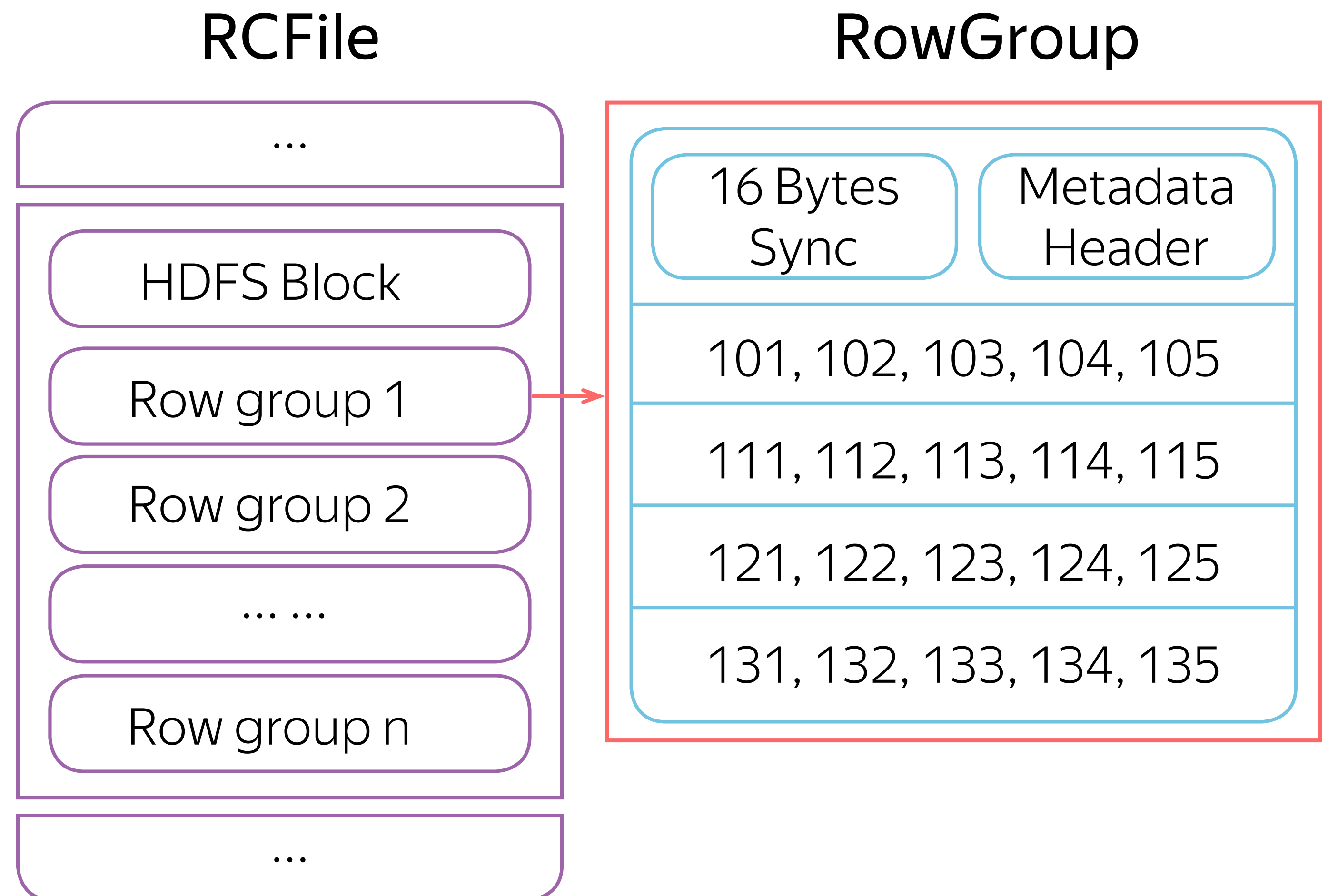
- › Speed

MODERATE TO GOOD, LESS I/O

- › Data types

- › Splittable

- › Extensibility



RCFile

- › Space efficiency

GOOD

- › Speed

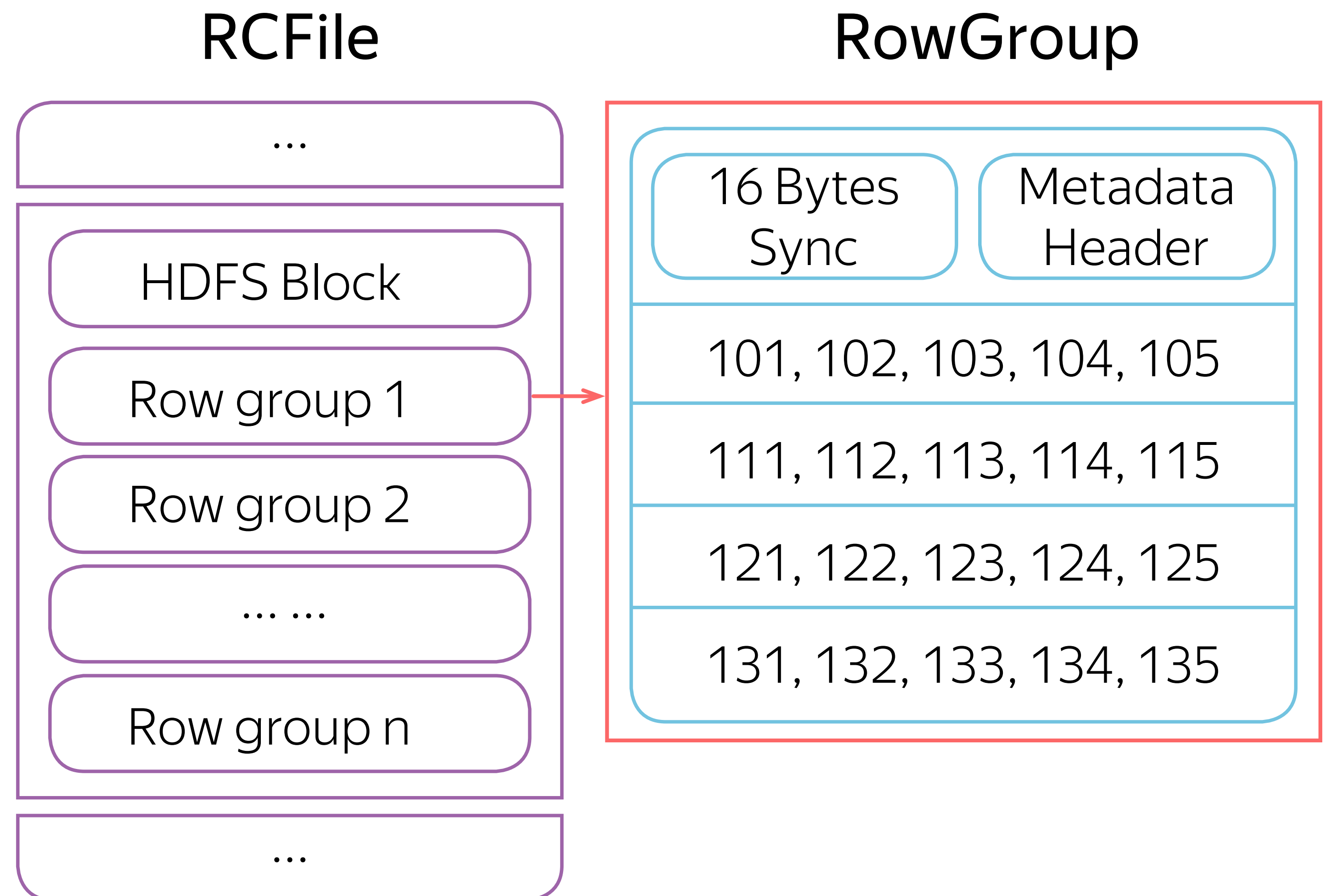
MODERATE TO GOOD, LESS I/O

- › Data types

