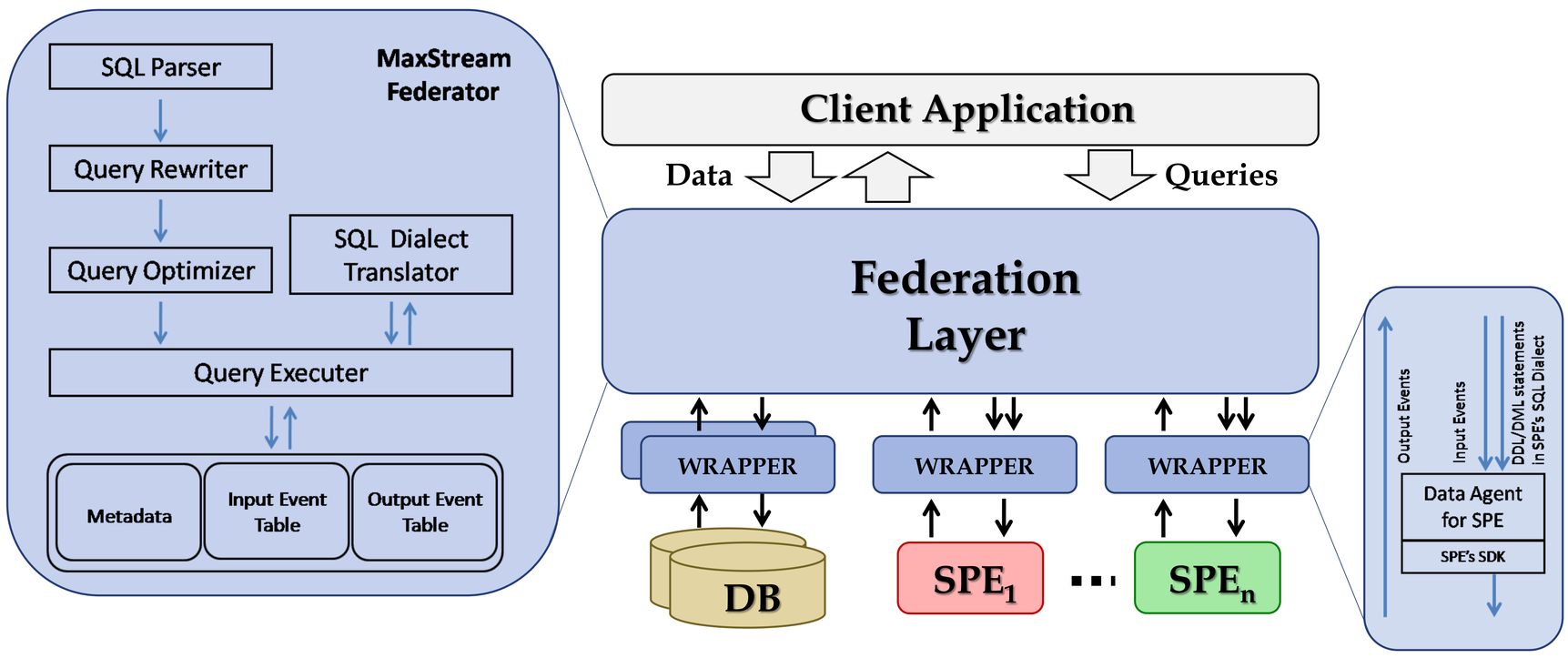
The MaxStream Federated Stream Processing System

I. Botan, Y. Cho, R. Derakhshan, N. Dindar, A. Gupta, L. Haas, K. Kim, C. Lee, G. Mundada,
M. C. Shan, N. Tatbul, Y. Yan, B. Yun, J. Zhang

