Zenoss®

WHITE PAPER

Zenoss Service Dynamics Architecture Overview

Service Assurance for the Hybrid Cloud Datacenter

IT infrastructure and delivery is undergoing a major evolution, marked by the transition to a service-based approach leveraging virtualization and cloud architectures. This new model challenges traditional assumptions, processes and tools. To capture its value, management processes and technologies must be re-examined throughout the operations lifecycle.

This paper discusses the management stack that supports delivery of IT-as-a-Service in a hybrid environment consisting of physical, virtual and cloud-based infrastructure, and how Zenoss Service Dynamics addresses the challenges of Service Assurance and Optimization for this new world of IT.

Introduction

IT is in the midst of a vast transformation from reactive management of discrete servers to planned service operations from a computing resource pool – the cloud. This transformation to the dynamic datacenter will affect every piece and process within enterprise IT.

Some of the changes are obvious yet surprising. There's a new fundamental operating system for enterprise servers – VMware's vSphere – running combinations of Windows and Linux workloads on precisely allocated hardware slices. A new server vendor has emerged – Cisco – with a radically simplified approach that unifies server and networking. In addition, the evolution of Cloud Platforms such as Amazon EC2 and OpenStack are presenting enterprise customers with sophisticated solutions for both outsourcing compute workloads to public cloud service providers or building private cloud offerings themselves. Enterprises can now rent temporary computing resources from a business that started by selling books online and integrate flexible capacity management with cloud vendors like Terremark and Hosting.com with their internal vSphere deployments or OpenStack clouds.

Changes to processes are more subtle. Organizational processes that supported a multi-week process for adding a new server are stressed to the breaking point by virtualization provisioning that takes only minutes. Change management is dramatically different, with VMware supporting libraries of pre-defined applications that can be provisioned rapidly and moved between physical servers automatically, while post-provisioning "customizers" like Chef, Puppet, and cfEngine ensure known-good configuration deployments.

The focus of this paper is the management software toolkit that supports the dynamic datacenter. Just as Dynamic Datacenters are built on a new operating system running on new server hardware, so the management software must change as well.

Zenoss customers tell us that traditional IT management software is challenged in this new world.

- IT is about the delivery of services, but most IT Management Software has traditionally been focused just on managing infrastructure
- Rapid commissioning, modification, and decommissioning of workloads makes the multiple discrete products that make up management frameworks expensive to maintain
- As tool administration falls behind change rates, customers find that they are attempting to manage devices that don't exist any longer and are failing to manage newly created devices
- Management paradigms and thresholds designed for dedicated physical Windows and Linux servers are simply wrong and cannot be trusted when it comes to shared workloads
- Without an automated mechanism to keep tools aligned with change, the drift between reality and the management model increases every day, reducing the ability of IT to confidently answer questions about performance and availability of business services
- Customers of IT demand increased visibility into the performance of their services and IT needs to provide individualized, customer-specific views of that performance. Every IT organization is now coming to grips with supporting multiple tenants of their IT infrastructure.
- In short, traditional IT management software has become too complex and too expensive. A simplified, cost-effective solution is needed.

Zenoss Service Dynamics is designed from the ground up to handle these challenges. In this paper we'll learn about the management stack that supports the dynamic datacenter, how the pieces fit together, and explore how the Zenoss architecture specifically addresses the challenges of this new world in IT.

Cloud Service Management Stack

You're probably familiar with the ITIL V3 Service Strategy of Design, Transition, and Operate, the similar Microsoft Operations Framework of Plan, Deliver, and Operate, or similar models. When you're managing the cloud – providing and customizing virtual machine images selected from a catalog, assuring and optimizing the executing workloads, and providing visibility into cost, problems, and change the processes will be familiar – but emphasis and tools will differ.

Three broad management categories make up the cloud management stack:

