

# A Change Data Capture System for SpazioDati

Thesis Presentation

*Daniele Parmeggiani*

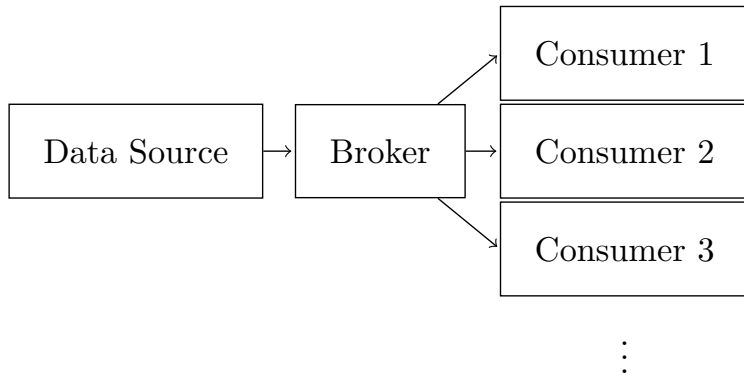
Università degli Studi di Trento

2021-03-15

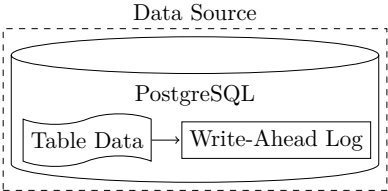
# Context and Scope



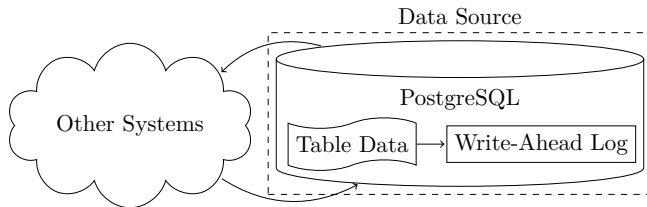
# Change Data Capture Systems



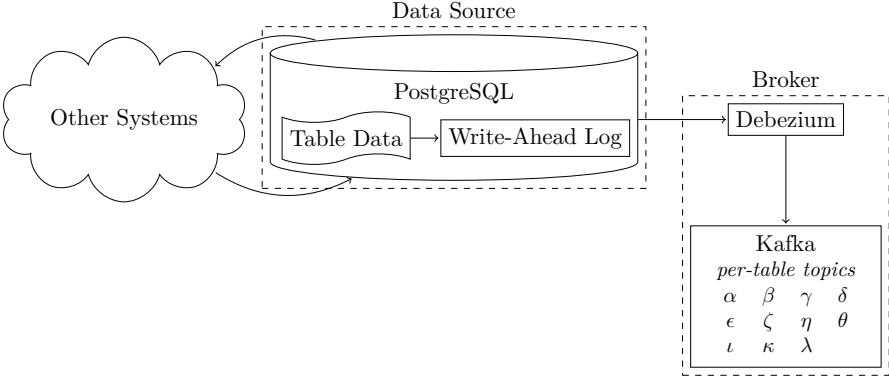
# Proposed CDC System Design



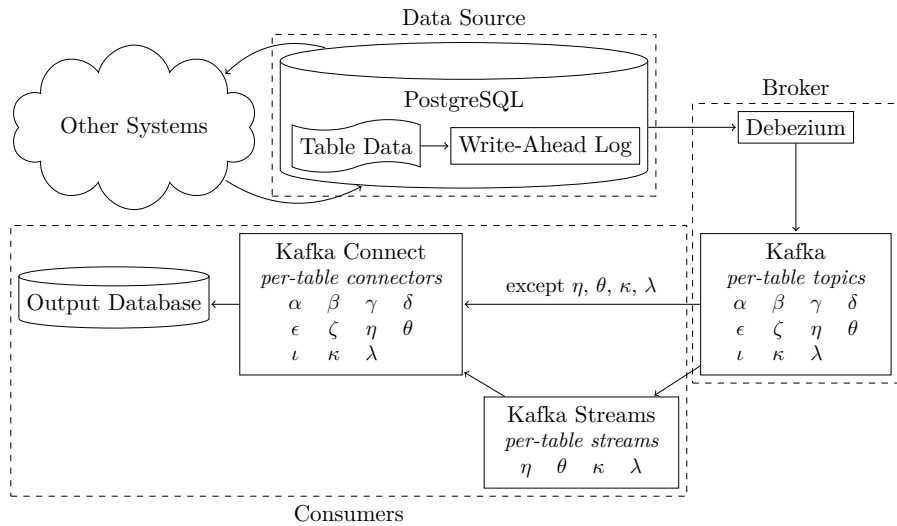
# Proposed CDC System Design



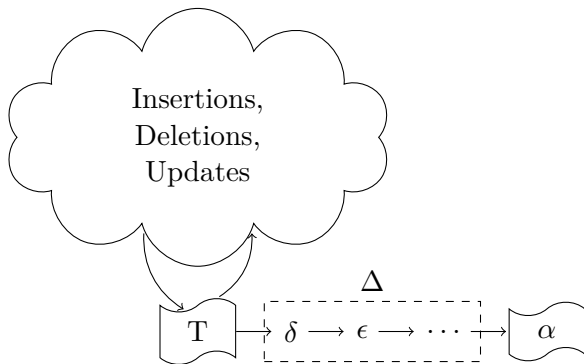
# Proposed CDC System Design



# Proposed CDC System Design



# Duality of Tables and Streams



$$\text{s.t. } \alpha = \sum_{\delta \in \Delta} \delta .$$

$$\Rightarrow \alpha = T .$$



- \* Plain:  $\text{dest}[\omega] := \pi_S(\text{source}[\omega])$
- \* Aggregated
- \* Time-traveled