Automatic Query Translation to Different SQL Dialects

Uniform Interface for Data and Queries



Management of both Streaming and Stored Data

Handling the Heterogeneity of Stream Processing Engines

Handling the Heterogeneity of SPEs

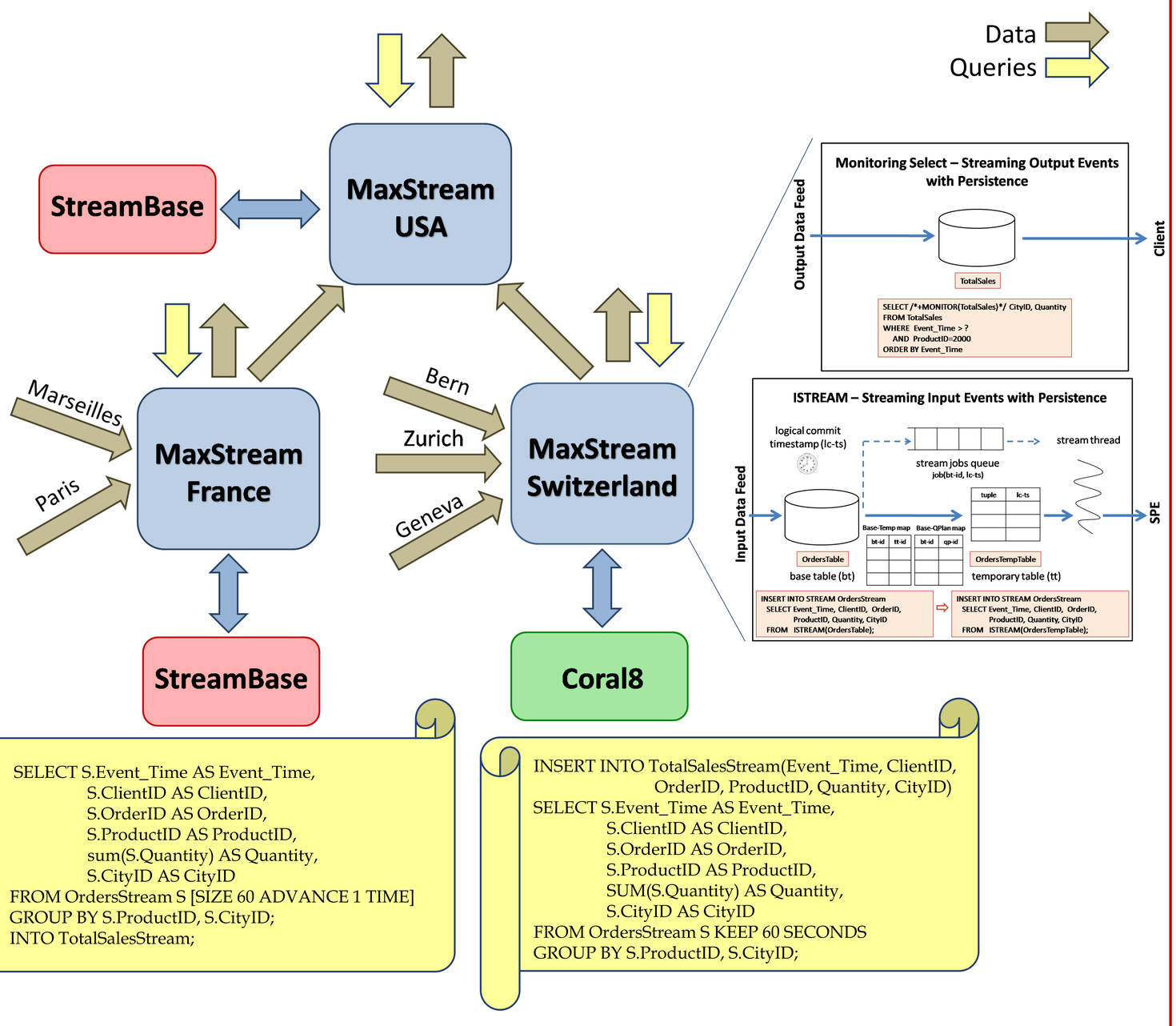
- Capability Differences: handling the differences in the type of queries that different SPEs can support
- Execution Model Differences: handling the differences in the internal query execution models of different SPEs
  - ➢ SECRET Model – A model that explains the windowing behavior of SPEs along five dimensions: **S**cope, **E**valuation, **C**ontent, **RE**port and **T**ick

Management of both Streaming and Stored Data

- ISTREAM: streaming input events through MaxStream in persistent or transient modes
- Monitoring Select: streaming output events through MaxStream in persistent or transient modes
- Native support for hybrid (stream-table) joins within MaxStream

Functionality: Business Monitoring Application Sales Map & Spikes

- Large international company with locations in Switzerland, France, and USA
- At each location, multiple sources of raw operational data: new orders, invoices, scheduling deliveries, etc.
- A different SPE installed at each site, keeping track of local aggregate sales volumes and unusual spikes on a minute-by-minute basis
- Company headquarters located in USA, monitoring the overall business