- Provision and Configure
- Assurance and Operations
- Support and Billing

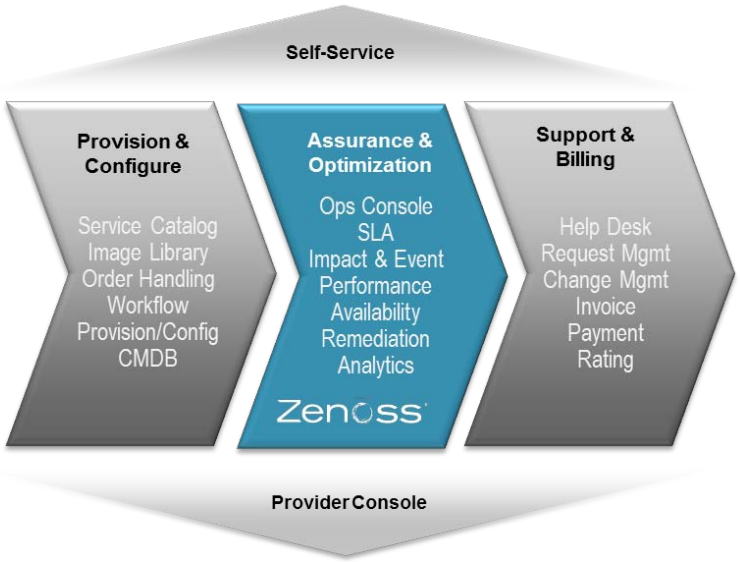


Figure 1: Cloud Management Stack

Provision and Configure

In cloud operations, customers expect the IT organization to provide them with immediate access to a library of pre-configured then customized system choices. They want to place an order with response coming in a predictable timeframe.

Although this sounds daunting to manage, it's actually far easier in practice than managing operating systems on physical servers. By operating the datacenter as a cloud, IT removes the variability that makes it so hard to manage in a physical server world. It's much easier to assure proper audit configuration, deployment of security patches, common file system layouts, etc. – the difficult aspects of change management are baked into the process from the beginning instead of tacked on afterwards and subject to exception after exception.

Five key functions make up the definition and activation category:

- Service Catalog
- Image Library
- Order Handling Workflow
- Resource Configuration
- Configuration Management Database

The **Service Catalog** is the set of pre-defined offerings that your cloud gives to your customers. The Service Catalog enables a customer to select the workload platform, operational characteristics, and service level they're willing to pay for. Typically, cloud providers offer several standard operating systems such as Windows Server or Red Hat Enterprise Server as workload platforms. A workload can operate in one of several environments that specify how much computing resource (CPU, memory, storage) is necessary and how robust the systems are that assure continuous availability (full or incremental backup and restore, capacity reservation within ESX, level of disaster recovery, etc.) Customers can often choose service levels such as daytime weekday, 24x7 operations, and selected availability and performance guarantees. In a Service Catalog, customers will see the cost implications of the choices they make, and will communicate their desires very precisely. Your organization won't be over- or under-purchasing compute resources and will operate more efficiently.

The **Image Library** is a set of pre-configured virtual machine images ready for deployment. These are the standard operating systems that support your workload offerings.

In the **Order Handling Workflow,** your customers communicate their needs to you. They select an offering from your Service Catalog, specify the timeframe they need it active for (no more virtual server sprawl!), and whether any customization beyond standard options is needed. You'll need to track orders and communicate activity back to the customer. In private cloud operations, the order handling system is often hidden from the end customer.

After a customer places an order, the **Resource Configuration** function takes the selected workload platform, operational characteristics, and service level options and establishes a base workload for the customer. After the workload is started, it can be customized using standard options from the **Configuration Management Database**. For example, a customer might need a MySQL database to support his application – you'll be assured that a known version of the database is deployed with proper patches and security settings.

An important part of the automation provided by Placement and Provisioning and Resource Configuration is appropriately adding the customer workloads into the Assurance and Optimization category.

Assurance and Optimization

Zenoss Service Dynamics is focused on the Assurance and Optimization category of cloud operations and provides a comprehensive implementation.

The **Console** provides a unified view of assurance and operations. Cloud operations consoles meet a combination of enterprise role-based security and service provider multi-tenant needs. Enterprises have traditionally defined multiple organizational roles limiting the actions a user in the role can perform. Service providers have been used to providing each customer a view of just the resources that the customer is paying for. In cloud-enabled organizations, both needs must be met simultaneously.

As far as customers are concerned, your ability to deliver against the **Service Level** they chose from your Service Catalog is paramount. Whether you characterize this mutual understanding as an expectation or a formal agreement is nearly irrelevant - you'll need to track and report against the agreed-to metrics in an easy-to-understand and consume way. This means providing separate views for each customer and workload and operating as if there are multiple distinct tenants using your common cloud resources.

To provide the service level metrics and meet the expectations, you'll need to monitor the workloads and the supporting infrastructure to detect issues that may impact the customer. The **Impact and Event and Management** collects device, application, and infrastructure information that identifies potential issues and evaluates it to determine which issues are critical. **Performance Monitoring** collects vital performance statistics at every layer of the infrastructure, evaluates it to determine whether anything should be fed into the Impact Management process, and stores it for long term analytics. These two functions provide reactive management to application, device, and infrastructure issues.

As issues develop, the **Remediation** function provides automatic correction where that is appropriate and notifies humans that action is necessary where manual intervention is required.

