Integrated Operations

From HUIT Architecture Advisory Group

Contents

[hide]

Integrated Operations

HUIT is a large organization with many different sub-groups that must work together to ensure that our services are reliable and perform up to expectations while at the same time keeping costs down. The objective of this section of this site is to outline technologies and operational procedures that should be used to help us achieve these goals.

Integration of information from different Git Repositories

As the title suggests, it is assumed that Git is the de facto standard for storage of information that will be used to configure our services and all the software and hardware that work together to support them. Since Git is a flexible tool we have to adopt some conventions about branching, code review, quality and other operational rules of the road to ensure smooth operation.

Check-in Requirements

Requirements for Git Repositories

Puppet Module Conventions

Puppet is a key component for the configuration of much of our infrastructure up through to and including the application layer. To ensure quality modules in our environment modules used for production must minimally meet the following criteria:

1. Module Fundamentals provides a good introduction to Puppet and the elements of a module. It also includes information and pointers to best practices for their construction.
2. There is also documentation about dependencies, versioning and other considerations to think about before publishing a module on Puppet Forge. Those rules represent the minimum set of considerations that must be applied before checking in a module in HUIT for use in a production environment. Note that it is not a requirement at this time that Modules be Published on Puppet Forge, only that they meet the publication standards.
3. Independent source code review - this is a common convention in many environments. It is noted here in recognition of the fact that Puppet modules are code and an independent code review is required before modules may be put into production or changed.

Common Infrastructure Scaffolding - The API

Have to flesh out more.

```
class iam_ldap_server (
  <parameters go here>
) {
  # 389_server is the module written by IAM
  class { '389_server':
    ensure  => 'present',
    require => Class['server'],
    <other parameters go here>
  }
  # server is the base server configuration
  # module written by Unix-Linux Systems
  class { 'server':
    <other parameters go here>
  }
}
```

haag_puppet.md hosted with ❤ by **GitHub** **view raw**

The Change Control Process

Incident Response

Incident response is a core function of any operations group; however, in an integrated operations group, incident response extends beyond the traditional role of mitigating operational degradations. An integrated operations group seeks not only to resolve problems when they occur, but also to foresee and prevent problems before they occur, and also to collaborate with developers and application owners in performing root cause analysis after problems have been resolved.

Prediction and Prophylaxis

say something about trend analysis, capacity planning, dealing with component failures that don't cause service degradation

Ad-hoc Remediation

say something about ITSM, monitoring that rapidly identifies the failed system component, fault tolerance, reverting to known good state, captured configurations, rapid deployment of replacement hosts

Incident Investigation

While 100% service uptime is a goal, failures occur. What we do, during and following the incident, can impact the likelihood of that or similar incidents taking place in the future.

While the Incident is Ongoing

Before a system or sub-system is restarted, as much data as possible should be collected from elements in the system that would be lost in a reboot. These data can be critical to the failure analysis. The data collection effort should be coordinated by the technical team to ensure that all aspects of the system have been investigated. This includes everything from the virtualization layer information (where appropriate) to the application layer software and all the infrastructure elements that connect them.

Enable additional logging, tracing and other data collection methods before restarting any components. Whether a 'smoking gun' has been found or not, the agreed log levels and additional data collection profile identified for the application after an incident should be enabled.

After the Incident is Resolved

After an incident that involved a customer facing service is resolved, a technical team should convene to perform a post-incident investigation. The goal of this investigation is to provide specific, concrete, actionable and as complete as possible answers to the following questions:

What service(s) were degraded during the incident?
: When an incident occurs, the extent/root cause of the degradation is not always obvious. Sufficient time for data collection and analysis must be provided to ensure a thorough analysis so that remediation actions are focused in those areas where they will have the best impact. Particular attention should be paid to the interrelationships and dependencies between elements in our environment.

