

Центры обработки данных



Конструкция вычислительных комплексов



Введение

Дата-центр (data center) — специализированное здание (площадка) для размещения серверного и коммуникационного оборудования и подключения к каналам сети Интернет.

В русскоязычной терминологии дата-центр получил устойчивое название центр хранения и обработки данных (ЦХОД) либо центр обработки данных (ЦОД).

Виды ЦОД:

- Крупные (5, 4 по ТИА 942)
- Средние (3, 2)
- Малые (1)



Международный стандарт ТИА-942

Различные уровни надежности центров характеризуются коэффициентом постоянной готовности - должен быть не менее 99,671% для серверной, при условии, что:

- отсутствуют резервные компоненты;
- невозможно проводить профилактические работы без остановки действующего процесса обработки данных;
- суммарное годовое время простоя 28,8 часов.

Требования, предъявляемые к ЦОД:

- Дублирование каналов связи
- Высокая надежность системы электроснабжения
- Поддержание климатических условий на оптимальном для работы оборудования уровне
- Защита помещений и ресурсов дата-центра от несанкционированного доступа
- Система порошкового пожаротушения
- Фальшполы
- Пыле- и влагозащищенность помещений.

Основные стандарты



Рисунок 1 - телекоммуникационный шкаф

Размещение оборудования должно быть:

- компактным и доступным,
- обеспечивать требуемый температурный режим
- ограничивать несанкционированный доступ
- гарантировать аккуратную укладку шнуров и соединительных кабелей

Высота:

один "юнит" (U) = 44,45мм
или 1.75"

Ширина:

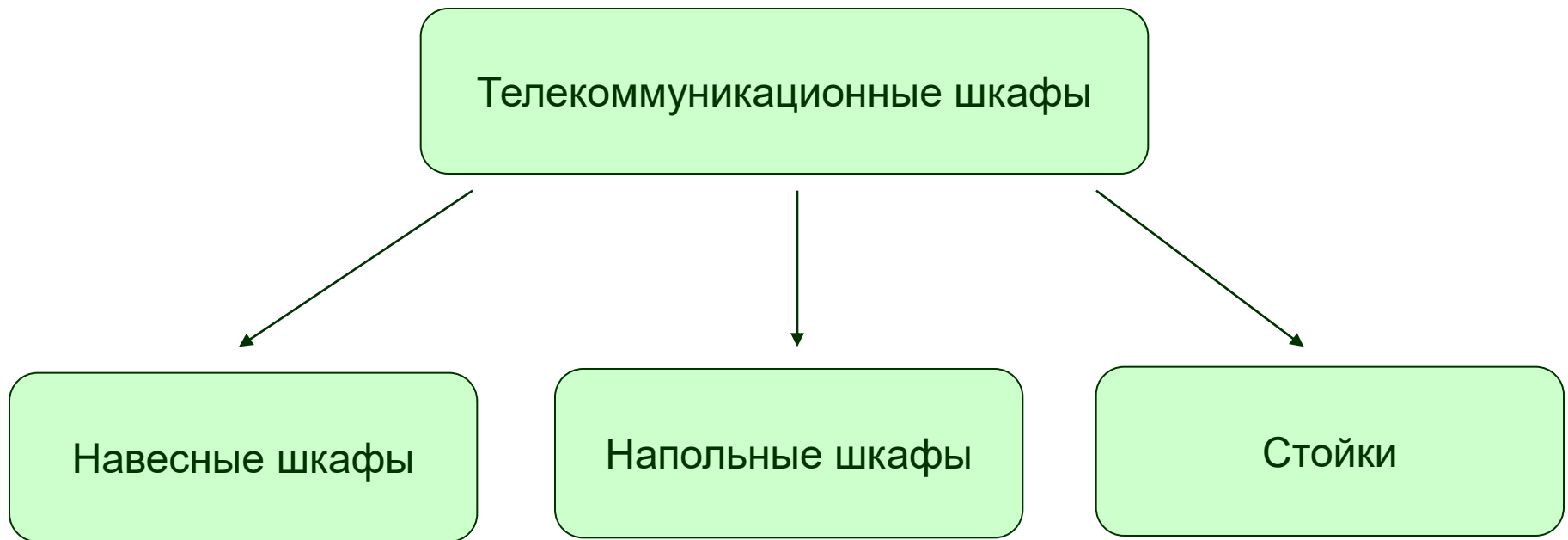
19 дюймов (19") = 482.6 мм

Телекоммуникационные шкафы

Монтажные конструктивы, предназначенные для
инсталляции различного телекоммуникационного
оборудования, называют телекоммуникационными
шкафами.

