Advanced Data Management for Data Analysis

Stefan Manegold

Data Management @ LIACS

Group leader Database Architectures
Centrum Wiskunde & Informatica (CWI)
Amsterdam

s.manegold@liacs.leidenuniv.nl
http://www.cwi.nl/~manegold/

ADM: Agenda

- 07.09.2022: Lecture 1: **Introduction**
- 14.09.2022: Lecture 2: **SQL Recap**

  (plus Assignment 1 [in groups; 3 weeks]: TPC-H benchmark)
- 21.09.2022: Lecture 3: **Column-Oriented Database Systems (1/6) - Motivation & Basic Concepts**
- 28.09.2022: Lecture 4: **Column-Oriented Database Systems (2a/6) - Selected Execution Techniques (1/2)**
- 05.10.2022: Lecture 5: **Column-Oriented Database Systems (2b/6) - Selected Execution Techniques (2/2)**

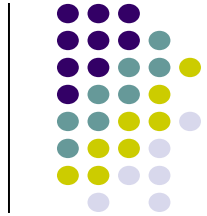
  (plus Assignment 2 [in groups; 3 weeks]: Compression techniques)
- 12.10.2022: Lecture 6: **Column-Oriented Database Systems (3/6) - Cache Conscious Joins**
- 19.10.2022: Lecture 7: **Column-Oriented Database Systems (4/6) - "Vectorized Execution"**
- ~~26.10.2022~~: ***No lecture!***
- 02.11.2022: Lecture 8: **DuckDB: An embedded database for data science (1/2) (guest lecture & *hands-on*)**

  (plus Assignment 3 [individual; 2 weeks]: Analysing NYC Cab dataset with DuckDB)
- 09.11.2022: Lecture 9: **DuckDB: An embedded database for data science (2/2) (guest lecture & *hands-on*)**
- 16.11.2022: Lecture 10: **Branch Misprediction & Predication**

  (plus Assignment 4 [individual; 2 weeks]: Predication)
- 23.11.2022: Lecture 11: **Column-Oriented Database Systems (5/6) - Adaptive Indexing**
- 30.11.2022: Lecture 12: **Column-Oriented Database Systems (6/6) - Progressive Indexing**

ADM: Literature

- **Column-Oriented Database Systems (2/6) - Selected Execution Techniques**
  - Compression
    - "Compressing Relations and Indexes". Goldstein, Ramakrishnan, Shaft. ICDE'98.
    - "Query optimization in compressed database systems". Chen, Gehrke, Korn. SIGMOD'01.
    - "Super-Scalar RAM-CPU Cache Compression". Zukowski, Heman, Nes, Boncz. ICDE'06.
    - "Integrating Compression and Execution in Column-Oriented Database Systems". Abadi, Madden, Ferreira. SIGMOD'06.
    - "Improved Word-Aligned Binary Compression for Text Indexing". Ahn, Moffat. TKDE'06.
  - Tuple Materialization
    - "Materialization Strategies in a Column-Oriented DBMS". Abadi, Myers, DeWitt, Madden. ICDE'07.
    - "Column-Stores vs Row-Stores: How Different are They Really?". Abadi, Madden, Hachem. SIGMOD'08.
    - "Query Processing Techniques for Solid State Drives". Tsirogiannis, Harizopoulos Shah, Wiener, Graefe. SIGMOD'09.
    - "Self-organizing tuple reconstruction in column-stores". Idreos, Manegold, Kersten. SIGMOD'09.
  - Join
    - "Fast Joins using Join Indices". Li and Ross. VLDBJ 8:1-24, 1999.



