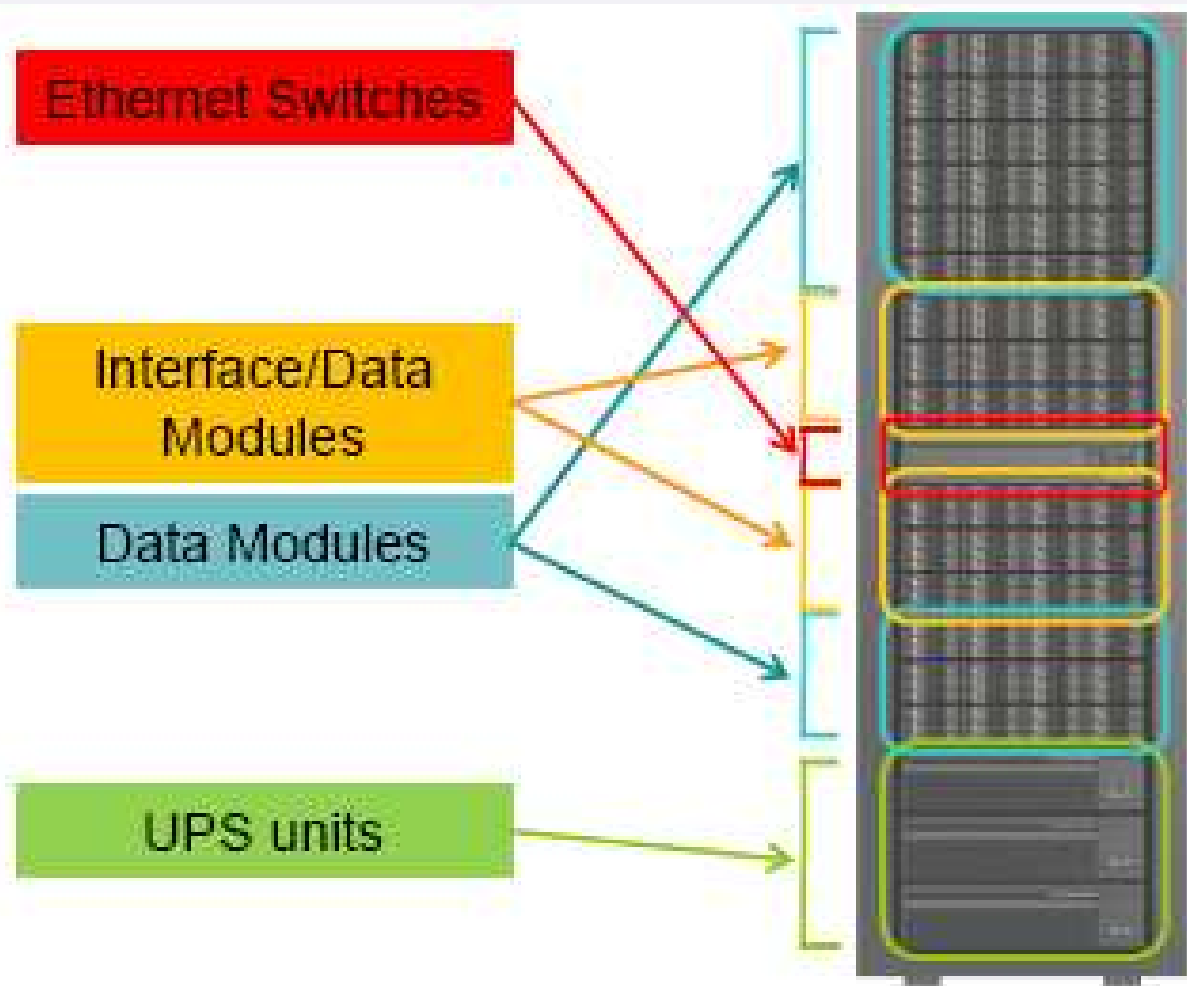


XIV: Архитектура и его ВОЗМОЖНОСТИ

Патрушев Александр
Product manager IBM
System p, Storage, System X.
e-mail: apatrushev@landata.ru

Обзор Архитектуры





Возможности и функционал

XIV характеризуется следующими характеристиками:

- iSCSI и FC интерфейсы
- Множественный доступ
- Управляющее ПО
 - GUI
 - CLI
- До 16000 snapshots
- Репликация на удаленную систему
- Простое управление емкостью
- Уведомление по e-mail, SNMP, SMS
- Нет единой точки отказа
- Отказоустойчивость, анализ сбоев, самовосстановление
- Обновление и техническое обслуживание без влияния на работу



Ключевые моменты архитектуры

- Глобальный параллелизм
 - Любая I/O активность задействует все компоненты системы.
- Балансировка нагрузки
 - Все диски и модули загружены равномерно, вне зависимости от характера нагрузки
- Самовосстановление
 - В случае выхода из строя любого компонента система автоматически перераспределит нагрузку и вернется на прежнюю производительность.
- Настоящая виртуализация
 - Физическая емкость и ее использование полностью скрыто от пользователей
- Thin Provisioning
 - возможность выделить требуемое место в нужный момент

Емкости и ресурсы



- 15 дисковых модулей
- - 9 модулей данных (1-3 и 10-15)
- - 6 модулей данных и интерфейсов
 - 24 порта 4Гб/с FC
 - 6 портов 1Гб/с iSCSI
- 180 SATA дисков (1/2 ТБ, 7200 RPM)
 - 12 дисков на модуль
 - 79/161 ТБ полезной емкости
- 2 внутренних Ethernet коммутатора
- 3 UPS
 - достаточно для 15 минут работы
- 15/21 процессора QC
- 120 GB памяти
- 250 Гб/с шина от диска до кэша
- Модуль технического обслуживания с модемом
- Автоматический коммутатор внешнего питания