*** Шкафы и телекоммуникационные стойки еще называют
рэковыми шкафами или рэковыми стойками (англ. rack).

Виды телекоммуникационных шкафов



1 Настенные шкафы



Рисунок 3 – Внешний вид настенных телекоммуникационных шкафов

2 Напольные шкафы

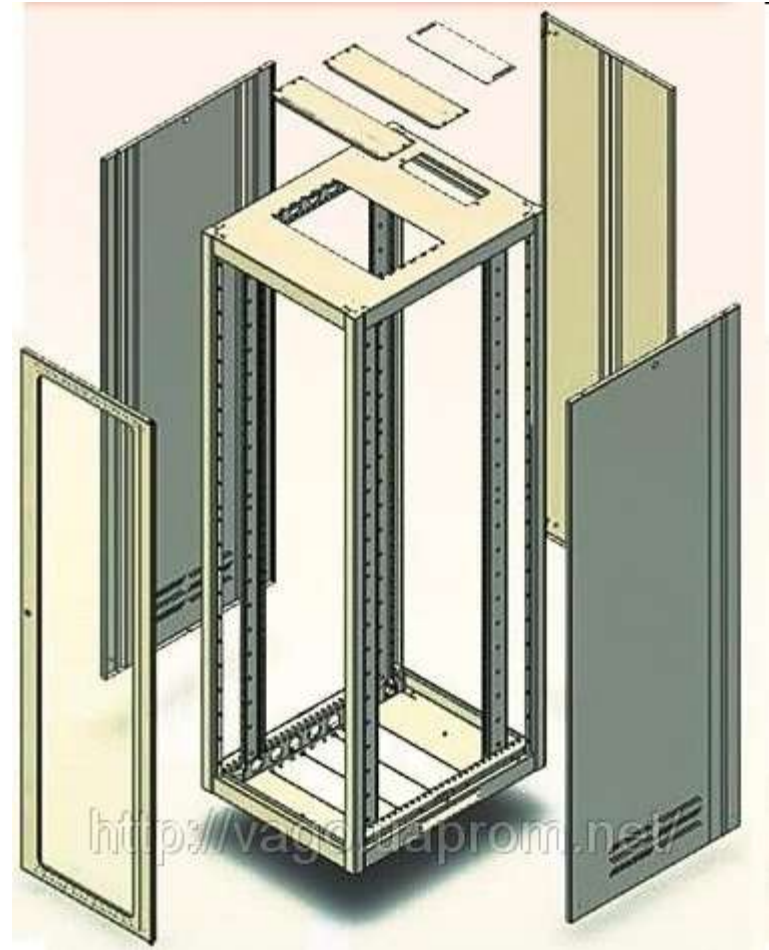


Рисунок 4 – Внешний вид напольных телекоммуникационных шкафов

3 Телекоммуникационные стойки



Рисунок 5 – Внешний вид телекоммуникационных стоек



Однорамная стойка



Двухрамная стойка

Рисунок 6 – Виды телекоммуникационных стоек

Монтируемое оборудование (серверы)

Виды:

1. стоечные (*rack*)
2. БЛЭЙД-серверы (*blade*)



Рисунок 7 – Серверы: rack(слева) и blade(справа)

Технологии коммутации узлов ВС

Fast Ethernet

Производители оборудования	Intel, CISCO, 3Com и др.
Показатели производительности	Пиковая пропускная способность - 100 Mbit/sec (12.5 MB/sec), полный дуплекс. В рамках MPI достигаются скорости порядка 6-7 MB/sec.
Примерные цены	Сетевой адаптер - \$15, 16-портовый коммутатор <u>Intel InBusiness Switch</u> - \$175
Программная поддержка	Драйверы для всех версий UNIX и <u>Windows NT</u> , протоколы TCP/IP, MPICH - бесплатная переносимая реализация MPI.
Комментарии	Преимуществами данной технологии являются: хорошая стандартизация и широкое распространение, а также низкие цены на оборудование и использование стандартных сетевых кабелей (UTP).

Реальные кластеры

- RHIC Computing Facility в Brookhaven National Laboratory, 2194 процессора Intel Pentium-III и Pentium-IV, Linux.

