

Platform monitoring

Platform monitoring

SamKnows uses Nagios, based on two external vantage points in London and New York, to monitor all its server infrastructure. This includes both the core SamKnows infrastructure (such as web servers, database servers, big data cluster, data collection servers etc.), and test servers.

We run both active and passive host-level checks; a large number of active service checks dependent on a server's role and health checks on our data.

Alerts are generated by Nagios and are managed by the SamKnows infrastructure team who respond to issues 24x7x365.

Failover, redundancy and resilience

SamKnows infrastructure is designed in a fault-tolerant manner to avoid single points of failures and is built with the intention that we should be able to lose any critical process, host or site and continue normal operation. In some cases, this is achieved at the application layer and in other cases we rely on load balancing techniques (either hardware load balancers or DNS based load balancing). However, SamKnows does require that multiple distinct instances of each server exist in separate physical locations in order to be able to guarantee the resilience of the platform. All of our critical services have provider-redundancy and geo-redundancy. Failover is automated in all instances where possible so as to react faster than a human can although the SamKnows infrastructure team would be notified to respond and deal with any failures to ensure continued high availability and redundancy is retained.

Our big data cluster (metric data), metadata database servers, data collection servers and global platform of web servers have even higher fault tolerances than a single server or service due to their size and critically.

More information on redundancy of individual services is mentioned throughout infrastructure documentation.

Checks

Some examples of our checks are listed below

Networking and host

- Reachability (Ping over IPv4 and IPv6)
- Latency
- Packet loss
- CPU utilisation
- Number of running processes
- Memory usage and free memory
- Disk usage and free disk space
- Network utilisation
- Firewall status
- Logged in users

Services & applications

- Checking APIs and Web Servers are actively reporting data over HTTPS
- Data ingestion and collection services are running successfully
- SamKnows test server client software is running and accessible
- Generic health checks for general services such as databases, hadoop platform services, web servers, caches
- Service observability metrics (e.g. Distributed systems communications or storage usage, database query performance)

Data monitoring

We also have monitoring in place designed to detect significant changes in the amount of data being ingested into our big data platform.

Application & API error monitoring

We also actively monitor errors that occur on APIs including APIs used to power SamKnows One and client-facing APIs. We do this using Sentry (for error monitoring) and an Elasticsearch cluster (for logging). Relevant development teams react to errors in order to address and bugs or issues as soon as possible, which may include escalations to infrastructure when it is due to a host or service outage.

Agent and test server monitoring

In addition to monitoring test servers from our monitoring vantage points, we also internally use SamKnows One analytics and alerting functionality in order to detect issues with measurement agents reporting themselves.

Measurement agents report on set test schedules and we monitor for missing data in order to identify issues that may be related to the measurement agents or test servers. This also allows us to monitor uptime of measurement agents and success of tests themselves (full integration of test servers and test clients).