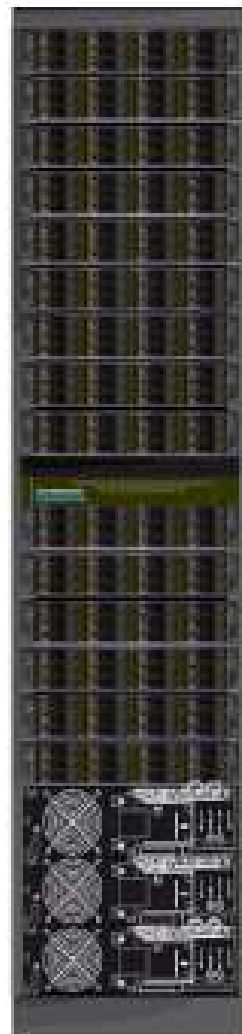


Поддерживаемые ОС

| Host Operating System | Minimum Supported Level |
|------------------------------|--------------------------------------|
| AIX | 5.3 TL7 |
| ESX | 3.0 |
| HP-UX | 11iv1 |
| Linux - CentOS | Enterprise Linux 4.6 |
| Linux - RHEL | Enterprise Linux 4.6 |
| Linux - SuSE | SLES 10 |
| Macintosh | OS X 10.4.10 |
| Power VM (Virtual IO Server) | VIOS 2.1.1 |
| Solaris | 9 |
| SVC | 4.3.0.1 |
| Windows | 2003 SP1 (Windows 2000 SP4 with RPO) |



Распределение компонент системы

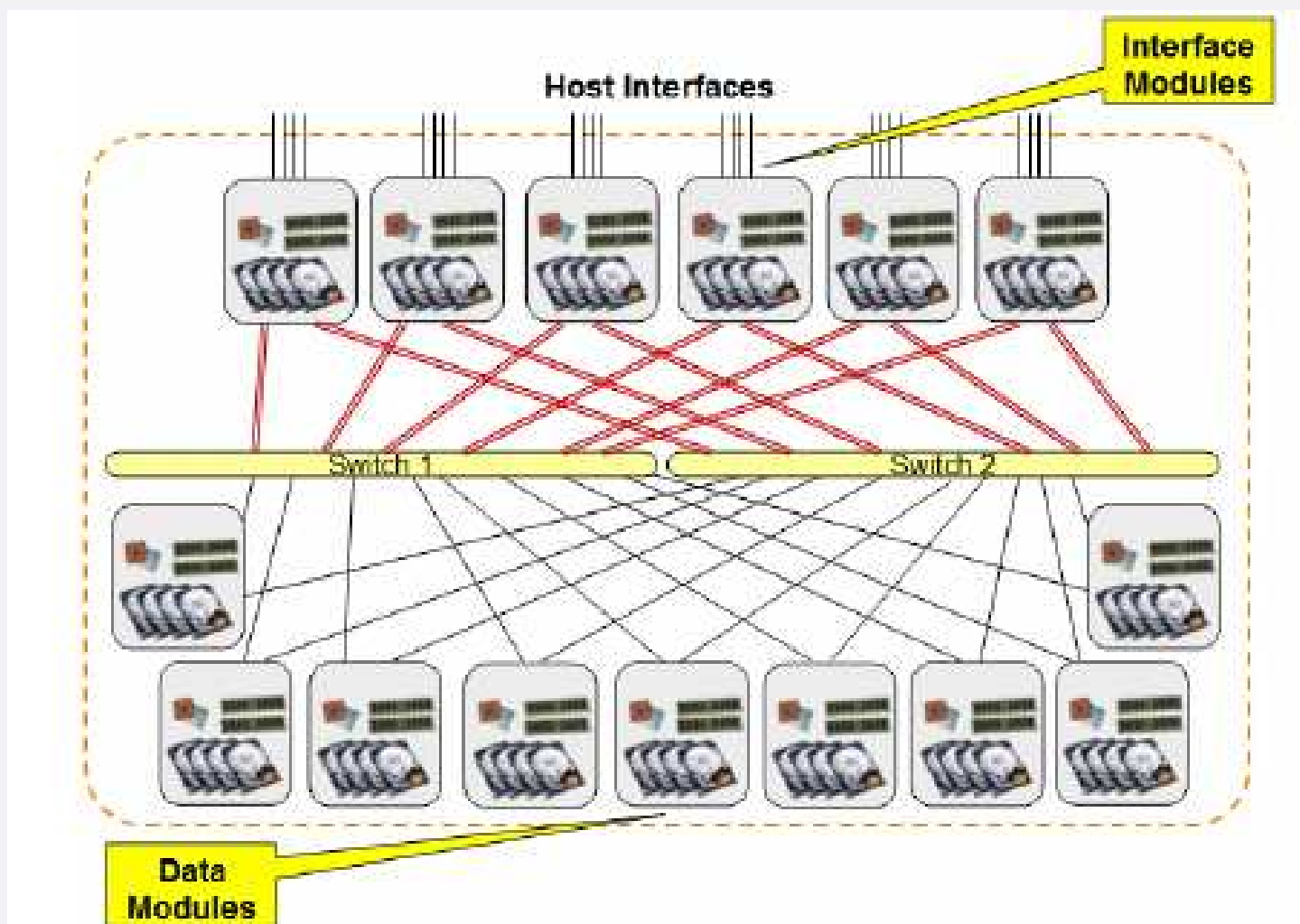


- Module 15 (Data)
- Module 14 (Data)
- Module 13 (Data)
- Module 12 (Data)
- Module 11 (Data)
- Module 10 (Data)
- Module 9 (Interface)
- Module 8 (Interface)
- Module 7 (Interface)
- Ethernet Switch Redundant PS
- Module 6 (Interface)
- Module 5 (Interface)
- Module 4 (Interface)
- Module 3 (Data)
- Module 2 (Data)
- Module 1 (Data)
- UPS 3
- UPS 2
- UPS 1



- Module 15 (Data)
- Module 14 (Data)
- Module 13 (Data)
- Module 12 (Data)
- Module 11 (Data)
- Module 10 (Data)
- Module 9 (Interface)
- Module 8 (Interface)
- Module 7 (Interface)
- Ethernet Switch - N2
- Ethernet Switch - N1
- Maintenance Module
- Module 6 (Interface)
- Module 5 (Interface)
- Module 4 (Interface)
- Module 3 (Data)
- Module 2 (Data)
- Module 1 (Data)
- UPS 3
- UPS 2
- UPS 1

