

МОДЕЛЬ ОРГАНИЗАЦИИ СИСТЕМЫ МОНИТОРИНГА НА БАЗЕ РАСПРЕДЕЛЕННОЙ СИСТЕМЫ ХРАНЕНИЯ БЛОЧНОГО ТИПА¹

А.В. Скатков, В.И. Шевченко, Е.Н. Машенко, Д.Ю. Воронин, В.С. Клепиков
(Севастополь)

Современные крупномасштабные системы мониторинга позволяют получать информацию о состоянии технических, экономических, природных [1] и многих других объектов, обеспечивают обработку и анализ получаемых данных, а также их хранение. Как правило, системы мониторинга представляют собой системы с радиальной, радиально-узловой или смешанной структурой. Такие системы в облачной вычислительной среде могут быть развернуты на базе технологий, применяемых при построении блочных систем хранения данных (СХД) [2]. Обобщенная структура СХД блочного типа для объектов мониторинга приведена на рис. 1.

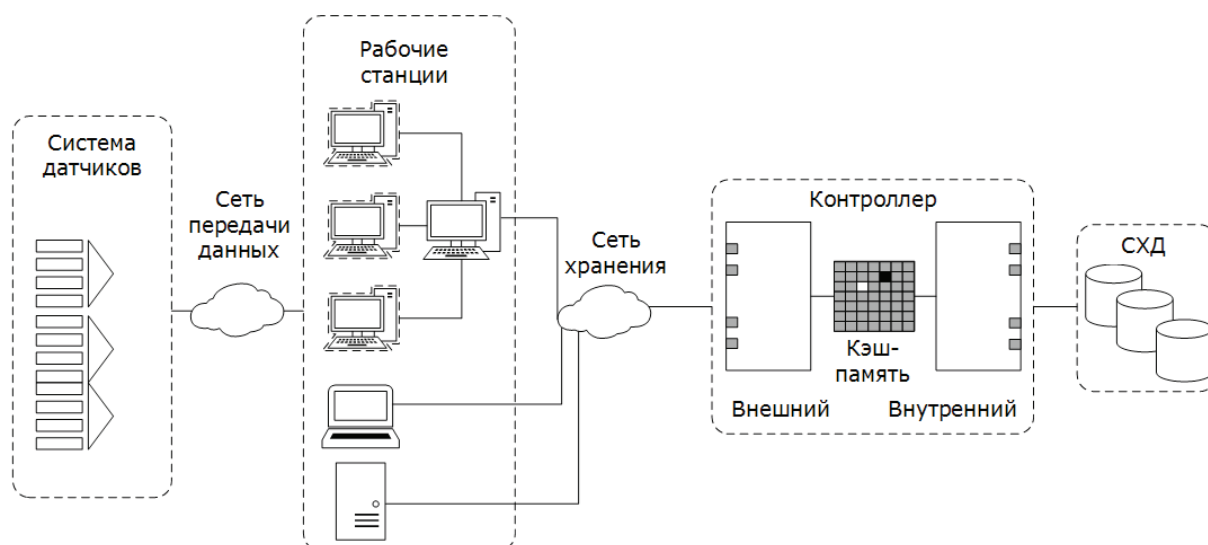


Рис. 1. Схема системы хранения данных для объектов мониторинга

Блочная СХД предоставляет вычислительным системам доступ на уровне блоков к томам хранения данных. В этой среде файловая система создается на базе клиентских вычислительных систем, а доступ к данным осуществляется через сеть на уровне блоков. Блочная система хранения состоит из одного или нескольких контроллеров и системы хранения. В свою очередь, контроллер СХД состоит из трех ключевых компонентов: клиентский интерфейс (внешние порты и контроллеры), кэш-память и серверная часть [3].

Ввиду сложности и масштабности подобных систем возникает целый ряд проблем, в частности, связанных с хранением и обработкой больших объемов данных, решением ресурсоемких задач по анализу получаемой информации, адаптацией системы под изменяющиеся условия окружающей среды и многие другие. Решением части этих задач может служить применение имитационных моделей для исследования проблем организации эффективной обработки данных.

