



Решение для хранения данных RAIDIX

Кластерная СХД на базе ПО RAIDIX и файловой
системы GPFS

Оглавление

Резюме	2
Введение.....	3
Задачи и решение.....	4
Архитектура решения.....	7
Настройка	8
Технические показатели.....	9
Результаты для бизнеса	9
О компании «Рэйдикс»	9

Резюме

Программное обеспечение для СХД RAIDIX позволяет построить высокопроизводительный кластер хранения из множества узлов. Так, путем интеграции с распределенными файловыми системами HyperFS или Intel Lustre RAIDIX обеспечивает горизонтальную масштабируемость действующей инфраструктуры до нескольких экзабайт и 64 узлов.

В данном документе рассматриваются техническое взаимодействие ПО RAIDIX и кластерной файловой системы GPFS, преимущества интегрированного решения для поддержки ресурсоемких приложений в индустрии мультимедиа и высокопроизводительных вычислениях, особенности архитектуры, конфигурации, а также функциональные характеристики GPFS и RAIDIX.

Введение

RAIDIX — это высокопроизводительная и надежная система хранения высокой плотности, работающая с высокими нагрузками в видеонаблюдении, медиаиндустрии, корпоративном секторе и других отраслях. СХД на базе ПО RAIDIX демонстрируют рекордную производительность при обработке сотен параллельных потоков, гарантируют целостность больших объемов данных и бесперебойную работу системы. СХД RAIDIX поддерживает работу в режиме кластера Active-Active прямо «из коробки» — без использования внешних компонентов.

В связке с высокопроизводительной кластерной системой GPFS (General Parallel File System — общая параллельная файловая система), разработанной компанией IBM, ПО RAIDIX поможет системному администратору организовать кластер хранения из множества узлов на базе стандартного серверного оборудования.

На сегодняшний день GPFS используется на многих суперкомпьютерах из списка 500 самых мощных. GPFS отличается от других кластерных файловых систем возможностью одновременного высокоскоростного доступа к файлам для приложений, выполняющихся на нескольких узлах кластера под управлением различных операционных систем, в том числе и на базе ОС RAIDIX. Помимо возможностей хранения данных, GPFS предоставляет инструменты для управления и администрирования GPFS-кластера и позволяет осуществлять совместный доступ к файловым системам с удалённых GPFS-кластеров.

Далее мы рассмотрим задачи и сценарии, при которых рекомендуется внедрение масштабируемой СХД под управлением RAIDIX с использованием файловой системы GPFS, а также вопросы конфигурирования и построения комплексной архитектуры решения.

Задачи и решение

В условиях лавинообразного роста объемов данных многие киностудии, телеканалы и компании, осуществляющие видеонаблюдение на своих объектах в режиме 24/7, нуждаются в высокомасштабируемых решениях. Зачастую речь идет о петабайтах информации и необходимости создания кластера хранения из нескольких узлов. При выполнении этой задачи администратор системы сталкивается с несколькими ограничениями традиционных файловых систем:

- метаданные и данные хранятся на одних и тех же разделах;
- файлы «размазываются» по разделу, возникают задержки доступа;
- отсутствует механизм, предотвращающий дефрагментацию;
- недостаточная масштабируемость по размеру, производительности, количеству файлов, вложенности папок и т.д.;
- «неродная» кроссплатформенность.

Ключевыми критериями для горизонтально-масштабируемого решения, помимо высокой производительности и низких задержек, являются:

- единое адресное пространство для нескольких кластеров хранения;
- одновременный доступ по различным протоколам;
- файловый и блочный доступ к одним и тем же данным.

Всем вышеперечисленным требованиям удовлетворяет файловая система GPFS. Система демонстрирует высокую производительность, позволяя получить одновременный доступ к данным с нескольких рабочих машин. Большинство существующих файловых систем разработаны для среды, включающей в себя один сервер, и добавление дополнительных файловых серверов в этом случае не приводит к улучшению производительности. GPFS выдает большую производительность

ввода/вывода благодаря объединению в страйп блоков данных с конкретных файлов, хранящихся на нескольких дисках, и параллельному чтению и записи этих блоков. Среди других возможностей GPFS — высокая доступность, поддержка гетерогенных кластеров, защита от катастроф, обеспечение безопасности, DMAPi, HSM и ILM.

Файл, который записывается в файловую систему, разбивается на несколько блоков определенного размера, менее 1 МБ каждый. Блоки распределяются между несколькими узлами файловых систем, что способствует высокой скорости чтения и записи для каждого файла (вследствие высокой суммарной пропускной способности множества физических дисков). При этом файловая система не обеспечивает отказоустойчивости: даже отказа одного диска будет достаточно для потери данных. Чтобы сохранить целостность информации, для узлов файловой системы используются RAID-контроллеры: множественные копии каждого блока пишутся на физические диски на отдельных узлах.

