Overview

This document describes a solution/system/algorithm for scalable and reliable capacity-management discovery, collection, and transfer using a dynamic proxy solution administered by a central console. Proxies can be added/cloned to reduce the number of physical and virtual systems collected at each individual proxy. Discovery is only performed once at a single proxy. Discovered entities are distributed to multiple proxies. The work of a non-functional proxy can be redistributed to the existing functional proxies.

Background

Cloud Environments can consist of a very large number (100s-1000s) of physical systems each of which running a large number (10s-100s) of virtual machines. Capacity management deals with performance and throughput (load) of physical and virtual systems. Discovery, data collection, and transfer of capacity-management data for the cloud is challenging due to the large number of dynamic elements. A successful capacity-management system must achieve the following objectives:

- The measurement must not impact what is being measured.
- It takes a small cloud to collect from a large cloud.
- The measurements must be made regardless of whether or not the physical and virtual machines are centrally managed.
- The measurement must not be impacted by outages of the collection infrastructure.

The challenge is in achieving all of these objectives.

Solution

A Central Capacity-Management Console that controls a list of proxy servers to perform discovery, collection, and transfer is an approach that effectively addresses the challenges with achieving the aforementioned objectives. Discovery of Physical Entities is done via
a single proxy. Splitting up the discovered entities among the available proxies is done at
the Central Console. Collection Requests for Capacity-Management Data for each proxy
are sent from the single proxy. Each proxy issues collection requests independently of the
Central Capacity-Management console. Each proxy persists the collected data. Data is
sent back to the Central Console for further analysis and long term persistence.

Discovery of physical and virtual systems is executed against the first proxy server on the
list. The proxy host contacts either a Central Management Cloud Console (Figure 1) or
each individual physical and/or virtual element (Figure 2). If the first proxy is non-
functional, the next valid proxy is used. A list of all valid proxies are validated to make
sure they are available and configured for collection. The list of valid discovered entities
is divided among the available proxies.

The discovery and collection approach insures scalability and reliability regardless of
whether or not systems are centrally managed. Spreading out the collection among
multiple proxies ensures that the measurement does not impact what is being measured.
The ability to distribute the work among multiple proxy hosts ensures reliable data
collection. If one or more proxy collection host is down, work is distributed to working
proxies.

The collection request is sent to each proxy host with the following:

- List of physical and/or virtual hosts
- Start time
- End time

If the proxy is configured to collect from a central cloud management console (Figure 1),
the data is collected from cloud management console. If the proxy is configured to collect
from each individual physical and/or virtual host, the data collected data is collected from
each individual element (Figure 2). The collected data is persisted on each proxy host for
the duration of the collection request. Additional processing may be performed on the
collected data and persisted.

After the collection duration has expired, the Collection Central Console issues a data
transfer request to each collection proxy host. The persisted data is transferred back to the
console and deleted from the proxy. The persistence of data on each proxy ensures
scalable data-collection methodology that would not overwhelm the console managing
the collection. Data on each console has an expiration date. If data is not transferred by
the expiration date, it is deleted. The data expiration on each proxy ensures reliable data
collection that will not fill up the available storage at each proxy host.
Drawings

Figure 1: Centrally Managed Cloud Environment Diagram

Figure 2: Non-Centrally Managed Cloud Environment Diagram