

# EMC Smarts Application Discovery Manager and Multisite Data Aggregation

**Abstract:** Without a complete overview, which includes a detailed as well as global representation, CIOs and their IT team may not have an accurate understanding of their infrastructure. The need for timely and accurate information regarding application behavior has moved from simply being an IT issue, to a matter of a company's ability to be competitive in the market. The lack of understanding of how systems run and their dependencies within the infrastructure reduces their ability to make effective business decisions for example to reduce software license expense, maintenance cost, increase employee productivity, reduce system downtime, or prevent loss of vital data.

Copyright © 2007 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

Sample Trademark page below.

EMC<sup>2</sup>, EMC, ApplicationXtender, Celerra, CentraStar, CLARAlert, CLARiiON, Connectrix, Dantz, Direct Matrix Architecture, DiskXtender, Documentum, EmailXtender, EmailXtract, HighRoad, Legato, Navisphere, PowerPath, RepliStor, ResourcePak, Retrospect, Smarts, SnapView/IP, SRDF, Symmetrix, TimeFinder, VisualSAN, and where information lives are registered trademarks and EMC ControlCenter, EMC Developers Program, EMC OnCourse, EMC Proven, EMC Snap, EMC Storage Administrator, Access Logix, ArchiveXtender, Automated Resource Manager, AutoSwap, AVALONidm, C-Clip, Celerra Replicator, Centera, CLARevent, CopyCross, CopyPoint, DatabaseXtender, Direct Matrix, DiskXtender 2000, EDM, E-Lab, EmailXaminer, Engenuity, eRoom, FarPoint, FLARE, FullTime, InfoMover, MirrorView, NetWin, NetWorker, OnAlert, OpenScale, Powerlink, RepliCare, SafeLine, SAN Advisor, SAN Copy, SAN Manager, SDMS, SnapSure, SnapView, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix DMX, and VisualSRM are trademarks of EMC Corporation.

All other brand names are trademarks or registered trademarks of their respective owners.

Part Number: S0079

- Executive Summary .....4**
- Aggregate Data from Multiple EMC Smarts Application Discovery  
Manager Appliances .....4**
  - Applications That Span Multiple Geographically Separated Data Centers..... 5
  - Security Measures That Prevent Installing the Same Appliance in Multiple Segments..... 5
  - Other Conditions That Prevent Connecting All Segments to the Same Appliance..... 5
- Aggregating Data From Multiple Application Discovery  
Manager Appliances .....5**
  - Deployment Considerations When Aggregating Multiple Application Discovery  
Manager Appliances..... 6
  - Network Connectivity Level .....7
  - Routing between the networks .....7
  - Network Address Translation (NAT).....7
  - Firewall and multiple routes..... 8
- Integration Points Built Into Application Discovery Manager for  
Data Aggregation .....8**
- Popular Methods for Aggregating Multiple Appliances .....9**
  - Alerts Aggregation ..... 9
  - Topology and CMDB Aggregation..... 9
  - Demand Aggregation..... 9
  - Aggregation Alternatives Comparison ..... 10
- Summary .....10**
- About EMC Smarts.....10**

## Executive Summary

Over the past decade, companies have made unprecedented capital investments in information technology with the goal of getting better information, cheaper, and faster. Businesses are still struggling to control their technology expenditures and searching for new ways to manage the complexity of the application environments they have created. Past IT investments in systems and network management tools while sufficient for understanding the availability and performance of the components, are not enough for this paradigm. The reason is quite simple: existing NSM and ESM tools are blind to application logic. Research from Giga shows that 70 percent to 80 percent of time is spent finding the root cause of downtime and 44 percent of services disruptions are caused by the application.

Additionally, IT organizations face challenges from an increased number of users who need new business solutions that are built on distributed technologies like .NET, J2EE and Web Services, which further complicate application architectures. With these complicated application deployments and distributed data centers, IT organizations need to find solutions that can discover, map, and track these complicated application architectures holistically, not in silos. Gaining visibility with scalable solutions that can identify, catalog, store and track the data of the configuration items in a federated configuration management database (CMDB) through aggregating this information from all data center locations is necessary for better management.

IT organizations that can gain this high-level visibility into their data centers with accurate, up-to-date, and complete information on all of their systems, can effectively direct their IT spending towards key business initiatives based on application demand while controlling costs and lowering expenditures allowing IT to do more with less.