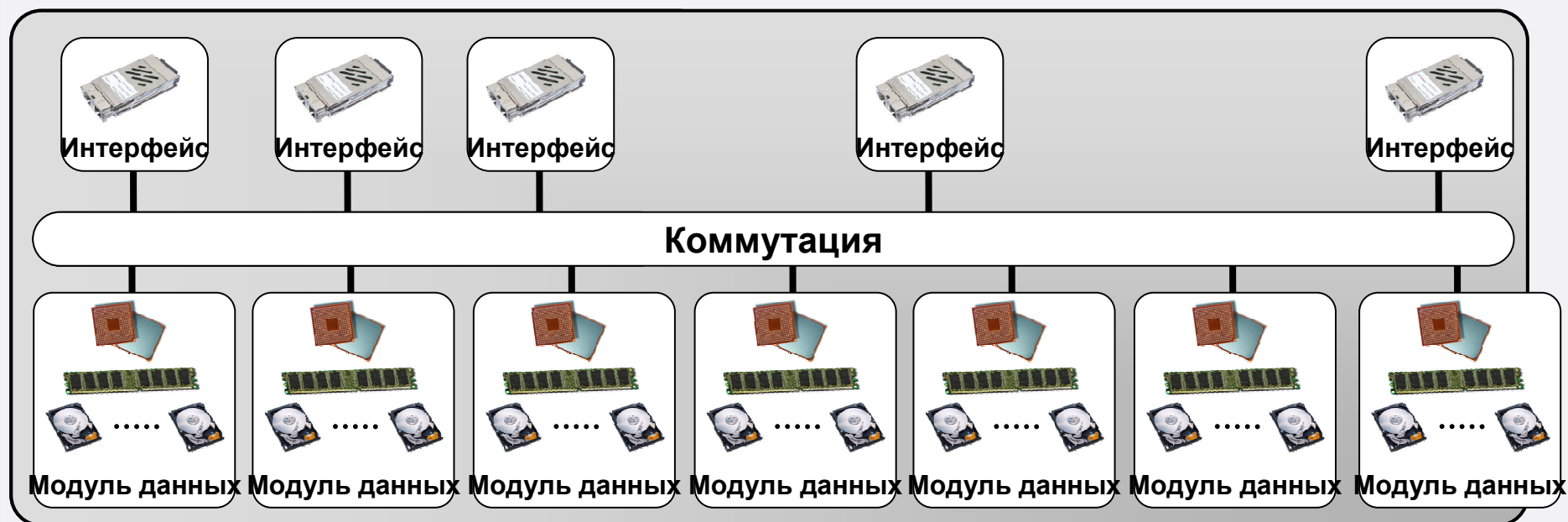
Внутренняя архитектура



Горизонтальный рост

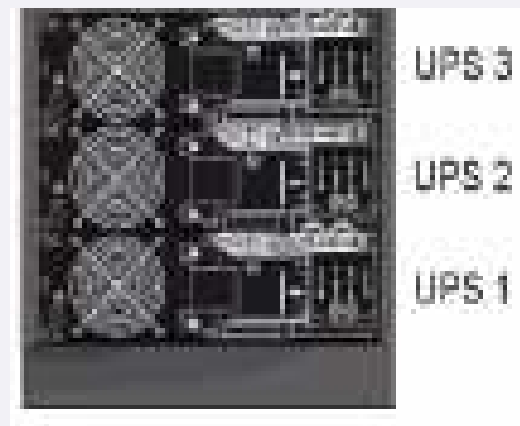


Горизонтальный
рост

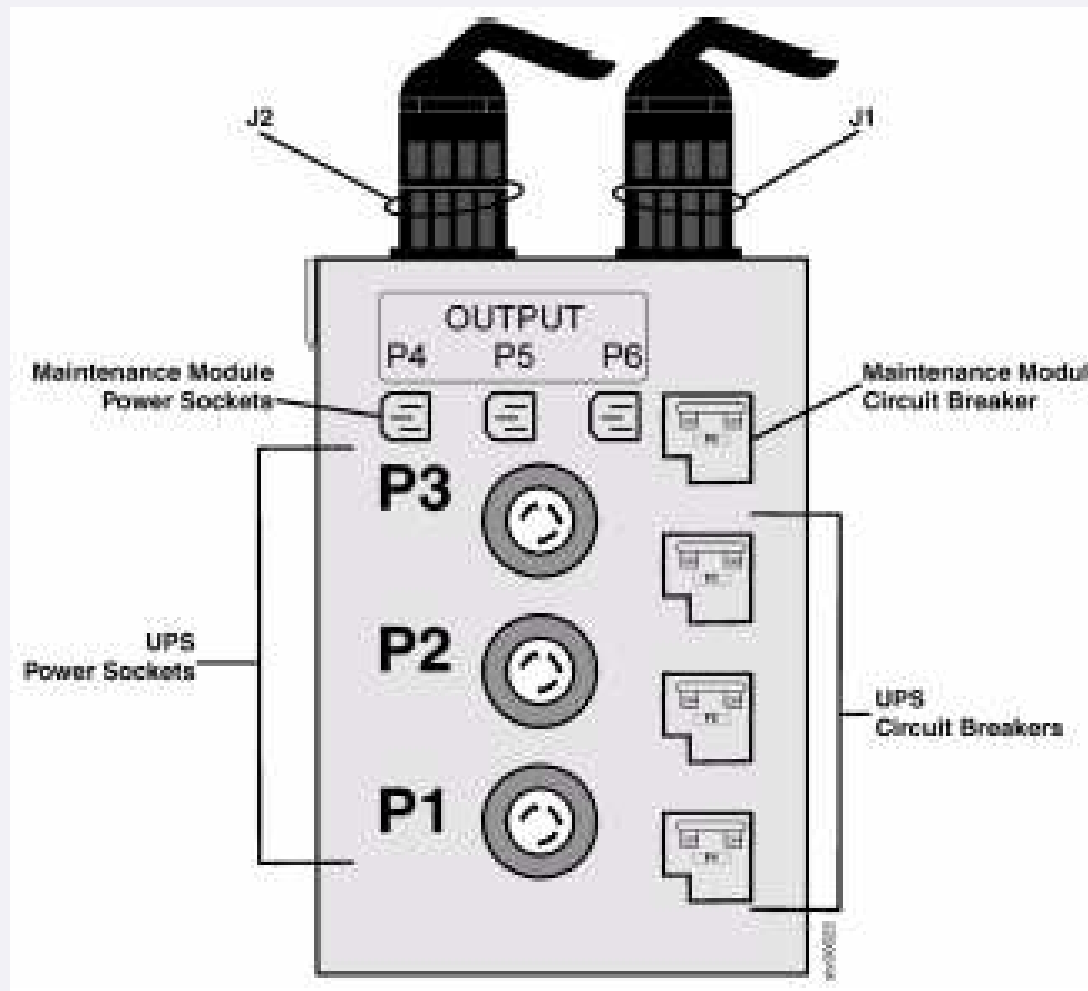


UPS

- 3 модуля UPS
- Каждый из них 3U высотой и мощностью 6kVA
- Поддержка работоспособности системы от 3.3 минут до 11.9 минут