Постановка задачи

Задана структура системы распределенного мониторинга объектов и ее параметры: предельное число мониторируемых узлов; средние значения времени между поступлениями запросов на обработку данных; средние значения времени обслуживания запросов; предельные значения времени обработки первичных данных от объектов мониторинга;

¹ Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 15-29-07936).

число уровней обработки данных; объем буферной памяти для хранения результатов предварительной обработки. Процесс сбора и обработки данных от мониторируемых объектов включает четыре основные стадии: 1) съем данных и передача данных от первичных систем мониторинга многомерных объектов (датчиков); 2) первичная обработка данных (фильтрация, классификация и кластеризация); 3) ввод пользовательских данных о процессе исследования; 4) архивирование полученной за рабочий период информации.

Задача заключается в построении имитационной модели процессов обработки данных и нахождении следующих характеристик: средней загрузки узлов обработки данных; среднего объема памяти, занимаемого данными; среднего числа пакетов данных, обработанных в системе в течение рабочего дня; среднего числа пакетов данных, получивших отказ из-за несоблюдения соглашения об уровне сервисов (SLA) между пользователями системы и провайдером ИТ-сервисов. Необходимо отразить в модели полный цикл обработки данных, включающий четыре вышеперечисленные стадии.

Концептуальная модель организации процессов обработки и хранения данных первичного мониторинга в распределенной сети хранения блочного типа

Процесс мониторинга многомерного объекта может быть представлен четырехфазной СМО с конечным числом состояний, где каждая из фаз соответствует отдельной стадии процесса сбора и обработки данных, описанных выше (рис. 2).

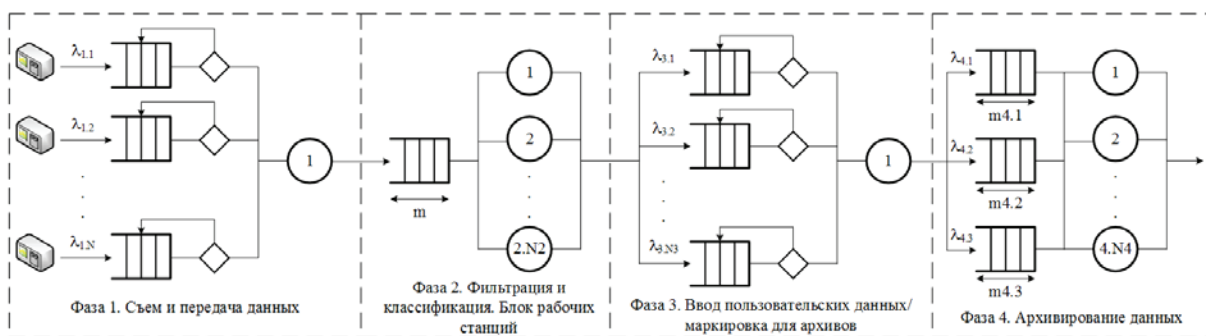


Рис. 2. Схема модели системы хранения данных для объектов мониторинга

На первой фазе обработки предполагаем, что число датчиков в системе мониторинга заранее определено и цикл обработки данных – последовательный, т. е информация со следующего датчика обрабатывается только после обработки данных предыдущего. В этом случае часть вычислительной инфраструктуры, отвечающая за текущий этап обработки данных, может быть представлена как одноканальное обрабатывающее устройство. Вторая фаза обработки соответствует этапу, когда переданные первичные данные мониторинга фильтруются и объединяются в кластер согласно классификационным признакам, указанным пользователями системы мониторинга (соответствует блоку рабочих станций на рис. 2). На этом этапе данные, принадлежащие одному кластеру, объединяются в пакеты и маркируются для дальнейшей обработки. Модель для второй фазы обработки – многоканальная СМО с ограниченной очередью. Третья фаза – данные последовательно обрабатываются пользователями и дополняются (модифицируются или остаются без изменений и получают статус архивирования по типам архива «рабочий массив»/«оперативный архив»/«долгосрочное хранение»). Используется один канал – сервер рабочих станций. Четвертая фаза – архивирование данных. Эта фаза может быть представлена многоканальной СМО с многомерным входящим потоком и ограниченной очередью. Объекты моделирования – сеть хранения и СХД. Многомерный входящий поток состоит из трех потоков (в рабочий массив – с большей интенсивностью; оперативный архив; долгосрочное хранение – самый тяжеловесный и редкий процесс).

