GHENT
UNIVERSITY

imec

FACULTY OF ENGINEERING AND ARCHITECTURE

DEPARTMENT OF INFORMATION TECHNOLOGY (INTEC)
RESEARCH GROUP IDLAB

GEDISTRIBUEERDE GEGEVENSVERWERKING

(E761040)

LAB SESSION 04 - Time Series DB

05/05/2025

GHENT UNIVERSITY

Why use time series DB?

- Optimized for time-stamped or time series data
- Support for dropping / archiving older data (retention policies)
- You care about ingesting speed and throughput as much as possible.
- Typical use cases: IoT, monitoring, sensor, data which is immutable

Time series DB: High cardinality issues

Cardinality often refers to the number of distinct elements in a single column.

Cardinality of the dataset = calculate the total number of unique combinations, the cardinality of each of the columns is multiplied.

Time series datasets tend to have high cardinality

Time series DB: High cardinality issues

	DIMENSIONS			MEASURES		
Timestamp	Source IP	Source Port	Protocol	Duration	Forward Packets	Backward Packets
06/12/2022 23:00:15	192.168.20.19	33118	6	36728	5	3
06/12/2022 23:00:51	192.168.20.48	53948	6	890408	17	16
06/12/2022 23:01:19	192.168.20.43	58752	6	3777984	13	13
06/12/2022 23:01:25	192.168.20.43	46084	6	8672390	17	16
06/12/2022 23:00:02	192.168.20.15	49156	17	116995737	40	0
06/12/2022 23:00:03	8.6.0.1	0	0	118298750	51	0
06/12/2022 23:00:04	192.168.20.13	49154	17	114999347	24	0
06/12/2022 23:00:04	192.168.20.12	49154	17	115000492	24	0
06/12/2022 23:00:04	192.168.20.42	37322	6	106719497	10	6
06/12/2022 23:00:03	54.217.52.155	443	6	119570085	12	13
06/12/2022 23:00:08	192.168.20.19	0	0	430015	2	0
06/12/2022 23:00:06	192.168.20.48	35582	6	112851237	25	13
06/12/2022 23:00:09	192.168.20.19	46048	6	90245130	4	4
06/12/2022 23:00:11	18.203.57.224	443	6	119999378	9	4
06/12/2022 23:00:13	192.168.20.42	35056	6	90254035	4	3
06/12/2022 23:00:03	192.168.20.10	47863	6	119997319	17	17

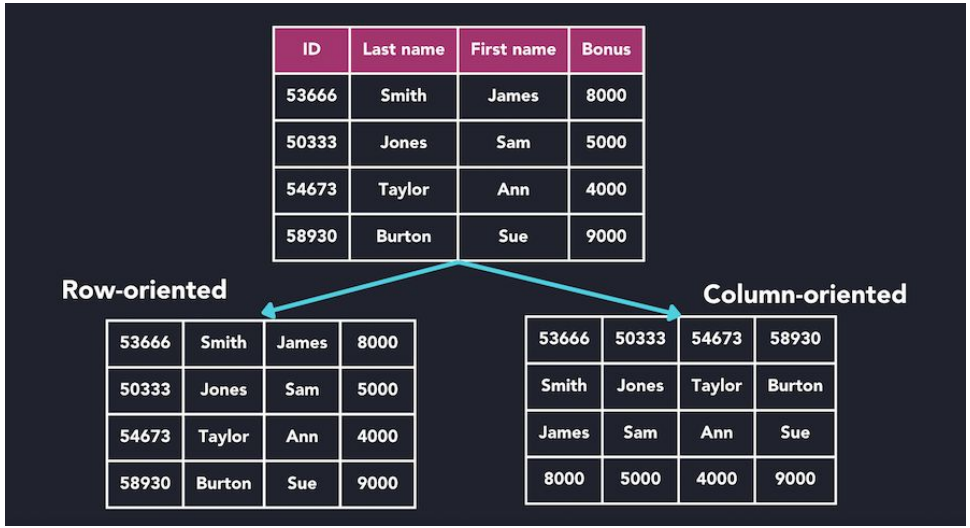
This network analytics table has high cardinality because the combination of unique values for Source IP, Source Port, and Protocol is very large.