Compression

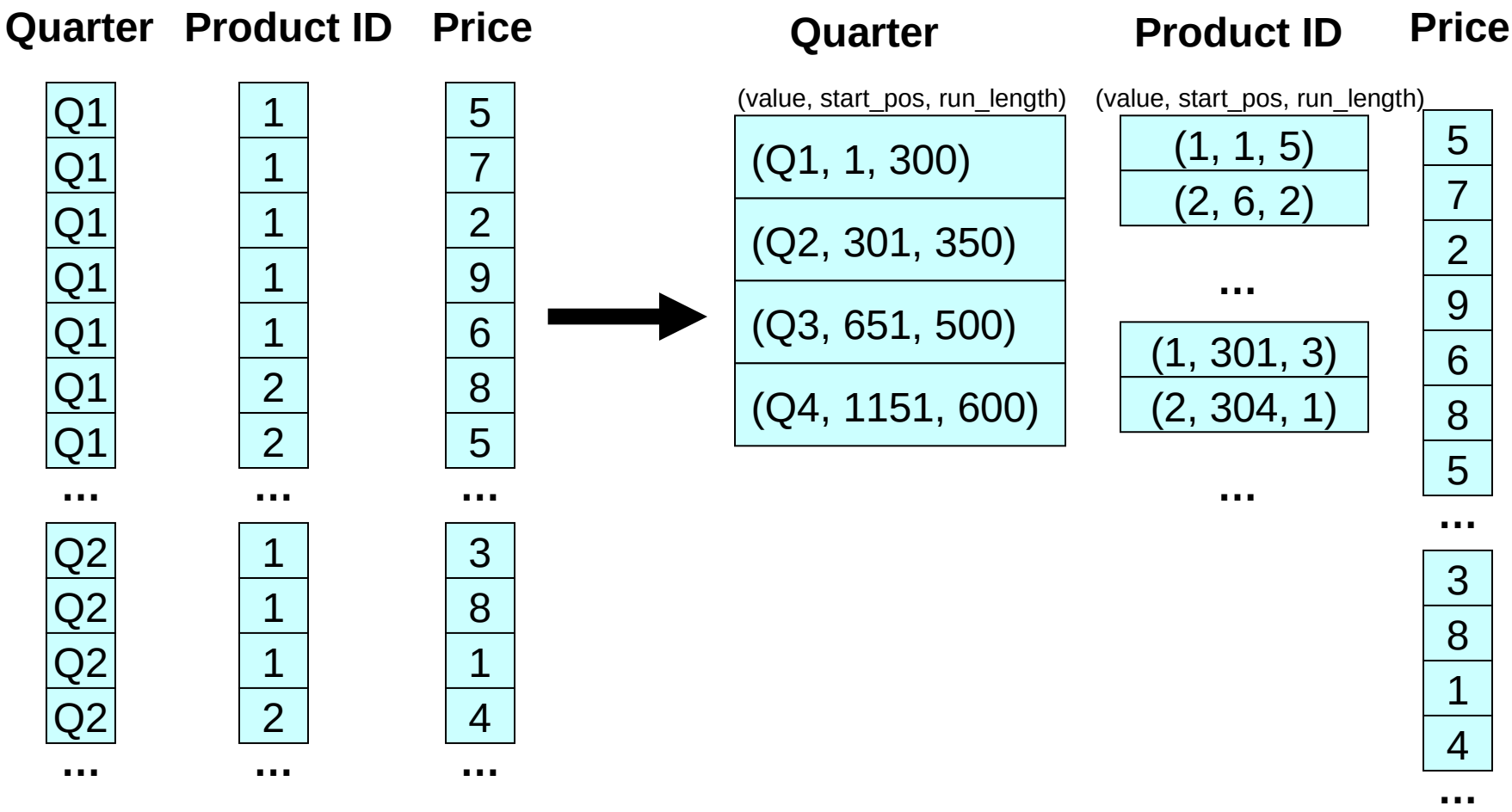
- Trades I/O for CPU
  - Smaller data size on storage

    => less data to read (or write)
  - Extra computation (CPU work) required to (de-)compress

- Increased column-store opportunities:
  - Higher data value locality in column stores
  - Techniques such as run length encoding far more useful
  - Can use extra space to store multiple copies of data in different sort orders

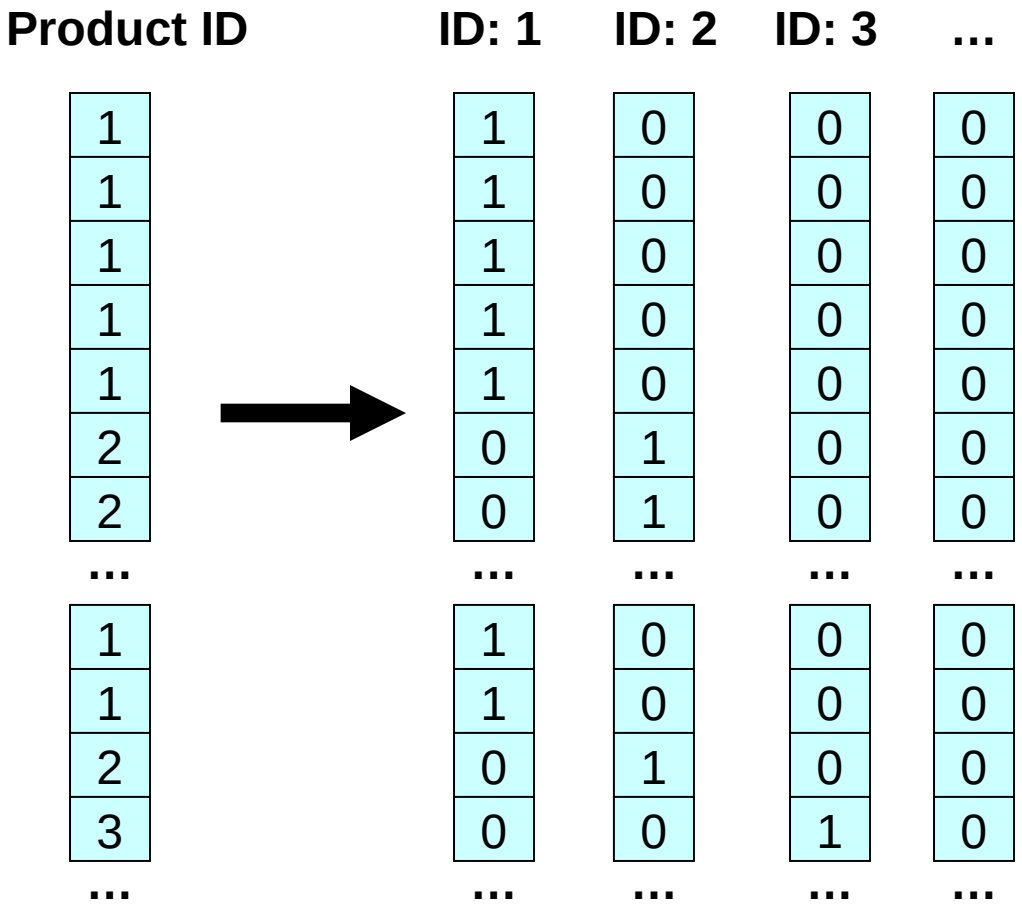
- Sample “light-weight” encoding techniques:
  - Run-length, bit-vector, dictionary, frame-of-reference, differential