ATS (Automatic Transfer System)



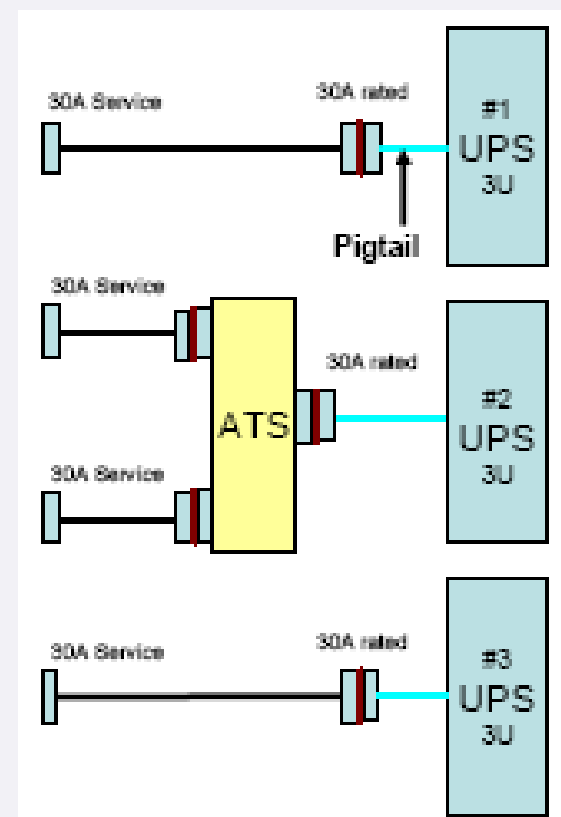
Возможное питание системы

Однофазное

- 2 60А кабеля
- 4 30А кабеля

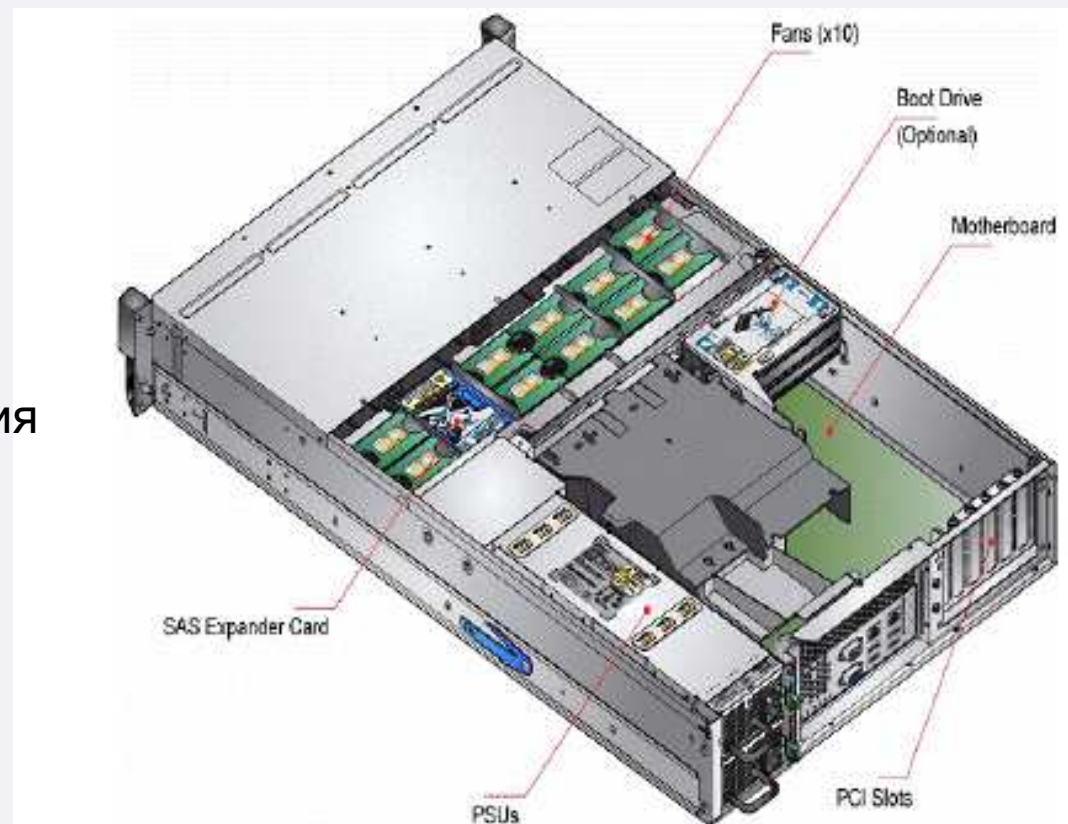
Трехфазное

- 2 60А кабеля
- 2 30А кабеля



Основные модули

- Размер 87.9 (2U)x48.3x70.7см
- Максимальный вес 30 кг
- Основные компоненты:
 - диски
 - системная плата
 - процессор (ы)
 - память (кэш)
 - плата управления
 - система охлаждения
 - флеш карта
 - резервированные блоки питания