For some issues such as a sudden failure of a power supply there is no advanced warning. For other issues there are trends that can be collectively understood and provide early warning of an impending issue. **Predictive Analytics** attempts to provide proactive analysis of data, searching for identification and remediation of issues before they reach critical state.

In cloud operations, the discipline of **Capacity Analytics** changes roles from an assessment of workload-driven resource requirements against the fixed amount available from a dedicated hardware platform. Reacting to the assessment can take weeks as new hardware is provisioned. Within the cloud datacenter, resource allocations can be altered in minutes and the reaction to changing workload needs should be just as fast. You'll need to communicate to customers the needs of their workloads and get confirmation that they're willing to bear any increased costs, or allow the resource needs to go unmet and application performance to suffer. You'll also need to provision enough resources for the overall datacenter to assure that peak load commitments can be met, which means understanding capacity at an aggregated level.

Support and Billing

The third category of cloud operations function is the area of Billing and Support, managing customer interactions for services in production.

As far as the support category, the primary benefit of cloud operations will be a reduction in cost driven by a reduction in complexity. The **Help Desk** is well understood, and is relatively unchanged in the cloud. People will still discover issues and want them tracked to completion. The **Request Management** function as well is little different – although the simplification that comes from with cloud operations should greatly reduce the overall workload. Traditional **Change Management** is also greatly eased by cloud simplification. Deployment of standardized operating systems, resource configuration, and virtual hardware means far fewer custom solutions to support.

The **Invoicing** category is quite different, however. Prior to the cloud, customers bought individual servers for individual applications. The equipment expense was recognized up front and depreciated over a number of years, while related expenses such as power and space were typically lumped into an overall IT budget. A Cloud operation brings new concepts to IT chargeback.

First, the IT organization needs to use metering to capture relevant information about resource usage. Metering data is collected by the resource monitoring function of service assurance and stored in association with individual customer services. For example, the time series of cpu reservation and actual usage on individual virtual machines are associated with a customer application. Second, **Rating** calculates the total customer cost of the metered service using the metered data and the charge model for the service. And finally, customer statements are distributed and for service providers integrated into the **Payment** management system.

Unified Service Assurance

A key tenet of delivering service assurance in a dynamic infrastructure is bringing the various forms of state and configuration information together, so that they can be leveraged as a whole in real-time to create relevant visibility and drive automation. Siloed management does not work in the cloud.

In addition, fragmented management environments can create more complexity than they remove, and they make it impossible to capture the operational efficiencies required to compete in the modern IT landscape.

For these reasons, Zenoss Service Dynamics brings together an end-to-end service assurance solution in a single, easy-to-deploy, highly-automated product. It's the only way to manage the hybrid cloud datacenter in real-time. The illustration below shows the key functional components that comprise Zenoss Service Dynamics. Together they work as a single, unified service assurance solution for services and applications deployed across physical, virtual and private/public cloud infrastructure.