Далее рассмотрим программную реализацию модели в среде имитационного моделирования AnyLogic на основе дискретно-событийного подхода.

Имитационная модель распределенной обработки и хранения данных

При разработке имитационной модели введены следующие предположения: на первой фазе обработки имитируется работа десяти источников данных первичного мониторинга (датчиков); интенсивность поступления запросов в систему и времена обработки запросов на всех четырех фазах распределены по экспоненциальному закону. Временные параметры модели могут задаваться пользователем перед началом эксперимента и меняться в соответствующем окне настроек в процессе моделирования.

При запуске разработанного программного продукта открывается окно презентации Simulation, в котором пользователь может задать параметры запуска модели. Они задаются в текстовых полях, доступны для редактирования и имеют интуитивно понятные подписи, говорящие о назначении каждого из них. При необходимости можно загрузить эти параметры из файла книги Excel. После настройки всех необходимых параметров будет доступно окно презентации Main. В этом окне презентации у пользователя доступно две области просмотра: режим отображения процессов моделирования (рис. 3) и режим отображения аналитики (рис. 4). В первом режиме на окне презентации отображены: потоковая диаграмма, включающая в себя все элементы модели, благодаря которой имеется возможность отследить движение всех запросов, сгенерированных источниками во время прогона; текстовые поля для изменения параметров модели; графические материалы, а именно круговые диаграммы уровня SLA и точечный график, отражающий динамику обработки данных по типам. Во втором режиме представлены две гистограммы, отображающие статистику по времени нахождения в системе запросов на обработку данных и элементы «Набор данных» и «Данные гистограммы», дающие возможность просмотреть точную информацию о заявках, прошедших потоковую диаграмму.

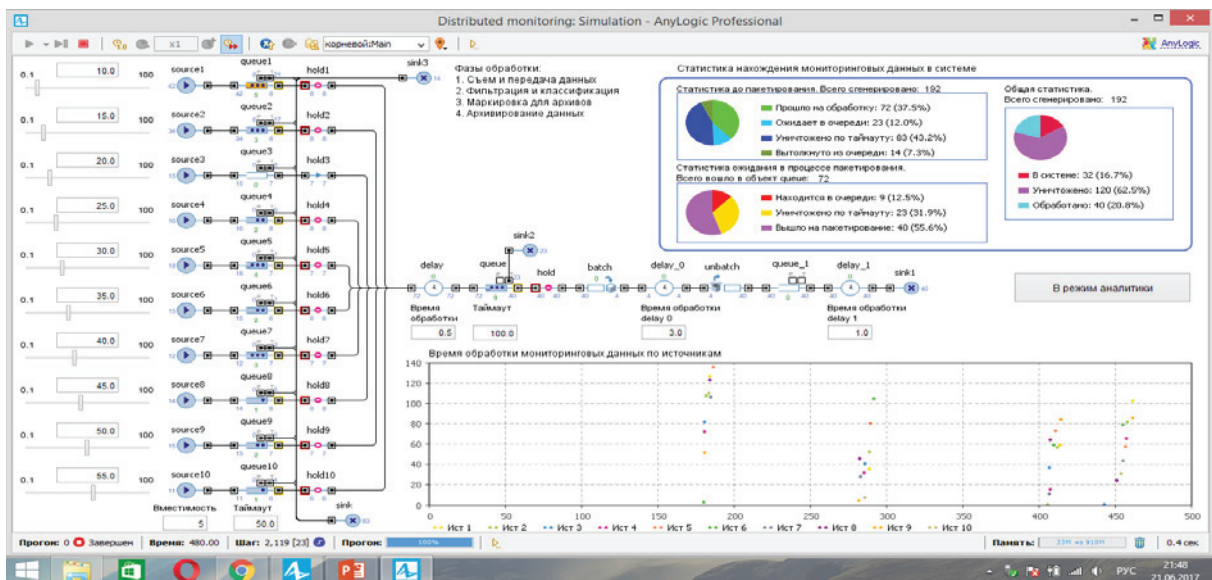


Рис. 3. Главное окно программы

Рассмотрим основные элементы главного окна программы более подробно. Объекты editbox представляют собой редактируемые текстовые поля, они отображают значения параметров модели, установленные в окне Simulation, также могут быть изменены пользователем в ходе моделирования. Текстовые поля, хранящие параметры для объектов source, ради удобства пользователя зависят от значения бегунков (объекты slider), им соответствующих, и меняют свое значение при их изменении. Каждое из полей в окне имеет