Run-length Encoding (RLE)



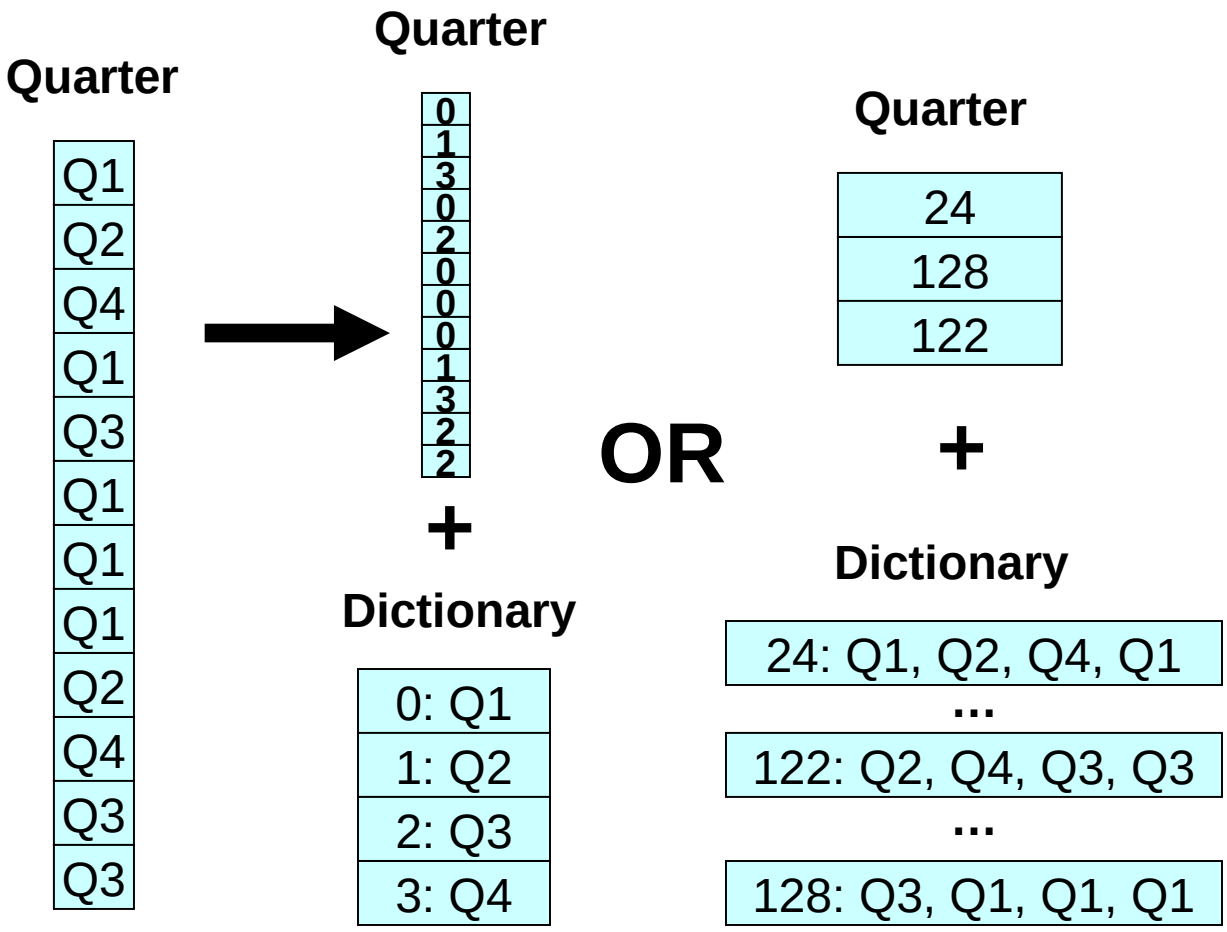
Bit-vector Encoding

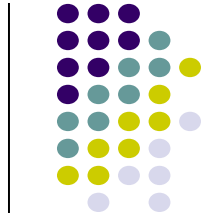
- For each unique value, v, in column c, create bit-vector b_v
  - $b_v[i] = 1$ if $c[i] = v$
- Good for columns with few unique values
- Each bit-vector can be further compressed if sparse