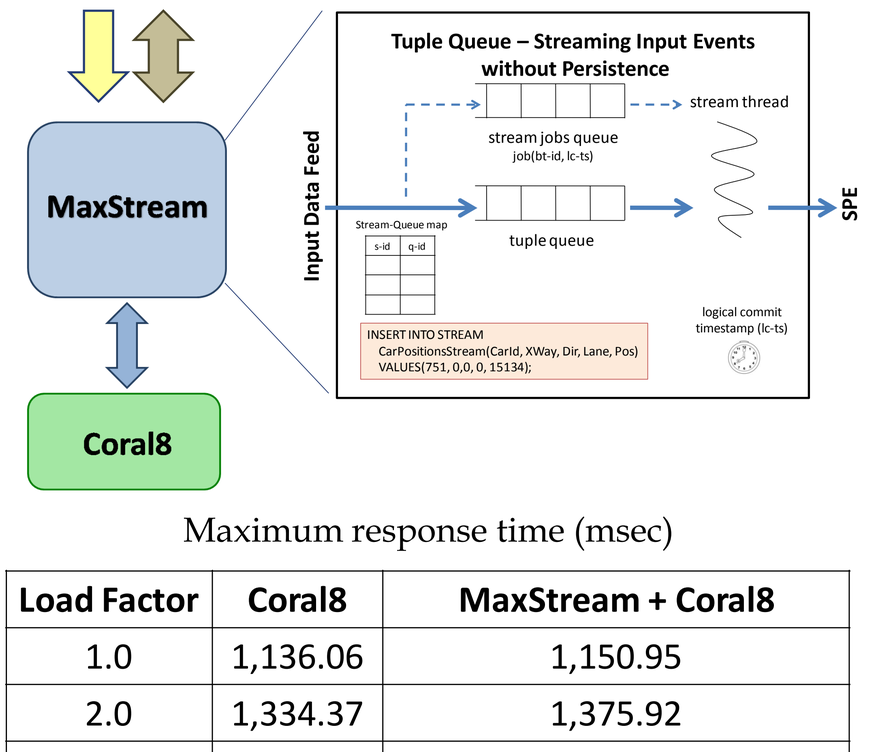



```
SELECT S.Event_Time AS Event_Time,
       S.ClientID AS ClientID,
       S.OrderID AS OrderID,
       S.ProductID AS ProductID,
       sum(S.Quantity) AS Quantity,
       S.CityID AS CityID
FROM OrdersStream S [SIZE 60 ADVANCE 1 TIME]
GROUP BY S.ProductID, S.CityID;
INTO TotalSalesStream;
```

```
INSERT INTO TotalSalesStream(Event_Time, ClientID,
            OrderID, ProductID, Quantity, CityID)
SELECT S.Event_Time AS Event_Time,
       S.ClientID AS ClientID,
       S.OrderID AS OrderID,
       S.ProductID AS ProductID,
       SUM(S.Quantity) AS Quantity,
       S.CityID AS CityID
FROM OrdersStream S KEEP 60 SECONDS
GROUP BY S.ProductID, S.CityID;
```

Performance : Linear Road Benchmark

- LRB simulates traffic on a set of highways determining variable tolling based on segment statistics and accident occurrences.
- Measure of the benchmark: L = the number of highways-worth of data that a given engine can handle while meeting the maximum response time constraint of 5 seconds for all queries.



Maximum response time (msec)

Load Factor	Coral8	MaxStream + Coral8
1.0	1,136.06	1,150.95
2.0	1,334.37	1,375.92
3.0	1,721.79	1,740.73
4.0	2,169.90	2,070.64
5.0	**5,105.93**	**5,416.07**

Number of alerts per time interval for L=5

Response Time Interval	Coral8	MaxStream + Coral8
[0, 1)	90.25%	82.68%
[1, 2)	8.45%	15.86%
[2, 3)	1.07%	1.15%
[3, 4)	0.17%	0.27%
[4, 5)	0.05%	0.04%
>= 5	0.01%	0.02%

This work has been supported in part by the following grants: Swiss NSF NCCR MICS 5005-67322, Swiss NSF ProDoc PDFMP2-122971/1, and ETH Zurich Enterprise Computing Center (ECC) SAP industrial partner grant DE-2008-022.

http://www.systems.ethz.ch/research/projects/maxstream/