

ВЫБОР ПАРАМЕТРОВ ДЛЯ ОЦЕНКИ ОТКАЗОУСТОЙЧИВОСТИ СЕРВЕРНОГО КЛАСТЕРА

Леонов М.А.

Омский государственный технический университет, Омск

Ключевые слова: серверный кластер, отказоустойчивость, надежность, мониторинг, производительность, метрики, распределенные системы.

Аннотация. Рассмотрены методы сбора данных для анализа состояния серверного кластера на основе мониторинга ключевых параметров и балансировки нагрузки. Предложен подход к сбору и анализу метрик, таких как использование CPU, RAM, сетевая активность и время отклика, с использованием современных инструментов. Описаны основные параметры, влияющие на устойчивость кластера, и предложены новые производные метрики для более глубокого анализа. Представлены примеры входных данных, построены графики и рассмотрено применение этих метрик в реальных сценариях. Результаты исследования демонстрируют эффективность предложенного подхода для повышения надежности распределенных систем.

SELECTION OF PARAMETERS FOR ASSESSING THE FAULT TOLERANCE OF A SERVER CLUSTER

Leonov M.A.

Omsk State Technical University, Omsk

Keywords: server cluster, fault tolerance, reliability, monitoring, performance, metrics, distributed systems.

Abstract. The methods of data collection for analyzing the state of a server cluster based on monitoring key parameters and load balancing have been examined. An approach to gathering and analyzing metrics such as CPU usage, RAM utilization, network activity, and response time using modern tools has been proposed. The main parameters affecting cluster stability are described, and new derived metrics for deeper analysis are suggested. Examples of input data are provided, graphs are constructed, and the application of these metrics in real-world scenarios is considered. The results of the study demonstrate the effectiveness of the proposed approach for enhancing the reliability of distributed systems.

Введение

Современные распределенные системы, такие как серверные кластеры, требуют высокой отказоустойчивости и надежности для обеспечения бесперебойной работы. В условиях высокой нагрузки и постоянных изменений в сетевой инфраструктуре критически важно отслеживать состояние серверов и балансировать нагрузку для предотвращения простоев и деградации производительности [1].

Ключевым инструментом обеспечения стабильности системы является мониторинг различных метрик, таких как загрузка процессора, потребление оперативной памяти, сетевая активность и время отклика. Однако, помимо базовых параметров, важную роль играют более сложные производные метрики, позволяющие глубже анализировать состояние системы и прогнозировать потенциальные проблемы. Рассматриваются методы мониторинга и

балансировки нагрузки, предлагаются новые метрики для оценки устойчивости кластера, а также приводятся примеры их применения.

Основные параметры серверного кластера

Разработчики клиент-серверных приложений и информационных систем разрабатывают новые методы хранения и кэширования пользовательских данных, чтобы уменьшить объем передаваемой по телекоммуникационным сетям информации и повысить скорость её доставки [2]. В стандартный набор параметров входят данные, полученные непосредственно с серверов, более детальное описание представлено в таблице 1.

Табл. 1. Параметры серверного кластера

Параметр	Описание	Влияние	Важность
<i>CPU</i>	Процент использования процессорных ресурсов	Высокая загрузка замедляет обработку запросов	Контроль предотвращает перегрузку
<i>RAM</i>	Объем используемой оперативной памяти	Недостаток приводит к использованию <i>swap</i>	Помогает избежать утечек памяти
Сетевая активность	Объем входящего и исходящего трафика	Высокая активность вызывает задержки	Выявление узких мест инфраструктуры
Время отклика	Время обработки запроса сервером	Долгое время отклика указывает на перегрузку	Оценка производительности
Доступность серверов	Время работы без сбоев	Низкая доступность приводит к простоям	Ключевой показатель отказоустойчивости
Насыщение ресурсов	Уровень загруженности системы	Высокая насыщенность приводит к деградации сервиса	Позволяет предсказать сбои и масштабировать ресурсы
Ошибки	Доля неудачных запросов	Высокий процент ошибок снижает доступность сервиса	Критический показатель надежности

В дополнение к базовым метрикам, предложены новые показатели для оценки отказоустойчивости и нагрузки серверного кластера:

1. Коэффициент отказоустойчивости (*Fault Tolerance Ratio, FTR*). Позволяет оценить устойчивость системы к сбоям.

$$FTR = \frac{N_{success}}{N_{total}}, \quad (1)$$

где $N_{success}$ – количество успешно обработанных запросов, N_{total} – общее количество запросов.