GHENT UNIVERSITY

QuestDB

- Open source TSDB
- Columnar storage model
- Supports time partitions

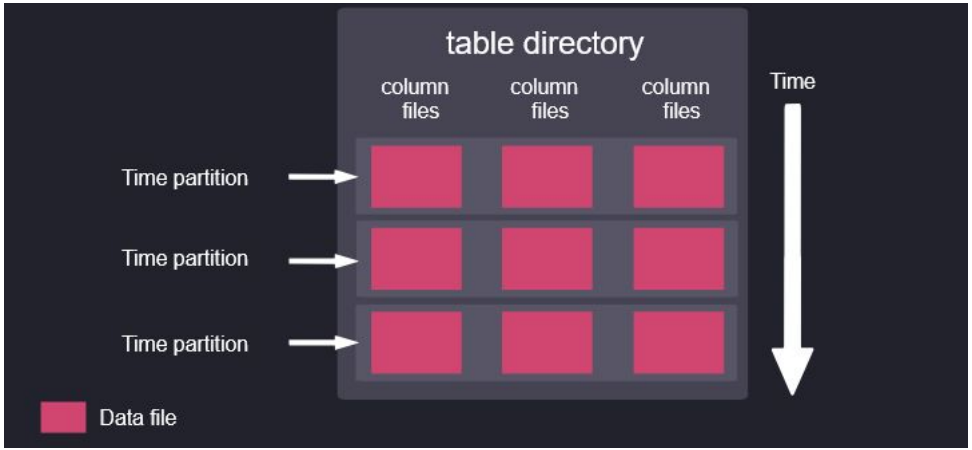
QuestDB: Columnar storage model



- Only retrieve the columns needed to solve the query
- Great for analytical queries: ““Calculate the average and total price for an order in the last month…”

GHENT UNIVERSITY

QuestDB: time partitions



- Technique that splits data in a large database into smaller chunks in order to improve the performance and scalability of the database system.
- QuestDB offers the option to partition tables by intervals of time. Data for each interval is stored in separate sets of files.

GHENT UNIVERSITY

Lab Session Structure

1. Store movie ratings and tags in QuestDB
   - Design table schema
   - Create a materialized view
   - Use Kafka Connect to ingest data
2. Work on project

Lab Session Structure: design table schema

- Time partitioning
- Symbols
- Deduplication

```
CREATE TABLE x AS (
  SELECT
    cast(x as int) i,
    - x j,
    rnd_str(5, 16, 2) as str,
    timestamp_sequence('2023-02-04T00', 60 * 1000L) ts
  FROM
    long_sequence(60 * 23 * 2 * 1000)
) timestamp (ts) PARTITION BY DAY WAL;
```

```
SHOW PARTITIONS FROM x;
```

index	partitionBy	name	minTimestamp	maxTimestamp	numRows
0	DAY	2023-02-04	2023-02-04T00:00:00.000000Z	2023-02-04T23:59:59.940000Z	1440000
1	DAY	2023-02-05	2023-02-05T00:00:00.000000Z	2023-02-05T21:59:59.940000Z	1320000

GHENT UNIVERSITY

Lab Session Structure: materialized view

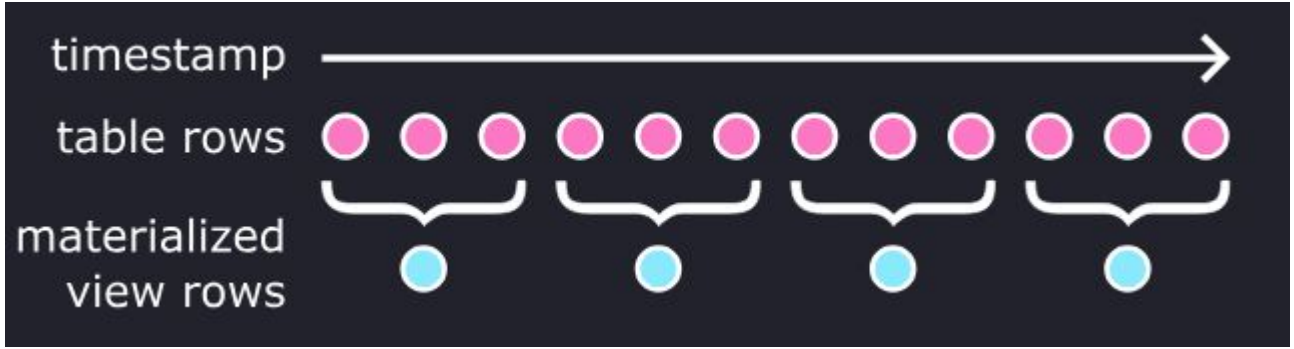
- Stores the pre-computed results of a query.
- Persist their data to disk.
- Efficient for expensive aggregate queries that are run frequently.

```
SELECT
  timestamp,
  symbol,
  side,
  sum(price * amount) AS notional
FROM trades
SAMPLE BY 1m;
```