## Aggregate Data from Multiple EMC Smarts Application Discovery Manager Appliances

In less than an hour, Application Discovery Manager passively and non-intrusively discovers the applications portfolio running in your network, the relationships, and dependencies between these applications and their supporting physical and logical resources. Application Discovery Manager automatically maps your active applications and their behavior using a non-intrusive approach that does not require you to install any agent on your production environment or provide the product with users and passwords. Just plug Application Discovery Manager into your network and in less than an hour, you will get an accurate and up-to-date map of your applications.

Application Discovery Manager is shipped as a self-contained appliance, which runs on a rack mountable IBM server. Application Discovery Manager guarantees to be completely passive by listening and inspecting packets on the network without scanning, spidering or probing the network. It connects to a special port in your switch called a “Monitoring Port” that provides a means to passively observe the traffic routed through the switch without modifying that traffic or sending new packets to your network. Monitoring ports are typically used for connecting sniffers or intrusion detection systems to networks. The “Monitoring Port” feature is supported by switch vendors (For example, Cisco Systems or Juniper Networks) and is guaranteed to have minimal impact on the switch performance.

Port mirroring (or SPAN in Cisco terminology) is the most popular method for local packet collection. Virtually all managed switches and some popular routers have the capability to mirror port traffic to specially designated “monitoring” port. It is also possible to mirror traffic from a few ports simultaneously.

Typical data centers use dozens of switches to connect servers at the physical layer. Application Discovery Manager is capable of handling multiple monitoring port sources simultaneously with a limitation of 115 input sources (as of May 2005) per single Application Discovery Manager appliance. A few challenges such as duplicate packets and clock synchronization, introduced when connecting Application Discovery Manager to multiple input sources simultaneously, are automatically handled by Application Discovery Manager.

However, in some cases it is impossible to physically connect monitoring ports from different switches into a single Application Discovery Manager device. Such cases may include the following scenarios:

## ***Applications That Span Multiple Geographically Separated Data Centers***

Applications may span multiple physical locations. For example, a “Sales Automation” application that has most of its servers located in the New York datacenter may have additional servers located in the San Francisco and Boston branches of the organization to provide faster response times. In this example, it might be desirable to have a single “picture” of the application that includes servers and their clients from all branches. Since the servers are physically located in different physical locations, it would be expensive to use the inter-branches bandwidth to transfer monitored traffic to a single Application Discovery Manager appliance.

In this scenario, a more efficient solution would include an Application Discovery Manager appliance in each of the branches and coupling that with the aggregation process as detailed later in this document.

## ***Security Measures That Prevent Installing the Same Appliance in Multiple Segments***

Some organizations have different security policies and requirements for the different subnets of their network. For example, a bank may require its on-line trading application to have stricter security policies and a greater sensitive level than its internal HR application. Security policies may limit connectivity between the two applications and connecting Application Discovery Manager to monitor both applications simultaneously may be considered as a security policy violation.

In this scenario, a more secure deployment would include multiple Application Discovery Manager appliances (one for each security domain) and a strictly controlled aggregation process. The synchronization process has complete control as to which data is aggregated from each system and where the data is stored.

## ***Other Conditions That Prevent Connecting All Segments to the Same Appliance***

In some data centers, it may be difficult to tunnel monitoring traffic of critical servers into a single location even though they physically located in the same area. Such limitations may be caused by physical location (different building or room), bandwidth considerations, switch configuration complexity, security considerations, firewalls, or any other limitations that prevent connecting all segments to be monitored into a single Application Discovery Manager device.

Depending on the limitation that prevents such deployment, alternative deployment approaches may help mitigate these limitations by allowing multiple Application Discovery Manager appliances to build a local CMDB that is later on aggregated into a master CMDB.

## **Aggregating Data From Multiple Application Discovery Manager Appliances**

The process of aggregating data from multiple Application Discovery Manager appliances and providing a single view of the aggregated data is called “Data Aggregation.” The process can be completed as a one-time operation or be automated as part of an integrated enterprise deployment solution. The process can also be used to import data from third party software into the Application Discovery Manager federated CMDB (F-CMDB).

