WIP: Unifying Tradeoffs in Parallel Data Systems

Xiaotian Yin, Jian Li
Futurewei Technologies
111 Speen Street, Framingham, MA 01701, USA
{xiaotian.yin, jian.li}@futurewei.com

Tingqiu Tim Yuan
Huawei Technologies
Shenzhen, China
ytq@huawei.com

ABSTRACT

Many fundamental challenges in Parallel Data Systems (PDS), and Information and Communication Technologies (ICT) in general, arise from limited resource or capability of communication, computation, storage, etc. In light of these mounting problems, the research community need to understand the essence of the digital world and possess a theoretical framework to guide the endeavor for novel solutions in PDS and ICT.

We propose a framework for studying various communication-computation-storage tradeoff problems in parallel data systems and ICT platforms, with supportive insights from a geometric point of view. We conjecture that there exists a fundamental theory unifying all these core ingredients of ICT. Both tradeoff and unification are indispensable support for parallel data systems and ICT, emphasizing the extrinsic difference and intrinsic similarity accordingly among all the core ingredients. Most of the existing research is on the tradeoff side, while unification remains mostly open. As sush, we propose a theoretical framework for parallel data systems and ICT, featuring a transition from tradeoff to unification, with an emphasis on geometry.

The proposal consists of four key components as follows:

Communication-Computation-Storage (CCS) Tradeoff Framework

- **Problem Model**: Define the class of problems under study, in terms of major components, common pipelines, basic operations, etc.
- **Tradeoff Space**: A 3-dimensional space consisting of the following axis, where each axis represent one factor for tradeoff and a quantitative measurement of cost:
  - L-Axis: Communication Cost
  - C-Axis: Computation Cost
  - R-Axis: Storage Cost

  The **feasible region** refers to the subspace of the whole tradeoff space, where points are achievable by certain tradeoff algorithms.
- **Tradeoff Algorithms**: Each tradeoff algorithm is a refinement of the abstract problem model, implementing the common pipeline with algorithm-specific details. Each tradeoff algorithm, with each parameter fixed, gives a point in the tradeoff space, whose coordinates are the cost of that algorithm along each dimension.
- **Tradeoff Optimality**: Results about the optimality of the tradeoff algorithms, in particular the boundary of the feasible region, called the **optimal tradeoff surface**, which represents what the best tradeoff algorithm can achieve.

The CCS tradeoff framework can be extended into a more general setting, called Multi-Party Tradeoff.

Multi-Party Tradeoff Framework

- **Problem Model**: The same as CCS framework.
- **Tradeoff Space**: Similar to CCS framework, except for having K (K>1) dimensions instead of 3 dimensions:
  - $X^{(1)}$-Axis, $X^{(2)}$-Axis, …, $X^{(K)}$-Axis,
- **Tradeoff Algorithms**: The same as CCS framework.
- **Tradeoff Optimality**: Similar to CCS framework, except for studying the **optimal tradeoff hyper-surface** instead of the optimal tradeoff surface.

Under the high-level abstraction of multi-party tradeoff framework, there are various lower level abstractions, such as Communication-Computation-Storage tradeoff framework, Computation-Communication tradeoff framework (2D), Communication-Computation-Storage-Security tradeoff framework (4D), etc.

The CCS tradeoff framework, as well as its extensions, can be interpreted by a geometric approach. For any K-party tradeoff framework, there are k individual optimal hyper-surfaces, each focusing on optimizing one dimension against the others. We prove that these hyper-surfaces coincide under the condition with Definition-1 and Theorem-1.

Definition-1 (Monotonic Hyper-Surface): Given a (k-1)-dimensional hyper-surface Σ embedded in a k-dimensional space Ω (k>2), Σ is monotonic if one of the following conditions holds: (1) k=2 and the curve Ω is either non-increasing or non-decreasing in any dimension; or (2) k>2 and fixing any dimension at an arbitrary valid value, the resulting (k-2)-dimensional hyper-surface is monotonic.

Theorem-1: For multi-party tradeoffs, the joint optimal hyper-surface exists if one of the individual optimal hyper-surface is monotonic.

Note that Theorem-1 only requires monotonicity but not strict monotonicity, where the latter is a stronger condition. On the other hand, it requires monotonicity but not convexity

Such a super optimal tradeoff corresponds to the coordinate origin of the whole tradeoff space, assuming every axis is normalized to start from its minimum valid value. We prove it with Theorem-2.

Theorem-2: Super Optimal Tradeoff is achievable if and only if the coordinate origin is within the feasible region.

Overall, we believe that a theoretical PDS (and ICT) foundation will come from a joint effort between a systematic tradeoff study and a unifying theory. As such, geometric theories and approaches will naturally play a fundamental role, which have already been demonstrated in the areas from theoretical physics to artificial intelligence.