УДК 004.043

Шарафутдинов Эмиль Михайлович, магистрант, Московский государственный технологический университет «СТАНКИН», г. Москва
Семячкова Елена Геннадьевна, кандидат технических наук, доцент кафедры информационных технологий и вычислительных систем, Московский государственный технологический университет «СТАНКИН», г. Москва

МОНИТОРИНГ KUBERNETES-КЛАСТЕРА: КОНЦЕПЦИЯ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДСИСТЕМЫ ОПОВЕЩЕНИЙ

Аннотация

В статье рассматривается процесс мониторинга Kubernetes-кластера, который играет важную роль в обеспечении надёжности и доступности сервисов в облачных и корпоративных средах. Одной из основных проблем мониторинга является неэффективная организация оповещений, приводящая к увеличению времени реакции на инциденты. Описаны этапы мониторинга Kubernetes и его технологические особенности. Выделены основные проблемы в организации оповещений: жёсткие пороги в правилах срабатывания, отсутствие механизмов подавления вторичных сигналов, отсутствие разделения по важности, неурегулированные правила направления уведомлений. Предложена концепция подсистемы интеллектуального оповещения, направленная на снижение трудоёмкости сопровождения Kubernetes-кластера за счёт уменьшения информационного шума и корректной доставки уведомлений. Рассмотрены существующие решения для реализации подсистемы оповещений в Kubernetes.

Annotation

The article examines the process of monitoring a Kubernetes cluster, which plays an important role in ensuring the reliability and availability of services in cloud and

enterprise environments. One of the main problems of monitoring is the inefficient organization of alerting, which leads to increased incident response time. The stages of Kubernetes monitoring and its technological features are described. The main problems in alert organization are identified, including rigid thresholds in alerting rules, the absence of mechanisms for suppressing secondary signals, the lack of differentiation by importance, and unregulated rules for routing notifications. A concept of an intelligent alerting subsystem is proposed, aimed at reducing the labor intensity of Kubernetes cluster maintenance by decreasing informational noise and ensuring proper notification delivery. Existing solutions for implementing an alerting subsystem in Kubernetes are also considered.

Ключевые слова: Kubernetes, мониторинг, контейнеры, телеметрические показатели, оповещения, события, кластер, инциденты, отклонения, динамичность среды, фильтрация, группировка, уведомления, панель мониторинга, подсистема оповещений, Prometheus, Alertmanager, Grafana Loki, VictoriaMetrics, анализ аномалий, прогнозирование, каналы доставки.
Keywords: Kubernetes, monitoring, containers, telemetry indicators, alerts, cluster, incidents, deviations, dynamic environment, filtering, grouping, notifications, dashboard, alerting subsystem, Prometheus, Alertmanager, Grafana Loki, VictoriaMetrics, anomaly analysis, forecasting, delivery channel.

Современные облачные и корпоративные сервисы требуют надёжных систем для мониторинга и управления ресурсами [1]. Одной из таких систем является Kubernetes, широко используемая для оркестрации контейнерных приложений. Надежность таких систем зависит от своевременного выявления отклонений в работе компонентов. В связи с ростом сложности и динамичности Kubernetes-кластеров, процесс мониторинга требует новых подходов для улучшения эффективности и снижения нагрузки на команды эксплуатации.

Мониторинг — это постоянное отслеживание состояния серверов, приложений и компонентов системы (кластера) для выявления отклонений и принятия мер. Система мониторинга собирает данные, например, использование ресурсов на рабочих узлах кластера, события на узлах и в важных приложениях, которые затем анализируются для формирования оповещений [3]. Для последующего рассмотрения процесса мониторинга целесообразно представить обобщённую схему Kubernetes-кластера (рис. 1).