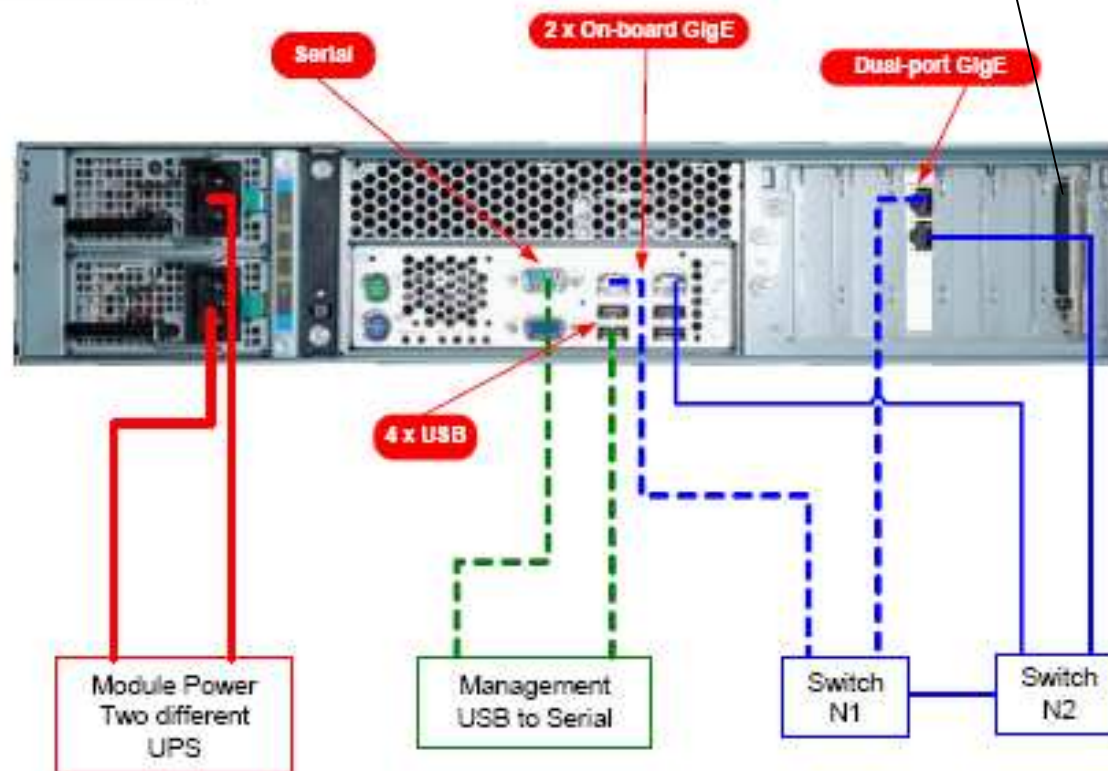
Модули данных

■ Основные компоненты

- Процессор 2.33GHz/12MB/1.33GHz]
- Память 8 Gb (8x1 fully buffered 667MHz)
- 4 Порта 1Gb Ethernet
- 4 Порта PCI-Express
- SAS адаптер



Data Module

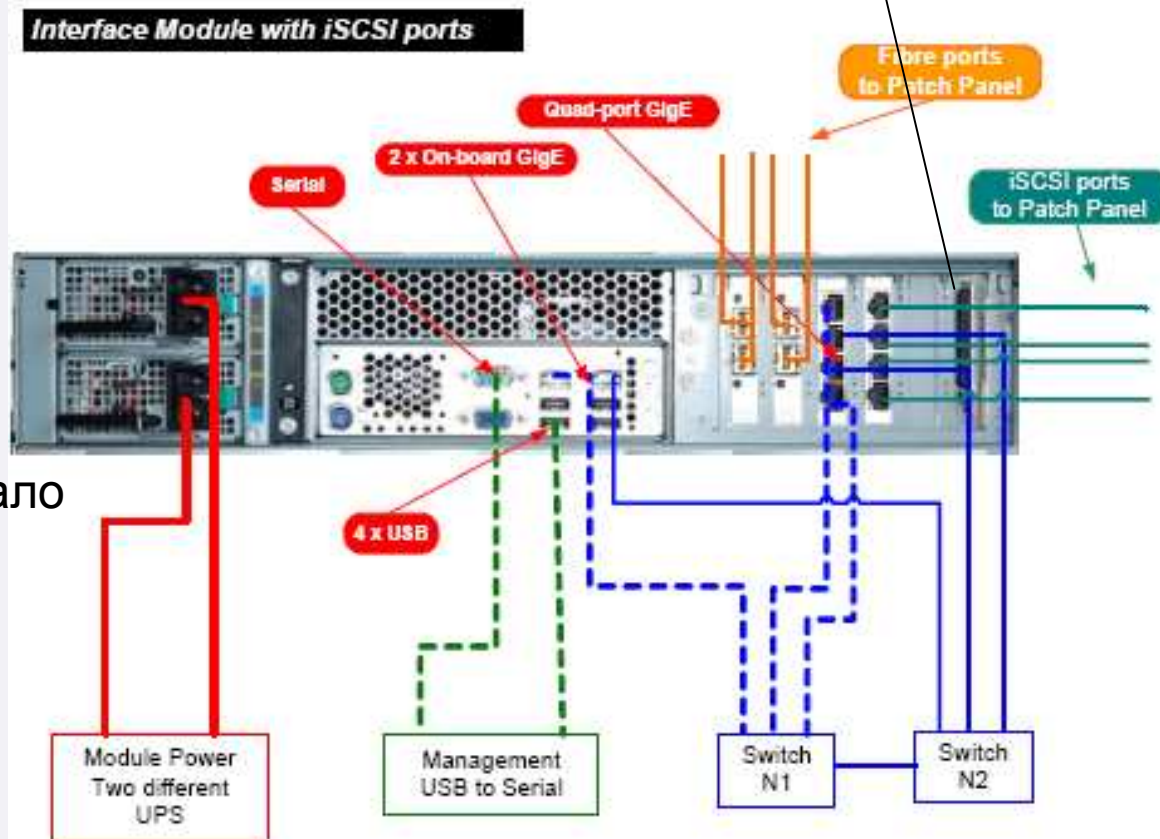


Интерфейсные модули

- Основные компоненты
 - Процессор (ы) 2.33GHz/12MB/1.33GHz]
 - Память 8 Gb (8x1 fully buffered 667MHz)
 - 4 Порта 1Gb Ethernet
 - 4 Порта PCI-Express
 - SAS адаптер
 - Дополнительная плата Eth
 - Дополнительные платы FC



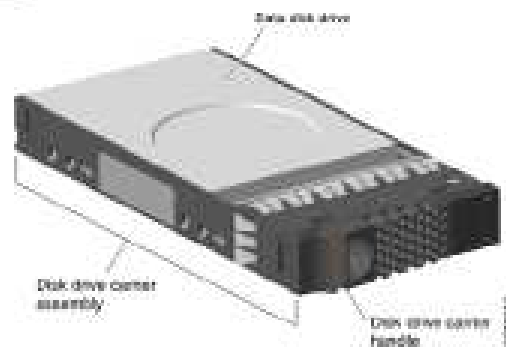
FC порты 1 и 3 – хост
 FC порты 2 и 4 - хост + зеркало
 или миграция