минимальное и максимальное значение, при попытке нарушения которых будет выбран ближайший к заданному значению минимум либо максимум.

Объекты `source1..10` – источники заявок, генерирующие их согласно времени между прибытиями, равному экспоненциальному вероятностному распределению `exponential(1.0/intens1..10)`. Здесь `intens1..10` задается при помощи текстовых полей либо слайдеров, находящихся слева от каждого из источников. Такая запись означает, что время генерации одной заявки в среднем равно параметру `intens1..10`, заданному в текстовом поле и бегунке. Является начальной точкой потока заявок в модели. При каждой генерации специально созданным параметрам заявки присваивается порядковый номер сгенерировавшего ее источника и модельное время, когда она была сгенерирована, в минутах.

Объекты `queue1..10` – очереди, содержащие заявки, сгенерированные каждым из источников до их дальнейшей обработки. Размер каждой очереди задается в соответствующем текстовом поле «Вместимость», а таймаут, для исключения заявок из очереди по истечении заданного времени, устанавливается в текстовом поле «Тайм-аут».

Объект `sink` – уничтожает поступившие заявки, которые были вытолкнуты из очереди из-за превышения времени ожидания, а также ведет подсчет количества уничтоженных заявок. Объект `sink3` – удаляет из модели заявки, которые вытолкнуты из очередей `queue1..10` за счет прибытия новых заявок в момент, когда очереди были уже заполнены.

