**Contribution**

Disco is an approach to distribute the aggregation of complex window types and aggregation functions.

Our main contributions are:

- ... avoids central raw data collection whenever possible.
- ... contains different window merging strategies to distribute common window types.
- ... scales linearly and reduces the network cost significantly compared to state-of-the-art.

**Current Stream Processing**

Cloud Data Center

Result for #1

{ 3 : 3 | 3 : 3 }

Intermediate Nodes

Smart Homes/Devices

**Architecture**

Root Node

Intermediate Nodes

Child Nodes

Window Creator

Window Merger

1. **Window Creator** creates slices and windows according to user-specified queries.
2. **Window Merger** merges incoming windows according to Disco’s strategies.
3. **Root Window Merger** performs final window merge.
4. **Root Window Merger** returns final window aggregate.

**Aggregation Handling**

Decomposable Aggregation

transfer partial aggregates instead of raw events

Holistic Aggregation

transfer slices as smallest unit instead of individual events

**Merging Strategies**

No merging for unique windows.

Merge context-free windows based on identical bounds.

Merge session windows based on global gaps.

**Performance Evaluation**

Disco's Scalability

distributed — linear scale up for both algebraic and holistic aggregation functions

centralized — significant reduction in network cost due to removed redundant raw data transfer

Key Results

Network Impact of 100 Million Events

<table>
<thead>
<tr>
<th>distributed</th>
<th>centralized</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>0.15</td>
<td>0.30</td>
</tr>
<tr>
<td>0.30</td>
<td>0.45</td>
</tr>
<tr>
<td>0.45</td>
<td>0.60</td>
</tr>
</tbody>
</table>

**Data Engineering Systems Group**
Lawrence Benson
Hasso Plattner Institute, Potsdam, Germany
E-Mail: lawrence.benson@hpi.de