What is the paper about?

The paper is about the process of monitoring a service through a system , and a vast explanation of how the monitoring system works at Google.

What is monitoring?
Monitoring is the process of “checking” a system, verifying it’s parameters and real-time work process, in order to maintain it to function properly, without bugs, problems and errors.

Why monitor a system in the first place?
You should monitor a system because firstly, it is easier to find a problem in case of malfunctions when you monitor the whole working process. Secondly, it helps you to be in touch with the long-term developing and working process of the system. It is much easier to understand what to change or add in your system to improve it.

Explain the 4 golden signals of monitoring.
Latency: it is about the time request takes to be sent. You should always be aware of the error latency, ergo in case a failed request causes a error to appear slowly, it sounds like your system doesn’t work efficiently.

Traffic: The measurement of the demand on your system. Ex: how many times was your website accessed.

Errors: The failed requests percentage.

Saturation: How loaded are the resources of the system.

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

Well it actually depends on each of your system individually but , according to Google you should use simple predictable and reliable rules that the catcher errors and malfunctions, the data collection application that is not really used should be removed and all the signals that are not exposed and used to an alert also should be removed. Also Google advises to use basic collection and aggregation of metrics in combination with alerting and dashboards for a good monitoring system.

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

As I understood the black box monitoring is actually solving the problems that are already happening so I would definitely use the white box. Why solve the problem when you can prevent it?

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

Because of the huge amount of email alert that they were receiving, the engineera were spending a lot of time solving minor problems or errors and missing other problems that caused crucial damage to the user interaction. They solved this problem by using the 75 percentile request latency and disabling the email alerst so that they could’ve had more time to fix the longer-term problems of the system.