

Each 6.1800 lecture will come with an outline. You can fill this in during lecture, after lecture, or not at all — it's entirely up to you how you use it. The goal of these outlines is to help you understand the main points that you should be taking away from each lecture. In some cases we will also include examples of things you should be able to do after each lecture.

*In the past, these outlines have proved to be an effective tool for studying for the exams. Note that the outlines are **not exhaustive**; there will be topics and nuances in lecture that aren't captured by the outline.*

Lecture 19: Distributed Transactions

- Today we're working with a system that has distributed data across multiple machines. What is the main problem that we're going to have to solve? (I.e., what becomes a problem now that we've introduced multiple machines that wasn't a problem before?)
- How do we deal with the fact that messages might be reordered by the network?
- What happens in a two-phase commit (2PC) protocol when there are no losses or failures?
 - What messages does the coordinator send to the machines? To the client?
 - What is the commit point?

We're going to spend most of this lecture talking about different failures that can occur during 2PC. You should understand how the protocol handles each of the following types of failures.

- Failures before the commit point. Note that in these cases, it is safe to abort the transaction.
 - Lost prepare message from the coordinator
 - Lost ACK for a prepare message
 - Worker failure before prepare
- Failures after the commit point. Note that it is *no longer safe to abort*.
 - Lost commit message
 - Lost ACK for a commit message
 - Worker failure before receiving commit
 - How do we ensure that workers complete the transaction?
 - Worker failure after receiving commit
- Coordinator failures
 - Coordinator failure before prepare
 - Coordinator failure after the commit point, but before writing DONE
 - Coordinator failure after writing DONE
- Broader issues
 - Why does this protocol have two phases instead of just one?
 - What are the performance issues with 2PC? What causes them?
 - What problem *haven't* we solved yet?