NETCOOL OPERATIONS INSIGHT: глубокое погружение в диагностику

С.Коломиец, технический эксперт IBM в России и СНГ

В статье рассматриваются принципы использования аналитического ПО компании IBM для сбора и обработки сообщений о неисправностях приложений и систем. Описывается проблемная область, способ решения и сценарии работы пользователя с аварийными сообщениями в контексте их статистических характеристик.

Трудоемкость поддержки современных приложений и систем растет с каждым годом. Сами приложения становятся все более распределенными и сложными, а несущая их нагрузку инфраструктура, широко используя технологии виртуализации и гибридных облачных технологий, принимает трудные для диагностики формы. В среднестатистическом ЦОД наблюдается стремительный рост объемов сопровождающих служебных данных и информации мониторинга. До 97% данных этого типа неструктурированны, как, например, логфайлы и стектрейсы. Многие ИТ-проблемы не могут быть решены без углубленного анализа. Когда эксперт сталкивается с проблемой производительности, то первое, к чему он обращается, это метрики (измеренные параметры производительности), аварийные сообщения и логфайлы. Метрики и сообщения позволяют ответить на вопрос "что происходит", а логфайлы – на вопрос "почему". На первый взгляд, все просто, но это не так. Часто поиск нужной информации в логфайлах сравним с поиском иголки в стоге сена. Это ручная кропотливая работа, отнимающая много времени.

Что касается измеряемых величин, то для решения некоторых проблем требуется сравнение поведения метрик во времени. Как в живом организме, скорость работы одной подсистемы влияет на скорость работы другой, метрики нескольких объектов могут коррелировать на определенных временных интервалах. Контроль поведения позволяет обнаружить аномалии на ранних стадиях развития аварийной ситуации и дает шанс ее предотвратить. Вот простой пример. Время отклика системы онлайн-услуг зависит от количества пользователей в данный момент времени, и это нормально. Мы можем описать характер этой зависимости, наблюдая ее в течение некоторого времени, а затем настроить мониторинг времени отклика в контексте его отклонения от нормали в зависимости от числа пользователей. Конечно, на практике, как и с логфайлами, с метриками тоже не все так просто. Их может быть много, и требуется вручную определять, что с чем сравнивать, какие взять пороговые значения, по какому математическому закону.

Год назад IBM впервые применила технологию мощной аналитической платформы IBM BigData для решения задач, связанных с мониторингом и диагностикой в ИТ. Тогда свет увидели два решения: SCA-LA (SmartCloud Analytics – Log Analysis) и SCA-PI (SmartCloud Analytics – Predictive Insights). Название первого говорит само за себя – это анализ текстов (логфайлов, стектрейсов и т.п.), а второе занимается анализом метрик. На момент выхода оба решения были вполне самостоятельными и не были тесно интегрированы с ПО мониторинга и управления неисправностями IBM. При выборе стратегии

интеграции принимался в расчет характер обрабатываемых данных. Так, если ПО управления сообщениями о неисправностях (Netcool/OMNIbus, Netcool/Impact) имеет дело с текстовыми данными, то SCA-LA логично интегрировать именно с ним. А мониторинг ресурсов ITM (IBM Tivoli Monitoring) и мониторинг транзакций TCAM (Tivoli Composite Application Manager) в большей степени работают с метриками и пороговыми значениями. Соответственно, они претендуют на интеграцию с SCA-PI.

Главные задачи SCA-LA – вобрать (в английском варианте используется термин ingest – проглатывать, усваивать) и оптимальным способом проиндексировать неструктурированные массы текстовых данных с целью их последующего консолидированного представления. SCA-LA работает со своей внутренней технологией хранения и обработки данных в формате DSV (Delimiter Separated Value). Не будучи СУБД, эта схема использует комбинацию резидентного и нерезидентного размещения данных в памяти и в большей степени пригодна для высокопроизводительного индексирования гигабайт текстовой информации. Чтобы лучше понять, какие блага приносит SCA-LA в классическое решение управления неисправностями IBM, напомним, как это решение работает.

Аварийные и информационные сообщения от самых разных источников приводятся к единому формату и поступают в OMNIbus ObjectServer. Там они обрабатываются, выводятся на экраны персонала, а утратив актуальность, отправляются на архивное хранение в отдельно стоящую дисковую базу данных для последующей отчетности. По сути, и ObjectServer, и внешняя база архива представляют собой базы данных. Обе они не слишком пригодны для постоянного индексирования, а ObjectServer к тому же содержит только актуальные сообщения – далеко в прошлое не заглянешь. Специалистам, работающим с большими объемами аварийной информации, давно не хватает инструмента быстрой интерактивной выборки сообщений по различным критериям и анализа их расположения во времени. Известно, что поверх внешней архивной базы работает генератор отчетов Tivoli Common Reporting. В нем можно создать шаблон отчета и получать результат по запросу или по расписанию, но при больших объемах не получается приемлемая скорость и реальная интерактивность.

