Kubernetes: What Needs Monitoring & Why
MONITORING CHALLENGES

Monitoring done right is still just a pipedream for most companies. And it’s a serious problem. Insufficient monitoring can have some serious consequences (some universal, others that are exemplified in Kubernetes):

- Without the right monitoring, operations can be interrupted.
- Your SRE team may be unable to respond to issues (or the right issues) as fast as needed.
- Monitoring management must reflect the state of clusters and workloads.
- Manual configuration increases availability and performance risks because monitors may not be present or accurate enough to trigger changes in key performance indicators (KPIs).
- Undetected issues may cause SLA breaches.
- Noisy pagers can result due to incorrect monitor settings.

The push towards observability - the ability to discover unknown unknowns - helps overcome these challenges and requires the use of the right observability tools to optimize monitoring capabilities.
WHAT TO MONITOR

With Kubernetes, you have to build monitoring systems and tooling to respond to the dynamic nature of the environment. Thus, you want to focus on availability and workload performance. One typical approach is to collect all of the metrics you can and then use those metrics to try to solve any problem that occurs. It makes the operators’ jobs more complex because they need to sift through an excess of information to find what’s important. At present, there’s also not a de facto standard for what to alert on, but there are monitors often created as Kubernetes best practice:

- How healthy are the deployments?
- Do we have security checks running actively?
- Are we checking if the instances we’re using are right sized?
- Are we checking the software we’re running against a list of known CVEs?

It is difficult to always know what to monitor ahead of time, so you need at least enough context to figure out what's going wrong when someone inevitably gets woken up in the middle of the night and needs to bring everything back online. Without this level of understanding, your team cannot parse what should be monitored and know when to grin and bear turning on an alert.

- Kubernetes deployment with no replicas
- Horizontal Pod Autoscaler (HPA) scaling issues
- Host disk usage
- High IO wait times
- Increased network errors
- Increase in pods crashed
- Unhealthy Kubelets
- Nginx config reload failures
- Nodes that are not ready
- Large number of pods that are not in a running state
- External-DNS errors registering records

HOW TO MONITOR CLUSTER WORKLOADS

Monitoring is primarily set up to focus on infrastructure. But teams are also concerned with how to monitor cluster workloads. Because a diverse set of stakeholders is involved in the process, you must determine who is responsible for what from both an infrastructure and a workload standpoint. For instance, you want to make sure the right people are alerted at the right time to limit the noise of being alerted about things that do not pertain to you.

Tooling must be flexible enough to meet complex demands, yet easy enough to set up quickly so that we can move beyond tier 1 monitoring (e.g., Is it even working?”). Tier 2 monitoring requires dashboards that reveal where security vulnerabilities are, whether or not compliance standards are being met, and targeted ways to improve.

Impact and urgency are key criteria that must be identified and assessed on an ongoing basis. Regarding impact, it is critical to be able to determine if an alert is actionable, the severity based on impact, and the number of users or business services that are or will be affected. Urgency also comes into play. For example, does the problem need to be fixed right now, in the next hour, or in the next day?
TOOLS FOR MONITORING KUBERNETES

As the Kubernetes industry matures, there are a number of tools that have been introduced to improve monitoring.

One thing to note, unless your core business is monitoring, you should leave the management of this cost sink to people who really and truly care about it because it’s their business. You will never make money from monitoring and instead of diverting your precious resources, you will be much better off using those resources for revenue generation.

**Prometheus**

*Prometheus*, an open source service monitoring system, offers a standardized way to collect and expose metrics in clusters and your workloads. Prometheus works well for recording any purely numeric time series. It fits both machine-centric monitoring as well as monitoring of highly dynamic service-oriented architectures. In a world of microservices, its support for multi-dimensional data collection and querying is a particular strength.

Prometheus is designed for reliability, to be the system you go to during an outage to allow you to quickly diagnose problems. Each Prometheus server is standalone, not depending on network storage or other remote services. You can rely on it when other parts of your infrastructure are broken, and you do not need to set up extensive infrastructure to use it.

**OpenMetrics**

*OpenMetrics*, an initiative sponsored by the Cloud Native Computing Foundation (CNCF), is an effort by the community to create an open standard for transmitting those metrics. Still in its inception, OpenMetrics is a working group to determine a standard for exposing metrics data, influenced by the Prometheus exposition format.
Datadog

Datadog is a monitoring and security platform for cloud applications. Its SaaS platform integrates and automates infrastructure monitoring, application performance monitoring and log management to provide unified, real-time observability of our customers’ entire technology stack. Datadog has already built out the systems and mechanisms for you to start monitoring on day one. It’s costly, but out-of-the-box you get a robust monitoring solution that does not require a lot of implementation effort. You can use Datadog to create metrics and then use their portal to view and manage monitors.

Astro

Astro is an open source software project built by Fairwinds to help you achieve better productivity and cluster performance. Built to work with Datadog, Astro watches objects in your cluster for defined patterns and manages Datadog monitors based on this state. As a controller that runs in a Kubernetes cluster, it subscribes to updates within the cluster. If a new Kubernetes deployment or other objects are created in a cluster, Astro knows about it and creates monitors based on that state in your cluster. Essentially, it provides a mechanism for dynamically creating and managing alerts in a way that Datadog can understand.

All you have to do is determine what you want to monitor and describe the types of monitors you want for different objects in a cluster. From there, Astro manages that state based on your Kubernetes cluster. Once you define how you want Astro to manage that state, monitoring is done by annotating objects in your cluster. As long as the monitors have the right annotations, they will automatically be created and routed to the right notification endpoints.
SUMMARY

Monitoring the health of infrastructure and applications is essential to every company—even if done poorly—even if, as in many teams, it is for the sole purpose of understanding the root cause of problems.

As teams standardize around Kubernetes, specific tools will become more and more universally accepted as standards. The focus can then shift from innovation velocity toward hardening. Once at the hardening stage, teams can then set standards for monitoring. These standards make it easier to monitor more of the right things. In the not-so-distant future, streamlined configurability and accessibility, along with straightforward tier 2 monitoring, will be the norm instead of the exception.

Although Kubernetes is designed to simplify container orchestration using many built-in features, it can be challenging to configure, deploy, and monitor. Teams that lack the inhouse expertise needed to use Kubernetes might consider a managed Kubernetes solution.

ADDITIONAL RESOURCES:

- Kubernetes Best Practices
- Fairwinds Kubernetes Managed Service
- Fairwinds Insights
WHY FAIRWINDS

Fairwinds is your trusted partner for Kubernetes security, policy and governance. With Fairwinds, customers ship cloud-native applications faster, more cost effectively, and with less risk. We provide a unified view between dev, sec, and ops, removing friction between those teams with software that simplifies complexity. Fairwinds Insights is built on Kubernetes expertise and integrates our leading open source tools to help you save time, reduce risk, and deploy with confidence.

WWW.FAIRWINDS.COM