

Methods of access to series of archived database snapshots

Koit Saarevet, 14 May 2025



Year 1



Year 2



Year 3

...



Year n

Objective:

Develop a solution for accessing a series of archived snapshots.

Research strategy

- a) Discover a full solution
- b) Find the closest one, then adapt it
- c) Invent from scratch

Algorithm 3 Addition of $D_{(i)}$ to $D_{(1)} \oplus \dots \oplus D_{(i-1)}$ (enhanced)

Require: Snapshot $D_{(i)}$

Require: $D^* = D_{(1)} \oplus \dots \oplus D_{(i-1)}$

Ensure: $D^* = D_{(1)} \oplus \dots \oplus D_{(i)}$

- 1: Get $D_{(i-1)}$ from D^*
- 2: Sort relations in $D_{(i)}$ s.t. $R_1 \prec R_2$ if $R_2 \triangleright R_1$
- 3: **for all** $R_{(i,j)} \in D_{(i)}$ **do**
- 4: **if** K_j is defined for schema $\mathcal{R}_j$ **then**
- 5: **for all** $r \in R_{(i,j)}$ **do**
- 6: **if** $r[K_j] \notin R_{(i-1,j)}[K_j]$ **then**
- 7: **insert** r **in** R_j^* **time** $[t_i, t_{max}]$
- 8: **else if** $r[K_j] \in R_{(i-1,j)}[K_j] \wedge r[\mathcal{R}_j] \notin R_{(i-1,j)}[\mathcal{R}_j]$ **then**
- 9: **update** $r_{(i-1)}$ **in** R^* **set** $E_i = t$.

```
1 CREATE PROCEDURE sync_databases(IN source_db VARCHAR(64), IN target_db VARCHAR(64))
2 BEGIN
3     DECLARE done INT DEFAULT 0; -- cursor control variable
4     DECLARE tbl_name VARCHAR(64);
5     DECLARE cur CURSOR FOR
6         SELECT TABLE_NAME
7         FROM INFORMATION_SCHEMA.TABLES
8         WHERE TABLE_SCHEMA = source_db;
9     DECLARE CONTINUE HANDLER FOR NOT FOUND SET done = 1;
10
11     SET FOREIGN_KEY_CHECKS = 0;
12
13     -- MariaDB uses @@timestamp to fill the row_start and row_end columns
14     -- Change it to use snapshot_time instead of the current clock time
15     SET @@timestamp = UNIX_TIMESTAMP(snapshot_time);
16
17     OPEN cur;
18     -- Loop through all tables in source_db
19     read_loop: LOOP
20         FETCH cur INTO tbl_name;
```

Solution: temporal DB

- Set up a blank DB in a temporal-capable engine, e.g. MariaDB
- Import each snapshot + its timestamp
- Use temporal relational algebra and temporal tools for queries

2010

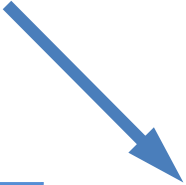
ID	Data1	Data2	Data3

2015

ID	Data1	Data2	Data3

2020

ID	Data1	Data2	Data3



ID	Data1	Data2	Data3	From	To

2010

ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod

ID	First	Last	Dept	From	To

2010

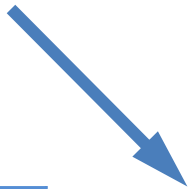
ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod



ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2038

2010

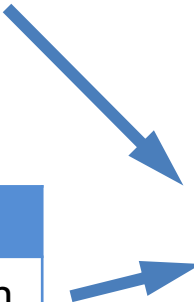
ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod



The diagram shows two blue arrows. One arrow originates from the 2010 table and points to the first row of the combined table. The second arrow originates from the 2015 table and points to the second row of the combined table.

ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2038

2010

ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod

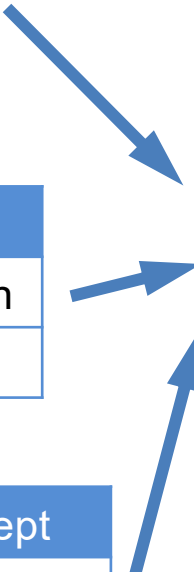


Diagram illustrating data transformation or migration. Three arrows point from the 2010, 2015, and 2020 tables to the combined table on the right.

ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038

2010

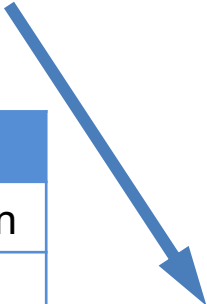
ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod



ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038
2	David	Fox	Sales	2010	2038

2010

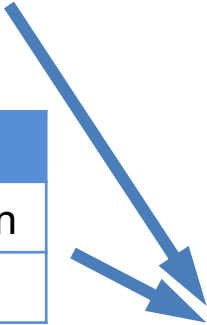
ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod



ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038
2	David	Fox	Sales	2010	2015

2010

ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod

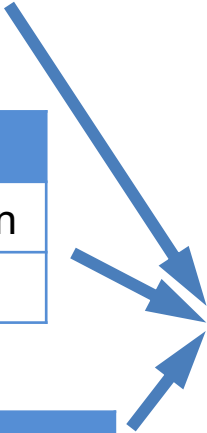


Diagram illustrating data transformation from three snapshots (2010, 2015, 2020) into a combined timeline table. Arrows indicate the mapping of data from the snapshots to the combined table.

ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038
2	David	Fox	Sales	2010	2015

2010

ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod

ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038
2	David	Fox	Sales	2010	2015

2010

ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod



ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038
2	David	Fox	Sales	2010	2015
3	Tom	Gray	Prod	2015	2038

2010

ID	First	Last	Dept
1	Jane	Doe	Admin
2	David	Fox	Sales

2015

ID	First	Last	Dept
1	Jane	Smith	Admin
3	Tom	Gray	Prod

2020

ID	First	Last	Dept
1	Jane	Grant	Admin
3	Tom	Gray	Prod

ID	First	Last	Dept	From	To
1	Jane	Doe	Admin	2010	2015
1	Jane	Smith	Admin	2015	2020
1	Jane	Grant	Admin	2020	2038
2	David	Fox	Sales	2010	2015
3	Tom	Gray	Prod	2015	2038



HAUDI

KALMISTUTE REGISTER

```
SELECT id, last_name, first_name, birth_date,  
death_date, burial_time, row_start, row_end  
FROM haudi_s0.deceased FOR SYSTEM_TIME ALL  
WHERE id IN  
(SELECT id  
FROM haudi_s0.deceased FOR SYSTEM_TIME ALL  
WHERE row_end < '2038-01-19')  
ORDER BY id;
```

id	last_name	first_name	birth_date	death_date	burial_time	row_start	row_end
325	Tüür	Al.-Der	1882-02-17			2021-01-07	2024-04-03
325	Tuur	Aleksander	1882-02-17		1958-11-04	2024-04-03	2038-01-19
546	Kanter	Hubert				2021-01-07	2024-04-03
546	Kanter	Herbert Heinrich	1908-02-21	1981-10-03	1981-10-10	2024-04-03	2038-01-19
742	Pärna	S.S.P.Laps				2021-01-07	2024-04-03
742	Pärna	S.s.p.laps Martin		1989-05-19	1989-05-24	2024-04-03	2038-01-19

Plans for the future

- Incremental archiving
- Schema versioning
- Temporal queries
- System-versioned snapshots
- More universal algorithm
- Prototype improvements
- Visualisation
- Comparison of methods

Thank you

http://mantagir.ee/thesis/Koit_Saarevet_MSc_Thesis.pdf



© European Union 2025

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