At a high level, the process of data aggregation from multiple appliances includes:

- Design of the aggregation process
- Creating scripts to support the aggregation process as designed in Step 1
- Automation of script execution where applicable

The following paragraphs discuss in details the different aspects of this process.

## ***Deployment Considerations When Aggregating Multiple Application Discovery Manager Appliances***

Since data aggregation between multiple Application Discovery Manager appliances may occur at multiple levels, you must carefully choose the right solution to match your requirements. As some solutions are more complicated to implement and may result in a longer deployment times, we suggest that you first answer the following questions prior to choosing an aggregation solution.

1. What do you need to deploy multiple Application Discovery Manager appliances?
  - a. My applications span multiple geographically separated data centers
  - b. My applications are in different security domains
  - c. Physical limitations do not allow me to connect monitoring ports to the same Application Discovery Manager product
2. What data would you like to aggregate?
  - a. Data discovered by Application Discovery Manager:
    - Alerts
    - CMDB data such as topology and CI information
    - Demand information
  - b. Data generated by third party tools
3. What operations do you expect to perform on the aggregated data?
  - a. Review of alerts only
  - b. Review of Topology and CI information
  - c. Review of Topology and CI information with demand analysis
4. How well are your monitored networks connected?
  - a. The networks do not communicate with each other and have no routing between them
  - b. The networks do not communicate with each other but have routing between them
  - c. There is connectivity between the networks but this activity is not critical for me
  - d. There is connectivity between the networks that is of high interest for me
5. What network technologies have you deployed between these networks?
  - a. There is a NAT between the networks
  - b. There is a VPN between the networks
  - c. There is a firewall between the networks
  - d. There are multiple routes available between the networks (such as failover switches or load balancing)

The following sections explain the consideration of common network technologies and their effect on data aggregation.

## ***Network Connectivity Level***

The level of connectivity between the monitored networks greatly affects the complexity of any deployment. On the one hand, we may decide to aggregate two separate networks that do not communicate with each other. In this case, the deployment is easy as either no special entity matching and filtering is needed or a very simple address translation would be required if the two separated network are private (10.1.1.\*) networks.

On the other hand, we may have two networks that highly communicate with each other and modeling of the communication between the two networks is important. In this case, we need to merge entities discovered by both devices that represent the same physical entity.

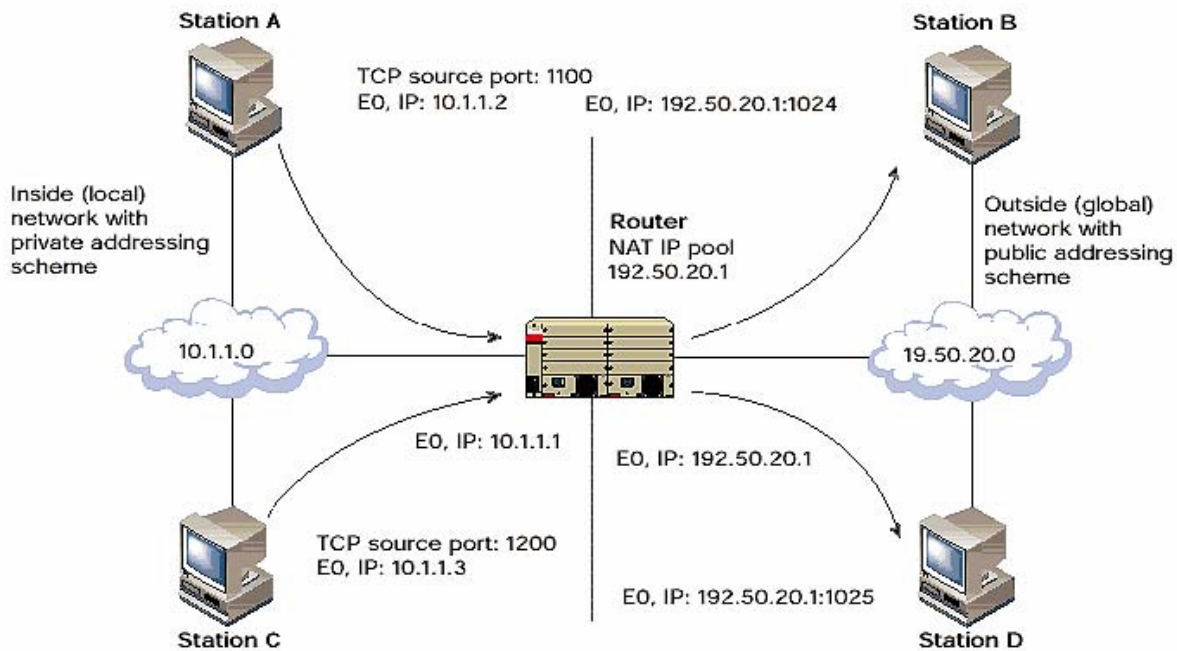
For example, an Application Discovery Manager appliance discovers an Oracle database on host 10.1.1.1 with 10 clients and a second appliance discovers the same Oracle database on the same physical host with 15 clients. The result of an aggregated view of both appliances may show this Oracle database with 15 to 25 clients, depending on whether the clients discovered by each appliance are the same physical clients or not.