Dictionary Encoding (DIC)

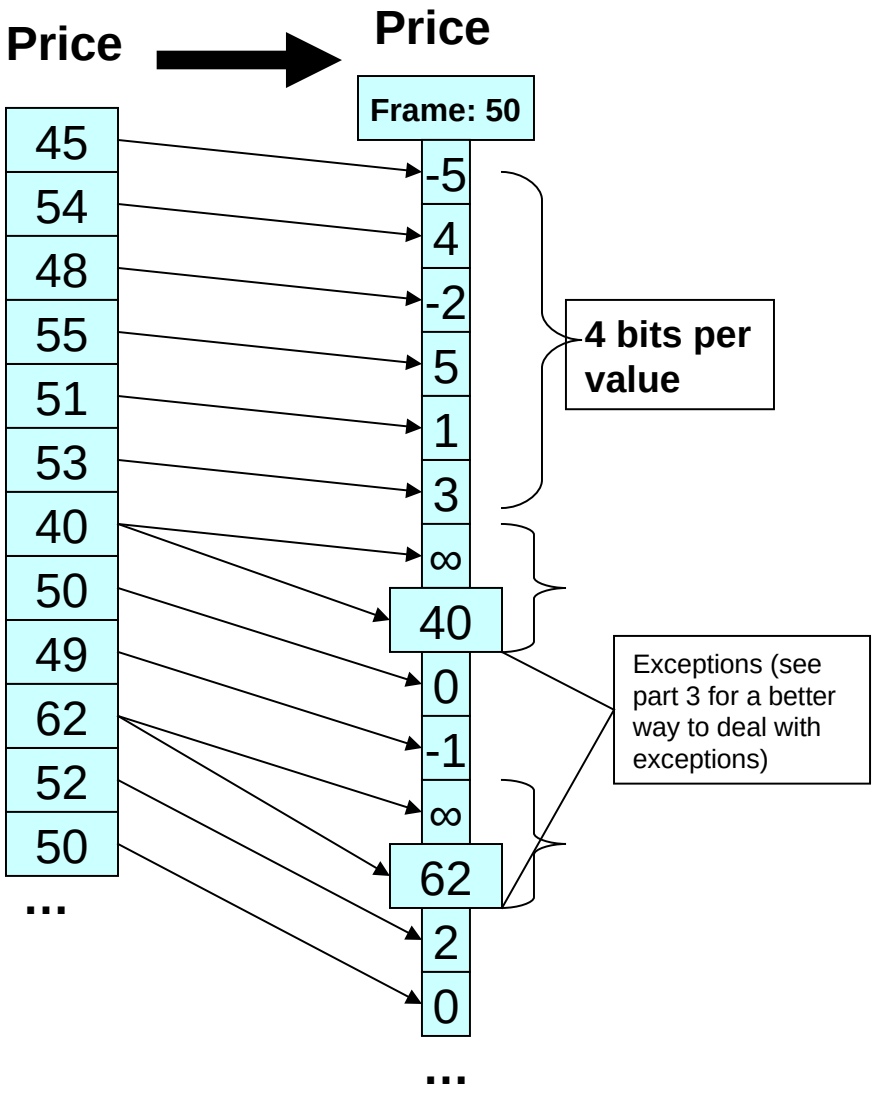
- For each unique value create dictionary entry
- Dictionary can be per-block or per-column
- Column-stores have the advantage that dictionary entries may encode multiple values at once





Frame Of Reference Encoding (FOR)

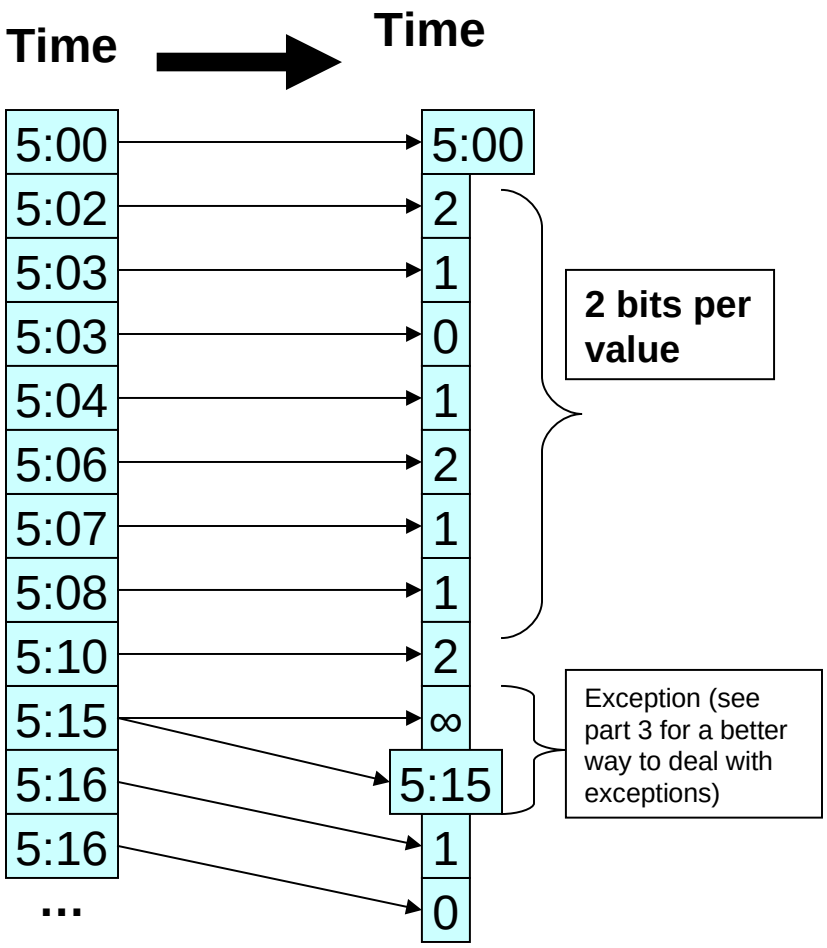
- **Encodes values as b bit offset from chosen frame of reference**
- **Special escape code (e.g. all bits set to 1) indicates a difference larger than can be stored in b bits**
  - **After escape code, original (uncompressed) value is written**



