Интеллектуальные модели построения экспертной системы, реализующие методы анализа диагностической информации о работе программно-определяемой СХД.

Дырночкин Александр

Введение

Системы хранения данных (СХД) корпоративного уровня в настоящее время являются сложными программно-аппаратными продуктами, которые могут включать в себя, кроме собственно носителей информации, одну или несколько управляющих ЭВМ, сетевую инфраструктуру, а также системное программное обеспечение, предоставляющее безопасный доступ к данным.

Обнаружение сбоев в процессе функционирования СХД является комплексной задачей, требующей анализа как программных, так и аппаратных компонентов системы, а также процесса их взаимодействия.

Актуальность задачи своевременного обнаружения сбоев в работе отдельных программных или аппаратных компонентов СХД и/или СХД в целом определяется тем, что её решение позволяет снизить или устранить вероятность деградации производительности СХД, временной потери доступа к пользовательским данным или потери данных.

Из этого следует, что одним из направлений повышения производительности СХД является автоматический мониторинг состояния СХД и своевременное реагирование на сбои и коллизии в ее работе.

Решение подобной задачи возможно посредством разработки моделей и методов интеллектуального мониторинга и анализа диагностической информации СХД в режиме реального времени с возможностью получения оценки в виде рекомендации в удобном для восприятия виде на основании экспертных знаний в рамках единой обучаемой программной платформы.

В настоящем обзоре предполагается проанализировать существующие подходы к обнаружению сбоев в области компьютерных систем, применяемые в существующих программных решениях для мониторинга и диагностики систем хранения данных, на основании чего выполнить обзор научных публикаций, предлагающих перспективные методы и средства, которые могут быть применены для эффективной реализации данных подходов.

Обзор научных публикаций

В статье [1] представлены исследования оценки автоматизированного анализа журналов для дальнейшего выпуска инструментов и тестов для легкого повторного использования. Во втором разделе статьи рассматриваются современные парсеры журналов и в третьем разделе сообщаются результаты сравнительного анализа существующих парсеров. Также рассматриваются методы синтаксических анализаторов, такие как: частые шаблоны,

кластеризация, эвристика и др. В общей сложности logparser содержит в себе 13 методов анализа журналов, среди них пять парсеров журналов (SLCT, LogCluster, LenMa, Drain, MoLFI).

В статье [2] выполнен обзор существующих программных средств, предназначенных для мониторинга состояния систем хранения данных, определены применяющиеся подходы к сбору, обработке и хранению данных. На основании проведенного анализа решаемых существующими программными средствами задач предложена типовая архитектура программного комплекса для обнаружения сбоев, описаны входящие в неё модули и характер их взаимодействия.

В статье [3] предметом исследования являются вопросы и особенности моделирования нагрузки на распределенную систему хранения данных. Исследование носит теоретический и аналитический характер и необходимо для дальнейшего экспериментального изучения. Проведен обзор литературы, анализ ситуации на рынке СХД. Определены особенности моделирования нагрузки на распределенную систему хранения данных. Исследование является начальным этапом разработки методов повышения производительности программно-определяемых систем хранения данных.

В статье [4] дано описание метода идентификации аномальных действий пользователей в корпоративных компьютерных систем на основе анализа лог-файлов. Предложенный метод основывается на кластеризации событий системного журнала алгоритмом IPLoM и построении матрицы счета событий для ее дальнейшего анализа с использованием методов машинного обучения.

В статье [5] анализируются алгоритмы автоматизации контроля состояний компьютерных систем средствами интеллектуального анализа неструктурированных данных системных журналов с целью обнаружения и диагностики аномальных состояний. На первом этапе осуществляется сбор логов с записями состояний системы и информации о выполнении процессов. На втором этапе используется парсер журнала для извлечения группы шаблонов событий, в результате чего необработанные журналы структурируются. На третьем этапе, после разбора журналов на отдельные паттерны, они дополнительно представляются в виде числовых векторов признаков (атрибутов). Совокупность всех векторов формирует матрицу признаков. На четвертом этапе матрица признаков используется для обнаружения аномалий методами машинного обучения для определения того, является ли новая входящая лог-последовательность аномальной или нет.

В статье [6] описана система, ключевой функциональностью которой является ее способность обрабатывать, собирать и анализировать большие объемы различных типов данных системных журналов. Данная система должна облегчить сбор журналов с разных узлов в сети. В статье объясняется предлагаемая система, которая собирает журналы с помощью Logstash, которая имеет возможность обрабатывать многие типы данных журналов, что помогает идентифицировать вредоносную активность в сети.