## ***Routing between the networks***

The process of aggregating data from multiple Application Discovery Manager appliances requires the transfer of data between multiple Application Discovery Manager appliances to a single location. Typically, the Application Discovery Manager management interface is configured to be part of a management network that is accessible across the organization. In such cases, no additional configuration is needed. While it is not mandatory that the monitored segments have a route path between them, it is mandatory that for Application Discovery Manager data aggregation, each of the Application Discovery Manager appliances are accessible from a single location.

## ***Network Address Translation (NAT)***

If NAT is used between two monitored segments, two Application Discovery Manager appliances might detect the same physical host with different IPs. An IP translation is required before aggregating the data from the two (or more) appliances. In case of static NAT, the aggregation script may transform the IPs based on the static NAT definition and an aggregated view may be easily achieved. In the more complex case of dynamic NAT, some other heuristics must be used to create a match between the entities found by each Application Discovery Manager appliance. Such heuristics may correlate the various host related information such as MAC address, operating system, hardware information and services running on the host to create the appropriate match. Usually Application Discovery Manager maintains the real server's IP in the aggregated view at the Master appliance and performs all translations on the "slave" Application Discovery Manager appliance that discovers the clients that access the server.



*NAT with Port Address Translation (PAT) of Global Addressing*

## **Firewall and multiple routes**

Since Application Discovery Manager is not dependent on the different routes of packets, multiple routes have no effect on the aggregation process.

Since firewalls may be configured to perform NAT (static or dynamic). In such a configuration, they may have the same effect as normal NAT has.

## **Integration Points Built Into Application Discovery Manager for Data Aggregation**

Application Discovery Manager can be integrated with any existing tools you already have in your datacenter and with other Application Discovery Manager™ appliances you have deployed. Depending on the outcome you are trying to achieve for your deployment, you can choose from a few integration points:

- Exporting Reports – any built-in report that ships with Application Discovery Manager can be exported to CSV, allowing the data to be used by most common reporting tools such as Crystal Reports or Excel.
- Monitoring module integration - The monitoring module is capable of sending emails via the SMTP protocol upon any user-defined policy event. Emails can be sent to any number of recipients and can be used for manual or automatic processing by any standard email application. The monitoring module can also run a user provided UNIX shell scripts upon such events. The shell script is provided with the entire alert's data, allowing Application Discovery Manager to integrate with Change Management or Help Desk applications.
- API - Application Discovery Manager has an extensive SDK that allows extraction and insertion of data from Application Discovery Manager's CMDB to / from an XML format. The type of data that is extracted or inserted varies from topology information such as inventory and dependencies to demand information such as baseline transactions of entities and their demand in a specific time range.
- Direct database connection – Application Discovery Manager's CMDB uses Oracle 10g as its underlying database. Standard Oracle ODBC and JDBC drivers can be used to directly access the internal CMDB model used by Application Discovery Manager.

You may choose any of the above integration points as the basis of your deployment. The integration points are listed in a low to high complexity order and a decoupled to tight integration level order. For most simple deployments, the first two options may be good enough. In case that a single Aggregated view of topology and demand is required, the more complex options 3 and 4 must be used.

A typical deployment defines a single Application Discovery Manager appliance as the master appliance into which all data is fed. Other Application Discovery Manager appliances are used as slaves on remote networks and are queried for newly discovered data occasionally.