"Compressing Relations and Indexes" Goldstein, Ramakrishnan, Shaft, ICDE'98

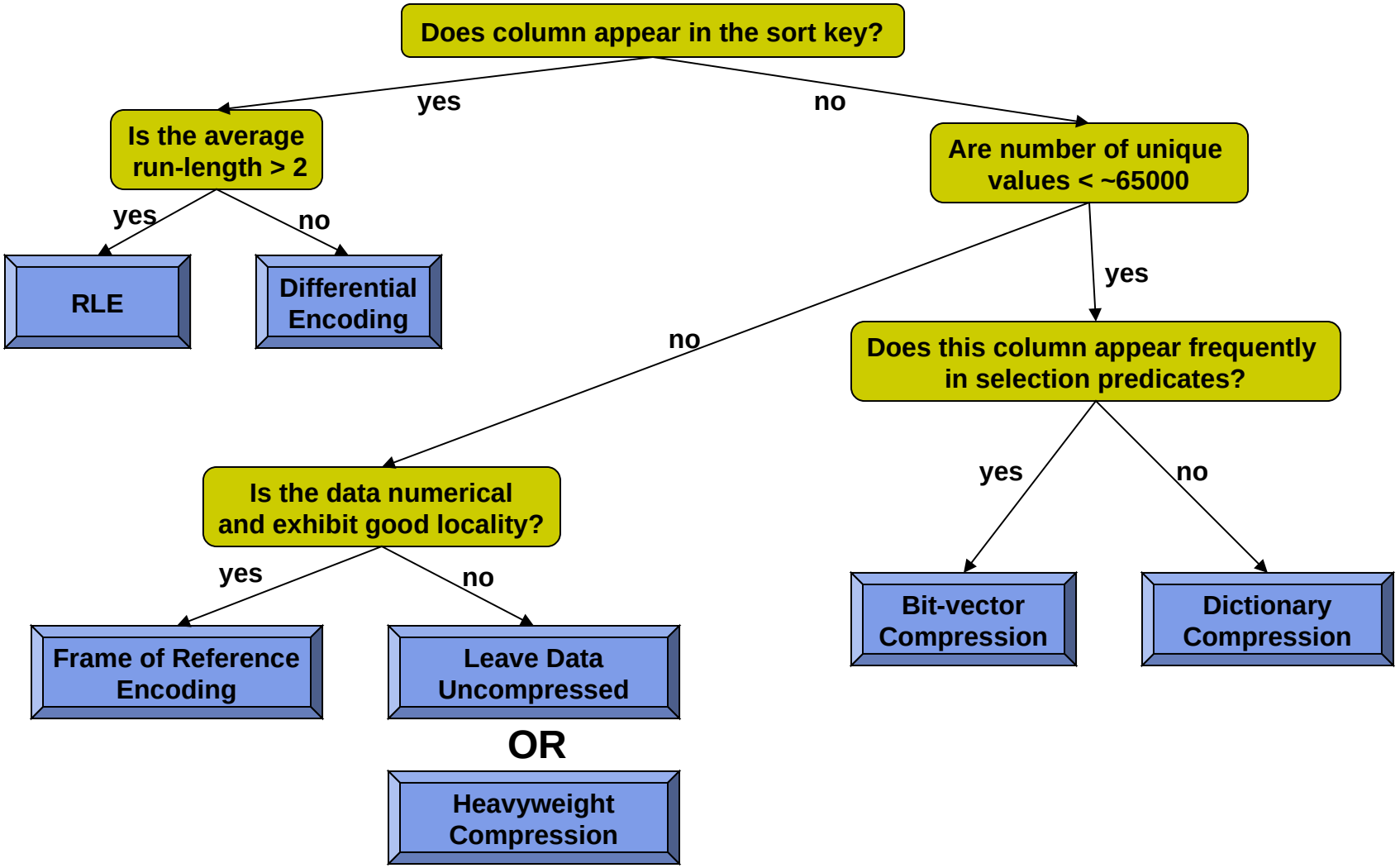
Differential Encoding (DIF)

- **Encodes values as b bit offset from previous value**
- **Special escape code (just like frame of reference encoding) indicates a difference larger than can be stored in b bits**
  - **After escape code, original (uncompressed) value is written**
- **Performs well on columns containing increasing/decreasing sequences**
  - **inverted lists**
  - **timestamps**
  - **object IDs**
  - **sorted / clustered columns**



"Improved Word-Aligned Binary Compression for Text Indexing" Ahn, Moffat, TKDE'06

What Compression Scheme To Use?