В статье [7] рассматриваются режимы работы и отказов, а также процесс обслуживания системы хранения данных с избыточностью данных. Предлагается модель обслуживания системы, в которой система подвергается процессу восстановления. Обслуживание системы с этими характеристиками не изучалось, и эти характеристики вносят новую модель надежности и обслуживания и новую проблему оптимизации. Такая проблема вызвана повышением доступности и эффективности системы хранения данных, использующей технологию резервирования данных. Приведено численное исследование, в котором параметры моделей задаются в соответствии с реальными данными и практическими соображениями.

В статье [8] предлагается методология анализа файлов журналов, основанная на разработке системы поддержки принятия решений. Целью этой методологии является сбор событий файла журнала с разных машин и сохранение необходимых данных в хранилище

данных для извлечения информации о безопасности, мониторинга, отчетности и получения оперативной информации и принятия более эффективных решений. Был рассмотрен конкретный пример использования файлов журналов веб-сервера.

В статье [9] исследуется полезность анализа файлов веб-журналов для улучшения использования цифровых репозиториев. Инструмент анализа журналов собирает информацию о посетителях из файлов журналов Интернета, обобщает их и дает более широкий обзор полезной статистики. Оценка предполагает, что анализ файлов журналов может дать пользователю обзор и характеристики репозитория, а также информацию о предыстории посетителей и, таким образом, улучшить использование цифрового репозитория.

В статье [10] представлен исследовательский подход к анализу файлов журналов организации. Цель анализа файла журнала - предоставить информацию о том, как клиент использует приложение (программное приложение), и обнаружить аномалии, которые не воспринимаются пользователем, чтобы эти аномалии можно было исправить до эскалации проблемы. Результат этого исследования направлен на определение архитектуры для поиска потоков программного обеспечения и применения методов сбора данных для измерения и получения знаний, которые будут записаны в систему.

Заключение

В ходе анализа актуальных тенденций в области диагностики систем хранения данных корпоративного уровня были рассмотрены наиболее распространенные программные системы управления и мониторинга СХД, анализа данных и программные средства общего назначения, а также рассмотрены научные материалы на тему методов анализа и обработки больших данных СХД. На текущий момент можно сделать вывод о том, что почти все рассмотренные актуальные программные средства предлагают некоторую функциональность для автоматического, автоматизированного или выполняемого администратором поиска неисправностей варьируемой сложности.

Список литературы

[1] Долгачев М. В., Москвичева К. С., Селезнева Е. А. Инструменты и тесты для автоматического анализа логов. // Тихоокеанский государственный университет. Хабаровск. Электронное научное издание «Ученые заметки ТОГУ». 2020. Т.11, No 1. С. 23 – 32.

[2] Успенский М. Б. Обзор подходов к обнаружению сбоев в системах хранения данных. // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2019. Т. 12. No 4. С. 145—158. DOI: 10.18721/JCSTCS.12412.

[3] Зарубин А. А., Савельева А. А., Швидский А. А. Моделирование нагрузки на распределенную систему хранения данных. 2018.

[4] Ефимова Ю. В., Гаврилов А. Г. Моделирование системы информационной безопасности на основе анализа системных журналов. // Инженерный вестник Дюна, No 6. 2019.

[5] Шелухин О. И., Рябинин В. С., Фармаковский М. А. Обнаружение аномальных состояний компьютерных систем средствами интеллектуального анализа данных системных журналов. // Вопросы кибербезопасности No 2 (26). 2018.

[6] Успенский М. Б. Обзор подходов к обнаружению сбоев в системах хранения данных. // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2019. Т. 12. No 4. С. 145—158. DOI: 10.18721/JCSTCS.12412.

[7] Zhu X., Wang J., Yuan T. (2019). Design and maintenance for the data storage system considering system rebuilding process.

[8] Azizi Yassine, Azizi, Mostafa, Elboukhari, Mohamed. (2021). Anomaly Detection from Log Files Using Multidimensional Analysis Model. 10.1007/978-3-030-73882-247.

[9] Mao Yue, Suleman Hussein, Williams Kyle, Paihama Gina. (2021). Analysing log files.

[10] C. Teixeira, J. B. de Vasconcelos and G. Pestana, "A knowledge management system for analysis of organisational log files,"2018 13th Iberian Conference on Information Systems and Technologies (CISTI), 2018, pp. 1-4, doi: 10.23919/CISTI.2018.8399229.