

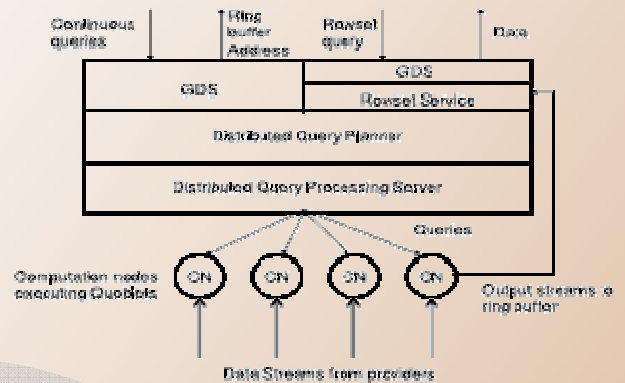
Calder: Enabling Grid Access to Data Streams

Nithya Vijayakumar, Ying Liu, Beth Plale
Indiana University, Bloomington

Calder - data stream access middleware

- Service interface enables programmatic access - the virtual piping of stream directly into a processing application.
- Query distribution that is sensitive to metrics such as minimized global network bandwidth consumption
- Approximation of query results under conditions of stream bursts
- Stream discovery
- Temporal and spatial aggregation operators that minimize CPU and network bandwidth consumption.

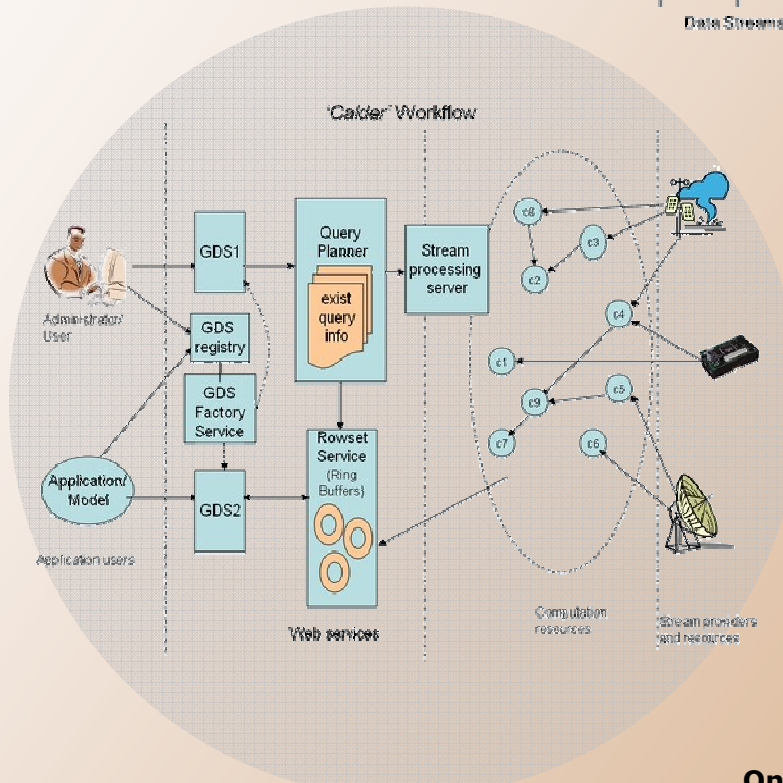
Calder Architecture



Continuous queries

```

SELECT * FROM
  METAR_EV,
  CAPS_EV
START
  2005-07-25 12:00:00
EXPIRE
  2005-07-25 20:00:00
RANGE
  10 minutes
    
```



Rowset Queries

- `getTuple (timestamp, ringBufferID)`
- `getRangeTuple (startT, endTS, ringBufferID)`
- `getMostRecent (lastRecent, num_events, ringBufferID)`

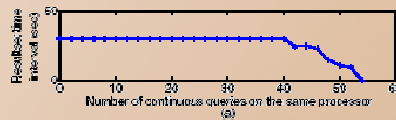
Calder Performance

Query deployment time – Time taken to deploy a new query in the system. Includes planning time, compilation time, distribution time and the time taken to setup ringbuffer.

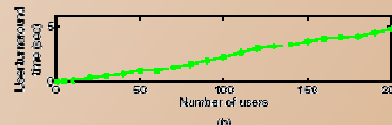
GDS Setup	Planning	Compile	Deployment	RD Setup
2951 us	15 us	24 us	2 us	24 us

Table 1. Query Deployment Time

Resultset time interval – Time interval of data stored in the ringbuffer at a time. Directly proportional to the number of events returned by the rowset query.



User turnaround time – Time taken by the rowset service to process a query. Affected by the number of active queries (and hence ringbuffers) in the system and the number of users handled by the rowset service at the same time.



Ongoing work

- Distributed stream query optimization
- Data stream approximation and accuracy management
- Secure data streams
- Provenance management for data streams


```
ERROR: undefined
OFFENDING COMMAND:

STACK:
```