Gigabit Ethernet

Производители оборудования	<u>Intel</u> , 3COM и др.
Показатели производительности	Пиковая пропускная способность - 1 <u>Gbit/sec</u> (125 <u>MB/sec</u>), полный дуплекс. В рамках TCP/IP достигаются скорости порядка 500 <u>Mbit/sec</u> (60 <u>MB/sec</u>), в рамках MPI - до 45 <u>MB/sec</u>
Примерные цены	Сетевой адаптер <u>Intel EtherExpress PRO/1000 MT</u> - \$50-\$80. Коммутатор <u>D-Link DGS-1016T 16-port UTP 10/100/1000</u> - \$900.
Программная поддержка	Драйверы для многих версий <u>UNIX</u> и <u>Windows NT</u> , протоколы TCP/IP.
Комментарии	Преимуществом данной технологии является совместимость и возможность плавного перехода с технологий <u>Ethernet/Fast Ethernet</u> .

Реальные кластеры

- Avalon в LANL - 140 процессоров Alpha, Linux, используются Fast Ethernet и Gigabit Ethernet.
- Кластер в компании GX Technology Corporation - 3264 процессора Sun U-2, Pentium-III и Pentium-IV.

InfiniBand

Производители оборудования	<u>InfiniBand Trade Association</u>
Показатели производительности	Пиковая пропускная способность каналов 10, 20, 30, 40 Gb/sec, латентность - 1,2 <u>мксек.</u>
Примерные цены	N/A
Программная поддержка	OFED - бесплатный пакет драйверов, библиотек, <u>служебного ПО</u> и MVARICH - реализация MPI, оптимизированная для <u>InfiniBand.</u>
Комментарии	<u>InfiniBand</u> предлагает удалённый прямой доступ в память (<u>remote direct memory access - RDMA</u>), позволяющий доставлять данные непосредственно в память процесса, не вовлекая системные вызовы. Данные могут передаваться 1-о,4-х и 12-ти кратной скоростью. Анонсирован стандарт QDR 12х, позволяющий <u>передавать</u> данные со скоростью 120 GB/sec.

Реальные кластеры

- СКИФ-МГУ.

Myrinet 2000 / Myrinet-10G

Производители оборудования	<u>Myricom</u>
Показатели производительности Myrinet-2000 (в скобках - для 10G)	Пиковая пропускная способность - 2 Gbit/sec (10 Gbit/sec), полный дуплекс. В рамках TCP/IP достигаются скорости порядка 1.7-1.9 Gbit/sec (9.6 Gbit/sec). Латентность - порядка 2 мксек. На MPI-приложениях латентность составляет около 10 мксек, скорость передачи данных - до 200 MB/sec (до 400 MB/sec на дуплексных операциях).
Примерные цены	Адаптер 10G-PCIE-8A-S+MX1 для шины PCI-Express - \$815
Программная поддержка	Драйвера для <u>Linux (Alpha, x86, PowerPC, UltraSPARC)</u> , <u>Windows (x86)</u> , <u>Solaris (x86, UltraSPARC)</u> и <u>Tru64 UNIX</u> . GM - интерфейс программирования на нижнем уровне. Пакеты HPVM (включает MPI-FM, реализацию MPI для <u>Myrinet</u>), BIP-MPI и др.
Комментарии	<u>Myrinet</u> является открытым стандартом. <u>Myricom</u> предлагает широкий выбор сетевого оборудования по сравнительно невысоким ценам. На физическом уровне поддерживаются сетевые среды SAN (<u>System Area Network</u>), LAN (CL-2) и оптоволокно. Технология <u>Myrinet-10G</u> использует кабели стандарта 10Gbit-Ethernet и карты от <u>Myricom</u> могут работать как 10Gbit-Ethernet карты. Технология <u>Myrinet</u> дает высокие возможности масштабирования сети и в настоящее время очень широко используется при построении высокопроизводительных кластеров.

Реальные кластеры

- ПАРИТЕТ (ИВВиБД) - 4 узла по 2 процессора Pentium II, RedHat Linux.
- HPC в University of Southern California - 756 узлов по 2 процессора Pentium-III и Pentium-IV, Linux RedHat 7.2.
- Cplant в Sandia - почти 1400 рабочих станций и серверов Compaq на базе Alpha, Linux.
- SuperMike 512 узлов с 2-мя процессорами Pentium Xeon 1.8 GHz, Linux RedHat 7.2.