BYTE STRINGS

- › Splittable

- › Extensibility



RCFile

- › Space efficiency

GOOD

- › Speed

MODERATE TO GOOD, LESS I/O

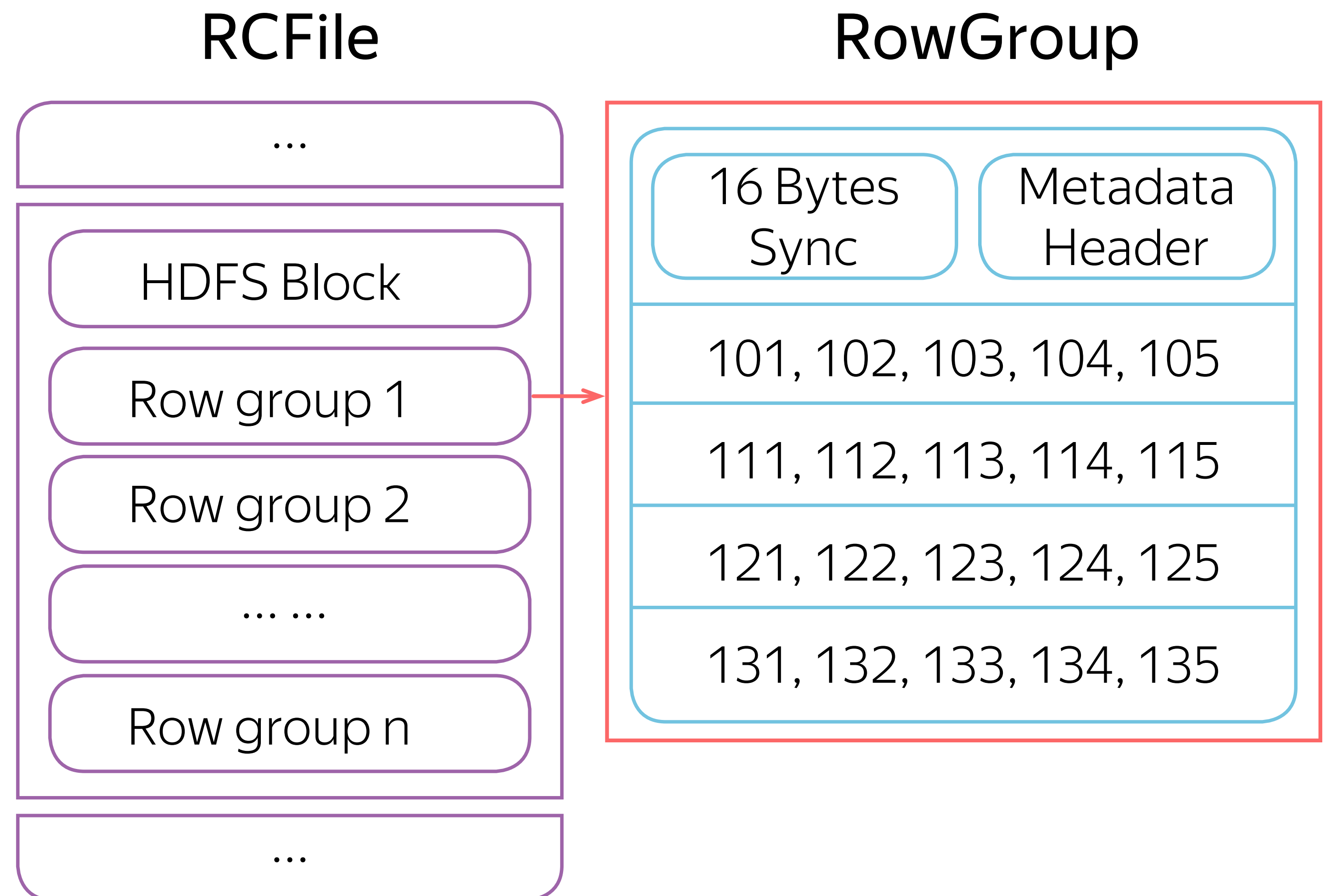
- › Data types

BYTE STRINGS

- › Splittable

SPLITTABLE

- › Extensibility



RCFile

