Fast Window Aggregate on Array Database by Recursive Incremental Computation

Li Jiang¹, Hideyuki Kawashima², Osamu Tatebe²

1: Google Dublin  2: University of Tsukuba

Window Aggregate over Array Database

[Insight] 1-D Incremental Computation and its Defect

- Naive method
  - For each window, scan all the elements inside it, accumulate and compute the aggregate results
- Problem: redundant calculation exists
- Considering adjacent windows
  - Large overlapping area
  - Few cells are different
- Incremental Computation works as stream processing

[Proposal] Recursive Incremental Computation

- Recursive Dimensionality Reduction
  - Keeping breaking a n-D window aggregate down to multiple smaller window aggregates.
  - Each level has its unique IC dimension.
  - Level 1: n-D task (the original window aggregate)
  - Level 2 (n-1)-D tasks
  - …
  - Level n: 1-D tasks

[Result] Experiment on Earth Science Benchmark

- A real application of earth scientific data analysis [1]
  - Earth Science benchmark [2]
  - Window aggregate operator used to reduce resolution of other analyzing tasks’ results on purpose of visualizing, comparison and further analysis.
- Data: NASA MODIS product
  - 45 MODIS files downloaded
  - Preprocessed, loaded into SciDB cluster
- Task: NDVI (Normalized Difference Vegetation Index)
  - Result: huge 2-D array
  - Average window aggregate to down-sample

[Conclusions]

1. Proposed a method for efficient calculating window aggregates in array databases.
2. Improvement is proportional to total window size W
   A) For a real earth application, recursive method achieved 10 times faster against SciDB method
   B) For a big window case in synthetic test, proposed method achieved 64 times faster

Code is on GitHib

https://github.com/ljiangjl/Recursive-IC-Window

---