Turning Disparate Data into Action

The process of consolidating data
to achieve organizational success

Visualize, Analyze, Manage, and Maintain Critical IT Networks

Turning Disparate Data into Action

Executive Summary

Integrating disparate infrastructure data sources in broadly distributed Information Technology (IT) enterprises has become quite challenging. The proliferation of point solutions designed to address very specific security and management requirements for IT assets has caused a deluge of administration screens and a great deal of disconnected data. Configuration and status information about IT assets and systems must be aggregated, correlated and reported on from enterprise systems in order to address the myriad operational and emergent questions which arise.

Contents

The Environment

The combination of cheaper computing, cheaper storage, SaaS and the continuously accumulating asset-base of software applications available is inexorably driving more and more of what we do and how we do it into the virtual world of electronic processing.

With today's technologies and tomorrow's expected advances, our dependence on our IT infrastructure will only continue to grow.

The result is that for all organizations, from governmental to industrial to educational institutions, the pace at which mission-critical dependence is shifting from mortar and brick to the IT network is constantly accelerating. What was once a self-contained tool touched by a few has grown into the single most important facility upon which we depend for virtually every aspect of our daily productive lives. Underlying this evolution/revolution is the simple principle of economics. Automation, simulation and emulation have always held the promise of drastically improving the efficiency and effectiveness of how tasks are performed. Counterbalancing the shift to automation has always been the cost (including capital and operational load) of building the computing tools and networks required to replace brick and mortar. With today's technologies and tomorrow's expected advances, our dependence on our IT infrastructure will only continue to grow.

The IT Management Challenge

Along with this progress, a long-standing challenge continues to expand and ultimately places organizations at significant risk of losing control over their own stability. This area is the operational management of the IT network (and the attached appliances) itself. IT organizations are squeezed by three compounding forces. First, the vital functions supported by the IT infrastructure network (e.g., billing, payroll, security, intellectual property development, and protection) are making the need for 7x24 availability non-negotiable. Second, the complexity of a typical communications-only network grows with the cube of the number of nodes or endpoints in the network. When one adds in the network's assets, this ratio can easily grow to the fourth or fifth power. Lastly, even though

computing and storage costs are becoming cheaper, the quantity of these resources used is growing at a faster rate so that the amount of money invested in the network continues to grow. Therefore, efficient investment in IT assets is an increasing concern, driving a growing need to optimize utilization through redeployment, timely retirement and other inventory/asset management activities.

Data Consolidation

Growing IT staff can be a short term solution to closing the gap. But clearly the combination of importance, rapidly expanding network complexity and growing investment means simple linear staff growth will not suffice. Beyond working harder, working *smarter* is required to stem the tide. By the mid 1980's, the first versions of the Information Technology Infrastructure Library (ITIL) saw the light of day which codified how to store all the data you might ever want to know about your IT network in a systematic data repository (relational database). Problem solved! Slay the technology by turning it on itself. Simply use the same computers and storage devices which made up the network to maintain an electronic record of everything that needs to be known.

Over time, the early ITIL specifications have been reviewed, refined and revised. Current generations of the specification exists with well in excess of 30 volumes of requirements to define how to unambiguously do what sounds like a simple task, namely “write down what you have.”

Data Discovery

Discovery Systems... are available which can prowl any given range of IP addresses and automatically discover any intelligent devices which exist within that address space.

In theory, the work of the ITIL specification body should have solved the problem. But, like many solutions, it also created (or perhaps focused attention on) more fundamental problems. Specifically, how will all that infrastructure data be entered in the first place? The task of gathering the data and entering it into the database was left as a task for the IT staff. However, the rapid growth in the number of elements involved, the amount of data which needs to be recorded for each element, and the pace at which the data dynamically changes (e.g., new operating systems, new applications, patch installation, network reconfiguration, need to track asset utilization rates) makes it impractical for humans to populate the entire central repository.

Again, engineers have turned the tables by building systems and capabilities to make the intelligent elements track their own configurations and populate the repositories autonomously. Known collectively as discovery systems, products/solutions are available which can prowl any given range of IP addresses and automatically discover any intelligent devices which exist within that address space or access other reported data. Once discovered, automated scripts are run to extract a myriad of parametric data from the device to characterize its capability, configuration, unique identity, and current state.

There are two classes of discovery systems: *agentless*, which perform all of their investigations without requiring the installation of any code or application; or *agent-based*, which push, or install, an application onto the target machines once they have been discovered through IP address searching. The latter allows the collection of a wider range of data, but conversely leaves a footprint on the destination asset.

Making Use of the Data

The combination of a single infrastructure repository tied with state-of-the art discovery tools provides the capability to build a database that can effectively be populated with every piece of information which could possibly be known about an IT network. Although on the surface it may seem that the problem is solved, in reality the hardest part of the job still lays ahead for the IT management organization. The dilemma is that all the discovered and logged information now held in the repository is highly fragmented and detailed, but the questions and problems of the IT operations teams are triggered by higher level users with application-level questions or concerns. Ironically, this last hurdle to get to action is only made increasingly higher by all the strides that the ITIL approach and discovery and reporting systems afford.

The scale of the detailed information in the repository can be quite overwhelming. For example, in an existing database store for a particular government agency, every faceplate requires over 300Kb of data to fully characterize all the items behind the faceplate, such as cable, wire, conduit, router, server, and power. Therefore, when an IT operations staff member asks a simple question about a faceplate, a simple answer of a data field is not possible.