2. Среднее время восстановления (*Mean Time to Recovery, MTTR*). Оценка эффективности процедур восстановления серверов после сбоя.

$$MTTR = \frac{T_{idle}}{T_{fault}}, \quad (2)$$

где T_{idle} – суммарное время простоя, N_{fault} – количество сбоев за промежуток времени.

$$LEC = \frac{\sum_{i=1}^n \frac{CPU_i + RAM_i + NET_i}{RT_i}}{n}. \quad (3)$$

3. Коэффициент эффективности распределения нагрузки (*Load Efficiency Coefficient, LEC*). Комплексный показатель, оценивающий балансировку нагрузки

с учетом не только нагрузки, но и отклика системы. Позволяет оценить не только нагрузку, но и насколько эффективно она распределяется между серверами.

Тестовые данные

Входные данные представляют собой данные серверов, собранные с помощью Prometheus. Данные были получены в рамках имитационного моделирования кластерного сервера с использованием балансировки методом Rounded Robbin [3]. Были зафиксированы 100 записей с временным интервалом – 1 минута. На рисунке 1 приведен график изменения коэффициента эффективности распределения нагрузки LEC с течением времени на основе полученных данных.

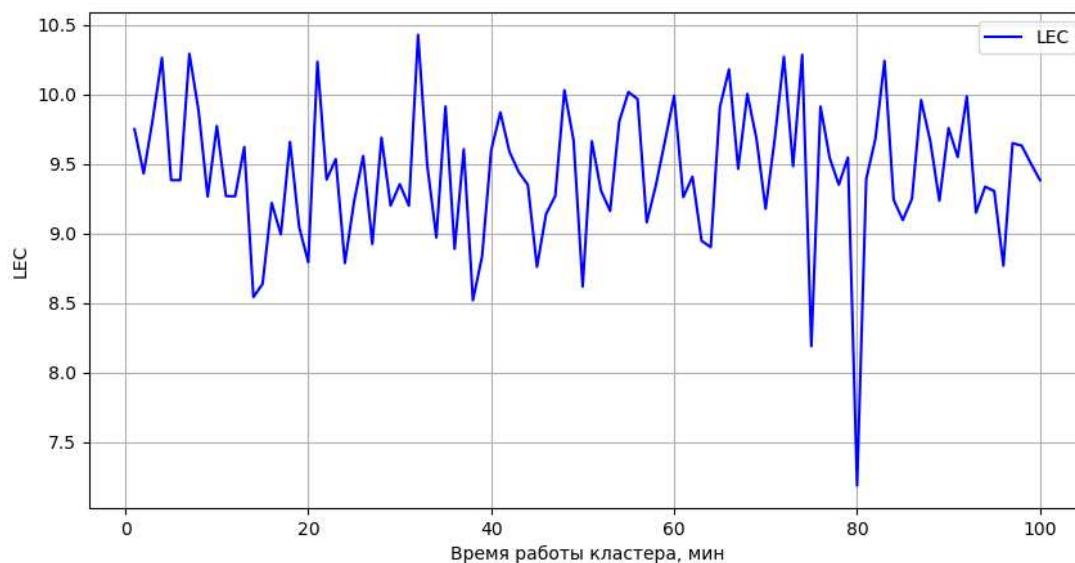


Рис. 1. Демонстрация изменения коэффициента LEC с течением времени

На графике видно стабильное изменение коэффициента эффективности распределения нагрузки (LEC) с одним резким пиком примерно на 80 выборке данных с сервера. Этот момент может свидетельствовать о перегрузке кластера – например, внезапном росте нагрузки или проблемах с балансировкой.

Для оценки устойчивости серверного кластера был проведен анализ распределения нагрузки между тремя серверами. В ходе эксперимента учитывались три ключевых параметра: загрузка процессора, использование оперативной памяти и сетевая активность. Особое внимание уделялось поведению системы в случае отказа одного из узлов.