"Super-Scalar RAM-CPU Cache Compression"
Zukowski, Heman, Nes, Boncz, ICDE'06

Heavy-Weight Compression Schemes

Algorithm	Decompression Bandwidth
BZIP	10 MB/s
ZLIB	80 MB/s
LZO	300 MB/s

- Modern disk arrays can achieve > 1GB/s
- 1/3 CPU for decompression ➔ 3GB/s needed

"Super-Scalar RAM-CPU Cache Compression"
Zukowski, Heman, Nes, Boncz, ICDE'06

Heavy-Weight Compression Schemes

Algorithm	Decompression Bandwidth
BZIP	10 MB/s
ZLIB	80 MB/s
LZO	300 MB/s

- Modern disk arrays can achieve > 1GB/s
- 1/3 CPU for decompression ➔ 3GB/s needed

➔ **Lightweight compression schemes are better**

➔ **Even better: operate directly on compressed data**

"Integrating Compression and Execution in Column-Oriented Database Systems" Abadi et. al, SIGMOD '06

Operating Directly on Compressed Data

- **I/O - CPU tradeoff is no longer a tradeoff**
- **Reduces memory–CPU bandwidth requirements**
- **Opens up possibility of operating on multiple records at once**

Examples

- **SUM_i(rle-compressed column[i]) ➔ SUM_g(count[g] * value[g])**
- **(country == "Asia") ➔ countryCode == 6**

strcmp **SIMD**

"Integrating Compression and Execution in Column-Oriented Database Systems" Abadi et. al, SIGMOD '06

Operating Directly on Compressed Data

Quarter

(Q1, 1, 300)
(Q2, 301, 6)
(Q3, 307, 500)
(Q4, 807, 600)

Product ID

1	2	3
...	...	...
1	0	0
0	0	1
0	1	0
1	0	0
1	0	0
0	0	1
0	1	0
1	0	0
0	0	1
0	1	0
0	0	1
...	...	...

```
SELECT ProductID, Count(*)
FROM table

WHERE Quarter = Q2

GROUP BY ProductID
```

"Integrating Compression and Execution in Column-Oriented Database Systems" Abadi et. al, SIGMOD '06

Operating Directly on Compressed Data

Quarter

(Q1, 1, 300)
(Q2, 301, 6)
(Q3, 307, 500)
(Q4, 807, 600)

Product ID

1	2	3
...	...	...
1	0	0
0	0	1
0	1	0
1	0	0
1	0	0
0	0	1
0	1	0
1	0	0
0	0	1
0	1	0
0	0	1
...	...	...

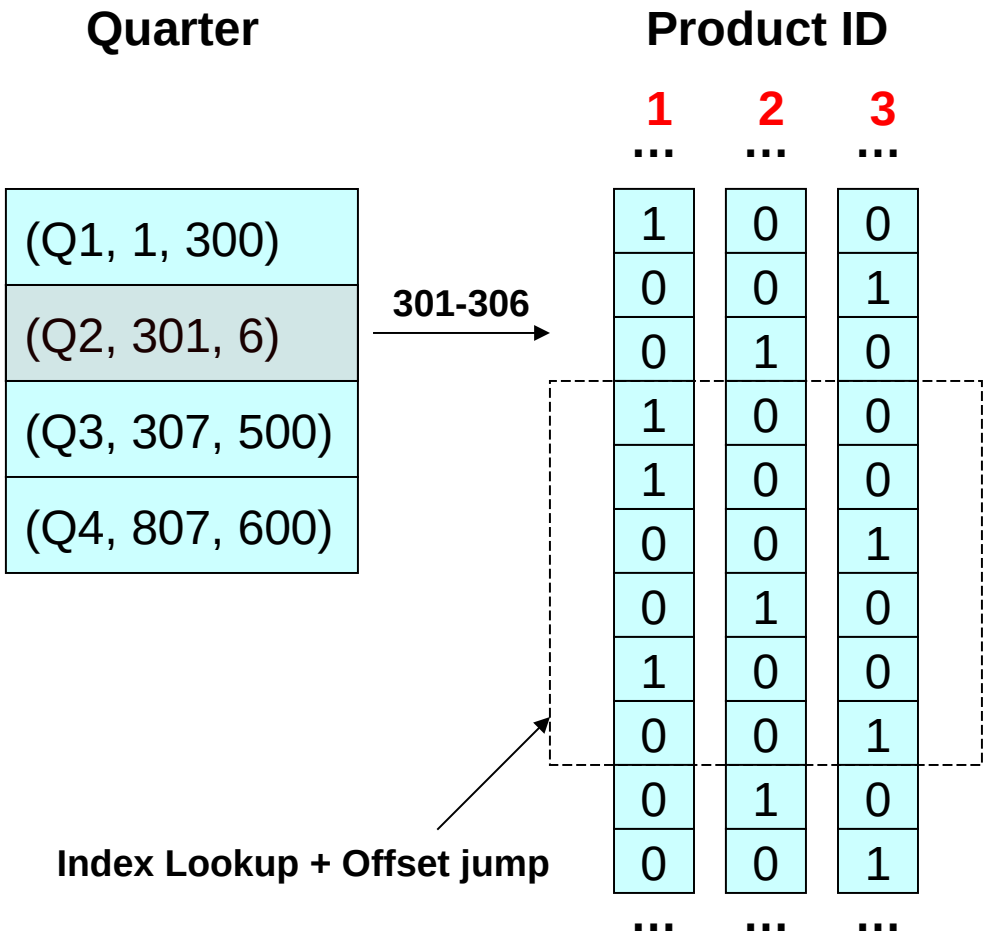
```
SELECT ProductID, Count(*)
FROM table

WHERE Quarter = Q2

GROUP BY ProductID
```

