**Logging and monitoring in the cloud**

- (The importance of monitoring) Monitoring is crucial for product reliability, providing real-time data about system performance and usage patterns. It enables continuous operations, trend analysis, and improved incident response. Transparency through postmortems builds trust with clients. Automated alerts and monitoring tools aid in debugging and ensuring continual improvement of products. Google's integrated monitoring tools will be covered later in the course.

- (Mesuring performance and reliability) The four golden signals that measure a system's performance and reliability are:

1. Latency: Measures how long it takes for a specific part of the system to return a result. It directly impacts user experience and can indicate emerging issues or capacity demands. Sample metrics include page load latency, query duration, and service response time.
2. Traffic: Measures the number of requests reaching the system. It indicates system demand, aids in capacity planning, and is essential for calculating infrastructure spend. Sample metrics include HTTP requests per second and the number of concurrent sessions.
3. Saturation: Measures how close the system is to its capacity limit. It focuses on constrained resources and indicates degrading performance as capacity is reached. Sample metrics include CPU utilization, memory utilization, and disk utilization.
4. Errors: Track system failures or issues, indicating potential failures, configuration problems, or service level objective violations. Sample metrics include the number of failed requests and 400/500 HTTP codes.

- Monitoring these golden signals helps ensure the system's health, performance, and reliability, allowing for timely identification and resolution of potential issues.

- (Understanding SLIs, SLOs, and SLAs)

+ SLIs (Service Level Indicators) are monitoring metrics that measure one aspect of service reliability, ideally reflecting user experience. They are expressed as the ratio of good events to all valid events.

+ SLOs (Service Level Objectives) combine SLIs with a target reliability, usually just short of 100 percent (e.g., 99.9% or three nines). SLOs should be specific, measurable, achievable, relevant, and time-bound to be effective.

+ SLAs (Service Level Agreements) are commitments made to customers about the maximum allowable downtime for systems and applications. They define the minimum service levels promised to customers and the consequences for breaking that promise.

To improve service reliability, all stakeholders must agree that SLIs, SLOs, and SLAs accurately represent user experience and use them as the primary driver for decision-making. Being out of SLO must have well-documented consequences, just as there are consequences for breaching SLAs. Strong executive support is essential for enforcing these consequences and driving change in development practices to meet SLO targets effectively.

- (Integrated observability tools) Google Cloud's integrated observability tools provide precise insights into the health and performance of applications and services. Signals, including metric, logging, and trace data, flow into Google Cloud Operations tools for visualization and analysis. The Logs Explorer dissects and analyzes logs, while health checks monitor uptime and latency. Error Reporting helps identify and analyze crashes. Automated alerts notify key personnel during incidents. Adhering to Service Level Objectives ensures reliability. Google Cloud offers comprehensive monitoring and diagnostics, surpassing on-premises environments. Upcoming videos will explore specific tools for monitoring, logging, error reporting, and debugging, essential for operations roles.

- (Monitoring tools) Monitoring in Google Cloud starts with signal data, and metrics are used to align measurements over time. Google Cloud collects over 1,000 metric streams by default, providing visibility into performance and health. Data scientists benefit from insights into BigQuery queries, while DevOps teams track CPU and memory utilization for containerized apps in Cloud Run. Custom metrics using OpenTelemetry can be added. Cloud monitoring ingests data from various sources, generating insights through dashboards, charts, and alerts.

- (Logging tools) Google Cloud's integrated logging tools, Cloud Logging, allow users to collect, store, search, analyze, and monitor log entries and events. Automated logging is integrated into various Google Cloud products. Logs can be exported to different destinations for analysis. Audit logs track user actions, agent logs collect data from instances, network logs offer network security telemetry, and service logs capture logs from deployed code. Cloud Logging provides centralized log management for Google Cloud projects.

- (Error reporting and debugging tools) Google Cloud offers error reporting and debugging tools:

1. Error Reporting: Counts, analyzes, and aggregates crashes in cloud services. Displays error details and allows alerts for new errors.
2. Cloud Trace: Collects latency data from distributed applications and provides near-real-time performance insights.
3. Cloud Profiler: Provides a complete CPU and heap picture of applications without performance impact. Supports various platforms and languages.