CERA: Complex Event Relational Algebra

Michael Eckert, Simon Brodt, Olga Poppe,
Steffen Hausmann, François Bry

Institute for Informatics, University of Munich, Germany

October 5, 2010

Outline

Motivation

CERA

Conclusion

Advantages of Transferring Relational Algebra to CEP

- Understandability
- Use of the results in DB research
- Queries over event and non-event data
- Abstract machine

Challenge in CEP: Unbounded Streams

Approach: Windows

- Stream is a *finite* relation at *each point of time*
- Quite ordernary relational algebra operators

Problem: Size of the window

Our proposal: CERA

- Stream is *one potentially infinite* relation
- Tailored variants of relational algebra operators restricted so that for each point of time it is sufficient to know the finite available part of the stream to compute the result up to that time point

Outline

Motivation

CERA

Conclusion

CERA Operators

1. Event matching Q
2. Natural join $\bowtie$
3. Selection σ
4. Projection π
5. Temporal join $\bowtie_{i \sqsupseteq j}$
6. Temporal anti-semi join $\overline{\ltimes}_{i \sqsupseteq j}$
7. Merging of time intervals μ
8. Renaming ρ
9. Event construction C

XChange^{EQ} Program

```
DETECT   fire { area { var A } }
ON and { event s: smoke {{ area {{ var A }} }},
         event t: temp {{ area {{ var A }}, value {{ var T }} }}
   } where { s before t, {s,t} within 1 min, var T > 40 }
END

DETECT   avg_temp { sensor { var S }, value { avg(all var T) } }
ON and { event m: temp {{ sensor {{ var S }} }},
         event i: timer:from-start-backward [ event m, 1 min ],
         while i: collect
                  temp {{ sensor {{ var S }}, value {{ var T }} }} }
END

DETECT   burnt_down { sensor { var S } }
ON and { event n: temp {{ sensor {{ var S }}, value {{ var T }} }},
         event i: timer:from-end [ event n, 12 sec ],
         while i: not temp {{ sensor {{ var S }} }}
   } where { var T > 40 }
END
```

Event Matching Q

```
DETECT   fire { area { var A } }
ON and { event s: smoke {{ area {{ var A }} }},
         event t: temp {{ area {{ var A }}, value {{ var T }} }}
   } where { s before t, {s,t} within 1 min, var T > 40 }
END
```

$$
\begin{aligned}
&C_{[\ fire\{area\{var\ \ A\}\}\ ]}(\\
&\quad \mu[[begin, end] \leftarrow s \sqcup t](\\
&\qquad \sigma[\max\{s.end, t.end\} - \min\{s.begin, t.begin\} \leq 1\ min](\\
&\qquad\quad \sigma[s.end < t.begin](\\
&\qquad\qquad \sigma[T > 40](\\
&\qquad\qquad\quad \mathsf{Smoke_s} \bowtie \mathsf{Temp_t})))))
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Smoke_s} &= \mathsf{Q}_{[\ s:\ smoke\{\{area\{\{var\ \ A\}\}\}\}\ ]}(\mathsf{E})\\
schema(\mathsf{Smoke_s}) &= \{s.begin, s.end, s.ref, A\}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Temp_t} &= \mathsf{Q}_{[\ t:\ temp\{\{area\{\{var\ \ A\}\}, value\{\{var\ \ T\}\}\}\}\ ]}(\mathsf{E})\\
schema(\mathsf{Temp_t}) &= \{t.begin, t.end, t.ref, A, T\}
\end{aligned}
$$

Merging of Time Intervals μ

```
DETECT   fire { area { var A } }
ON and { event s: smoke {{ area {{ var A }} }},
         event t: temp {{ area {{ var A }}, value {{ var T }} }}
   } where { s before t, {s,t} within 1 min, var T > 40 }
END
```

$$
\begin{array}{l}
C_{[\ fire\{area\{var\ \ A\}\}\ ]}(\\
\quad \mu[[begin, end] \leftarrow s \sqcup t](\\
\quad\quad \sigma[\max\{s.end, t.end\} - \min\{s.begin, t.begin\} \leq 1\ min](\\
\quad\quad\quad \sigma[s.end < t.begin](\\
\quad\quad\quad\quad \sigma[T > 40](\\
\quad\quad\quad\quad\quad \mathrm{Smoke}_{s} \bowtie \mathrm{Temp}_{t})))))
\end{array}
$$

$$
\begin{aligned}
begin &= \mathbf{min}(s.begin, t.begin) \\
end &= \mathbf{max}(s.end, t.end)
\end{aligned}
$$

Event Construction C without Aggregation

```
DETECT   fire { area { var A } }
ON and { event s: smoke {{ area {{ var A }} }},
         event t: temp {{ area {{ var A }}, value {{ var T }} }}
   } where { s before t, {s,t} within 1 min, var T > 40 }
END
```

$$
\begin{aligned}
&C_{[\ \mathit{fire}\{\mathit{area}\{\mathit{var}\ \ A\}\}\ ]}(\\
&\quad \mu[[\mathit{begin},\mathit{end}] \leftarrow s \sqcup t](\\
&\quad\quad \sigma[\max\{s.\mathit{end}, t.\mathit{end}\} - \min\{s.\mathit{begin}, t.\mathit{begin}\} \leq 1\ \mathit{min}](\\
&\quad\quad\quad \sigma[s.\mathit{end} < t.\mathit{begin}](\\
&\quad\quad\quad\quad \sigma[T > 40](\\
&\quad\quad\quad\quad\quad \mathrm{Smoke}_s \bowtie \mathrm{Temp}_t)))))
\end{aligned}
$$

$C_{[h]}(R) = E'$ such that $\forall r \in R$ an event $e \in E'$ represented as the data term annotated with the time interval $[r(\mathit{begin}), r(\mathit{end})]$ is constructed. The data term results from substituting each free variable X of h by $r(X)$.