# Aggregated Tables

$$\begin{aligned} \text{dest}[\eta] &:= \gamma_{\text{user}, \text{day}, \text{class}, \text{count}(\ast) \rightarrow n, \sum(\text{cost}) \rightarrow C} ( \\ &\quad \sigma_{\text{isFinal}(\ast)}(\text{source}[\eta]) \\ & ) \end{aligned} \qquad \begin{aligned} \text{dest}[\theta] &:= \gamma_{\text{user}, \text{day}, \text{count}(\ast) \rightarrow n} ( \\ &\quad \sigma_{\text{isFinal}(\ast)}(\text{source}[\theta]) \\ & ) \end{aligned}$$

$$\begin{aligned} \text{dest}[\kappa] &:= \gamma_{\text{user}, \text{day}, \text{class}, \text{subclass}, \sum(\text{cost}) \rightarrow C} ( \\ &\quad \sigma_{\text{class}=\text{'c'}}(\text{source}[\kappa]) \\ & ) \end{aligned} \qquad \begin{aligned} \text{dest}[\lambda] &:= \pi_{\text{dest}[\alpha].\text{id}, \text{day}, \text{class}, n} ( \\ &\quad \text{dest}[\alpha] \bowtie \gamma_{\text{email}, \text{day}, \text{class}, \text{count}(\ast) \rightarrow n} ( \\ &\quad \quad \text{source}[\lambda] \\ & )) \end{aligned}$$

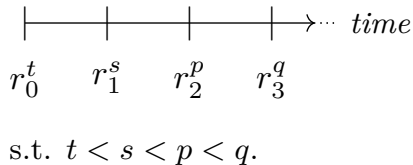


Figure: Representation of changes to a row  $r$ .

Was part of official PostgreSQL distributions, up to version 12.

$r$	$validity$	$r$	$validity$	$r$	$validity$	$r$	$validity$
0	$[t, +\infty)$	0	$[t, s)$	0	$[t, s)$	0	$[t, s)$
		1	$[s, +\infty)$	1	$[s, p)$	1	$[s, p)$
				2	$[p, +\infty)$	2	$[p, q)$
						3	$[q, +\infty)$

Table: Representation in  $\text{dest}[\omega]$  of the changes outlined in Figure 1.

# Time Travel Mapping

$$\text{source}[\omega] \rightarrow \text{INSERT } r_i^t \xRightarrow{\tau} \text{INSERT } r_i \text{ s.t. validity} := [t, +\infty) \quad (1)$$

$$\text{source}[\omega] \rightarrow \text{UPDATE } r_i^t \xRightarrow{\tau} \left\{ \begin{array}{l} 1. \text{ UPDATE } r_{i-1} \text{ s.t. validity} \cap (-\infty, t) \\ 2. \text{ INSERT } r_i \text{ s.t. validity} := [t, +\infty) \end{array} \right\} \quad (2)$$

$$\text{source}[\omega] \rightarrow \text{DELETE } r_i^t \xRightarrow{\tau} \text{UPDATE } r_{i-1} \text{ s.t. validity} \cap (-\infty, t) \quad (3)$$

- \* uniqueness (including primary-key),
- \* conditional (including non-null), and
- \* referential integrity (foreign-key).

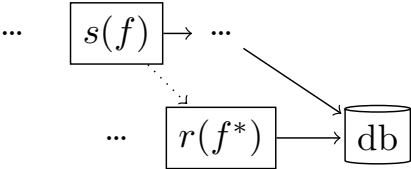
# On Constraints

- \* uniqueness (including primary-key),
- \* conditional (including non-null), and
- \* referential integrity (foreign-key).

## Assumption 1

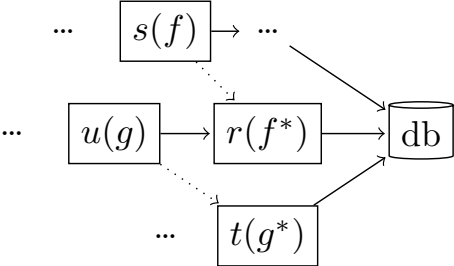
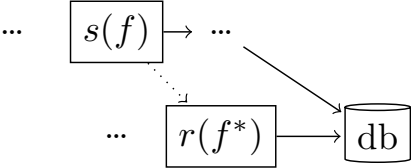
source[\*] is consistent.

# Foreign Key Constraints





# Foreign Key Constraints



*Thank you for your attention.*