В эксперименте серверная нагрузка распределялась равномерно между тремя узлами. Однако на 50-й временной итерации был смоделирован отказ одного из серверов (Сервер 2). После отключения данного узла его нагрузка перераспределилась между оставшимися серверами. Таким образом, было проведено исследование реакции системы на сбой и способности кластерной архитектуры адаптироваться к изменяющимся условиям. На графике (рис. 2) видно, что до 50-й итерации загрузка процессоров трех серверов колебалась в диапазоне 40-60%. Однако после отказа второго сервера (красная пунктирная линия) его нагрузка снизилась до нуля, а оставшиеся сервера (Сервер 1 и Сервер 3) начали компенсировать недостающие вычислительные ресурсы. Их загрузка возросла, что подтверждает перераспределение нагрузки в системе.

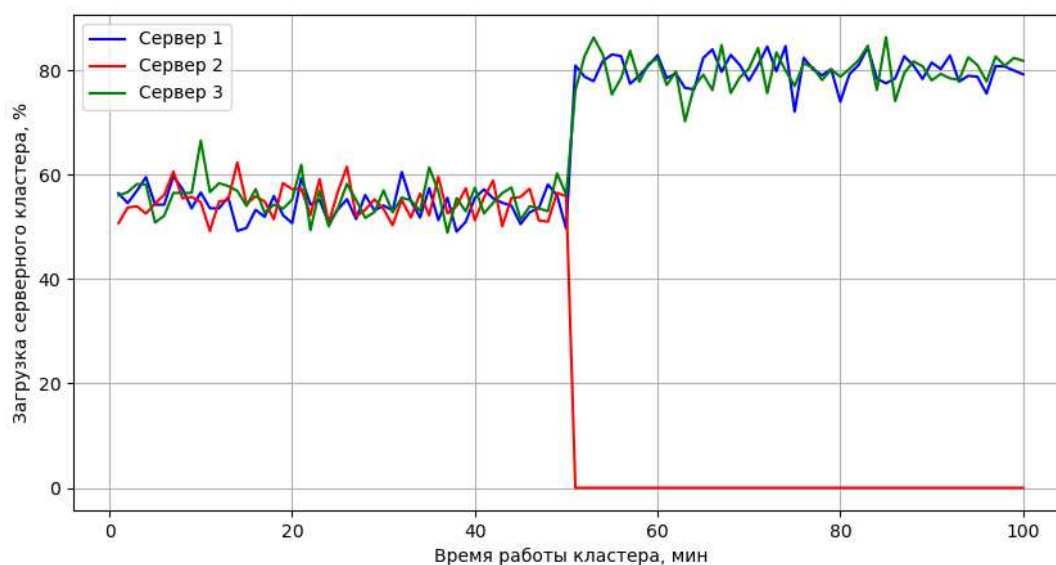


Рис. 2. Распределение нагрузки при выходе из строя одного из серверов

Заключение

Предложенный метод анализа состояния серверного кластера на основе мониторинга ключевых параметров и новых производных метрик позволяет более точно оценивать его отказоустойчивость и надежность. Разработанные показатели, такие как коэффициент эффективности распределения нагрузки (*LEC*), помогают выявлять узкие места в инфраструктуре и прогнозировать потенциальные сбои.

Анализ на основе временных рядов показал, что система работает стабильно, однако выявленный резкий пик указывает на возможные проблемы с балансировкой нагрузки или нехватку ресурсов в определенные моменты времени. Использование представленных методов и визуализации метрик на практике позволит оптимизировать работу серверного кластера, минимизировать время простоя и повысить общую производительность распределенных систем.

Список литературы

1. Гридин В.Н., Анисимов В.И., Васильев С.А. Методы повышения производительности современных веб-приложений // Известия ЮФУ. Технические науки – 2020. – № 2. – С. 193-197.
2. Викулов Е.О., Леонов Е.А., Денисова Л.А. Автоматизированное распределение больших объемов данных высоконагруженных // Динамика систем, механизмов и машин. – 2014. – № 3. – С. 146-149.
3. Денисов О.В., Викулов Е.О. Распределение данных в информационной системе с помощью сервера-балансира // Прикладная математика и фундаментальная информатика. – 2019. – Т. 6, №4. – С. 46-57.

Сведения об авторе:

Леонов Михаил Андреевич – аспирант.