"Integrating Compression and Execution in Column-Oriented Database Systems" Abadi et. al, SIGMOD '06

Operating Directly on Compressed Data



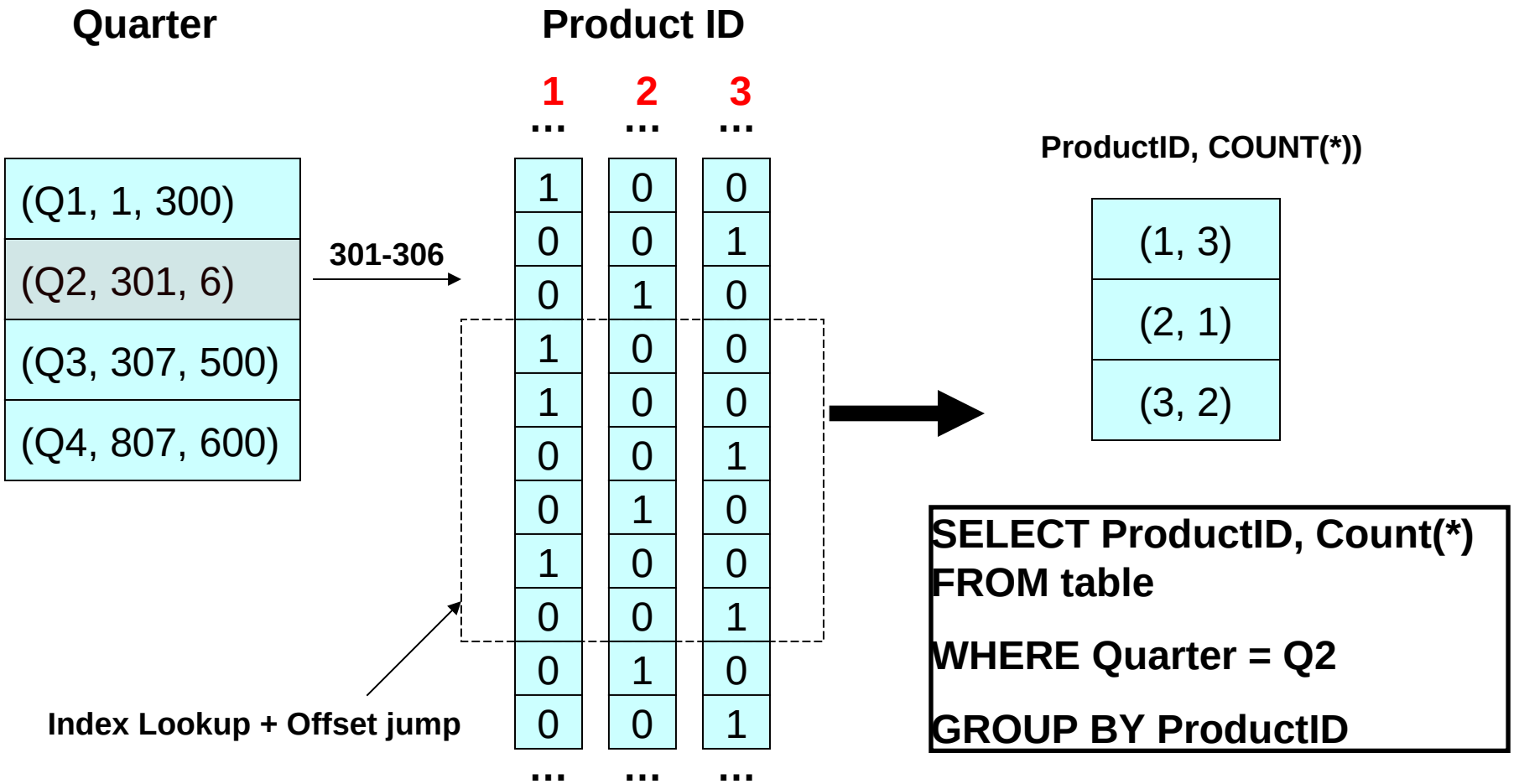

```
SELECT ProductID, Count(*)
FROM table

WHERE Quarter = Q2

GROUP BY ProductID
```