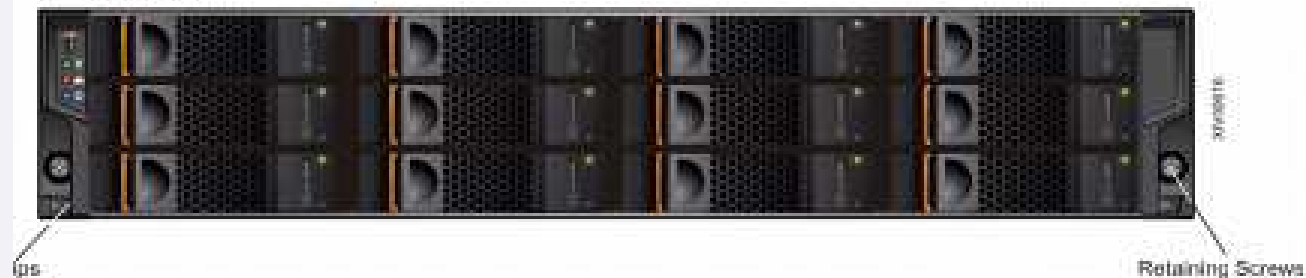
SATA диски

- SAS интерфейс 3Гб/с
- Поддержка NCQ (Native Command Queuing)
- 32MB кэш
- Поддержка RVS (Rotation Vibration Safeguard)

