Contribution ID: 24

Make Your Data FABulous

Tuesday, August 27, 2019 10:40 AM (40 minutes)

The CAP theorem is widely known for distributed systems, but it's not the only tradeoff you should be aware of. For datastores there is also the FAB theory and just like with the CAP theorem you can only pick two:

* Fast: Results are real-time or near real-time instead of batch oriented.

* Accurate: Answers are exact and don't have a margin of error.

* Big: You require horizontal scaling and need to distribute your data.

While Fast and Big are relatively easy to understand, Accurate is a bit harder to picture. This talk shows some concrete examples of accuracy tradeoffs Elasticsearch can take for terms aggregations, cardinality aggregations with HyperLogLog++, and the IDF part of full-text search. Or how to trade some speed or the distribution for more accuracy.

Presenter: KRENN, Philipp (Elastic) **Session Classification:** Plenary