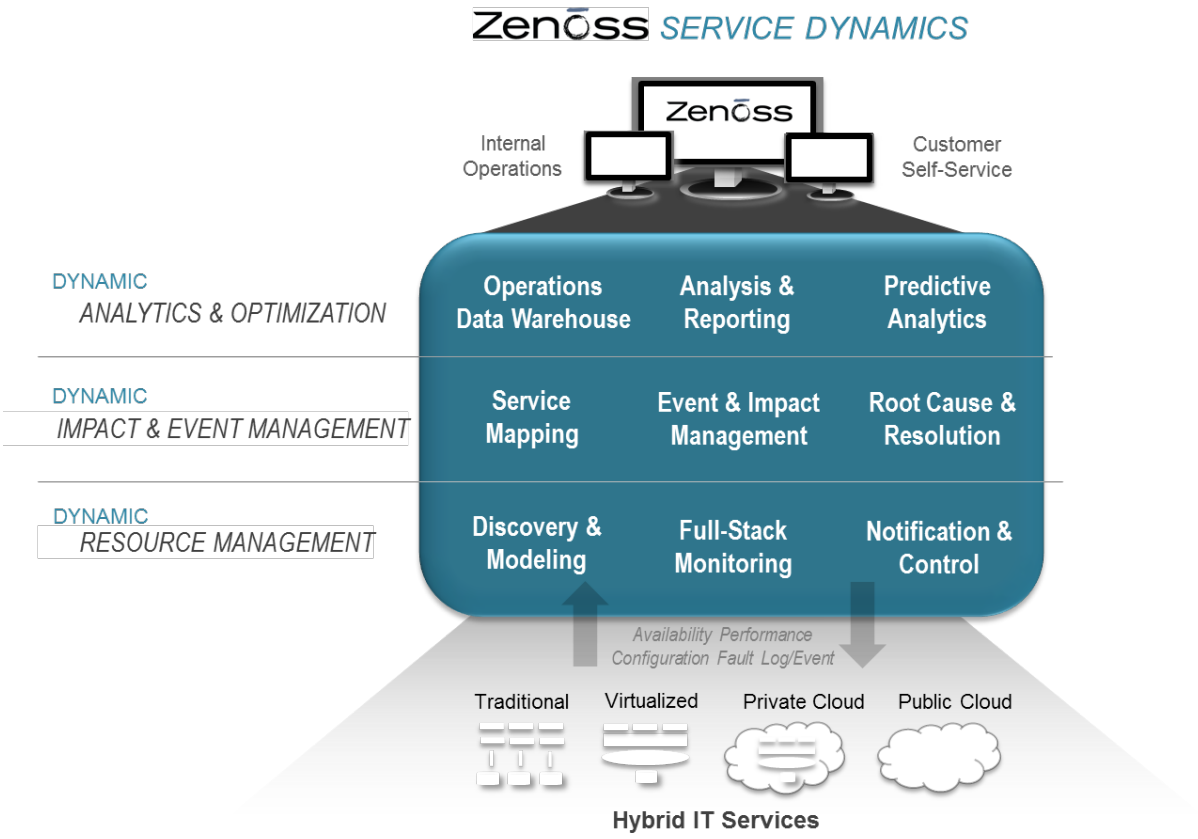


Figure 2: Zenoss Service Dynamics Functional Overview

Hybrid IT Services: Manage Seamlessly Across Physical, Virtual and Cloud

Large organizations will move their IT operations into the cloud over a period of several years. New technologies replace existing technologies very slowly. Since the fundamental business objective of IT is to provide computing resources to enable business services, you'll need to manage physical and virtualized devices, operating systems, networking devices, hypervisors, and cloud platforms for many years.

Zenoss Service Dynamics provides management across all of these technologies. You'll be able to manage one business service running completely on physical hardware, another provided by a third party cloud platform, and services composed of mixed physical, virtual, and cloud elements without having to change your management strategy.

Real-Time Service Model: The Key to Managing in a Dynamic Environment

The core issue in the cloud datacenter is keeping up with its rapid rate of change. When a customer order for a new IT service is automatically fulfilled and placed into production, there's no time to manually update tools or brains.

At the heart of Zenoss Service Dynamics is unified, real-time understanding of the entire IT environment including resources, services, relationships, dependencies, state and configuration. With this understanding, Zenoss is able to dramatically simply the service assurance process from cradle to grave with template-driven resource monitoring, automated impact and root cause analysis, and pre-built analytics.

Unlike traditional configuration databases that are updated in batch mode, the Zenoss model is maintained in near real-time through a series of discovery and modeling techniques that tap into the stream of configuration changes as they happen across the physical, virtual and cloud-based infrastructure. Like every other aspect of the product, this model can be easily extended through its open API.

Without this real-time understanding, it is simply impossible to manage in an automated way, and to produce relevant analytics in a dynamic environment. Without these, you can't capture the value of the cloud.