Relative Timer Events

```
DETECT   avg_temp { sensor { var S }, value { avg(all var T) } }
ON and { event m: temp {{ sensor {{ var S }} }},
         event i: timer:from-start-backward [ event m, 1 min ],
         while i: collect
                  temp {{ sensor {{ var S }}, value {{ var T }} }} }
END
```

$$
\begin{aligned}
&C_{[\ avg\_temp\{sensor\{var\ \ S\},value\{avg(all\ \ var\ \ T)\}\}\ ]}\big(\\
&\quad \mu[[begin,end] \leftarrow m \sqcup i \sqcup w](\\
&\qquad (\mathrm{Temp}_{\mathrm{m}} \bowtie X_{[\text{ i: from-start-backward[m,1 min] }]}(E)) \bowtie_{i \sqsupseteq w} \mathrm{Temp}_{\mathrm{w}}))
\end{aligned}
$$

$$
\begin{aligned}
i(begin) &= m(begin) - 1\ \ min \\
i(end) &= m(begin)
\end{aligned}
$$

Temporal Join $\bowtie_{i \sqsupseteq j}$

```
DETECT   avg_temp { sensor { var S }, value { avg(all var T) } }
ON and { event m: temp {{ sensor {{ var S }} }},
         event i: timer:from-start-backward [ event m, 1 min ],
         while i: collect
                  temp {{ sensor {{ var S }}, value {{ var T }} }} }
END
```

$$
\begin{aligned}
&C_{[\ avg\_temp\{sensor\{var\ \ S\},value\{avg(all\ \ var\ \ T)\}\}\ ]}(\\
&\quad \mu[[begin, end] \leftarrow m \sqcup i \sqcup w](\\
&\qquad (\mathrm{Temp}_{\mathrm{m}} \bowtie X_{[\text{ i: from-start-backward[m,1 min] }]}(E)) \bowtie_{i \sqsupseteq w} \mathrm{Temp}_{\mathrm{w}}))
\end{aligned}
$$

$R \bowtie_{i \sqsupseteq j} S$ returns all tuples of $R \bowtie S$ satisfying the additional condition $i.begin \leq j.begin \wedge j.end \leq i.end$, where $\{i.begin, i.end\} \subseteq schema_{time}(R)$ and $\{j.begin, j.end\} \subseteq schema_{time}(S)$.

Event Construction C with Aggregation

```
DETECT   avg_temp { sensor { var S }, value { avg(all var T) } }
ON and { event m: temp {{ sensor {{ var S }} }},
         event i: timer:from-start-backward [ event m, 1 min ],
         while i: collect
                  temp {{ sensor {{ var S }}, value {{ var T }} }} }
END
```

$$\mathrm{C}_{[\ avg\_temp\{sensor\{var\ \ S\},value\{avg(all\ \ var\ \ T)\}\}\ ]}(\mu[[begin,end]\leftarrow m\sqcup i\sqcup w]((\mathrm{Temp_m}\bowtie \mathrm{X}_{[\text{ i: from-start-backward[m,1 min] }]}(\mathrm{E}))\bowtie_{i\sqsupseteq w}\mathrm{Temp_w}))$$

$\mathrm{C}_{[h]}(R) = E'$ groups the tuples of R by event references, time attributes and data attributes and computes the values of aggregation functions for each group $R_i \subseteq R$. $\forall R_i \subseteq R$ an event $e \in E'$ represented as the data term annotated with the time interval $[r(begin), r(end)], r \in R_i$, is constructed. The data term results from substituting each aggregation function of h by the computed value and each free variable X of h by $r(X)$.

Temporal Anti-semi Join $\overline{\ltimes}_{i \sqsupseteq j}$

```
DETECT   burnt_down { sensor { var S } }
ON and { event n: temp {{ sensor {{ var S }}, value {{ var T }} }},
         event i: timer:from-end [ event n, 12 sec ],
         while i: not temp {{ sensor {{ var S }} }}
   } where { var T > 40 }
END
```

$$
\begin{aligned}
&\mathrm{C}_{[\ burnt\_down\{sensor\{var\ \ S\}\}\ ]}(\\
&\quad \mu[[begin, end] \leftarrow n \sqcup i \sqcup v](\\
&\qquad \sigma[T > 40](\\
&\qquad\quad (\mathrm{Temp_n} \bowtie \mathrm{X}_{[\text{ i: from-end[n,12 sec] }]}(\mathrm{E}))\overline{\ltimes}_{i \sqsupseteq v} \mathrm{Temp_v})))
\end{aligned}
$$

$R \overline{\ltimes}_{i \sqsupseteq j} S$ returns R with those tuples r removed that have a "partner" in S, i.e., $\exists s \in S$ that agrees on all shared attributes with $r \in R$ and whose timestamps $s(j.begin)$, $s(j.end)$ are within the time bounds $r(i.begin)$, $r(i.end)$, where $\{i.begin, i.end\} \subseteq schema_{time}(R)$ and $\{j.begin, j.end\} = schema_{time}(S)$.

Outline

Motivation

CERA

Conclusion

Future Work

- Minimal set of orthogonal operators
- Reordering of operators for optimization
- More flexible time attributes and temporal relations
- More flexible grouping and aggregation
- Non-event data
- Graph data
- Actions and states
- Provenance

CERA in the Context of the Group Research