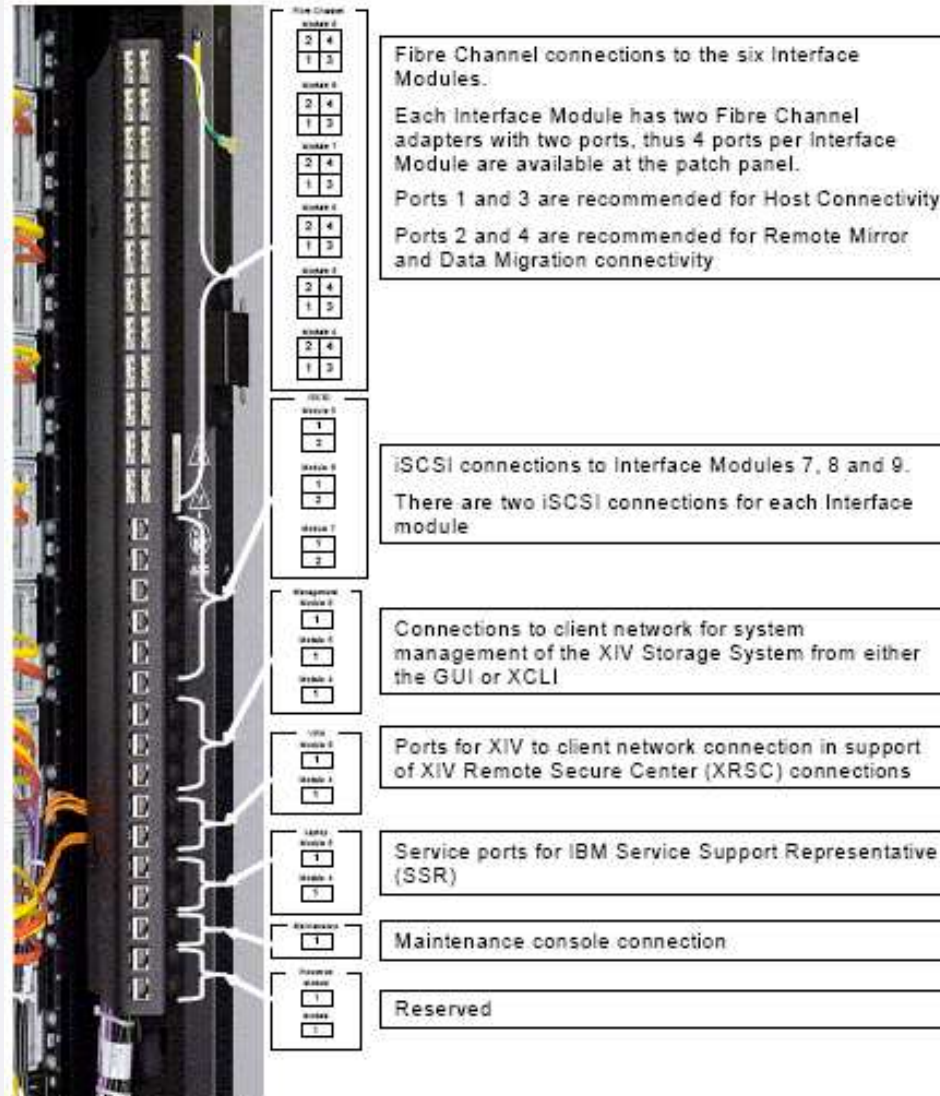
SATA disk drives



Module front view



Коммутационная панель





Внутренняя коммутация

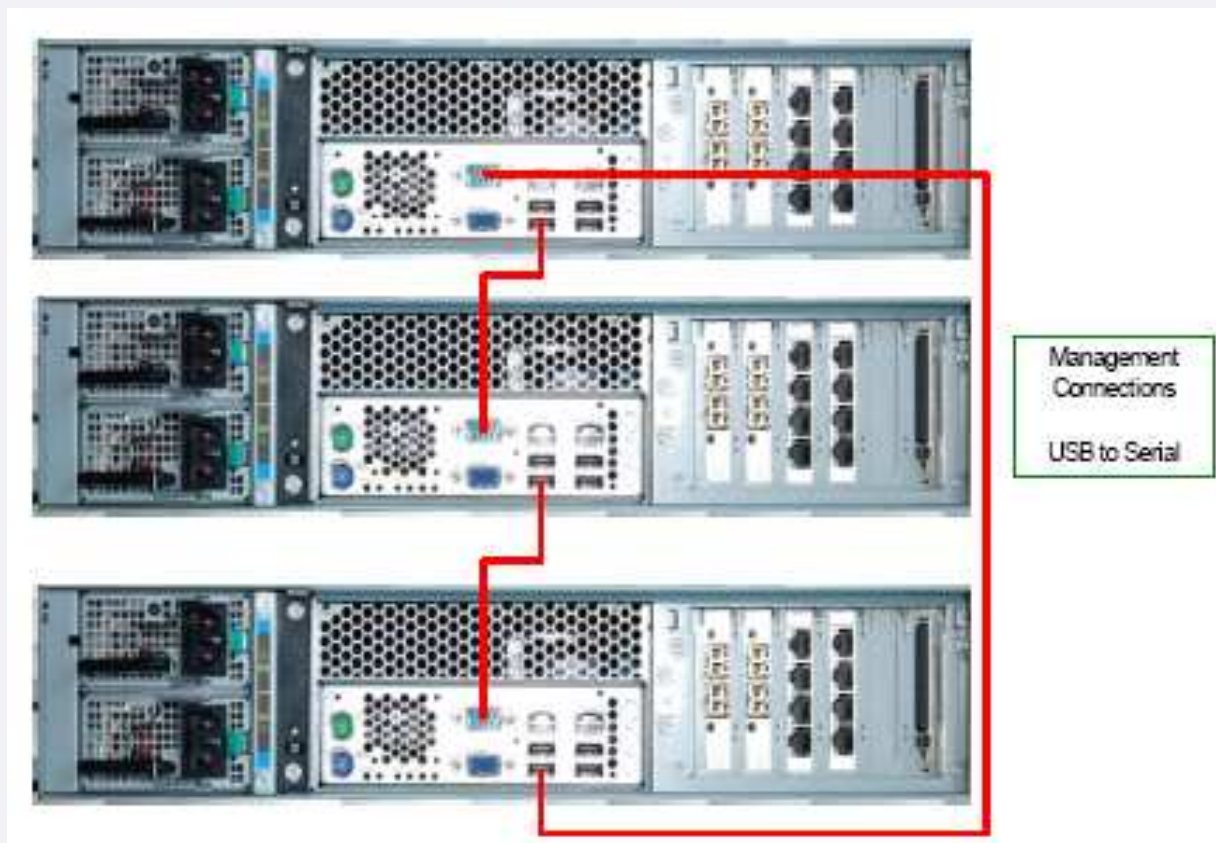
Два свитча:

- 48 портов каждый
- резервированные блоки питания
- возможность установки 10 Гб SFP

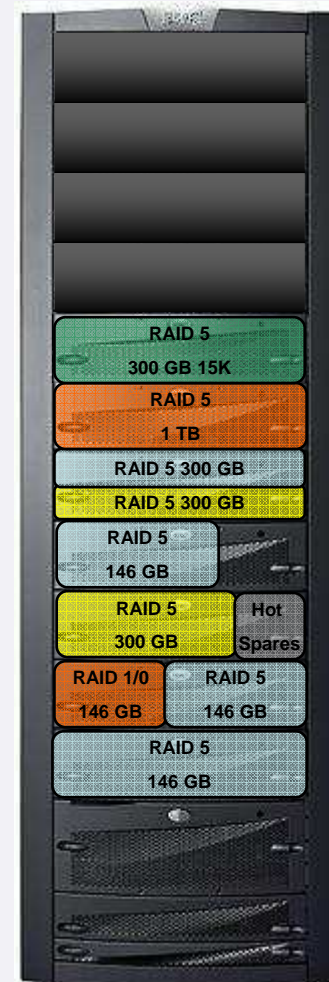
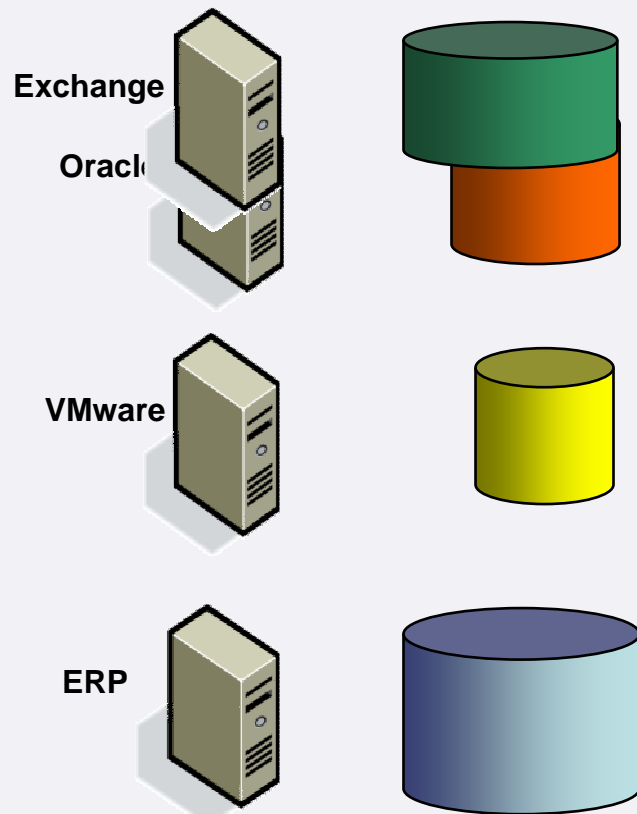


Резервная внутренняя коммутация

- Всегда группируются по 3
 - 1-3
 - 4-6
 - 7-9
 - 10-12
 - 13-15

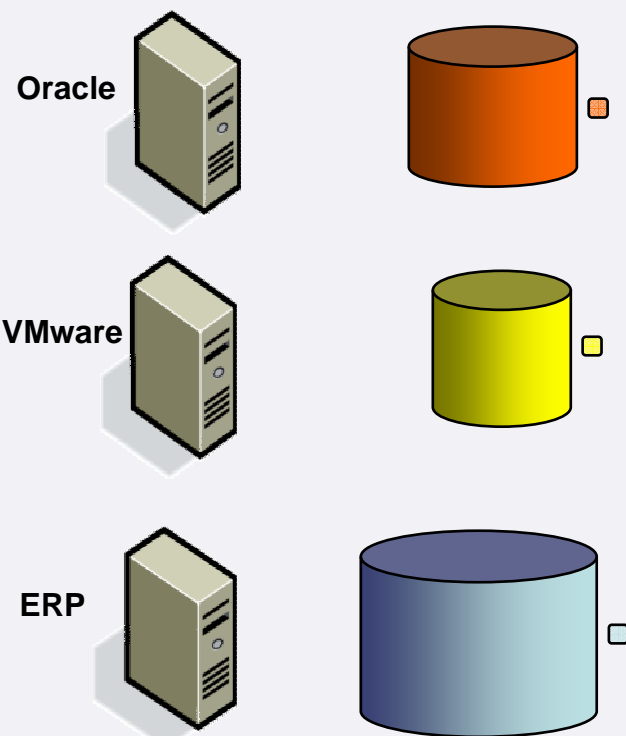


Распределение данных в обычной системе



- Оборудование
 - Disks, Tiers,
 - Spares, RAID
- Масштабирование ограничено и не
- Performance
- Узкие места и нехватка
 - Требует настройки
- Производительности требует анализа, снижает
- Производительность системы
- Для увеличения производительности необходима переконфигурация