Crossing the Hurdle to Action

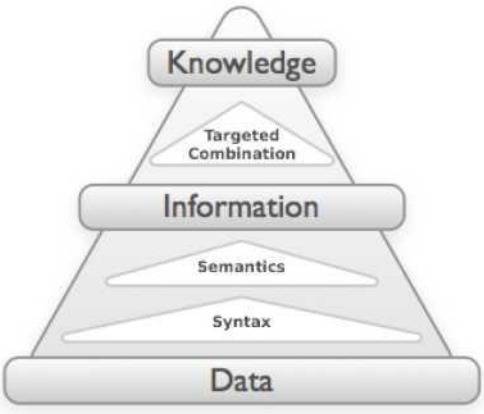
To allow the raw data to drive action, two steps are actually needed.

Step One: Aggregate

First, the flat data needs to be combined so higher level conclusions can be drawn from the data and made useful. This fusing of myriad points into a higher level concept has been abstracted and formalized for study by the scientific and engineering community for many years and is known as Knowledge Discovery and Data Mining (KDD).

The second stage of KDD transforms information into knowledge.

The figure below is a high level view of the specific steps and methodologies used in KDD. The process involves two levels of mappings. The first stage completes the task started in the discovery/consolidation phase by adding a rich set of syntax rules and semantic structures around the data to turn it into information. In any real world situation, there will be multiple origins of the data in the repository and, therefore, an alignment function needs to be added to explicate the various data forms, which need to be recast or normalized into a unified construct. This normalization deals more with meaning rather than representation. The key is these syntax and semantic processes are static in the sense that this transform can be done up front or out of the context of a specific question or investigation which needs to be done by IT operations.



The second stage of KDD transforms information into knowledge. This may sound like a philosophical distinction, but it is a very real one. As shown in the diagram, this transformation is done by targeted combination, which implies that the target or question being asked is known. Therefore it needs to be done in-line with the context of the questions being asked, not up front as with syntax and semantics. For many systems, a number of pre-configured

questions can be defined to allow some of the knowledge building to be done offline, but any useful system needs to provide a mechanism for users to add and extend their questioning in real time and adapt to such extensions in real time.

Step Two: Correlate

Although knowledge may not be a void philosophical concept, it is still too abstract in and of itself to drive action. This brings us to the second step - communication. Ultimately, no knowledge derived from a management system is of use until it can be communicated to the staff member who needs to take action. Here, visualization is the key. Graphics, maps and pictures are the richest and most efficient means to communicate information and insight to people (they're worth a thousand words, right?). Therefore, the final piece of the chain to action is to translate and present the insights and knowledge gleaned from the repository into visualization (whether they're geographic, blueprint-based, schematic, or even table-driven for numbers) to help people get the message clearly, quickly and accurately. Beyond making use of standard computer graphic tools, two items must be added. The system needs to fuse true spatial location information with the discovered data to allow the system to show where something is (e.g., what floor, what building, how many meters up/down, latitude/longitude, etc.). Secondly, the tool can best support the communications process if the drawings allow the users to explore or drill-down for detailed information. This can allow a great deal of information to be offered and analyzed in context without drowning the viewer/user with noise.

Graphics, maps and pictures are the richest and most efficient means to communicate information and insight to people.

Summary

The most important and valuable asset of any organization is ultimately how people understand and operate their businesses. Taking disparate infrastructure data and molding it into usable knowledge for actionable tasks is the key to operational success.

"Know where to find the information and how to use it - that's the secret of success." - Albert Einstein

For More Information

info@planetassoc.com
+1.201.693.8700
www.planetassoc.com

Planet Associates products provide a natural and intuitive path to help organizations turn infrastructure data into action. Read more...

The Planet IRM Suite

Unlike many other offerings, the Planet approach has always been to focus our energies and innovations on knowledge transformation and visually communicating end of the journey, while remaining agnostic towards specific data being collected in terms of range, manner of collection or archiving approach.

Visual

Core to our value proposition is our focus on presenting visualized output to our users. By centrally focusing on being the entry and storage tool to capture and track what items are in the network, where they are located and how they are connected, we provide a consistent and centralized way to uniquely see the network. Although the required information can be entered by field-based electronic input (singularly or bulk-loaded), we also provide a CAD drafting-style graphical entry capability to allow users to draw what they see. We additionally provide GIS-enabled support in our drawings to allow the outputs to provide accurate latitude and longitude location identification, plus the fusing of satellite and standard roadmap information to further allow the users to relate the data to the world in which they live.

Drillable

Planet IRM provides visual documentation, which allows users to start at a higher level view of their infrastructure and drill-down with mouse clicks to see in a building, a rack or across the ocean though an international leased line. Mouse-overs and comment pop-ups are included in our patented iiPDF technology to simultaneously reduce the clutter on the screen while providing virtually limitless access to the various snippets of information which the system can ultimately make available when desired.

Agnostic

Although Planet allows IT data to be manually entered into the tool, it can also build (using Dynamic Data Reconciliation DDR or Discovery Correlation Engine DCE) virtually any raw data from any source (be it human or machine-driven as with discovery systems, service support or network management tools). Simple dialog-driven screens allow users to connect to data from multiple sources and map to the unified syntax and semantic model of the underlying Planet IRM schema easily. Once these semantic links are defined in Planet IRM, all the data provided can be seamlessly leveraged by its standard knowledge generation and visualization engines. Planet Associates Inc. develops, licenses and supports the Planet IRM family of infrastructure relationship management software products. Planet IRM is uniquely capable of total enterprise network asset consolidation, incorporating Inside and Outside Cable Plant, Data Center, WAN, and User/Desktop Management, and all interconnectivity.