SCI (Scalable Coherent Interface)

Производители оборудования	Dolphin Interconnect Solutions и др.
Показатели производительности	Для продуктов <u>Dolphin</u> : пиковая пропускная способность - 10 GB/sec, полный дуплекс, в рамках MPI достигается около 700 MB/sec. Аппаратная латентность - 0.2 мксек, в рамках MPI - порядка 1,4 мксек.
Примерные цены	На начало 2008г. фиксированных цен не анонсировано.
Программная поддержка	Драйверы для Linux, Windows NT, Solaris. <u>ScaMPI</u> - реализация MPI компании <u>Scali Computer</u> для систем на базе SCI. <u>SISCI API</u> - интерфейс программирования нижнего уровня.
Комментарии	SCI - стандартизированная технология (ANSI/IEEE 1596-1992). Кроме стандартной сетевой среды, SCI поддерживает построение систем с разделяемой памятью и когерентностью кэшей. На коммуникационной технологии SCI основаны кластерные системы компании <u>SCALI Computer</u> , системы семейства <u>hpcLine</u> компании <u>Siemens</u> , а также cc-NUMA-сервера <u>Data General</u> и <u>Sequent</u> . Технология SCI использовалась для связи гиперузлов в системах <u>HP/Convex Exemplar X-class</u> . Иллюстрации

Реальные кластеры

- Monolith cluster on NSC - 200 узлов по 2 процессора Intel Xeon 2.2 GHz.

Энергоснабжение ЦОД



Рисунок 8 – Источники бесперебойного питания

Охлаждение ЦОД

Этапы проектирования охлаждения ЦОД:

- Техническое задание
- Разработка проектного решения
- Выбор концепции кондиционирования
- Трёхмерное математическое моделирование

Уровни охлаждения ЦОД:

