

**РЕФЕРАТ ПО ДИСЦИПЛИНЕ
«ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ, СИСТЕМЫ
И СЕТИ»**

**КЛАСТЕРНАЯ
АРХИТЕКТУРА:
КОНЦЕПЦИЯ,
ОСОБЕННОСТИ
ОРГАНИЗАЦИИ**

Логинов Андрей, А-13-08

Введение

- ◎ *Кластер* — группа компьютеров, объединённых высокоскоростными каналами связи и представляющая с точки зрения пользователя единый аппаратный ресурс.
- ◎ Один из первых архитекторов кластерной технологии Григорий Пфистер дал кластеру следующее определение: «**Кластер** — это разновидность параллельной или распределённой системы, которая:
 1. состоит из нескольких связанных между собой компьютеров;
 2. используется как единый, унифицированный компьютерный ресурс».

Обычно различают следующие основные виды кластеров:

- 1) отказоустойчивые кластеры
(High-availability clusters, HA, кластеры высокой доступности)
- 2) кластеры с балансировкой нагрузки
(Load balancing clusters)
- 3) вычислительные кластеры
(Computing clusters)
- 4) grid-системы

Они будут рассмотрены чуть позже.

- ◎ История создания кластеров неразрывно связана с ранними разработками в области компьютерных сетей. Одной из причин для появления скоростной связи между компьютерами стали надежды на объединение вычислительных ресурсов. Появилась операционная система Hydra («Гидра») для компьютеров PDP-11, на этой основе создали кластер (Питтсбург, шт. Пенсильвания, США, 1971).
- ◎ Тем не менее, только около 1983 г. были созданы механизмы, позволяющие с лёгкостью пользоваться распределением задач и файлов через сеть, по большей части это были разработки в Sun Microsystems.



PDP-11

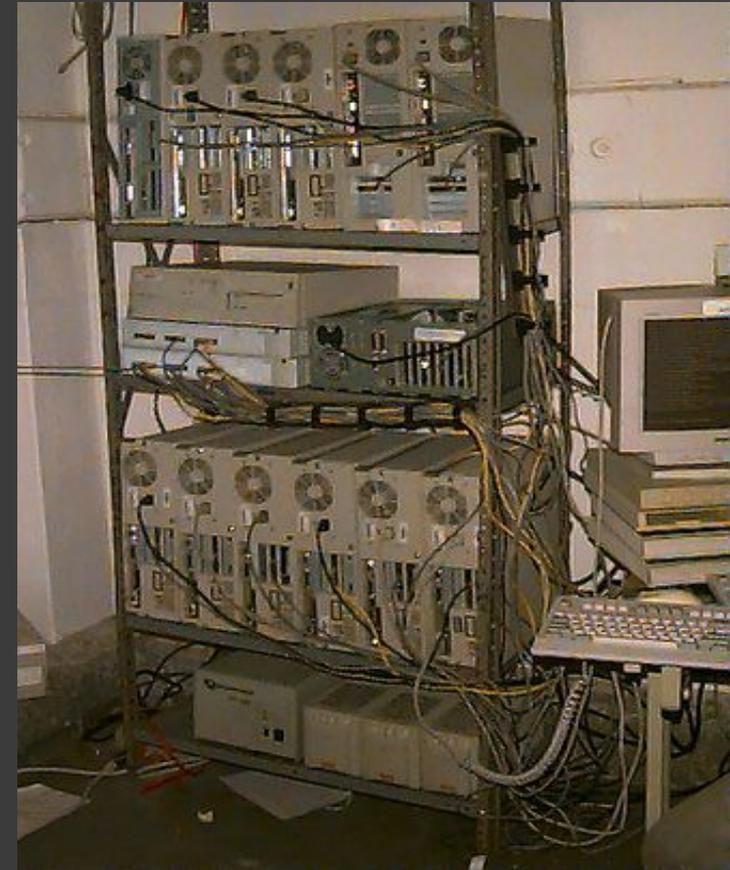
- История создания кластеров из обыкновенных персональных компьютеров во многом обязана проекту Parallel Virtual Machine. В 1989 г. это программное обеспечение для объединения компьютеров в виртуальный суперкомпьютер открыло возможность мгновенного создания кластеров. В результате суммарная производительность всех созданных тогда дешёвых кластеров обогнала по производительности сумму мощностей «серьёзных» коммерческих систем.
- Создание кластеров на основе дешёвых персональных компьютеров, объединённых сетью передачи данных, продолжилось в 1993 г. силами Американского аэрокосмического агентства (NASA), затем в 1995 г. получили развитие кластеры **Beowulf**, специально разработанные на основе этого принципа. Успехи таких систем подтолкнули развитие grid-сетей.

Beowulf

- ◎ ***Beowulf (Beowolf)*** — кластер, который состоит из широко распространённого аппаратного обеспечения, работающий под управлением операционной системы, распространяемой с исходными кодами
- ◎ Особенностью такого кластера является масштабируемость, то есть возможность увеличения количества узлов системы с пропорциональным увеличением производительности.

Преимущества Beowulf систем

- стоимость системы гораздо ниже стоимости суперкомпьютера;
- возможность увеличения производительности системы;
- возможность использования устаревших компьютеров, тем самым увеличивается срок эксплуатации компьютеров;
- широкая распространённость, а значит и доступность, аппаратного обеспечения.



Первый Beowulf кластер

- Проект *Beowulf* начался летом 1994 года сборкой в GSFC 16-процессорного кластера (на процессорах 486DX4/100MHz, 16MB памяти и 3 сетевых адаптера на каждом узле, 3 "параллельных" Ethernet-кабеля по 10Mbit)
- Далее в GSFC и других подразделениях NASA были собраны другие, более мощные кластеры. Например, кластер *theHIVE* (Highly-parallel Integrated Virtual Environment) содержит 64 узла по 2 процессора Pentium Pro/200MHz, 5 коммутаторов Fast Ethernet. Общая стоимость этого кластера составляет примерно \$210 тыс.



