6.824 Spanner FAQ

Q: What is time?

A: 13:00:00.000 April 7th 2020 UTC is a time. Time advances at a steady rate of one second per second.

Q: What is a clock?

A: A clock is an oscillator driving a counter. The oscillator should tick at a steady known rate (e.g. one tick per second). The counter must initially be synchronized to a source of the correct time; after that, the oscillator's ticks cause the counter to advance the time. If the oscillator ticks at precisely the right rate, the counter will advance in precise synchrony with the correct time. In real life, the oscillator ticks at a varying and somewhat incorrect frequency, so the counter gradually drifts away from the correct time.

Q: What time is it?

A: In order to provide a notion of univeral time (UTC), a bunch of government laboratories individually maintain highly accurate clocks, and compare with each other to produce a consensus on what time it is. The current time is broadcast in a variety of ways, such as WWV, GPS, and NTP, so that ordinary people can know the official time. Those broadcasts take a varying and hard-to-predict amount of time to reach listeners, so you never know the exact time.

Q: What is an atomic clock?

A: A highly stable oscillator. There are two main technologies that go by the name "atomic clock": rubidium clocks and cesium clocks. Both exploit changes in the state of the outer electron, which involve specific quanta of energy and thus wavelength. One can tune a signal generator to precisely that wavelength by watching how excited the electrons are. An atomic clock is just the oscillator part of a clock; on startup, it must be synchronized somehow to UTC, typically by radio broadcasts such as GPS.

Q: What kind of atomic clock does Spanner use?

A: Sadly the paper doesn't say. Rubidium clocks are typically a few thousand dollars (e.g. https://thinksrs.com/products/fs725.html). Rubidium clocks drift by perhaps a few microseconds per week, so they need to be re-synchronized to UTC (typically by GPS) every once in a while. Cesium clocks cost perhaps $50,000; the HP 5071A is a good example. A cesium clock doesn't drift. Of course, any one clock might fail or suffer a power failure, so even with perfect cesium clocks you still need more than one and the ability to synchronize to UTC. My guess, based on price, is that Spanner uses rubidium clocks that are synchronized with GPS receivers.

Q: How does external consistency relate to linearizability and serializability?

A: External consistency seems to be equivalent to linearizability, but applied to entire transactions rather than individual reads and writes. External consistency also seems equivalent to strict serializability, which is serializability with the added constraint that the equivalent serial order must obey real time order. The critical property is that if transaction T1 completes, and then (afterwards in real time) transaction T2 starts, T2 must see T1's writes.

Q: Why is external consistency desirable?

A: Suppose Hatshepsut changes the password on an account shared by her workgroup, via a web server in a datacenter in San Jose. She whispers

the new password over the cubicle wall to her colleage Cassandra. Cassandra logs into the account via a web server in a different datacenter, in San Mateo. External consistency guarantees that Cassandra will observe the change to the password, and not, for example, see a stale replica.

Q: Could Spanner use Raft rather than Paxos?

A: Yes. At the level of this paper there is no difference. At the time Spanner was being built, Raft didn't exist, and Google already had a tuned and reliable Paxos implementation. Have a look at the paper Paxos Made Live by Chandra et al.

Q: What is the purpose of Spanner's commit wait?

A: Commit wait ensures that a read/write transaction does not complete until the time in its timestamp is guaranteed to have passed. That means that a read/only transaction that starts after the read/write transaction completes is guaranteed to have a higher timestamp, and thus to see the read/write transaction's writes. This helps fulfil the guarantee of external consistency: if T1 completes before T2 starts, T2 will come after T1 in the equivalent serial order (i.e. T2 will see T1's writes).

Q: Does anyone use Spanner?

A: It's said that hundreds of Google services depend on Spanner. The paper talks about its use by Google's advertising system. Google's Zanzibar Authorization system uses Spanner. It's offered as a service to Google's cloud customers in the form of Cloud Spanner. The CockroachDB open-source database is based on the Spanner design.