Assignment 1: Design space analysis

Overview

Assignment group size: 1 (Individual)

In this assignment, you will practice the kind of high-level design analysis previously done in class (MapReduce (Week4-Wed.html), Aurora (Week5-Mon.html), Spark (Week5-Wed.html)). Your analysis will have a more specific structure than those done in class, however.

This assignment differs from how you may have been asked to read and analyze papers for more research-oriented classes:

- The papers describe systems in *active production use* at large scale. They are not research prototypes demonstrating a concept. As active production systems, they will draw on many ideas rather than exemplify a single idea. The papers describing them may not have the precise focus of a research paper.
- You will not consider how innovative the ideas are compared to prior work.
- You will focus on the *tradeoffs between competing needs* made by the designers.
- You will skip the sections where the authors justify their claims empirically. (This is more to keep the assignment manageable, though there can be real problems relating a specific research result to your practical needs.)

> **The emphasis of this assignment is how system architects, faced with a large-scale existing need, balanced competing design goals to address that need.**

In many cases, they were forced to design a new system because no available system at the time could handle their latency or throughput objectives.

To the extent practical, set aside your "research-reading" habits from assignments in other courses and exercise the rather different skills required for this one.

What you will gain from this assignment

This assignment develops a skill useful in your professional career. When your team admits that you have an unmet need, you are faced with a "buy / modify / build" choice amongst many contending products, both commercial and open-source. Some choices will address your need with minimal effort, others will do so with extreme effort, while still others will not ever solve your problem. Making a good choice—there will likely not be a "best" choice—will have an enormous effect on the quality of your work life. Anlyses of this type are a key component of making a good choice amongst the alternatives.

What you will do

Select any **one** paper from the list at the end of this page. Read it with the following questions in mind:

1. What was the key problem they were addressing? Latency? Throughput? Durability? Note that they might describe these concepts in terms different from those used in class.
2. Who or what is the immediate source of the data coming in to this system? Another system? Users? If users, what is their role?
3. Who or what consumes the data from this system? Another system? Users? If users, what is their role?
4. Does the system store data persistently (long-term) or does it only keep temporary copies of data?

The analysis focuses on tradeoffs in the design, so skip the following sections:

1. Previous work / Related work
2. Evaluation / Experiments
3. API / Interface (though you may need to refer back to this to understand some of the sections that you do read)

You're welcome to come back and read these sections later, if you're interested, but I recommend focussing on the pertinent sections while doing this assignment.

Some suggestions about reading for this assignment:

- First skim the article, noting the sections to read and which ones to skip. Note any questions about what is going on.
- **Do not expect to understand it all the first time through**.

- Focus on the larger structure and ignore the small points. This assignment is about the *design*, not the details.
- If they use a term that you do not understand, particularly a concept or product name that is unfamiliar to you, just note it down but do not look it up immediately. Only look it up later if understanding it seems essential to understanding the design.
- Focus on the relationship between the parts, not the details of the parts.

The challenging part of this assignment is staying fixed on the overall design rather than the details, seeing the forest rather than individual trees.

Structure of the analysis

Your analysis must be short and to the point: two–three pages long. It should have the following structure:

1. What does the system do? In one or two paragraphs, describe its purpose.
2. What is the context? How does this system fit into the larger processes of the organization? What activities does it support? (1 paragraph)
3. What is the most important goal for the design? Low latency (median? high percentiles?), throughput, durability, fault tolerance something else that we have discussed in class? (This section and the next should constitute the bulk of your paper.)
4. What property or property did the designers sacrifice to reach that goal? What was the key tradeoff? (Note that authors sometimes do not want to state this too clearly. You might have to tease it out.)
5. What is an example use case for which this system would not meet its goals, or for which other goals would be more important? (1–2 paragraphs)

Submission

Length: 2–3 pages.

Total points: 10

Submit a PDF of your analysis to CourSys (https://coursys.sfu.ca/2019sp-cmpt-756-g1/+a1design).

List of papers

Choose one of the following:

- Spanner: Becoming a SQL System (https://ai.google/research/pubs/pub46103)
- Cassandra: A Decentralized Structured Storage System (http://dl.acm.org.proxy.lib.sfu.ca/citation.cfm?id=1773912.1773922)
- Bigtable: A Distributed Storage System for Structured Data (http://dl.acm.org.proxy.lib.sfu.ca/citation.cfm?id=1298475)
- F1: A Distributed Database that Scales (https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/41344.pdf)
- MillWheel: Fault-tolerant Stream Processing at Internet Scale (http://dl.acm.org.proxy.lib.sfu.ca/citation.cfm?id=2536229)