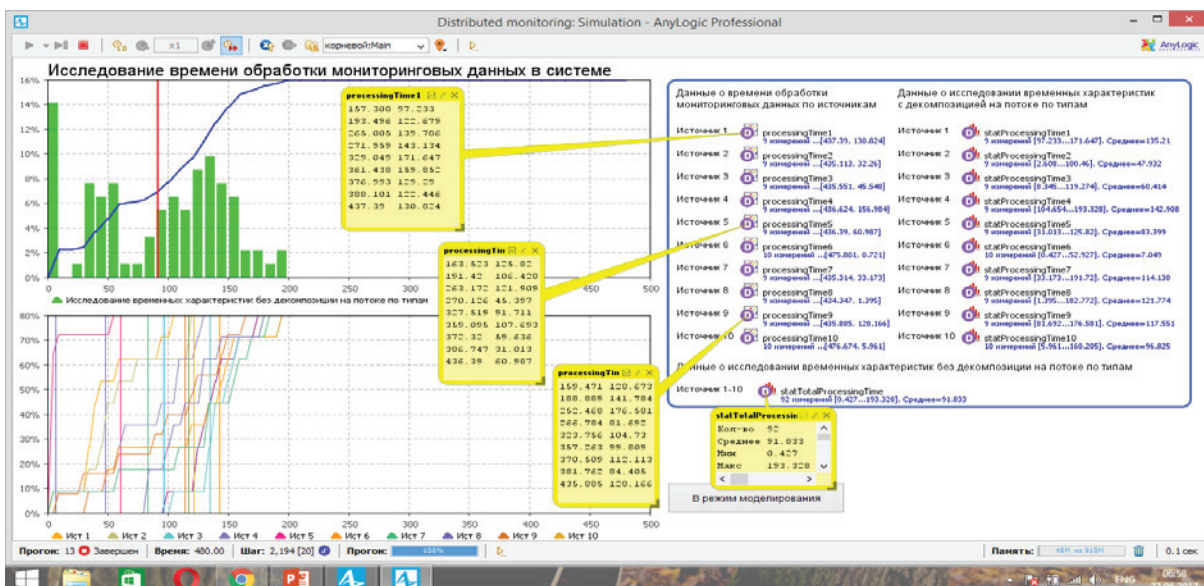


Рис. 4. Окно режима аналитической отчетности для моделируемой системы

Объекты `hold1..10` – блокируют выход заявок из соответствующих очередей, пока не сформирован пакет данных (финальная стадия фазы 1 обработки данных). Пакет данных должен содержать по одной заявке из каждого источника. Как только заявка проходит через блок `hold` и переходит на обработку, он блокируется, останавливая прохождение новых заявок от определенного источника. При формировании пакета из 10 заявок элементы `hold1..10` будут разблокированы.

Объект `delay` – имитирует процесс обработки заявок, поступающих через объект `hold`. Время задержки заявок зависит от номера источника, сгенерировавшего заявку, и распределяется по экспоненциальному закону. При этом за основу берется задержка, установленная в текстовом поле «Время обработки», находящемся под объектом `delay`, затем она увеличивается на эмпирический коэффициент, зависящий от номера источника заявок.

Система элементов queue, sink2, hold и hold1..10 обеспечивает пакетирование десяти заявок, каждая из которых сгенерирована отдельным источником, а также уничтожение заявок, которые слишком долго ожидают пакетирования, тем самым достигается сбор пакета актуальных заявок. Время, которое заявка является актуальной, согласно требованиям SLA, задано тайм-аутом объекта queue.

В режиме аналитической отчетности (рис. 4) отображены гистограммы с результатами исследований времени обработки мониторинговых данных в системе, а также подробные данные о времени пребывания заявок в системе и статистика по этим данным. На гистограмме «Исследование временных характеристик без декомпозиции на потоки по типам» отображены плотность вероятности и функция распределения на 48 интервалах, каждый из которых равен 10 минутам модельного времени, а также среднее значение времени обработки для запросов без разделения по типам на основе объекта «Данные гистограммы» statTotalProcessingTime. На гистограмме «Исследование временных характеристик с декомпозицией на потоки по типам» можно выделить мониторинговые данные отдельно по какому-либо из источников, кликнув на подпись источника в легенде, удалить данные или скопировать их в буфер обмена для переноса в другую программную среду.

Круговые диаграммы «Статистика нахождения мониторинговых данных в системе» дают четкое представление об эффективности работы системы с заданными параметрами. На диаграмме можно увидеть отношение обработанных мониторинговых данных к уничтоженным и к тем, которые на момент остановки модели остались в системе. На отдельной диаграмме отображена статистика до пакетирования, благодаря которой можно наглядно увидеть отношение прошедших на обработку заявок к тем, которые ожидают в очереди, уничтожены по тайм-ауту либо уничтожены из-за ограничения буфера хранения.

Заключение

Рассматриваемая имитационная модель носит прикладной характер и позволяет провести исследование полного цикла процессов обработки данных для систем мониторинга, работающих с разнородными пакетами данных, имеющих ограничение на время обработки и состав каждого пакета. Перспективами дальнейших исследований являются: подбор параметров для построенной модели на основе анализа статистических данных по функционированию распределенных систем обработки и хранения данных мониторинга крупномасштабных объектов, в том числе и в облачных инфраструктурах; исследование полученных характеристик качества обработки данных в рамках процессов поддержки принятия решений по управлению качеством ИТ-сервисов для распределенных систем мониторинга.

Литература

1. **Хмельницкая И.В.** Методологические подходы к проведению мониторинга крупномасштабных социально-экономических систем // Экономика и управление. 2013. № 1 (33). С. 100–103.
2. **Мащенко Е.Н., Шевченко В.И.** Исследование процессов организации эффективного хранения данных систем мониторинга на базе распределенной системы хранения блочного типа // Экологическая, промышленная и энергетическая безопасность – 2017: материалы научно-практической конференции, 15 сент. 2017 г. / под ред. Ю. А. Омельчук, Н. В. Ляминой, Г. В. Кучерик. Севастополь: СевГУ, 2017. С. 876–978.
3. От хранения данных к управлению информацией. 2-е изд. СПб.: Питер, 2016. 544 с.