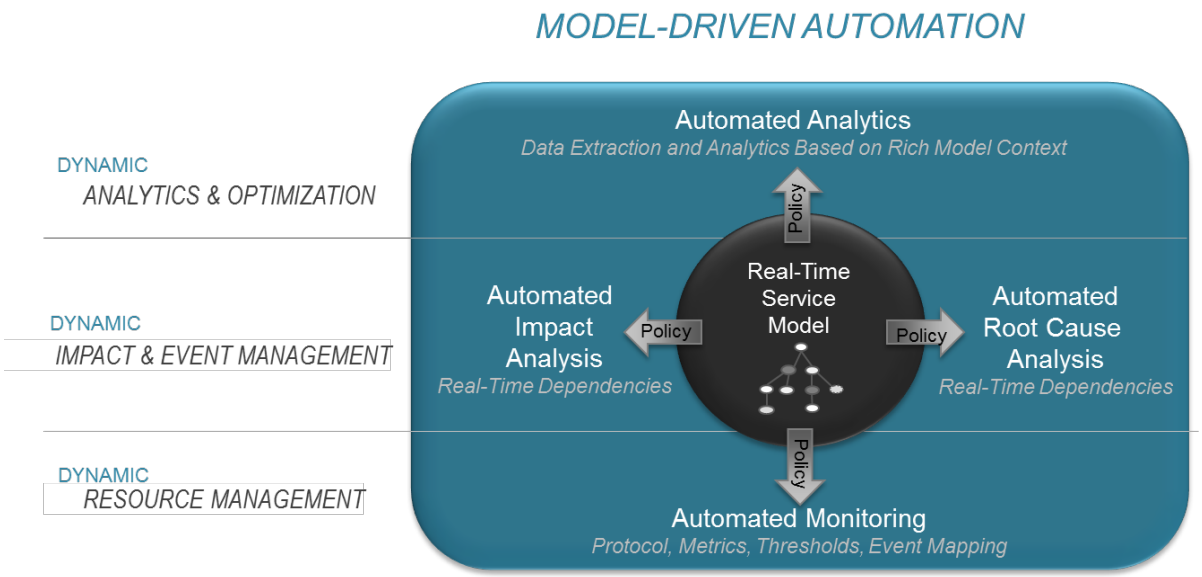


Figure 3: Zenoss Real-Time Service Model

Dynamic Resource Management

The foundation of service assurance in the hybrid cloud datacenter is unified, cross-domain resource monitoring and control that brings together configuration, performance, availability, fault, event and log information across the physical, virtual and cloud-based infrastructure and applications, and enables automated actions to be performed at the resource level. Zenoss Service Dynamics delivers this capability on a scalable, open platform that is easy to extend, and is able to track dynamic elements and relationships as they evolve in near real-time.

Key feature areas include:

- **Discovery & Modeling** – Automatically maintain real-time inventory and configuration details for the entire IT environment; includes real-time relationship tracking of dynamic relationships common in virtualized and cloud-based infrastructures.
- **Full-Stack Monitoring** – Unify and automate performance, availability and event monitoring of networks, servers, storage and applications across your physical, virtual and cloud-based environments with a single, model-driven, horizontally-scalable, extensible collection platform.
- **Notification & Control** - Rich alerting and remediation framework allows you to either be notified via email, text or pager based on user-specified policies or to take direct automated action to address a problem in real-time.

Dynamic Impact & Event Management

Perhaps the hardest challenge in assuring service in a hybrid datacenter is maintaining a real-time perspective on service health, and linking health issues to the underlying infrastructure in a reliable, simple and cost-effect way. Here in particular, legacy approaches to impact management and root cause analysis simply break down due to the shared and dynamic nature of virtualized and cloud-based infrastructure.

The diagram below illustrates how Zenoss Service Dynamics transforms state information from the resource manager into a stream of events that are processed in near real-time by its service impact and root cause analysis engine, leveraging the Real-Time Service Model. This processing feeds a service health dashboard and generates service events that are used for alerting, trouble-shooting, and to initiate real-time automation and service remediation. The result is real-time service level awareness, rapid triage and closed-loop automation that thrives in the dynamic, hybrid cloud environment. We'll go into more detail on how this works below.