Scans the entire dataset -> slower as the dataset grows

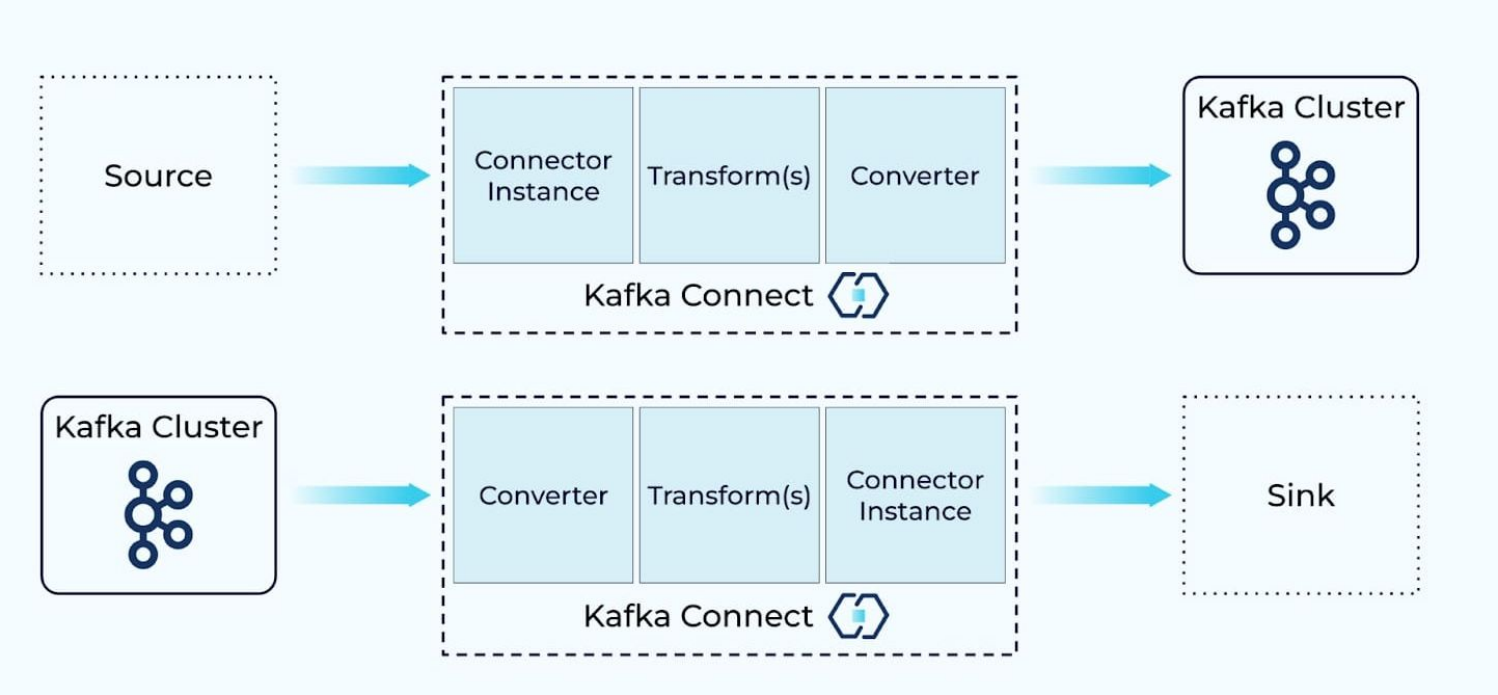
Lab Session Structure: materialized view

As you add new data to the base table, the materialized view will efficiently update itself. You can then query the materialized view as a regular table without the impact of a full table scan of the base table.



GHENT
UNIVERSITY

Lab Session Structure: Kafka Connect



Lab Session Structure: Kafka Connect

Use the REST API to submit connector config, start, stop and monitor workers / tasks.

```
name=questdb-sink
client.conf.string=http::addr=localhost:9000;
topics=example-topic
table=example_table

connector.class=io.questdb.kafka.QuestDBSinkConnector

# message format configuration
value.converter=org.apache.kafka.connect.json.JsonConverter
include.key=false
key.converter=org.apache.kafka.connect.storage.StringConverter
value.converter.schemas.enable=false
```

Tips: use json for the connector config

https://github.ugent.be/GDV/lab-time-series-db