Amoeba: Aligning Stream Processing Operators with Externally-Managed State

Antonis Papaioannou [1# *] and **Kostas Magoutis** [1]

[1] Institute of Computer Science – FORTH

Presenting and Corresponding author: Antonis Papaioannou, email: papaioan@ics.forth.gr
* Corresponding author: Antonis Papaioannou, email: papaioan@ics.forth.gr

ABSTRACT

Scalable stream processing systems (SPS) often require external storage systems for long-term storage of non-ephemeral state. Such state cannot be accommodated in the internal stores of SPSes that are mainly geared for fault tolerance of streaming jobs, lack externally visible APIs, and their state is disposed of at the end of such jobs. Recent research has pointed to scalable in-memory key-value stores (KVS) as an efficient solution to manage external state. While such data stores have been interconnected with scalable streaming systems, they are currently managed independently, missing opportunities for optimizations, such as exploiting locality between stream partitions and table shards, as well as coordinating elasticity actions. Both processing and data management systems are typically designed for scalability, however coordination between them poses a significant challenge.

In this work we describe Amoeba, a system that dynamically adapts data-partitioning schemes and/or task or data placement across systems to eliminate unnecessary network communication across nodes. Our evaluation using state-of-the art systems, such as the Flink SPS and Redis KVS, demonstrated 2.6x performance improvement when aligning SPS tasks with KVS shards in AWS deployments of up to 64 nodes.

Amoeba has demonstrated the ability to analyze in real-time nearly 500,000 information records per second using 64 Amazon Web Services Virtual Machines, and further scalability is possible and expected by scaling deployment resources even further. Reaching such levels of throughput through the scalability benefits made possible by Amoeba will have economic and social impact in areas such as social networking, surveillance operations for counter-terrorism activities, financial trading, fraud detection and other stream processing applications. We are currently in the process of demonstrating Amoeba-related technology in the domain of processing financial data. Overall, Amoeba enables real-time event-processing for data analytics applications more efficiently. Coupled with rapid elasticity, it allows such applications to absorb steep spikes of communal activity expected during times of emergency, an important societal concern in recent years.

This work is funded by the Hellenic Foundation for Research and Innovation through the STREAMSTORE faculty grant (Grant ID HFRI-FM17-1998).