128-процессорный Beowulf кластер NASA



Современный Beowulf кластер

Классификация кластеров.

Кластеры высокой

ДОСТУПНОСТИ

- Создаются для обеспечения высокой доступности сервиса, предоставляемого кластером. Избыточное число узлов, входящих в кластер, гарантирует предоставление сервиса в случае отказа одного или нескольких серверов.
- Строятся по трем основным принципам:
 - 1) с **холодным резервом** или активный/пассивный. Активный узел выполняет запросы, а пассивный ждет его отказа и включается в работу, когда таковой произойдет.
 - 2) с **горячим резервом** или активный/активный. Все узлы выполняют запросы, в случае отказа одного нагрузка перераспределяется между оставшимися.
 - 3) с **модульной избыточностью**. Применяется только в случае, когда простой системы совершенно недопустим.

Кластеры распределения нагрузки

- Принцип их действия строится на распределении запросов через один или несколько входных узлов, которые перенаправляют их на обработку в остальные, вычислительные узлы. Первоначальная цель такого кластера — производительность, однако, в них часто используются также и методы, повышающие надёжность.
- Программное обеспечение может быть как коммерческим (OpenVMS, MOSIX, Platform LSF HPC, Solaris Cluster Moab Cluster Suite, Maui Cluster Scheduler), так и бесплатным (OpenMosix, Sun Grid Engine, Linux Virtual Server).

Вычислительные кластеры

- Для вычислительных кластеров существенными показателями являются высокая производительность процессора в операциях над числами с плавающей точкой (flops) и низкая латентность объединяющей сети
- Вычислительные кластеры позволяют уменьшить время расчетов, по сравнению с одиночным компьютером, разбивая задание на параллельно выполняющиеся ветки, которые обмениваются данными по связывающей сети. Одна из типичных конфигураций — набор компьютеров, собранных из общедоступных компонентов, с установленной на них операционной системой Linux, и связанных сетью Ethernet, Myrinet, InfiniBand или другими относительно недорогими сетями (это и есть Beowulf).
- Специально выделяют высокопроизводительные кластеры (Обозначаются англ. аббревиатурой HPC Cluster — High-performance computing cluster).

Системы распределённых вычислений (grid)

- Такие системы не принято считать кластерами, но их принципы в значительной степени сходны с кластерной технологией. Главное отличие — низкая доступность каждого узла, то есть невозможность гарантировать его работу в заданный момент времени (узлы подключаются и отключаются в процессе работы), поэтому задача должна быть разбита на ряд независимых друг от друга процессов. Такая система, в отличие от кластеров, не похожа на единый компьютер, а служит упрощённым средством распределения вычислений. Нестабильность конфигурации, в таком случае, компенсируется большим числом узлов.

Масштабируемая архитектура класса

- Кластеры должны создаваться таким образом, чтобы объем операций управления доступом к ресурсам и данным не изменился с увеличением объемов томов, иначе все эти увеличения приведут к росту кластеров. Поскольку при добавлении новых компьютеров в кластер управление доступом к дискам и данным усложняется, простейшим способом решения такой проблемы является организация кластеров, использующих расширяемые симметричные мультипроцессорные системы (SMP). Такой подход позволяет кластеру, независимо от его архитектуры, расширяться за счет простого добавления микропроцессоров в уже установленные SMP.
- Корректно организованные масштабируемые кластеры дают массу преимуществ. Технологические изменения, модернизация и даже обслуживание могут осуществляться без прекращения работы.

Самые производительные

- Дважды в год организацией TOP500 публикуется список пятисот самых производительных вычислительных систем в мире, среди которых в последнее время часто преобладают кластеры. Самым быстрым кластером является IBM Roadrunner
- **Roadrunner** — суперкомпьютер в Лос-Аламосской национальной лаборатории в Нью-Мексико, США. Разработан в расчёте на пиковую производительность в 1,042 петафлопса (на июнь 2011)



- Roadrunner работает под управлением Red Hat Enterprise Linux совместно с Fedora и управляется по xCAT. Он занимает приблизительно 12 000 кв. футов и весит 226 тонн. Энергопотребление — 3,9 МВт. Вступил в строй в июне 2008 года. Стоимость IBM Roadrunner составила 133 миллиона долларов.



IBM Roadrunner

- Министерство энергетики планирует использовать RoadRunner для расчёта старения ядерных материалов и анализа безопасности и надёжности ядерного арсенала США. Также планируется использование для научных, финансовых, транспортных и аэрокосмических расчётов.



IBM Roadrunner

Отечественные кластеры

- Кластер "**СКИФ К-500**", созданный в рамках совместной российско-белорусской программы "СКИФ" стал второй из созданных у нас компьютерных систем, вошедших в список TOP500
- Основой кластера стали 128 процессоров Intel Xeon с тактовой частотой 2.8 ГГц, которые объединены в 64 двухпроцессорных вычислительных узла. Кластер располагает 128 Гб оперативной памяти, а его пиковая производительность составляет 716.8 гигафлопс.



СКИФ К-500

- Созданный в рамках проекта кластер **СКИФ К-1000** является самой мощной из всех вычислительных систем, установленных на территории России, СНГ и Восточной Европы.
- «СКИФ К-1000», 288-узловой кластер на базе 576 64-разрядных процессоров AMD Opteron
- Пиковая производительность кластера составляет 2.5 терафлопа, реальная производительность на тесте linpack — 2.032 терафлопа.



СКИФ К-1000