

City University of New York (CUNY)

## CUNY Academic Works

---

Open Educational Resources

Hunter College

---

2020

### CSCI 49378: Lecture 11: Logging, Monitoring, and Advanced Topics

Bonan Liu  
*CUNY Hunter College*

NYC Tech-in-Residence Corps

[How does access to this work benefit you? Let us know!](#)

More information about this work at: [https://academicworks.cuny.edu/hc\\_oers/22](https://academicworks.cuny.edu/hc_oers/22)

Discover additional works at: <https://academicworks.cuny.edu>

---

This work is made publicly available by the City University of New York (CUNY).  
Contact: [AcademicWorks@cuny.edu](mailto:AcademicWorks@cuny.edu)

# Logging, Monitoring & Advanced Topics

Bonan Liu

Tech-In-Residence Member, Hunter College, CUNY

Spring 2020



# Disclaimer

*The content of this presentation is being provided for educational and informational purposes only. The views, thoughts, and opinions expressed in this presentation belong solely to the author, and not necessarily to the author's employer.*

*The content of this presentation is not endorsed by the author's employer.*

- Site Reliability Engineering
  - Deployment
  - Logging
  - Monitoring
- Advanced Topics
- Course Review

Site reliability engineering is method of service management where software engineers run production systems using a software engineering approach.

## Traditional Engineering Phases:

- Development
- Operations (Ops)

## Site Reliability Engineering:

- 50% traditional operations
- 50% improvements to existing systems

Site Reliability Engineering starts from design phase of development.

### Goals of site reliability engineering:

- Use engineering approach to solve operational problems
- Maximum development velocity while maintain health service
- Plan changes and capacity
  - Change management
  - Demand forecast and resource management
  - Performance and Efficiency
- Emergency response

## Key metrics for site reliability engineering:

- Service Level Agreement (SLA)
- Service Level Objective (SLO)
- Uptime
  - 99.9%
  - 99.99%
  - 99.999%
- Performance
  - Throughput
  - Latency: P99, P95, P90 and P50



### Key metrics for site reliability engineering:

- Service Level Agreement (SLA)
- Service Level Objective (SLO)
- Uptime
  - 99.9%
  - 99.99%
  - 99.999%
- Performance
  - Throughput
  - Latency: P99, P95, P90 and P50

### Key metrics for site reliability engineering:

- Health of the deployment
  - Time to production
  - Number of changes in a deployment
- Error Budget
- Resource Management
  - Footprint
  - Usage
  - Headroom

## Principles for deployment management

- Isolation between production and deployment environment
  - Production
  - Staging
  - Dev
  - Single-box

## Principles for deployment management

- Incremental deployment with strict policies
  - Error budget control
  - Cooling period
- Fully automated deployment with minimum human involvement
  - Major workflow should be fully automated
  - Humans involvement should be banned in general

## Principles for deployment management

- Always plan for rollback situation
  - Rollback as a step of deployment
  - All changes should have a rollback plan
- Special consideration for data deployment

## Different Logs

- By storing format
  - Structured Logs
  - Text Logs
- By severity type
  - Debug Logs
  - Info Logs
  - Warning Logs
  - Error Logs

## Different Logs

- By generation source
  - System Logs
  - Application Logs
- By purpose
  - Audit Logs
  - Transparency Logs (AXT)

## Key challenge in logging design

- Storage space
- Security
- Retention period
- Proper level of logging
  - Issue of logging too much
  - Issue of logging too little
- Efficiency of querying logs



## Demo: Google Cloud Logging Explorer

Monitoring refers to a set of tooling for engineers to observe, evaluate and manage the cloud-based services, applications, and infrastructure.

Alerting refers to the mechanism to notify the engineers the problem in cloud-based services, applications and infrastructures.

- Source of the monitoring data
  - Logs
  - Time-series based metrics
  - Probers
- Time series data
  - Timestamp-Value tuple
  - Tuples are put in to different buckets
- View of the monitoring data
  - Aggregation of the value in different buckets

## Key challenge of monitoring design

- Monitoring is not real-time. It's near real-time.
- Storage challenge
  - Sampling rate
  - Downgrade over the time
  - Retention period
- Usefulness of monitoring

## Key challenge of alerting design

- Alerting delivery failures
  - Email
  - Text
  - Instant Message
  - Phone Call
  - IRC
- Alerting at the right threshold

Demo: Google Cloud Alerting Policies.

TLA+ is a high-level language for modeling programs and systems - especially concurrent and distributed ones.

TLA+ is an **algorithm** language, not a **programming** language.

TLA+ has two components:

- Models in specification language
  - Possible initial variables (states)
  - The relationship between current state with next state(s)
- An open-sourced toolbox



Cloud Key Management Service is a fully managed service to generate and manage cryptographic keys.

Options to store sensitive (or even insensitive) data:

- Plain text
- Simple hashing function (for example, MD5 or SHA1)
- Hash with an encryption key (for example, HMAC)
- Hash with an encryption key per user

## Basic Concepts:

- Key Rings
- Keys
- Key versions
- Key rotation

Demo: Key generation and key rotation.

Cloud ML, Cloud AI and Cloud AutoML are hosted machine learning solutions for customized use cases.

Two major problems in machine learning:

- Regression
- Classification

Machine learning problem could be solved by different models.

Two major procedures in machine learning:

- (Offline) training
- (Online) inference

The computational resources required for two procedures are different.

Cloud Machine Learning provides easy model defining, powerful model training and fast inference service.

Cloud AutoML provides model training from small sample of data set.

Cloud AI provides a set of integrated APIs for machine learning problems.



- Distributed System Concepts
  - Hardware/Software/Logical elements
  - Communication
  - Key Challenges
  - CAP

- Distributed System Patterns
  - Primary/Replica
  - Master/Worker
  - Sharding
  - Consistent Hashing
  - Two-Phase Commit

- Replication
- Consistency
- Synchronization
- Distributed File System
- Distributed Applications

- Cloud Computing Concepts
  - IaaS
  - PaaS
  - FaaS
  - SaaS
- Different Cloud Types

- Cloud Computational Products
  - GCE - Virtual Machines
  - GKE - Containers
  - GAE - Applications
  - GF - Functions

- Cloud Storage Products
  - GCS - Blob Storage
  - Cloud SQL
  - Cloud Spanner
  - Cloud Datastore
  - Cloud Bigtable

- Cloud Networking
- Cloud Monitoring
- Cloud Logging
- Cloud KMS
- Cloud Machine Learning

CSCI 49378 delivers two parts: distributed system and cloud computing. Hopefully you learnt:

- Concepts of distributed systems and cloud computing.
- Established patterns of high-performance and large-scalable systems.
- Major cloud computing product to build a service rapidly.
- Basic knowledge of system design interview.



*Site Reliability Engineering: How Google Runs Production Systems.*  
Chapter 1 & 2. <https://landing.google.com/sre/books/>.