- › Space efficiency

GOOD

- › Speed

MODERATE TO GOOD, LESS I/O

- › Data types

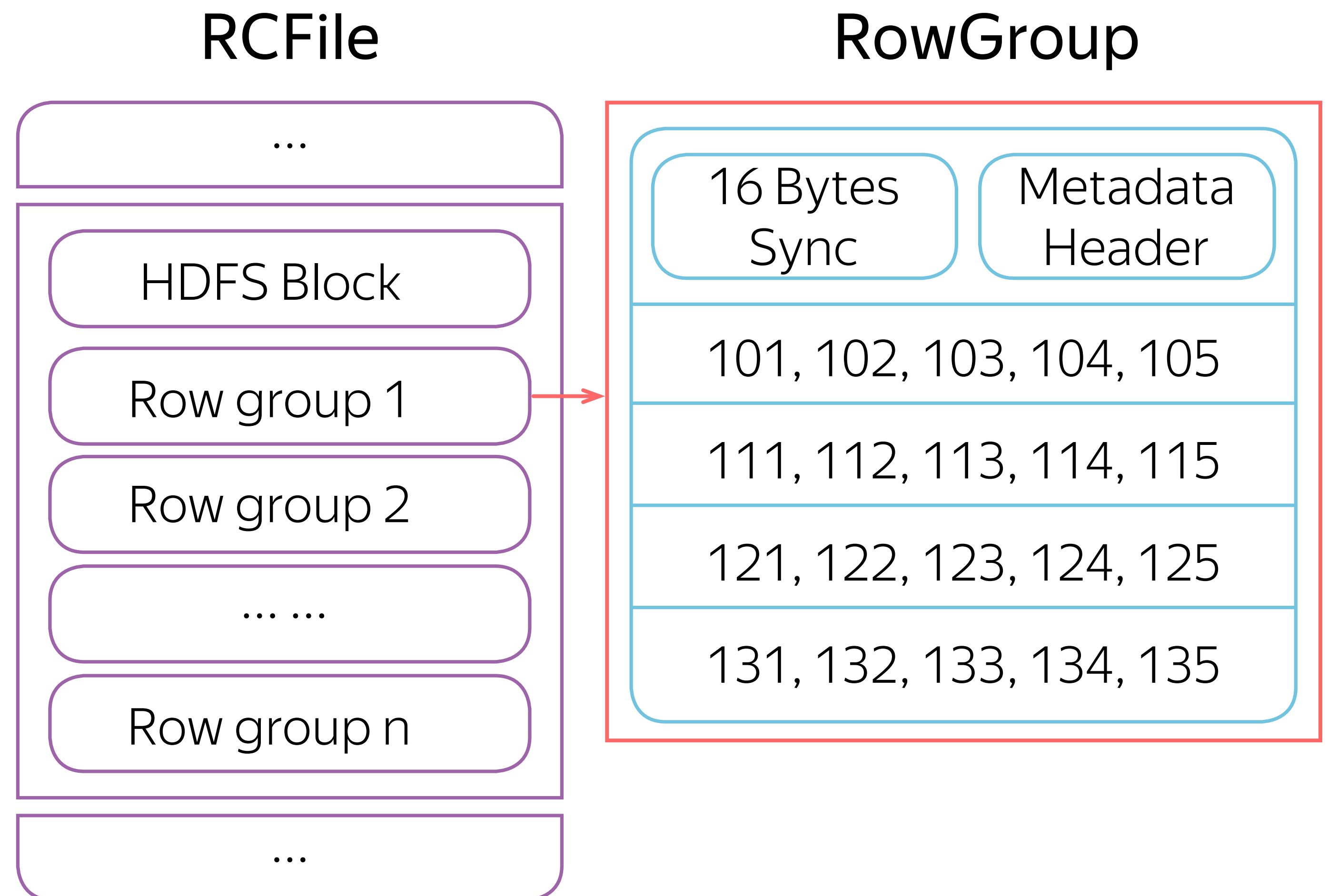
BYTE STRINGS

- › Splittable

SPLITTABLE

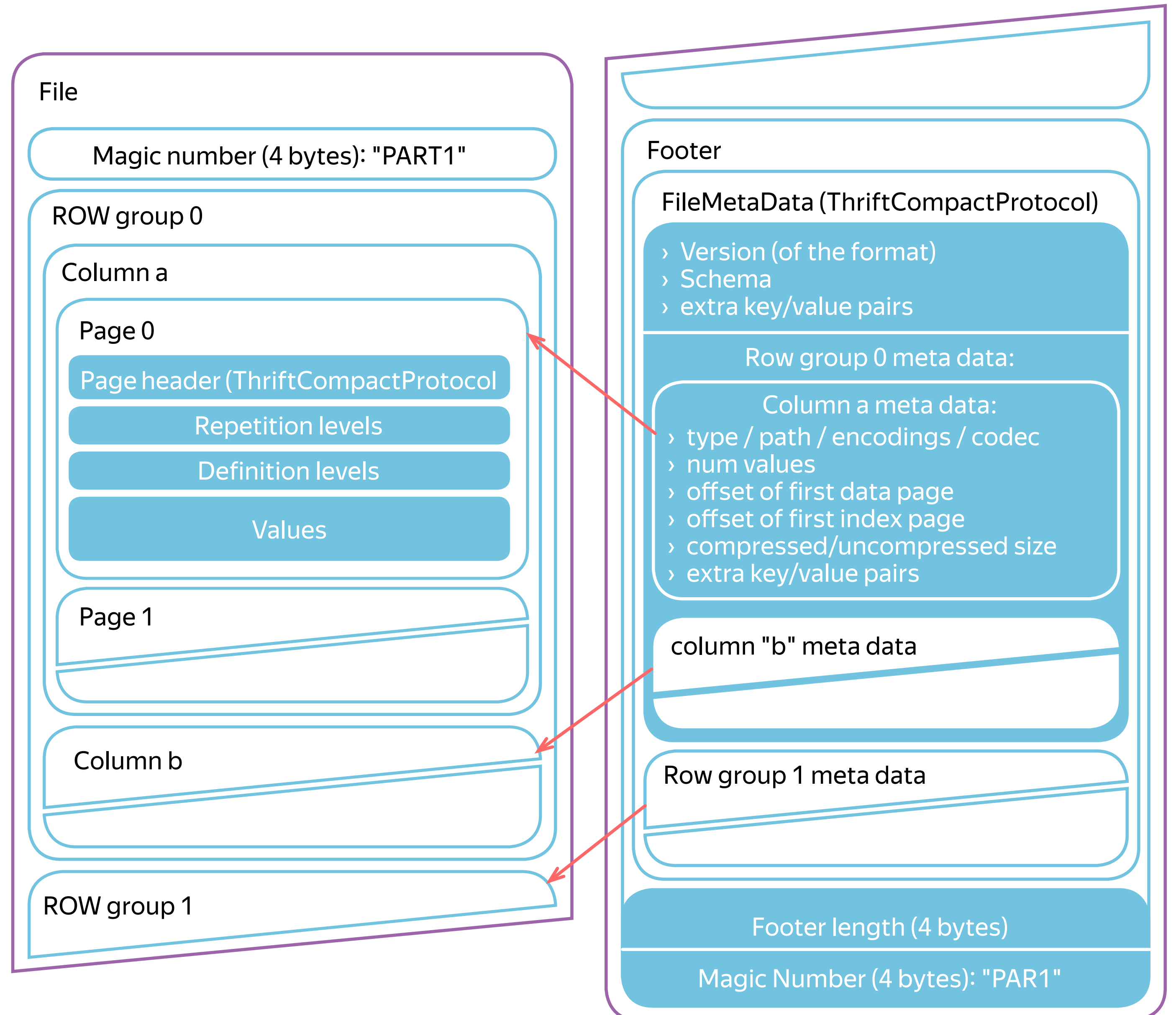
- › Extensibility

NO



Parquet

- › The most sophisticated columnar format in Hadoop
- › Collaborative effort by Twitter & Cloudera
- › Supports nested and repeated data
- › Exploits many columnar optimizations (such as predicate pruning, per column codecs)
- › Optimizes write path



Conclusion

- › Binary formats are efficient in coding data
 - › SequenceFile is a reasonable choice for Java users
 - › Avro is a good alternative for many use cases
 - › RCFile/ORC/Parquet are best for “wide” tables and analytical workloads
- › Next video
 - › Compression

BigDATAteam