## Popular Methods for Aggregating Multiple Appliances

### ***Alerts Aggregation***

Alerts aggregation is done by using the native capabilities of Application Discovery Manager™ monitoring module. Alerts from multiple appliances and locations are sent as Email, SNMP traps, or custom scripts into any route-able destination in your data center.

For example, if you are using IBM Tivoli TEC, Application Discovery Manager appliances would send all alerts to the TEC console from wherever they are installed.

Integrating the alerts into a single central Application Discovery Manager™ (“Application Discovery Manager Master”) is also possible by using the API or Direct DB Connection integration points. These two integration points can query Application Discovery Manager’s CMDB for new alerts in an automated fashion and feed the results into the “master” appliance.

### ***Topology and CMDB Aggregation***

The Topology and CMDB information stored in Application Discovery Manager’s CMDB can be aggregated using the Application Discovery Manager API or direct database access. EMC highly recommends the usage of the API over direct database access.

In a Topology and CMDB aggregation process, topology and CMDB data is periodically extracted from one or more Application Discovery Manager appliances and fed into the master Application Discovery Manager appliance. During this process, some processing of the extracted data must be performed prior to its insertion to the master appliance. Such processing may include NAT translations, filtering of data or any other data manipulation required for to create complete picture of the data center.

The Application Discovery Manager SDK comes with a command line API used to perform such extraction and insertion. The “dm\_query” script is used for extracting information and “dm\_feed” script is used for inserting new information. Any XML processing program may be used to perform any kind of processing of the data prior to inserting it. A common processing method is to use XSLT to filter the output of “dm\_query.”

It is also possible to extract topology and CMDB information to any other third party tool that accepts XML data such as IBM Tivoli TBSM or HP Open View. Data can also be imported from these third party applications into the Application Discovery Manager CMDB.

### ***Demand Aggregation***

The most complex integration involves aggregation of demand found by multiple Application Discovery Manager appliances or any other third party software into the master Application Discovery Manager appliance. During this process, some processing of the demand metrics can be done to eliminate duplicate counts, time zone adjustments, handling of NAT translations, filters, etc. A common processing method is to use XSLT to filter and manipulate the output of “dm\_query.”

Demand aggregation is considered a more complex integration than Topology and CMDB since it includes a much larger time based data set and may take considerably more time to perform. The same scripts used in Topology aggregation (“dm\_query” and “dm\_feed”) are used in the demand aggregation process.

## Aggregation Alternatives Comparison

	What data needs to be aggregated		
	Alerts	Topology and CMDB	Demand information
<b>Purpose of aggregation</b>			
A single aggregated alerts console	<input checked="" type="checkbox"/>	No need	No need
A single view of the data center with no single demand information location	No need	<input checked="" type="checkbox"/>	No need
A single view of topology, CMDB, and demand information	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

## Summary

Without a complete overview, which includes a detailed as well as global representation, CIOs and their IT team may not have an accurate understanding of their infrastructure. The need for timely and accurate information regarding application behavior has moved from simply being an IT issue, to a matter of a company’s ability to be competitive in the market. The lack of understanding of how systems run and their dependencies within the infrastructure reduces their ability to make effective business decisions for example to reduce software license expense, maintenance cost, increase employee productivity, reduce system downtime, or prevent loss of vital data.

## About EMC Smarts

EMC Smarts plays a crucial role in managing your information infrastructure by automating the discovery, understanding, and mapping of the complex relationships that exist among business processes, applications, and the IT infrastructure.

With EMC Smarts solutions, organizations gain the visibility needed to accelerate and increase ROI on their highest-priority IT service and cost management initiatives. Offering the easiest, most-comprehensive solution in the industry, EMC Smarts technology allows organizations to:

- Accelerate ITIL and CMDB standardization
- Reduce costs—up to 25 percent in the first year
- Maximize resource utilization
- Mitigate risks and ensure business continuity
- Enhance business agility and IT service delivery by accelerating and simplifying initiatives that support business service management and data center automation

### Getting Started

To learn more about how the EMC Smarts Solution for Data Center Audits—and other EMC Smarts solutions—can positively impact your business and IT operations, contact your local EMC or EMC Smarts sales representative, or visit our websites at [www.EMC.com](http://www.EMC.com) and [www.smarts.com](http://www.smarts.com).