What failure(s) of system component(s) caused the service degradation?
: "Failure" in this case may mean something other than complete loss of function; for example, a database cluster that performs poorly under load may cause service degradation or outage, even if it is technically still responding to requests. This portion of the investigation should be focused on root cause analysis, for example, if a web service is degraded because a web application server performed poorly under load, and that poor performance was due to poor performance of the backend database cluster under load, and that poor performance was due to poor performance of the filesystem on which the database files resided, and that poor performance was due to the filesystem being delivered via NFS on a network without jumbo frames enabled, and that jumbo frames were not enabled on this network because of the difficulty in scheduling a switch downtime in order to make the configuration change, then it is important both that the investigation discover the root cause and also that the investigation capture the entire chain of dependencies to aid in decision-making.

Why did the failure(s) result in service degradation?

> Every component of a system will fail under some given set of conditions. From this it follows that in order to reduce the frequency of service degradation, a service platform should be designed with redundancy and fault tolerance of as many components as possible. As with the previous question, it is important that the investigation proceed until it uncovers root causes; for example, if an application is deployed on a standalone server instead of in a distributed configuration, it is important to discover whether this deployment exists because of budgetary considerations, or whether the application in question cannot be deployed in a distributed configuration, or whether there is some other cause.

What steps should be taken, and by whom, to reduce the likelihood of a similar degradation occurring in the future?

> This is the most important question in the post incident investigation; but without the previous questions this one cannot be answered. A successful post-incident investigation produces concrete, actionable findings and recommendations. These findings and recommendations should be presented to management at the conclusion of the post-incident investigation; the point of this presentation is that it highlight areas where improvement is needed. These areas of improvement alone, or in combination may include, allocation of additional resources for servers etc., operational procedures, improvements in tools used to operate the environment, etc.

The Technical Team

In order for the team to answer the questions above, they need data collected during and after the outage. A preliminary meeting may be held to review steps already taken to take to collect data, and make recommendations for adjustment of the data collection approach that will allow definitive answers to the above questions if needed. That would be followed up with another meeting once sufficient data has been collected to answer the questions.

The team should consist of technical representatives of as many functional groups as necessary to perform a meaningful investigation; however, the number of team members should be kept as small as possible. The team should independently investigate the problem and recommend solutions.

Team are organized along service lines; a team will exist for each of the main services such as PIN, iSites, Payroll, PeopleSoft, etc. When formed, these teams are permanent but are dormant except when a major incident requires activity. These teams have the following additional attributes:

1. They are self governing - that is they meet as they feel appropriate.
2. They appoint their own team leader for coordination and self-goverance.
3. Team members may call on other technical resources in HUIT to help identify and remediate root causes for incidents.
4. The team leader is responsible for coordinating with the Service Management area.

Reporting and Analysis

The objective of all this work is to identify root causes and recommend actions that would prevent further similar (and potentially other) failures. It is not expected that a 'report' is filed for each incident, but sufficient information must be captured that clearly identifies the chain of events to the root cause with specifics abut how to avoid this. In the instance of a repeat incident, the previous recommendations should be included. This information should also be made available for the Service Management system, even if in abbreviated form. Over time this can become a knowledge base that support people may use to triage problems.

The frequency of update to the Service Management Team shall be not less than monthly.

1. An on-line (searchable) repository that final reports are recorded in.
2. Quarterly reviews of common causes of failures.
3. Executive Jim/Sara sponsorship of meetings and process.

Data Collection

The teams identified above are in the best position to identify specific data collection items for:

1. Default data collection in a normal state production environment.
2. Recommendations for data collection/logging when searching for root causes.

The teams are responsible for identification of these profiles and recording them so that operations teams can put them into effect as required. These profiles should as a minimum conform to manageability standards appropriate to the systems in their domain. For details on the management information that should be made available and collected from elements in the environment see the section on Manageability Standards and Requirements.

Retrieved from "https://wikis.fas.harvard.edu/huitarch/Integrated_Operations"