1. На уровне зала
2. На уровне рядов
3. На уровне отдельной стойки

1 Охлаждение на уровне зала

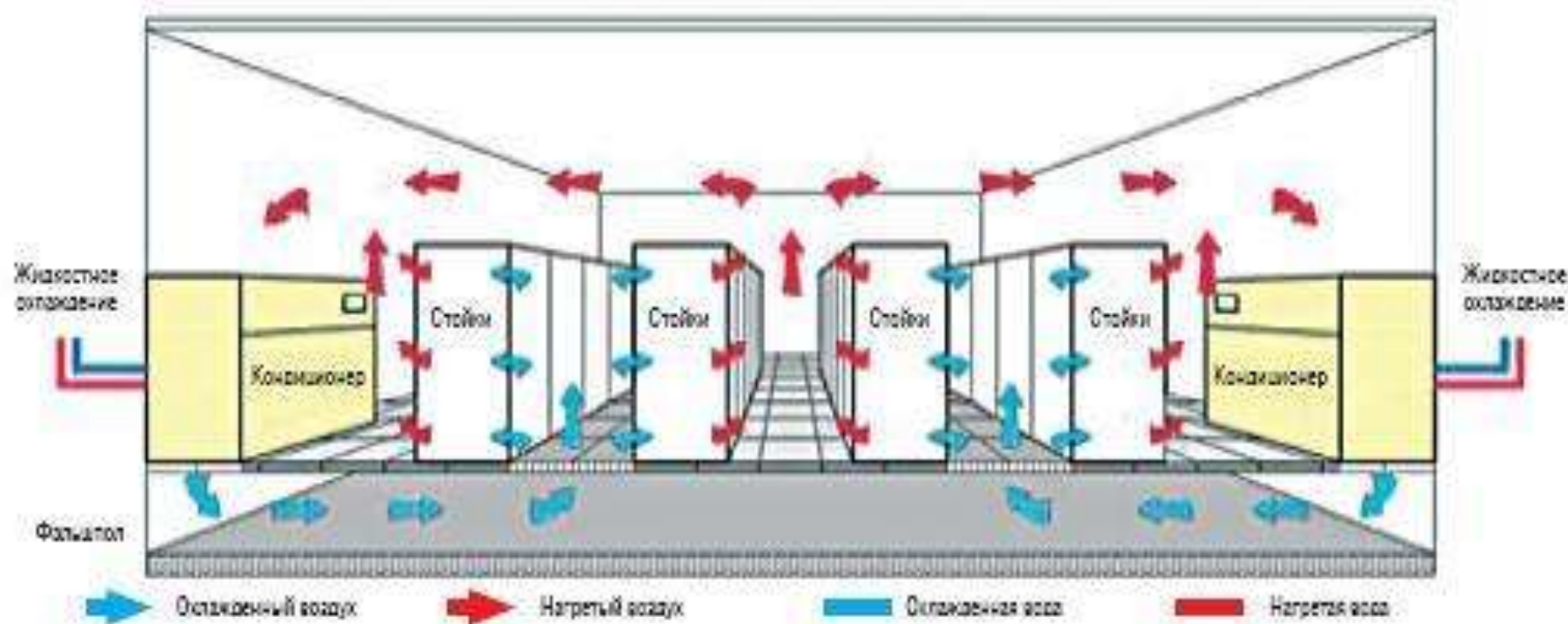


Рис. 1. Схема охлаждения на уровне зала с использованием фальшпола

Рисунок 9 – Охлаждение на уровне зала с использованием фальшпола

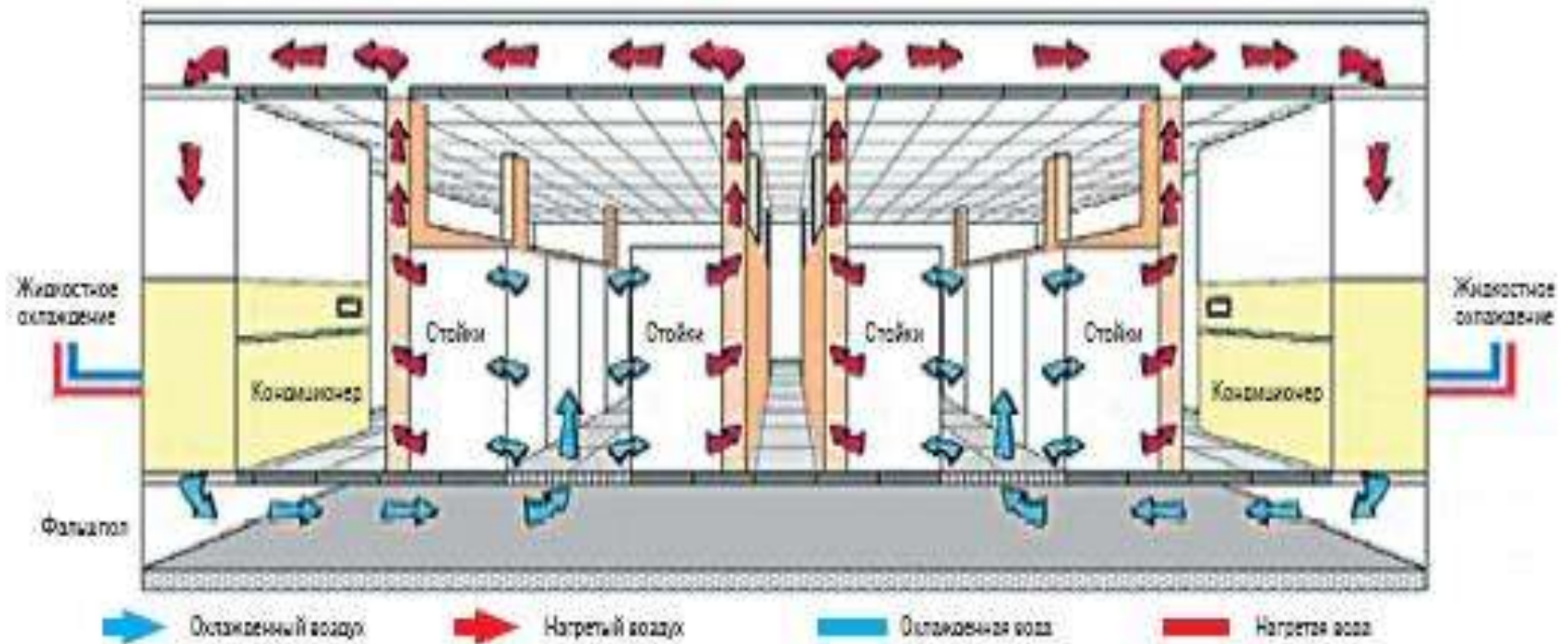


Рис. 2. Схема охлаждения на уровне зала с использованием фальшпола и системы вытяжной вентиляции

Рисунок 10 – Охлаждение на уровне зала с использованием фальшпола и вытяжной вентиляции