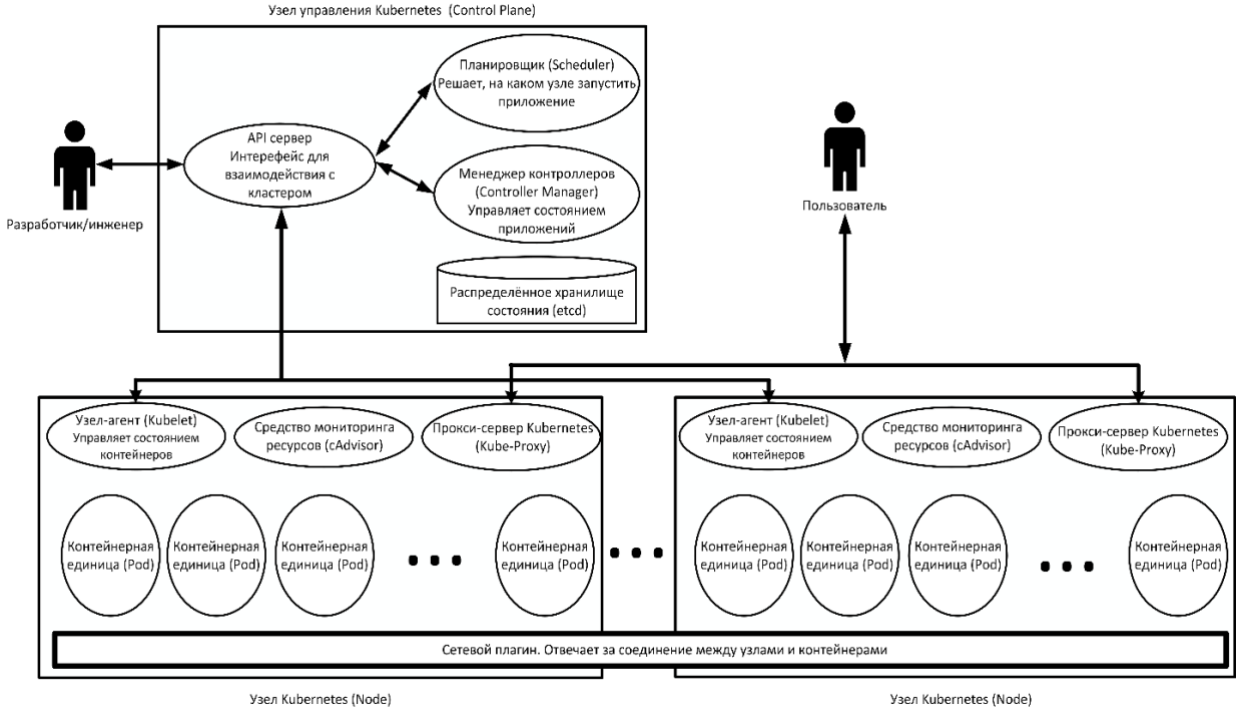

Рис. 1. Обобщённая схема Kubernetes-кластера

Сам процесс мониторинга состоит из нескольких этапов:

1. Возникает отклонение, при котором значения показателей работы приложения или инфраструктурных компонентов выходят за пределы нормы.
2. Инструменты мониторинга фиксируют необходимые показатели и события. Полученные данные обрабатываются и сохраняются.
3. Выполняется проверка условий, при которых должно формироваться оповещение.

4. Производится обработка оповещений, если настроены правила (группировка, фильтрация, подавление).
5. Оповещения направляются в соответствующие каналы взаимодействия, чтобы специалисты могли своевременно отреагировать на ситуацию.
6. Дополнительно текущее состояние системы визуализируется с помощью панелей мониторинга.
7. После восстановления штатной работы оповещение переводится в завершённое состояние.

Корректность сбора данных, скорость и форма доведения информации о проблеме до ответственных специалистов определяют эффективность процесса мониторинга. Значение имеет не только сам сбор телеметрических показателей, но и организация обработки оповещений. Совершенствование этой части процесса позволит уменьшить нагрузку при сопровождении Kubernetes-кластера и повысить качество его обслуживания.

Технологические особенности Kubernetes, которые влияют на организацию мониторинга:

1. Kubernetes – динамичная среда. Контейнеры приложений (поды) постоянно создаются, удаляются и перемещаются между рабочими узлами (нодами). Поэтому в данном случае нужно использовать метки для отслеживания приложения, а не уникальные идентификаторы.
2. При перераспределении рабочей нагрузки система мониторинга должна адаптироваться для корректного отслеживания новых целей.
3. В микросервисной архитектуре сбой одного компонента может вызвать проблемы в работе других. Мониторинг должен учитывать зависимости между сервисами и подавлять вторичные сигналы.
4. С увеличением объема данных возрастает риск перегрузки системы. Необходимо оптимизировать хранение данных и ограничить избыточность меток, чтобы не вызвать излишнюю нагрузку на инфраструктуру.

Как отмечалось ранее, на эффективность мониторинга напрямую влияет качество оповещений. Ошибки в их настройке могут вызвать информационный шум и увеличить трудозатраты на реагирование [2, 3, 4, 5].

