

Problem: It takes too long time to figure out what is wrong when system goes down.

Suggestion 1: Using automatic management and provisioning

Automatic provisioning and installation.

- Provisioning and installation manually will be costly and error-prone.
- Small configuration differences make problem determination much more difficult.

Configuration changes made in production must produce an audit log record.

- Frequently scan all servers to ensure their current state matches the intended state.
- Helps catch install and configuration failures, detects server misconfigurations early, and finds non-audited server configuration changes.

Recover at the service level.

- Handle failures and correct errors at the service level, not in lower software levels.
- For example, build redundancy into the service rather than depending upon recovery at the lower software layer.

Fail services regularly.

- Take down data centers, shut down racks, and power off servers.
- Regular controlled brown-outs will go a long way to exposing service, system, and network weaknesses.
- Without production testing, recovery will not work when called upon.

Problem: The dev team only found system down after customers called the CEO directly.

Suggestion 2: Using auditing, monitoring and alerting

Configurable logging.

- Support configurable logging that can optionally be turned on or off as needed to debug issues.
- Having to deploy new builds with extra monitoring during a failure is very dangerous.

Track all fault tolerance mechanisms.

- Track every time a retry happens, or a piece of data is copied from one place to another, or a machine is rebooted or a service restarted.
- Avoid fault tolerance hiding little failures, minimize possibility for them to become big failures.

Audit all operations, keep historical data and record all significant actions.

- Every time somebody does something, especially something significant, log it.
- Historical performance and log data is necessary for trending and problem diagnosis.
- Having important action record helps immensely in debugging problems.