Sonatype

Sonatype CLM "Hybrid SaaS" Architecture

Combining the benefits of SaaS & On Premise

WHITE PAPER

About Sonatype

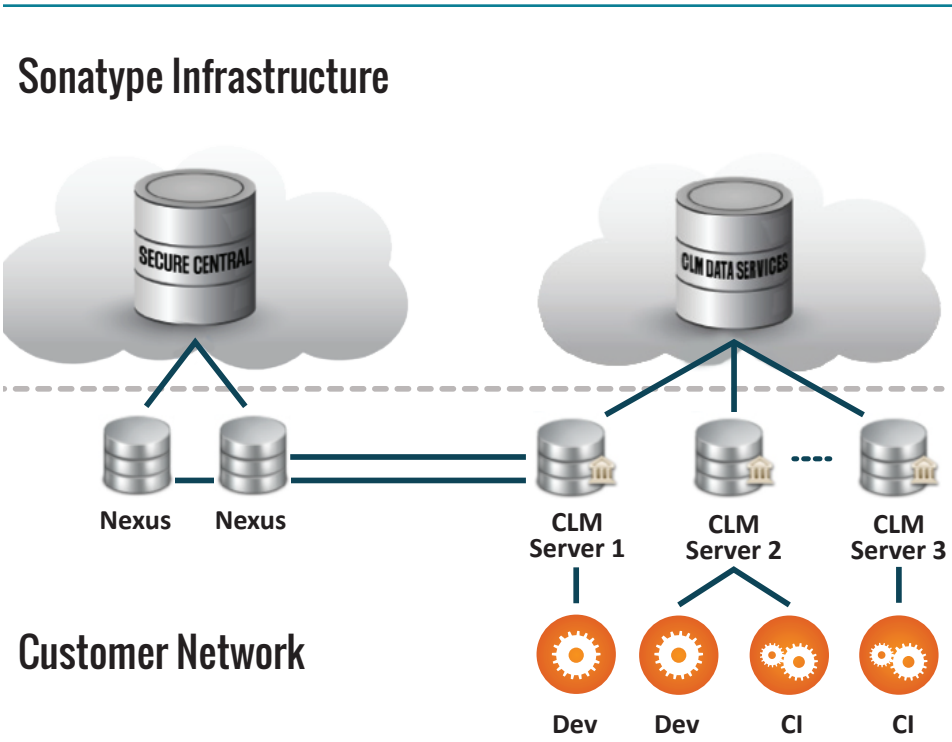
Sonatype has been on the forefront of creating tools to manage, organize, and better secure components since the inception of the Central Repository and Maven in 2001. Today, over 70,000 companies download over 8 billion components every year from the Central Repository, demonstrating the explosive growth in component-based development. Today's software ecosystem has created a level of complexity that is increasingly hard to manage. Partnering with application developers, security professionals and the open source community, Sonatype has introduced a way to keep pace with modern software development without sacrificing security. We call it Component Lifecycle Management (CLM), the new platform for securing the modern software supply chain.

We believe that to achieve application security, the approach has to be simple to use, integrated throughout the lifecycle and ensure sustaining trust. With CLM we're improving the visibility, management and security of component-based development across the entire lifecycle. Together with our customers, we're ushering in a new era of application security.

Sonatype CLM (Component Lifecycle Management) is a new solution built to secure the modern software supply chain. CLM tracks usage, enforces policy and prevents the use of flawed components throughout the software supply chain. Since CLM plays a critical role in your infrastructure, you need to trust that the CLM will meet the needs of your users - it needs to perform reliably, quickly and be constantly available to all regardless of location. At the same time, it needs to be easy to manage with minimal infrastructure and operational costs. CLM is architected to accomplish this by utilizing the advantages of the cloud and on-premise deployment.

CLM Architecture Utilizes Hybrid, SaaS Strategy

The Sonatype CLM platform operates as a "distributed SaaS" – a hybrid topology yielding the benefits of SaaS-based deployments (e.g. streamlined software updates, reduced infrastructure and operational costs) but with the privacy and controls typically associated with deployments within corporate networks. In this model Sonatype operates, on its infrastructure, the CLM Data Services backend, which is the source of FOSS component metadata, event notifications (e.g. the disclosure of new component security vulnerabilities) and CLM platform software updates.