2 Охлаждение на уровне рядов

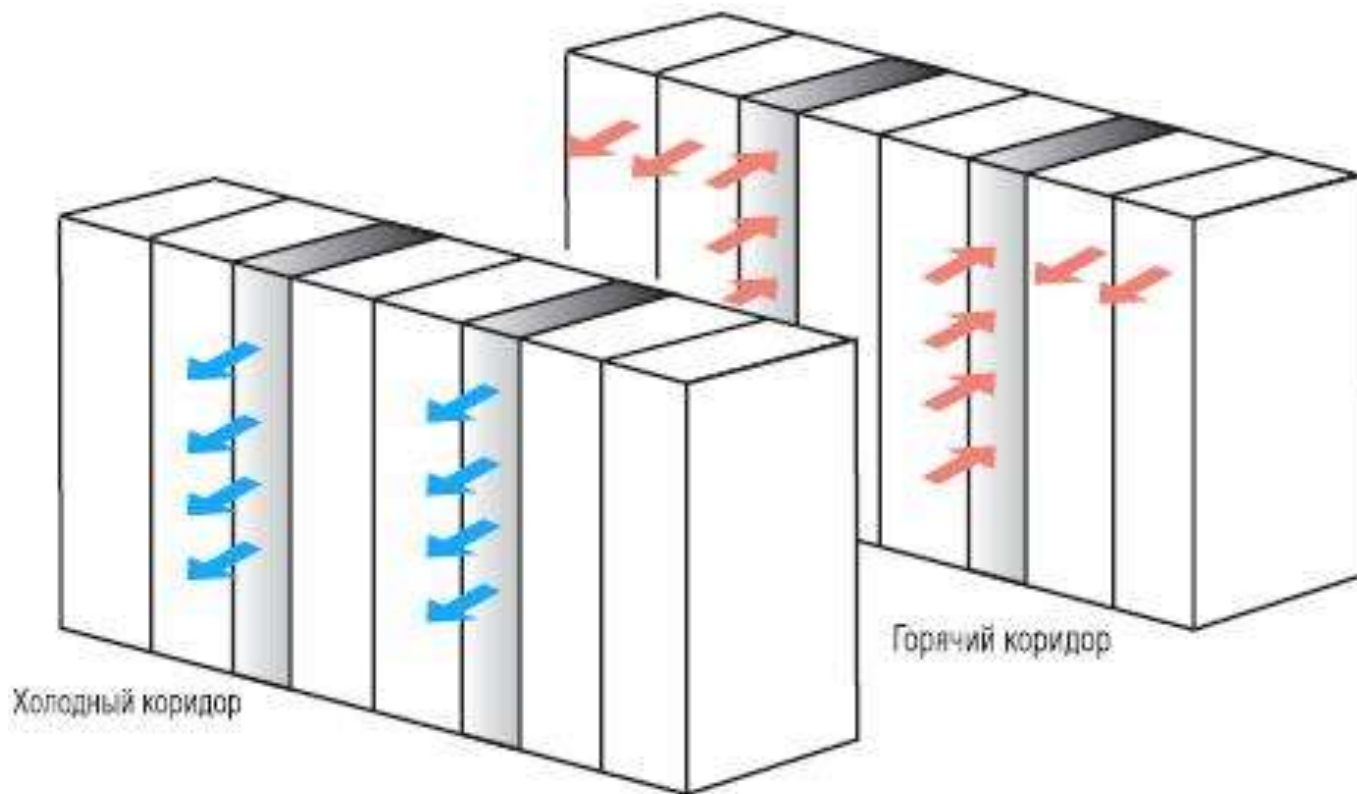


Рис. 3. Схема охлаждения ЦОД на уровне ряда

Рисунок 11 – Охлаждение на уровне рядов

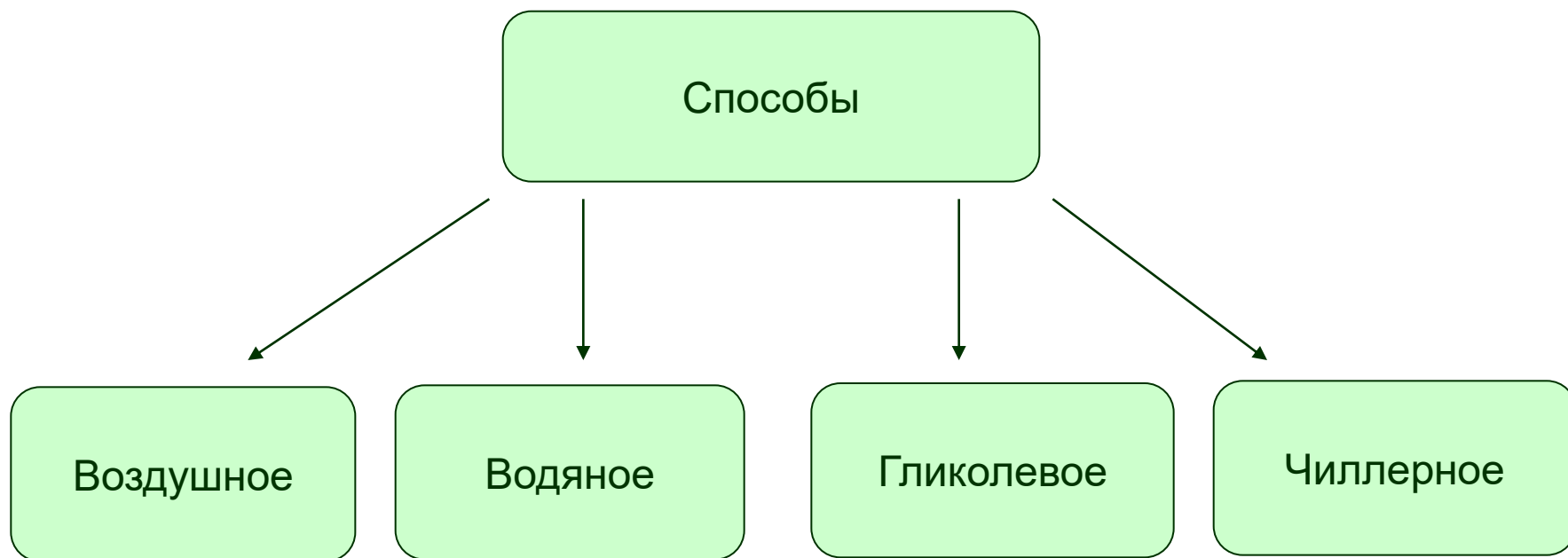
3 Охлаждение на уровне стоек



Рис. 4. Иллюстрация работы замкнутой системы кондиционирования на базе решения APC RACS

Рисунок 12 – Схема охлаждения на уровне стойки(на базе APC RACS)