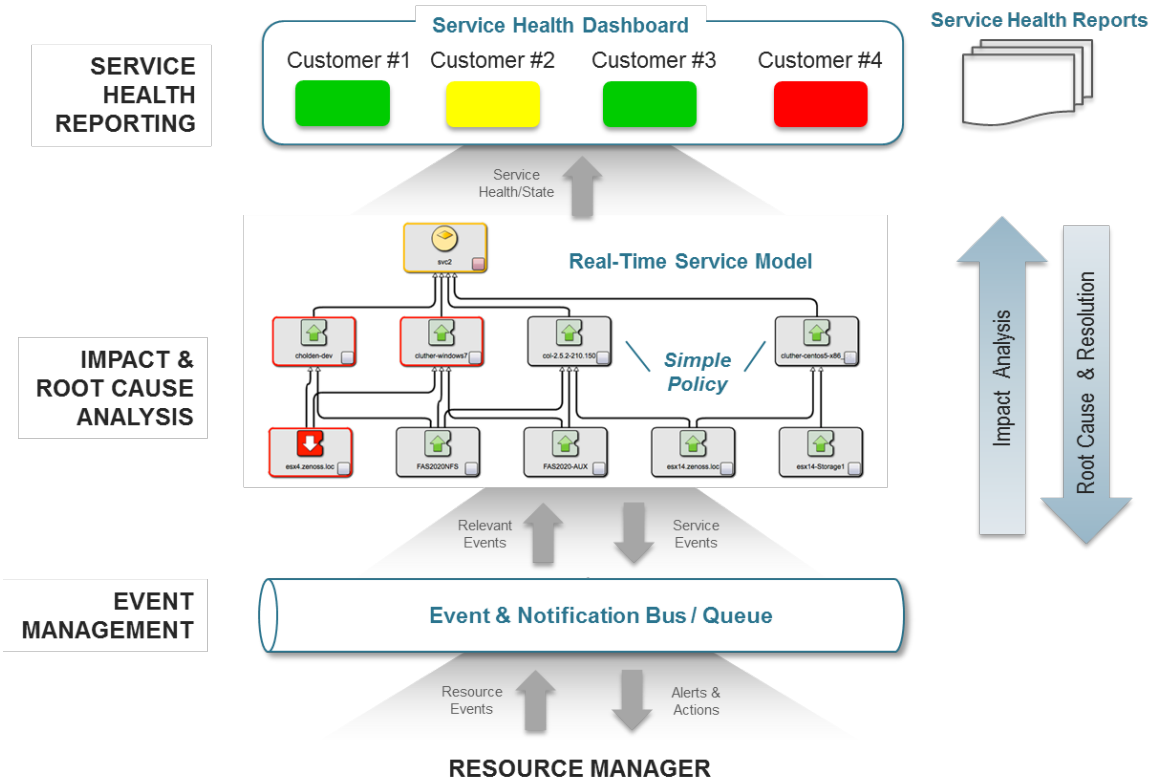


Figure 4: Zenoss Impact & Root Cause Analysis

- **Dynamic Service Modeling** - Zenoss Service Dynamics maintains a real-time service model, and automatically discovers infrastructure dependencies. Service constructs can be easily defined based on logical business constructs to define the infrastructure groupings supporting a specific application service. For example, a CRM application service might require Email, Web, and Database services to be present in order to operate. These logical definitions define the collection of services required for the CRM service to be considered functional. Once this logical service hierarchy is defined, the application or OS instances delivering the service functions are associated and Zenoss Service Dynamics takes care of the rest. Using advanced dependency modeling capabilities, Zenoss Service Dynamics will pull in all relevant infrastructure elements. VM partitions, blades, chassis, storage, network interfaces, and a wide variety of device components are all automatically discovered and mapped into the relevant service dependency graphs.

- **Dynamic Impact Analysis** - Identifies which services are affected by conditions in supporting components or infrastructure. For example, a failing fan in a Cisco UCS chassis might result in dozens of virtual machines being moved to new virtual hosts. With Zenoss impact analysis, IT operations can determine which business services will be affected by the fan failure and can plan corrective action to minimize service level disruptions.

- **Dynamic Root Cause Analysis -** Allows IT operators to quickly identify the specific events most likely to be the cause of a service impacting condition. In complex IT environments, it is not uncommon for a single component failure to cause a cascade of failures, resulting in an event storm totaling thousands of individual events. Zenoss Service Dynamics includes a proprietary **Confidence Ranking Engine** built on top of Impact Analysis to quickly triage these events and identify where IT resources should be applied to correct these types of situations. This algorithm filters impact events based on a variety of criteria including severity of the event, service graph depth, and the number of graph branches affected by an event. This ranking algorithm allows IT operators to target resources to address events deemed the most likely cause of a service failure or degradation. Real world deployments of Zenoss Service Dynamics have validated the effectiveness of the Service Impact framework by demonstrating significant event reduction and highly accurate identification of root cause events.

- **Simple, Modular Policy: "Policy Gates"** - Traditional impact managers require complex top down rule sets to be defined which require either static IT infrastructure or a detailed understanding of all possible infrastructure configurations to identify service impact or determine root cause. This approach fails in dynamic virtualized or cloud datacenters due to the need to maintain hard dependencies on named infrastructure elements. In contrast, Zenoss Service Dynamics uses Policy Gates to define impact rules on an element-by-element basis, and rolls up impact results through the current service model to reach conclusions. The design premise of this system is that the state of any given element in a service graph is determined by analyzing the state of the immediate children of that element. A change in the state of a given element is propagated to the parents of that element, causing the parents to evaluate their own state using their own Policy Gate configurations. The net result of this approach is that functions such as event aggregation, filtering, de-duplication and masking are provided automatically, eliminating the need for highly specialized skills to write impact rules and dramatically reducing human effort in event processing.

- **Unified, Scale Event Management** – Aggregate and manage events for your entire IT environment with a next generation event management system that provides automated event normalization and enrichment, and is easily extended, integrated and scaled through an embedded message bus. Zenoss Service Dynamics is capable of processing in excess of 1,500 events per second with a single event processor. Field deployments have shown the system is capable of quickly parsing through event storms scaling to thousands of events in seconds resulting in just a handful of events after processing through Service Impact and Confidence Ranking.

