A Change Data Capture System for SpazioDati

Thesis Presentation

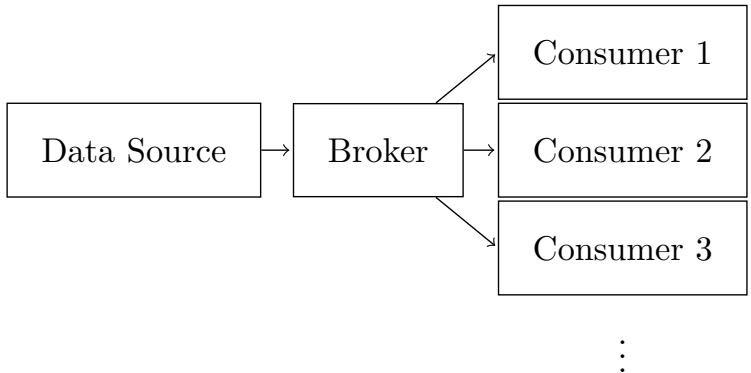
Daniele Parmeggiani

Università degli Studi di Trento

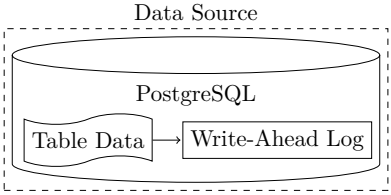
2021-03-15

Context and Scope

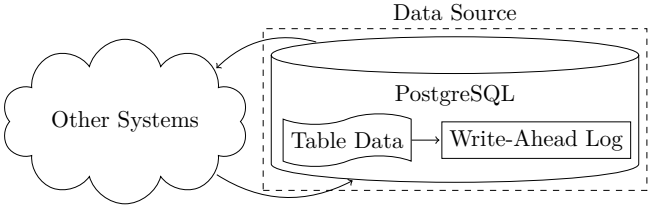
Change Data Capture Systems



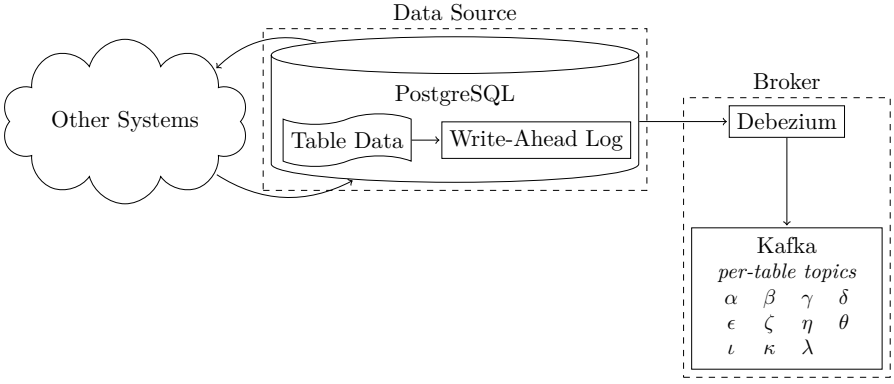
Proposed CDC System Design



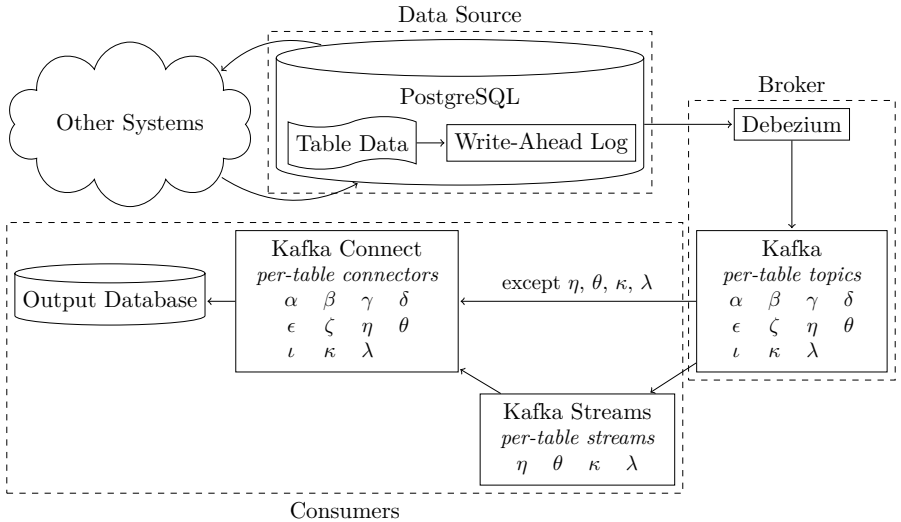
Proposed CDC System Design



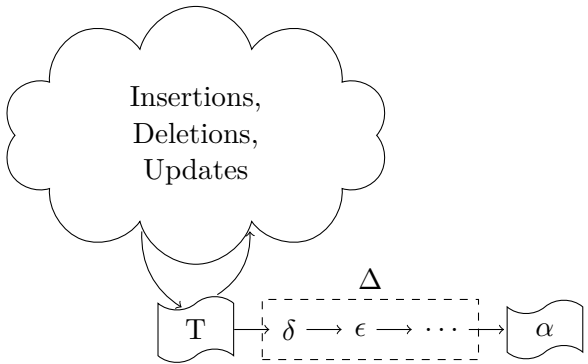
Proposed CDC System Design



Proposed CDC System Design



Duality of Tables and Streams



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

$$\Rightarrow \alpha = T \ .$$

Classes of Tables

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

Aggregated Tables

$$\text{dest}\,[\eta] := \gamma_{\text{user, day, class, count}(*)\to n,\ \sum(\text{cost})\to C}\Big(\sigma_{\text{isFinal}(*)}\big(\text{source}\,[\eta]\big)\Big)$$

$$\text{dest}\,[\theta] := \gamma_{\text{user, day, count}(*)\to n}\Big(\sigma_{\text{isFinal}(*)}\big(\text{source}\,[\theta]\big)\Big)$$

$$\text{dest}\,[\kappa] := \gamma_{\text{user, day, class, subclass},\ \sum(\text{cost})\to C}\Big(\sigma_{\text{class}=\texttt{'c'}}\big(\text{source}\,[\kappa]\big)\Big)$$