Способы охлаждения



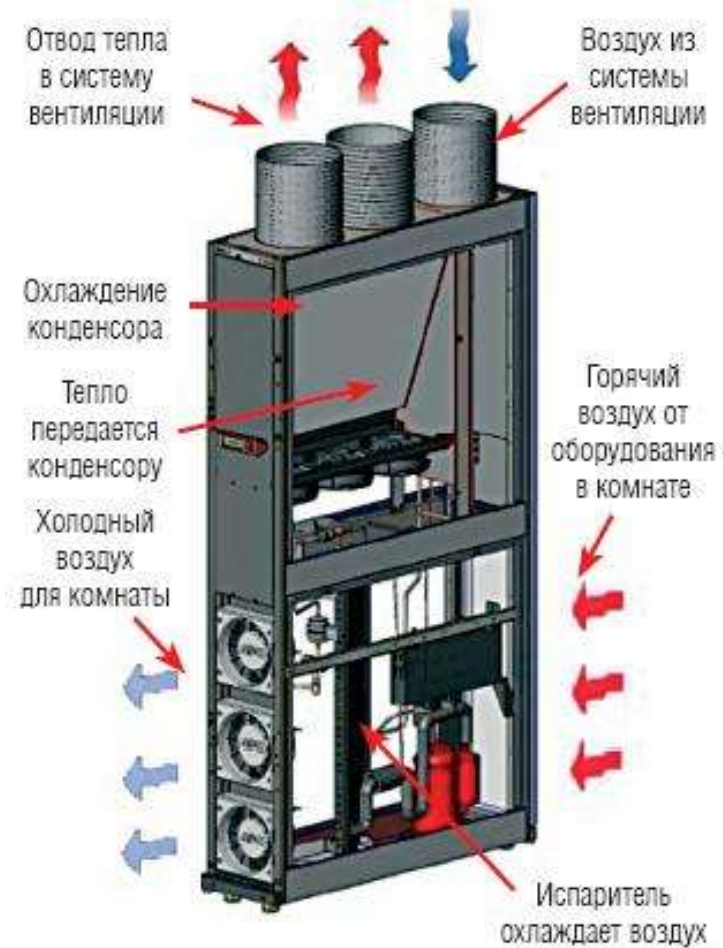


Рис. 5. Схема работы кондиционера APC InRow SC

Рисунок 13 – Схема работы воздушного кондиционера (APC InRow SC)



Рис. 8. Устройство для распределения чиллерной воды — CDU

Рисунок 14 – Устройство для распределения чиллерной воды

Экономия на охлаждении

Рекомендации по экономичному использованию систем охлаждения:

- Заделка отверстий в фальшполе
- Установка бланкирующих панелей
- Координация работы систем кондиционирования
- Улучшение движения воздуха под фальшполом
- «Горячие» и «холодные» проходы
- Установка датчиков температуры
- Герметизация «горячих» и «холодных» проходов

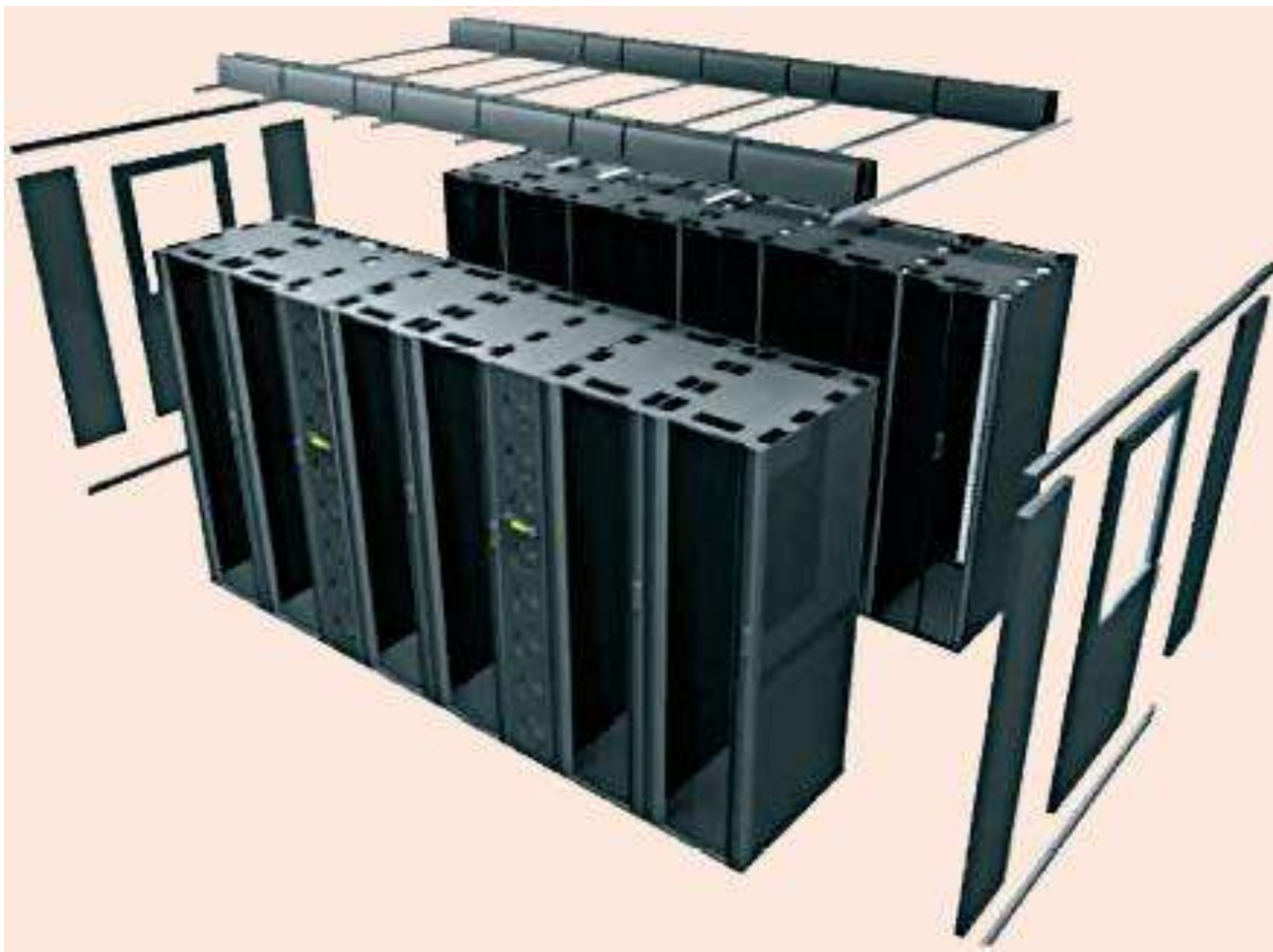


Рисунок 15 – Иллюстрация герметизации проходов

*****Это интересно*****

Технология Cryogel: охлаждение дата-центра с помощью
ледяных шариков



Диаметр шарика = 103 мм



Ёмкость с гликолем