К основным проблемам при организации оповещений относятся:

1. Неполный набор правил, который не учитывает особенности инфраструктуры.
2. Жёсткие пороги в правилах срабатывания, не предусматривающие изменение нагрузки.
3. Формирование оповещений по вторичным событиям, а не по первопричине.
4. Отсутствие группировки сообщений.
5. Неопределённые уровни важности событий в оповещениях.
6. Неурегулированные правила направления уведомлений.
7. Отсутствие периодов тишины, когда оповещения не приходят.
8. Перегруженные сообщения в оповещениях.

В свою очередь, эти проблемы приводят к игнорированию оповещений, увеличению времени реакции на инциденты, пропуску значимых событий, и, как следствие, снижается качество сопровождения.

Разработка и внедрение подсистемы интеллектуального оповещения позволит улучшить процесс мониторинга, путём снижения уровня шума и ускорения реакции на важные события в Kubernetes-кластере. Эта подсистема входит в состав общей системы мониторинга. Сводные положения об области её применения приведены в таблице 1.

Таблица 1. Область применения подсистемы оповещений

Аспект	**Охватывается**	**Не охватывается**
Типы событий	Инциденты и отклонения, которые выявляются на основе показателей и правил: недоступность компонентов, увеличение числа ошибок, ухудшение производительности, нехватка	Анализ полных журналов событий (логов) и трассировок, которые содержат записи о действиях и ошибках в системе, а также бизнес-показателей, которые не представлены в виде данных для мониторинга.

	ресурсов; события, происходящие в системе управления Kubernetes.	
Уровни важности	Критические и высокие события; предупреждения (при необходимости).	Информационные рассылки и уведомления без эксплуатационной ценности.
Среды эксплуатации	Производственная среда (prod); при наличии — стенд предэксплуатации.	Разработческие и локальные среды.
Каналы доставки	Корпоративный мессенджер (например, Telegram) и механизмы для автоматического обмена данными между системами (веб-хуки) для интеграций.	Электронная почта, SMS, голосовая телефония.
Потребители	Дежурные инженеры эксплуатации, ответственные разработчики/дежурные по сервисам.	Конечные пользователи и внешние клиенты.
Функции обработки	Группировка, фильтрация, подавление вторичных сигналов, маршрутизация, периоды тишины, повторные доставки.	Автоматическое исправление, пост-инцидентный разбор.

Рассмотрены несколько популярных подходов для реализации подсистемы оповещений в Kubernetes. Среди них комплекс Prometheus и Alertmanager, Grafana Loki и VictoriaMetrics.

1. Комплекс Prometheus и Alertmanager эффективен для мониторинга с жёсткими порогами в правилах срабатывания. Ограничен в гибкости [6, 7, 8, 9].
2. Grafana Loki работает с текстовыми данными и анализирует ошибки, которые невозможно отследить по числовым показателям. Минусы подхода - высокая ресурсоёмкость и сложность в настройке [10].
3. VictoriaMetrics использует анализ аномалий и прогнозирование для динамичных оповещений. Требует тщательной настройки [11, 12].

С помощью совместного применения нескольких подходов можно компенсировать ограничения каждого из них и тем самым достичь максимальной результативности подсистемы оповещений.

Таким образом, мониторинг Kubernetes-кластера является важным условием обеспечения устойчивой работы и доступности приложений, которые в нём развёрнуты. При этом важной проблемой остаётся организация оповещений: их неправильная настройка повышает нагрузку на специалистов эксплуатации и затрудняют своевременное реагирование на инциденты. Проведённый анализ показывает, что снижение этой нагрузки возможно за счёт внедрения подсистемы интеллектуального оповещения. Реализация такой подсистемы позволит повысить практическую ценность мониторинга и сократить трудозатраты при решении инцидентов.

Литература

1. Бабичев С. Л., Коньков К. А. Распределённые системы: учебное пособие. — М.: Юрайт, 2021. — 320 с. [Электронный ресурс]. URL: https://urait.ru/book/raspredelennye-sistemy-445188 (дата обращения: 14.03.2026). — Режим доступа: по подписке.
2. Grafana Documentation [Электронный ресурс]. URL: https://grafana.com/docs/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
3. Prometheus Documentation [Электронный ресурс]. URL: https://prometheus.io/docs (дата обращения: 14.03.2026). — Режим доступа: свободный.
4. Grafana Labs. Observability Survey Report 2023: Key Findings. — Grafana, 2023 [Электронный ресурс]. URL: https://grafana.com/observability-survey/2023/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
5. Prometheus Authors. Alertmanager Configuration (Telegram) [Электронный ресурс]. URL: https://prometheus.io/docs/alerting/latest/configuration/ (дата обращения: 14.03.2026). — Режим доступа: свободный.

