Tempest: Vector Engine Lights Up Cloud Telemetry Analytics

Min Wei
Microsoft

Cloud Telemetry Analytics

- Telemetry data
  - User and Server metrics
  - 10s of millions of events of per second
  - 10s of TBs per day
  - Over 50% of the DC network traffic
  - Exceeds Wall Street Tick Data
- Analytics
  - Programmatic monitoring queries
  - Near real time interactive and ad hoc analysis of code regression and configuration changes
  - 100s of Experiments/flights per week
  - Data schema is very dynamic
    - Number of columns can be in 1000s

Technical challenges

- OLTP+OLAP
  - Mostly append, occasional updates
- Performance absolute matters
  - Data doubles every a few months
  - Sub-second query response time
  - Query time range spans over weeks
- COGS matters
  - 1000s machines are too expensive
  - Need to squeeze every bit of machine resource
  - 1:100 ration infrastructure vs feature engineering

Tempest

- Vector database
  - OLTP in memory
  - OLAP disk and memory
  - Single threaded per process
  - Integrated with native local file system
  - Run 10s processes per machine
- APL like primitives
  - Designed for data shaping and transformations
  - Beyond SQL in-database analytics
  - Time series queries
- Distributed computation
  - Map-Reduce, Tree
- Integrated with interactive UX
  - Data intellisense
  - JavaScript language integration

Conclusion

- Initial rolling out to Azure 1st party services
- Vector algorithm thinking
  - productivity gain
  - Resource efficiency gain
- Unified OLTP+OLAP database let a small team to support a big engineering org.
- High performance data computing
  - Cheaper and faster than generic solutions
  - Limited by memory, i/o and network bandwidth