В 2014 году для радикального решения задачи IBM выпустил решение Netcool Operations Insight. Разработчики IBM пришли к следующему

варианту интеграции. Посредством XML-шлюза (eXtensible Markup Language), специального модуля OMNIbus, аварийные сообщения из ObjectServer переписываются с преобразованием формата в сервер SCA-LA, накапливаясь за значительные промежутки времени. Там они индексируются и после этого готовы к обработке внутренними приложениями SCA-LA. На сервер SCA-LA вместе с пакетом интеграции устанавливаются три специализированных для данных OMNIbus приложения. Контекстный вызов этих приложений SCA-LA из веб-интерфейса OMNIbus замыкает круг интеграции. Вывод результатов запросов осуществляется также через веб-интерфейс.

Какие же новые возможности появляются у специалиста в связи с появлением Operations Insight? Из списка аварийных сообщений по щелчку мышкой он может:

- показать сообщения от узла. Выбрав одно или несколько сообщений из списка в OMNIbus, можно вызвать интерфейс SCA-LA со всеми сообщениями от этих узлов за требуемый период времени (15 мин, 1 ч, день, неделя, месяц, год, произвольно выбираемый календарный интервал) от момента их прихода. Выводятся тексты самих сообщений;
- показать похожие сообщения. Подбираются не только сообщения от этого узла, а все похожие, то есть имеющие тот же тип (проблема/решение), группу (тематика) и критичность. Также выводятся сами сообщения;
- показать ключевые слова и количество сообщений по ним. Приложение берет все выбранные в списке OMNIbus сообщения и генерирует из них список ключевых слов. По начальной установке оно находит эти слова в полях сообщений Summary (краткое изложение), Node (имя узла) и AlertGroup (тип). Используя эти слова в конструкции фильтра запроса, оно выводит все сообщения, в которых упоминаются эти ключевые слова, за требуемый интервал. В интерфейсе SCA-LA выводится интерактивная сводка о том, сколько сообщений по каждому ключевому полю найдено;
- показать панель диаграмм по узлу. За указанный промежуток времени и для всех имен узлов в выбранных сообщениях, приложение генерирует панель из восьми диаграмм: тенденция изменения критичности; всплески количества по темам; всплески количества сообщений по узлам; горячая область по сочетанию "узел–тема"; распределение по критичности; пять самых "говорящих" тем; пять самых "говорящих" узлов; горячая область по сочетанию "тема–критичность";
- показать панель диаграмм по всем сообщениям OMNIbus за день. Приложение генерирует панель из восьми диаграмм, как в приведенном выше варианте, но берет все сообщения OMNIbus за последний день. В отличие от показа панели диаграмм по узлу, эта панель поддерживает дальнейший диалог. Двойной щелчок мышью на любой диаграмме формирует отчет с соответствующими ей аварийными сообщениями.

Можно не только быстро посмотреть, как развивалась ситуация, и сделать выводы, но и получить стратегическое решение с помощью RCA (Root Cause Analysis). Настройка и выверка автоматических механизмов анализа первопричины неразрывно связана с большим количеством разнообразных выборок сообщений из истории. Netcool Operations Insight может служить хорошим подспорьем в этой работе. Созданные с его помощью правила корреляции позволят уменьшить количество симптоматических сообщений в интерфейсе оператора и сфокусируют его внимание на наиболее вероятных причинах.

На оперативных выборках преимущества от интеграции со SCA-LA не заканчиваются, есть еще одно интересное применение. Допустим, есть подозрение, что некоторые ситуации повторяются с определенным периодом. Можно проверить закономерности появления сообщений в архиве внешней базы в определенные дни, часы и минуты. Эти закономерности могут служить бесценной подсказкой при определении их зависимости от других сообщений или от проводимых мероприятий или процессов в инфраструктуре. Практически без участия человека SCA-LA может в фоновом режиме проанализировать гигабайты сообщений на предмет возможной повторяемости и представить результат в виде панели диаграмм. SCA-LA также может индексировать и предоставлять выборки из документов технической поддержки и в случае необходимости выводить их на интерфейсы SCA-PI с графиками метрик.

В заключение можно отметить, что Netcool Operations Insight, пожалуй, единственное решение IBM в области управления неисправностями, реализующее высокоразвитый алгоритм аналитической обработки текста аварийных сообщений и обнаружения их периодичности. Вскоре предстоит тестирование Netcool Operations Insight на действующих OSS-системах (Operation Support System) лучших российских заказчиков IBM в телекоммуникационной отрасли. ■