- **Automated Service Remediation** – Automatically respond to conditions in your environment to ensure service levels are maintained in real time; enables critical cloud operations capabilities including auto-scaling of cloud-based applications and real-time workload bursting across private to public clouds

Dynamic Analytics and Optimization

The final step of the service assurance lifecycle is historical analysis and planning. Only with deep, cross-domain analytics can you perform capacity planning and drive optimization of the environment. Zenoss Service Dynamics enables this through its integrated analytics capabilities that directly leverage the real-time service model and all state information from the Resource, Impact and Event Management modules. Leveraging a powerful, open business intelligence engine, Zenoss analytics capability provides a scalable and rich analytics platform that addresses tenant reporting needs, management dashboards, capacity planning and insight for optimization. Specific capabilities include:

- **Turnkey Operations Data Warehouse** – Automatically aggregates and normalizes configuration, performance and event history for your entire IT environment across physical, virtual and cloud-based infrastructure and applications
- **Unified Historical Analytics** – Understand utilization and health trends across your IT entire infrastructure including tenant-based consumption and availability reporting; Gain deep, timely insight through drag-and-drop dashboards, out-of-the box reports and powerful ad-hoc analytics all available through a multi-tenant web portal
- **Predictive Analytics** – Forecast capacity needs and anticipate availability problems through predictive trending that allows you to visualize and proactively manage upcoming operational issues and infrastructure requirements

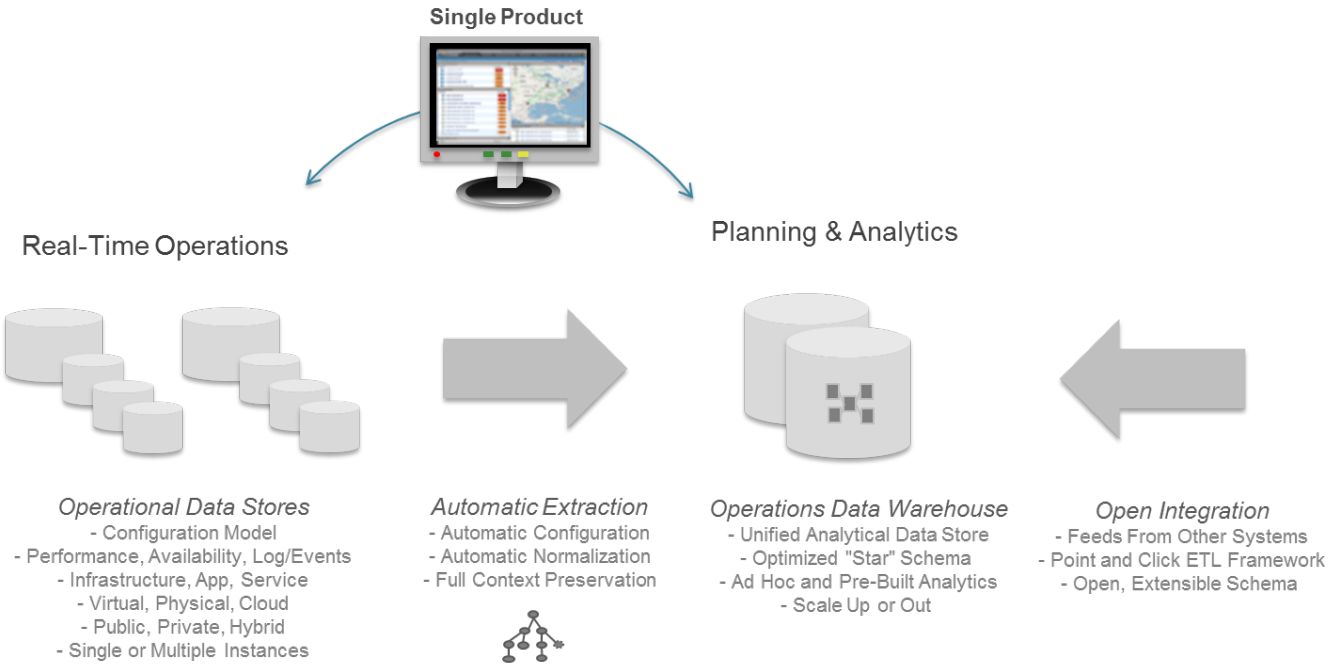
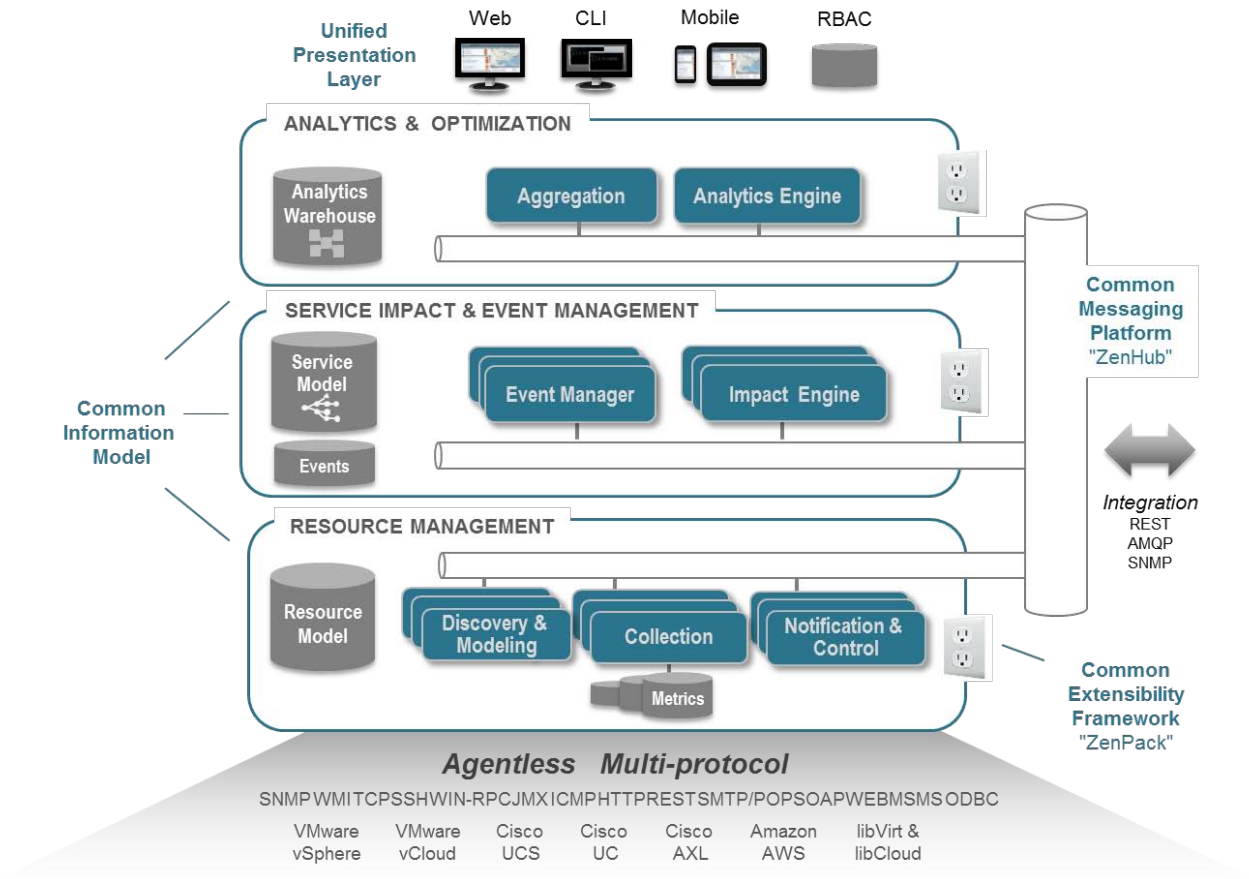