Файловая система GPFS поддерживает следующие функции и возможности:

- Распределенные метаданные, в том числе дерево каталогов. Файловая система не имеет единого «контроллера каталогов» (directory controller) или «сервера индексного поиска» (index server).
- Эффективная индексация записей в каталоге для каталогов большого объема. Многие файловые системы ограничены определенным количеством файлов в одном каталоге (как правило, 65536). GPFS не имеет подобных ограничений.
- Распределенное управление блокировками. Полная поддержка семантики файловой системы Posix, включая блокировку для исключительного доступа к файлам.
- Гибкая работа с разделами (Partition Aware). Отказ сети может привести к разделению файловой системы на две или больше группы узлов, при этом группы смогут «видеть» только собственные узлы. Такую ситуацию возможно отследить с помощью протокола HeartBeat: в случае разделения файловая система остается функциональной для самого большого сформированного раздела. Таким образом, даже в режиме деградации файловой системы часть машин будет продолжать работать.

Обслуживание файловой системы может осуществляться в режиме онлайн. Большинство изменений в файловой системе (добавление новых дисков, сбалансированное распределение данных по дискам) может производиться в штатном режиме. Таким образом, файловая система обеспечивает более высокий уровень доступности в суперкомпьютерном кластере.

ПО RAIDIX может быть использовано как платформа для развертывания GPFS — в качестве отдельного узла хранения или узла хранения в сочетании с NSD-сервером. Связка RAIDIX и NSD является экономичным решением, так как позволяет сократить расходы на оборудование и использовать одни и те же аппаратные мощности для хранилища и серверов GPFS.

Архитектура решения

ПО RAIDIX возможно сконфигурировать для работы в режиме двухконтроллерного кластера (рис. 1).

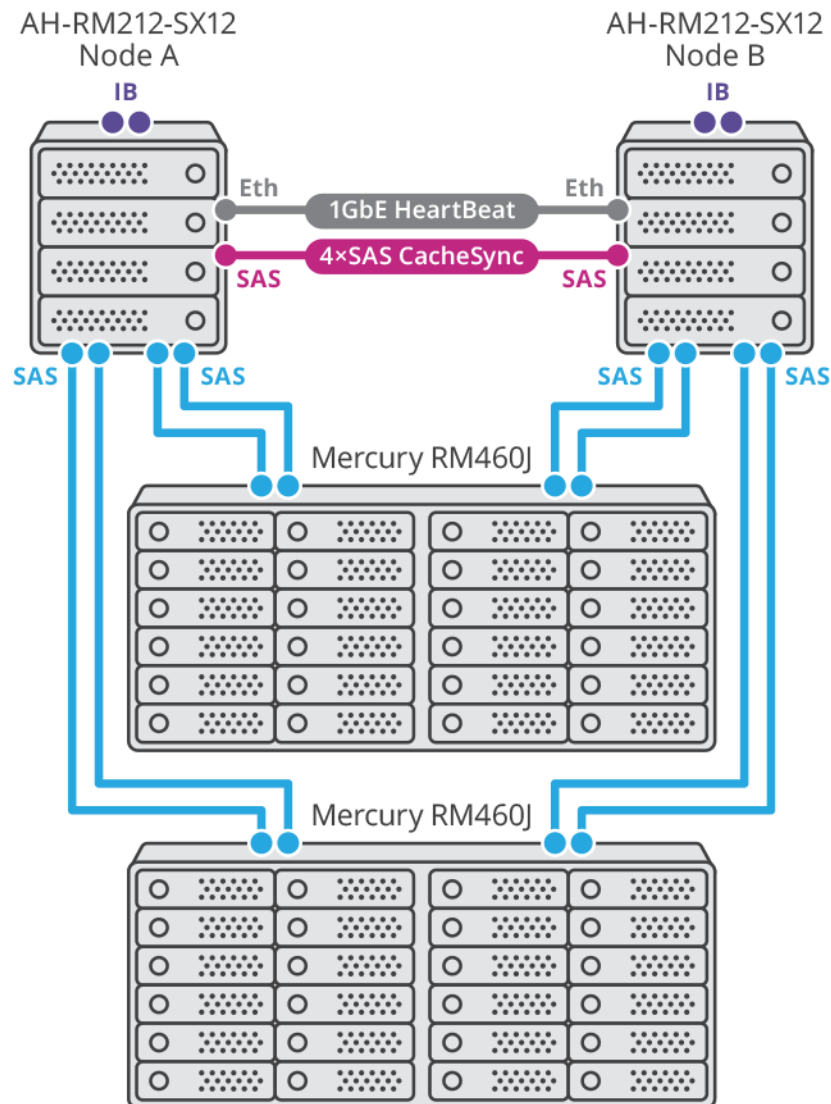


Рис. 1. Двухконтроллерная архитектура RAIDIX

При этом будут использованы два отдельных узла для двухточечных соединений:

- 1 GbE для контрольного сигнала (HeartBeat)
- 4 x SAS 12G для синхронизации кэша на запись (CacheSync).