6. Prometheus Authors. Alertmanager — Overview [Электронный ресурс]. URL: https://prometheus.io/docs/alerting/latest/alertmanager/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
7. Prometheus Authors. Alerting — Overview [Электронный ресурс]. URL: https://prometheus.io/docs/alerting/latest/overview/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
8. Grafana Labs. Configure notification policies (routing) [Электронный ресурс]. URL: https://grafana.com/docs/grafana/latest/alerting/configure-notifications/create-notification-policy/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
9. Grafana Labs. Configure silences [Электронный ресурс]. URL: https://grafana.com/docs/grafana/latest/alerting/configure-notifications/create-silence/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
10. Grafana Labs. Loki — Alerting and recording rules [Электронный ресурс]. URL: https://grafana.com/docs/loki/latest/alert/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
11. Grafana Labs. Dynamic alerting and metric forecasts (Grafana Cloud) [Электронный ресурс]. URL: https://grafana.com/docs/grafana-cloud/machine-learning/dynamic-alerting/ (дата обращения: 14.03.2026). — Режим доступа: свободный.
12. VictoriaMetrics. Anomaly Detection & Alerting Setup (vmanomaly + vmalert) [Электронный ресурс]. URL: https://docs.victoriametrics.com/anomaly-detection/guides/guide-vmanomaly-vmalert/ (дата обращения: 14.03.2026). — Режим доступа: свободный.

Literature

1. Babichev S. L., Konkov K. A. Distributed Systems: textbook. — Moscow: Yurayt, 2021. — 320 p. [Electronic resource]. URL:

https://urait.ru/book/raspredelennye-sistemy-445188 (date of access: 14.03.2026). — Access mode: subscription.

2. Grafana Documentation [Electronic resource]. URL: https://grafana.com/docs/ (date of access: 14.03.2026). — Access mode: open access.
3. Prometheus Documentation [Electronic resource]. URL: https://prometheus.io/docs (date of access: 14.03.2026). — Access mode: open access.
4. Grafana Labs. Observability Survey Report 2023: Key Findings. — Grafana, 2023 [Electronic resource]. URL: https://grafana.com/observability-survey/2023/ (date of access: 14.03.2026). — Access mode: open access.
5. Prometheus Authors. Alertmanager Configuration (Telegram) [Electronic resource]. URL: https://prometheus.io/docs/alerting/latest/configuration/ (date of access: 14.03.2026). — Access mode: open access.
6. Prometheus Authors. Alertmanager — Overview [Electronic resource]. URL: https://prometheus.io/docs/alerting/latest/alertmanager/ (date of access: 14.03.2026). — Access mode: open access.
7. Prometheus Authors. Alerting — Overview [Electronic resource]. URL: https://prometheus.io/docs/alerting/latest/overview/ (date of access: 14.03.2026). — Access mode: open access.
8. Grafana Labs. Configure notification policies (routing) [Electronic resource]. URL: https://grafana.com/docs/grafana/latest/alerting/configure-notifications/create-notification-policy/ (date of access: 14.03.2026). — Access mode: open access.
9. Grafana Labs. Configure silences [Electronic resource]. URL: https://grafana.com/docs/grafana/latest/alerting/configure-notifications/create-silence/ (date of access: 14.03.2026). — Access mode: open access.

10. Grafana Labs. Loki — Alerting and recording rules [Electronic resource]. URL: https://grafana.com/docs/loki/latest/alert/ (date of access: 14.03.2026). — Access mode: open access.
11. Grafana Labs. Dynamic alerting and metric forecasts (Grafana Cloud) [Electronic resource]. URL: https://grafana.com/docs/grafana-cloud/machine-learning/dynamic-alerting/ (date of access: 14.03.2026). — Access mode: open access.
12. VictoriaMetrics. Anomaly Detection & Alerting Setup (vmanomaly + vmalert) [Electronic resource]. URL: https://docs.victoriametrics.com/anomaly-detection/guides/guide-vmanomaly-vmalert/ (date of access: 14.03.2026). — Access mode: open access.