Seminar 2: Monitoring DS - https://sre.google/sre-book/monitoring-distributed-systems
Learn from Google how they monitor their services.
Keywords: monitoring, white-box, saturation, resolution, *whatever word you didn't understand* etc.

- **What is the paper about?**

This article investigates what system monitoring is, including its plusses and minuses.

- **What is monitoring?**

Monitoring is the process of gathering and analyzing quantitative information about a system, such as query time, service demand, and so on.

- **Why monitor a system in the first place?**

Monitoring is critical since it helps us to identify defects in system communication, identify bottleneck services, and understand how to scale the system appropriately using empirical data.

- **Explain the 4 golden signals of monitoring.**
    - **Latency** – A bug might result in increased time of response.
    - **Traffic** – Detecting either amount of requests or the amount of data transferred in streaming connections.
    - **Errors** – Unusual rate of errors might indicate a poor quality of service.
    - **Saturation** – Calculating the load on the service might be indicative about scaling opportunities. Especialy this would be useful for notifying about database running out of space.

- **According to the paper, how do you do monitoring? What is important? Exemplify.**

It's critical to choose and examine data. For example, a typical error is averaging out the duration each request in a 5-second span. This is a chance for a large processing time to be hidden among several small ones. A better method would be to divide them into time intervals and count how many were between [0.1 and 0.3], [0.3 and 0.7], and so on.

- **What approach would you use for your lab: White-box or Black-box monitoring? Why?**

The optimum option is a mix of the two. It's useful to have a summary of the system's availability for the client (black-box) as well as a visual of what's going on within (white-box).

- **What happened with Bigtable SRE and how did they "fix" the situation?**

When the SLO was approaching, email notifications were sent, and when the SLO was exceeded, paging alerts were sent. Both sorts of warnings were generating an excessive amount of engineering effort: the team had to spend a substantial amount of time triaging the alerts to discover the few that were truly actionable, and we frequently missed the problems that genuinely harmed customers because there were so few of them.
To fix the problem, the team used a three-pronged approach: while working hard to enhance Bigtable's performance, we temporarily reduced our SLO objective to the 75th percentile request latency. They also turned off email notifications since there were so many that diagnosing them was impossible.