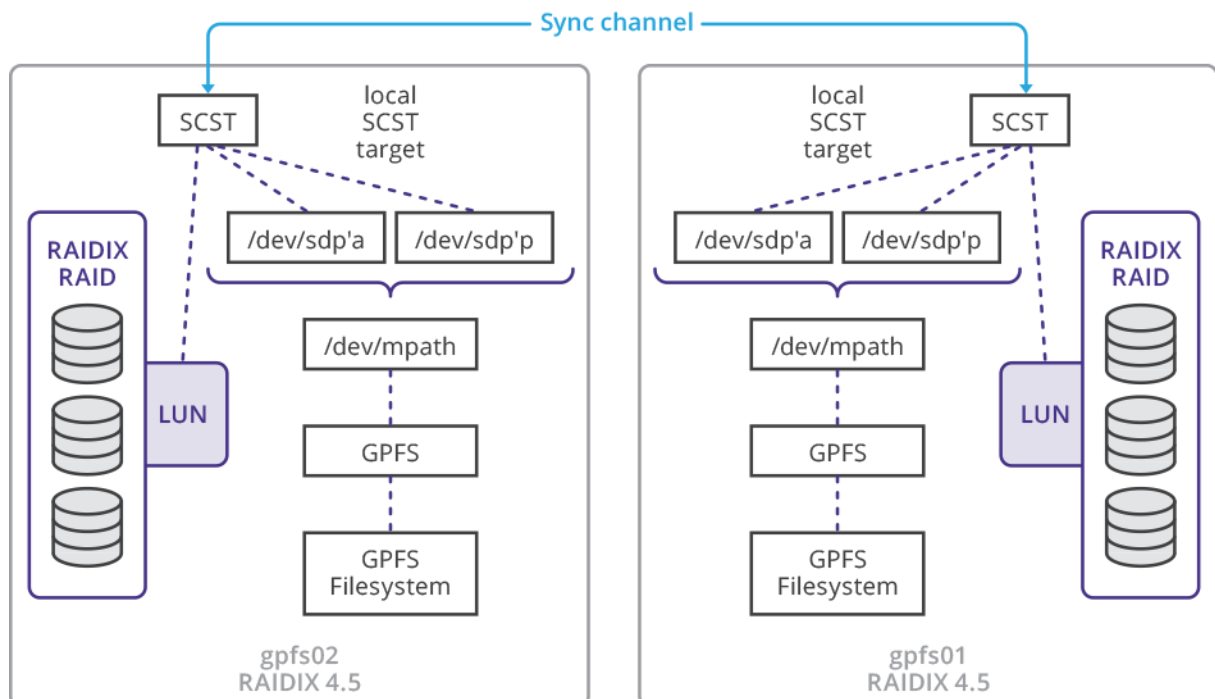
Оба узла соединяются с дисковыми полками JBODs через 4 порта SAS 12G для каждого узла. В качестве аппаратного интерфейса следует использовать Infiniband 100GbE.

При необходимости подключения большого количества клиентов — без установки на них клиентского программного обеспечения — возможно использование шлюзов файлового доступа (NAS Gateway), через которые осуществляется работа с данными.

Настройка

RAIDIX устанавливается на оба узла AH-RM212-SX12 и конфигурируется в соответствии с руководством администратора. Пакеты GPFS также должны быть установлены на оба узла, управляемые операционной системой RAIDIX. Пошаговую инструкцию можно найти в технической [документации GPFS](#).

Существует возможность перебросить тома с одного узла RAIDIX на другой через канал синхронизации и настроить GPFS, используя устройства mpath.



Технические показатели

По результатам проведенных тестов с IOzone и fio-test суммарная производительность на back end'e SAS может достигать 12-13 ГБ/с на последовательное чтение при записи блоками 1 МБ. Таких показателей можно достичь с помощью 120 жестких дисков NL-SAS. Эти жесткие диски распределяются по 12 массивам RAID6i (8D+2P).

Результаты для бизнеса

Интегрированное решение на базе RAIDIX и GPFS позволяет использовать множество узлов хранения (СХД), динамически распределяя информацию между ними и балансируя нагрузку. Архитектура решения позволяет добавлять к системе новые узлы хранения по требованию – без необходимости переносить данные и менять конфигурацию системы.

Технология RAIDIX в сочетании с файловой системой GPFS удовлетворяет высочайшим требованиям по скорости и отказоустойчивости, обеспечивает одновременную работу с данными с нескольких рабочих станций. Использование RAIDIX позволяет горизонтально масштабировать действующую инфраструктуру без простоев и снижения производительности и минимизировать расходы на апгрейд оборудования при создании кластеров хранения.

О компании «Рэйдикс»

Компания «Рэйдикс» (www.raidix.ru) (осн. в 2009 году) — ведущий поставщик систем хранения данных. Системы RAIDIX поставляются во многие страны мира. Используя собственную, запатентованную в России и США, технологию помехоустойчивого кодирования и обширную научную базу, компания предлагает отечественное решение для управления отдельными серверами СХД и построения масштабируемых высокопроизводительных кластеров из множества узлов хранения.