All other technology is deployed within your network, protecting your intellectual property. Information exchanged with the Sonatype SaaS is limited to hashes of binary components used to retrieve corresponding metadata, that is license, security, popularity, version and other associated component information. All of your specific data such as application and organizational policies or proprietary component metadata is stored within your network. The CLM Server (depicted in Figure 3.1) takes feeds from the CLM Data Services backend to enable analysis and reporting on components used within your application portfolio. The CLM Server acts, among other things, as a proxy between all of the various integration points (e.g. IDE, CI, binary repository) deployed within your environment and the CLM Data Services SaaS run by Sonatype. This provides three important benefits:

- Your intellectual property, such as source, object or executable code or application specific governance policy is kept within your network and never transmitted to Sonatype.
- Data is updated automatically on a near real-time basis with the Data Services SaaS. The frequency of updates is a function of underlying data types. Immediate updates (e.g. new security vulnerabilities) are achievable through event notifications sent from SaaS to subscribing CLM Servers and passive updates are made through local cache expiration and revalidation/refresh with varying expiration windows.
- Overall system performance is optimized by caching data locally, providing extremely fast application and component analysis against timely data.

Since the CLM Server acts as an intermediary between all integration points and caches component metadata locally, the traffic pattern between your system and Sonatype is analogous to the traffic between existing Nexus users and the Central Repository today – or for that matter, any other tooling that interacts with the Central Repository.

CLM Enforcement Points

The CLM Server works in conjunction with an array of enforcement points across the entire application lifecycle, including the following:

- IDE (Integrated Development Environment)
- CI (Continuous Integration) Server
- Binary Repository Manager (e.g. Nexus)
- Ad Hoc, API-based Integration

By having multiple enforcement points along the various stages of an application lifecycle combined with a single consistent policy in effect for the application, actions that are contextually appropriate for the point of enforcement can be applied. There is no need to have an IDE policy that differs from the CI policy. Instead, the actions taken in response to policy violations can vary.

For example, it is perhaps acceptable for an unapproved open source component to be used in a development context while it is undergoing review. One policy constraint for the application is, in effect, "all components must be on the approved list." In the IDE context, the action may be simply to warn the developer that the component is not yet approved. However, if an attempt were made to release the component from a Nexus staging repository, the action would likely be to fail the release process.

All of these enforcement activities occur within the confines of your network as the individual enforcement (integration) points work in conjunction with the CLM Server. The resulting data that is collected as part of this is captured and used for reporting purposes as well as auditing. It is never outside of your control.

Please also see the descriptions given in schedule 4 (capacity planning and geographic deployment topology) and schedule 5 (product functionality) for additional information about the platform architecture.

Architecture Deployment Optimizes Capacity and Performance

Sonatype CLM is designed to meet the needs of any size organization. The CLM supports multiple deployment topologies based on your individual requirements.

Information such as breadth of integration points, build environment complexity, build rates, peak usage based on geographic organizational distribution, rollout scope and timing and network resources and other environmental constraints all influence the ideal target enterprise deployment architecture.

Scaling Options

The Sonatype CLM platform supports multiple scaling options. Local scale out (n+1) is achieved by adding nodes within a given (local) network environment. For optimal performance and availability, a minimum of three local nodes per location (a cluster) is recommended for redundancy, with appropriate DNS and/or hardware-based load balancing for inbound request routing under failure/maintenance situations. Additional nodes per location are added to support incremental scale and maintain high performance. Geographic redundancy is possible, and recommended for distributed teams where practical for performance reasons and also for overall fault tolerance.

Resiliency and Redundancy

As outlined in the scaling options above, both local and geographic scale out options are possible, in arbitrary combinations. Given the critical nature of the infrastructure leveraging services from the CLM platform, both local and geographic redundancy is recommended, though not required. This topology protects against multiple, simultaneous failure modes, allowing for continuity of operations with degraded performance.