$$\text{dest}\,[\lambda] := \pi_{\text{dest}[\alpha].\text{id, day, class, }n}\Big(\text{dest}\,[\alpha] \bowtie \gamma_{\text{email, day, class, count}(*)\to n}\big(\text{source}\,[\lambda]\big)\Big)$$

Time Travel

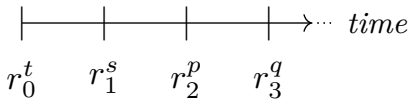


Figure: Representation of changes to a row r.

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

Validity Column

r	$validity$
0	$[t, +\infty)$

r	$validity$
0	$[t, s)$
1	$[s, +\infty)$

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

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

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

Time Travel Mapping

$$\text{source}\,[\omega] \to \texttt{INSERT}\ r_i^t \overset{\tau}{\Rightarrow} \texttt{INSERT}\ r_i\ \text{s.t. validity} := [t, +\infty) \tag{1}$$

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

$$\text{source}\,[\omega] \to \texttt{DELETE}\ r_i^t \overset{\tau}{\Rightarrow} \texttt{UPDATE}\ r_{i-1}\ \text{s.t. validity} \cap (-\infty, t) \tag{3}$$

On Constraints

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

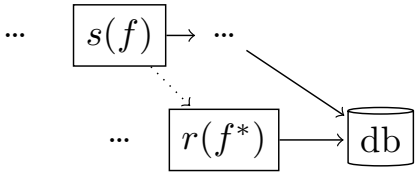
On Constraints

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

Assumption 1

source $[*]$ is consistent.

Foreign Key Constraints



Foreign Key Constraints