Figure 5: Zenoss Analytics and Optimization

Architecture Overview

Designing an end-to-end service assurance platform that can address the functionality, scalability and extensibility requirements of the hybrid cloud datacenter is a major challenge. Zenoss Service Dynamics achieves this through a layered approach that has been built from the ground up to take on the challenges of the cloud era. The illustration and discussion below highlights key aspects of the architecture that distinguish it from other platforms.



- **Unified Design** - End-to-end service assurance and analytics capability designed from the ground up as one product on a common architecture.
- **Horizontal Scaling** - Scale your deployment to manage hundreds of nodes from a single server to 100K nodes in globally distributed configuration leveraging low-cost hardware. Scale as needed to manage your elastic infrastructure.
- **Agentless, Multi-protocol** - Agentless collection and control platform that leverages a suite of secure access methods, management APIs and synthetic transactions to instrument the full stack at scale without the need for proprietary agents
- **Open Integration & Extensibility Framework** - Rapidly extend, customize and integrate with other management tools, leveraging our open architecture and "ZenPack" plug-in framework. Leverage a global community of extension developers and partners.

Summary

IT infrastructure and delivery is undergoing a major evolution, marked by the transition to a service-based approach leveraging virtualization and cloud architectures. This new model challenges traditional assumptions, processes and tools. To capture its value, management processes and technologies must be re-examined throughout the operations lifecycle.

Zenoss Service Dynamics uniquely addresses the service assurance challenges this new, dynamic world of IT that spans physical, virtual and cloud-based infrastructure.