"Integrating Compression and Execution in Column-Oriented Database Systems" Abadi et. al, SIGMOD '06

Operating Directly on Compressed Data



ADM: Literature

- **Column-Oriented Database Systems (2/6) - Selected Execution Techniques**
  - Compression
    - "Compressing Relations and Indexes". Goldstein, Ramakrishnan, Shaft. ICDE'98.
    - "Query optimization in compressed database systems". Chen, Gehrke, Korn. SIGMOD'01.
    - "Super-Scalar RAM-CPU Cache Compression". Zukowski, Heman, Nes, Boncz. ICDE'06.
    - "Integrating Compression and Execution in Column-Oriented Database Systems". Abadi, Madden, Ferreira. SIGMOD'06.
    - "Improved Word-Aligned Binary Compression for Text Indexing". Ahn, Moffat. TKDE'06.
  - Tuple Materialization
    - "Materialization Strategies in a Column-Oriented DBMS". Abadi, Myers, DeWitt, Madden. ICDE'07.
    - "Column-Stores vs Row-Stores: How Different are They Really?". Abadi, Madden, Hachem. SIGMOD'08.
    - "Query Processing Techniques for Solid State Drives". Tsirogiannis, Harizopoulos Shah, Wiener, Graefe. SIGMOD'09.
    - "Self-organizing tuple reconstruction in column-stores". Idreos, Manegold, Kersten. SIGMOD'09.
  - Join
    - "Fast Joins using Join Indices". Li and Ross. VLDBJ 8:1-24, 1999.

ADM: Agenda

- 07.09.2022: Lecture 1: **Introduction**
- 14.09.2022: Lecture 2: **SQL Recap**

  (plus Assignment 1 [in groups; 3 weeks]: TPC-H benchmark)
- 21.09.2022: Lecture 3: **Column-Oriented Database Systems (1/6) - Motivation & Basic Concepts**
- 28.09.2022: Lecture 4: **Column-Oriented Database Systems (2a/6) - Selected Execution Techniques (1/2)**
- 05.10.2022: Lecture 5: **Column-Oriented Database Systems (2b/6) - Selected Execution Techniques (2/2)**

  (plus Assignment 2 [in groups; 3 weeks]: Compression techniques)
- 12.10.2022: Lecture 6: **Column-Oriented Database Systems (3/6) - Cache Conscious Joins**
- 19.10.2022: Lecture 7: **Column-Oriented Database Systems (4/6) - "Vectorized Execution"**
- ~~26.10.2022~~: ***No lecture!***
- 02.11.2022: Lecture 8: **DuckDB: An embedded database for data science (1/2) (guest lecture & *hands-on*)**

  (plus Assignment 3 [individual; 2 weeks]: Analysing NYC Cab dataset with DuckDB)
- 09.11.2022: Lecture 9: **DuckDB: An embedded database for data science (2/2) (guest lecture & *hands-on*)**
- 16.11.2022: Lecture 10: **Branch Misprediction & Predication**

  (plus Assignment 4 [individual; 2 weeks]: Predication)
- 23.11.2022: Lecture 11: **Column-Oriented Database Systems (5/6) - Adaptive Indexing**
- 30.11.2022: Lecture 12: **Column-Oriented Database Systems (6/6) - Progressive Indexing**

ADM: Literature

- **Column-Oriented Database Systems (2/6) - Selected Execution Techniques**
  - Compression
    - "Compressing Relations and Indexes". Goldstein, Ramakrishnan, Shaft. ICDE'98.
    - "Query optimization in compressed database systems". Chen, Gehrke, Korn. SIGMOD'01.
    - "Super-Scalar RAM-CPU Cache Compression". Zukowski, Heman, Nes, Boncz. ICDE'06.
    - "Integrating Compression and Execution in Column-Oriented Database Systems". Abadi, Madden, Ferreira. SIGMOD'06.
    - "Improved Word-Aligned Binary Compression for Text Indexing". Ahn, Moffat. TKDE'06.
  - Tuple Materialization
    - "Materialization Strategies in a Column-Oriented DBMS". Abadi, Myers, DeWitt, Madden. ICDE'07.
    - "Column-Stores vs Row-Stores: How Different are They Really?". Abadi, Madden, Hachem. SIGMOD'08.
    - "Query Processing Techniques for Solid State Drives". Tsirogiannis, Harizopoulos Shah, Wiener, Graefe. SIGMOD'09.
    - "Self-organizing tuple reconstruction in column-stores". Idreos, Manegold, Kersten. SIGMOD'09.
  - Join
    - "Fast Joins using Join Indices". Li and Ross. VLDBJ 8:1-24, 1999.