Communication Protocols

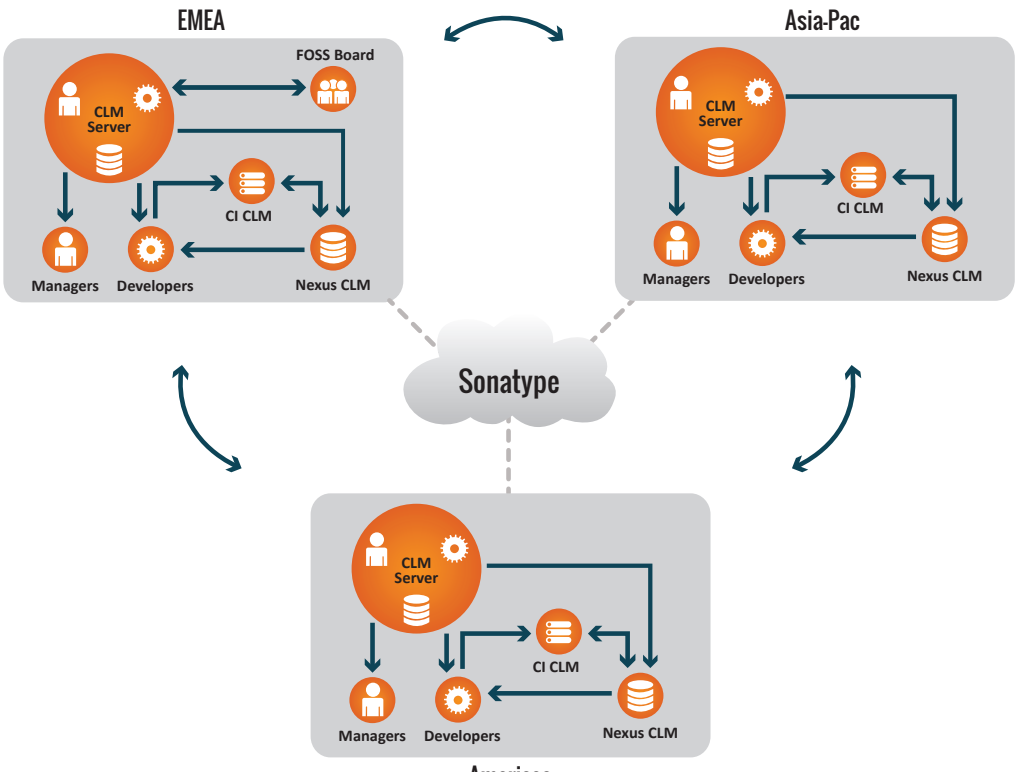
All system/service traffic is TCP-based with configurable port settings with the exception of the mandatory use of port 443 (with TLS) in communicating with the CLM Data Services SaaS operated by Sonatype. Traffic between CLM Servers, both local and geographically deployed, requires multiple TCP ports, all of which are configurable.

Architecture Deployment Example

To illustrate the flexibility of the CLM architecture, let's look at an example. This example is relatively complicated since it supports a distributed development organization that is managing a large number of projects with enforcement points throughout the entire development lifecycle.

- Key environmental requirements for this example:
- Geographically distributed development in America, EMEA, Asia Pacific
- Centralized FOSS board in EMEA
- Support for thousands of active application projects
- Integration or enforcement points across IDE, CI and Repository Manager
- Support for horizontal scalability (scale out)
- Localized resiliency for site specific performance
- Overall geographic fault tolerance
- Take advantage of existing infrastructure load balancing capabilities
- Secure (private), network connectivity between geographically deployed CLM Server instances

High-Level Deployment Architecture



Additional considerations that will help optimize the CLM deployment:

- Situate each CLM server geographically proximate to clients (IDE, CI, Nexus integration points) that are requesting component/application analysis – more specifically, having high-bandwidth, low latency connectivity between these endpoints is ideal
- Use of multi-core, multi-socket CPU architectures can yield near linear improvements in scan processing performance per core with adequate disk and memory resources
- High bandwidth, low latency connectivity between each CLM server and the Sonatype Data Services SaaS ensures rapid request servicing for metadata and improves overalll application and component analysis times

Sonatype Experts Will Team with You to Design an Optimal Strategy

Sonatype experts will work with you to identify your expected usage patterns and key environmental requirements. A deployment architecture will be designed to meet those needs. This architecture design will account for your geographical, performance, scalability, resiliency, network and hardware needs.

The architecture planning service includes the use of a capacity planning spreadsheet for sizing the CLM server deployment. The capacity plan is based on factors such as number of builds, number of stages, number of developers, number of application inventory assessment, etc. Using these factors, Sonatype will make recommendations about the server hardware required to meet your capacity needs.

In addition to helping design your deployment architecture; Sonatype has a number of other service offerings that can help expedite your overall CLM strategy and implementation. These include help optimizing your Nexus Repository Manager implementation, working with your Open Source Review Board to define a policy that accurately represents your risk tolerance, and designing a quick start program to improve your use of open source, proprietary and open source components.

Summary

The Sonatype CLM is designed to support flexible deployment models that will meet your current and future needs. Given the critical nature of CLM, the architecture supports extreme performance and scalability requirements along with the ability to address high availability with resiliency support.

Sonatype "distributed SaaS" combines the simplicity of SaaS with the security and control of on-premise deployment. Since Sonatype manages the SaaS CLM Data Services backend, software updates are streamlined, and your infrastructure and operational costs are reduced. Since your key application information never leaves your corporate network, you still have the privacy and control necessary to secure your environment.

Sonatype complements this advanced architecture approach with expertise that helps you get up and running quickly. Sonatype will work with you to design a deployment architecture based on your individual needs.

Sonatype Inc. · 12501 Prosperity Drive, Suite 350 · Silver Spring, MD 20904 · 1.877.866.2836 · www.sonatype.com