Распределение данных в XIV

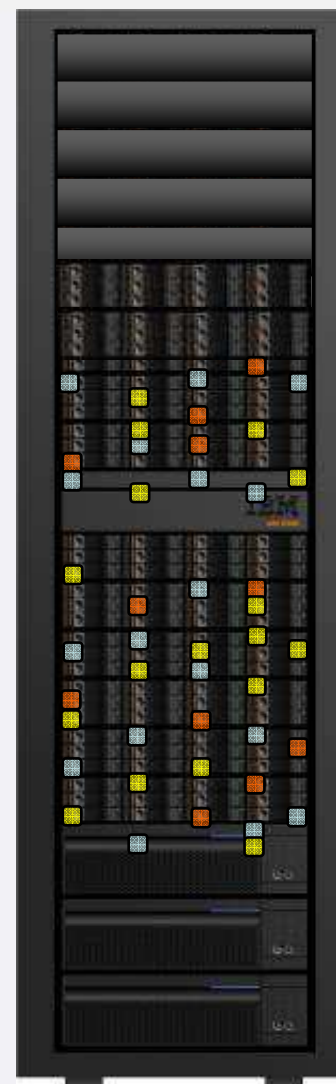
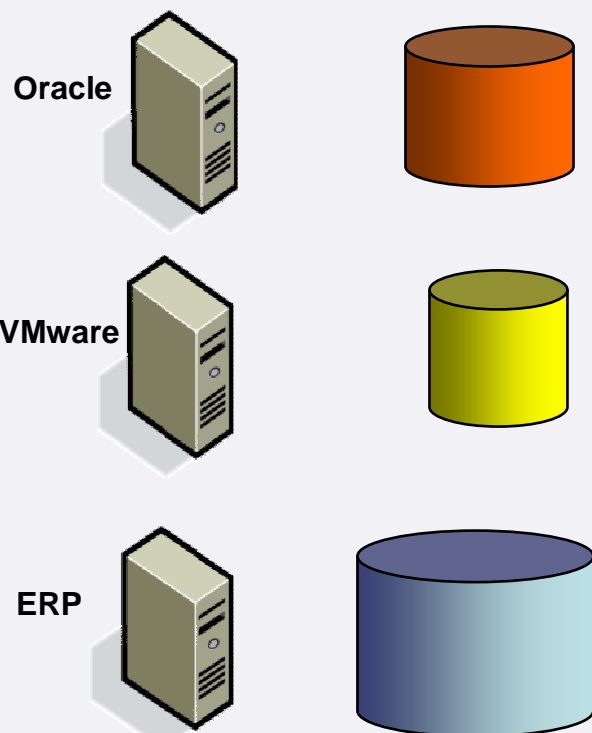


- Каждый раздел распределяется по всем дискам
- Данные разбиваются на блоки в 1 МБ
- Использование дисков оптимально при любом сценарии
- Не требует никакого вмешательства
- Не требует настройки
- Нет узких мест



Балансировка нагрузки в XIV

Добавление модулей



Добавляем модули

Данные перераспределяются автоматически

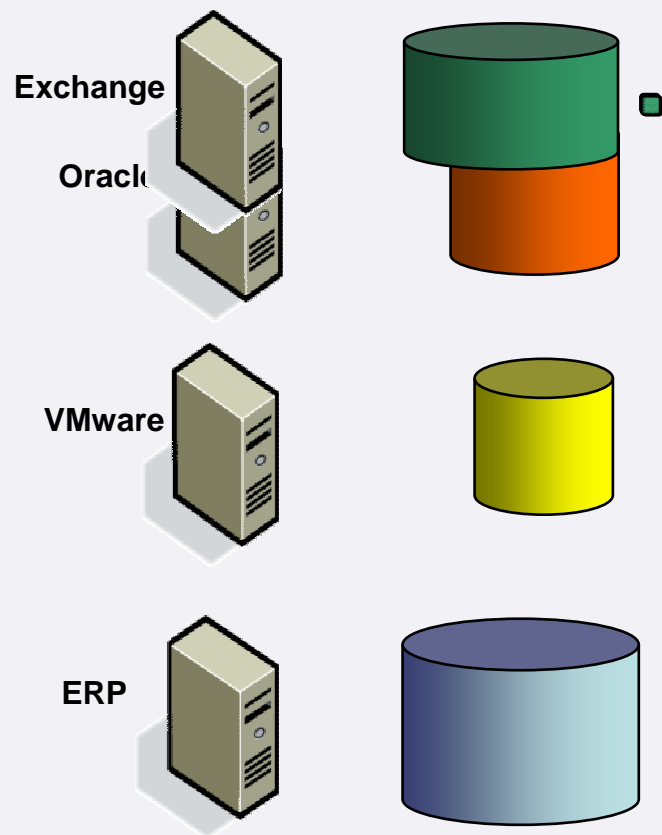
Нет воздействия на производительность

Не требуется вмешательства

Не требуется настройка

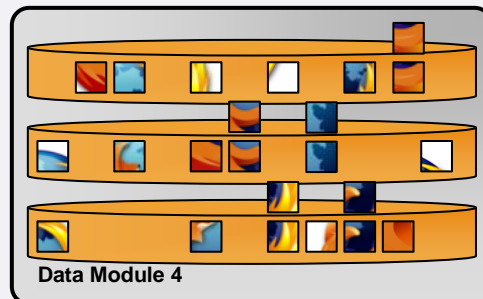
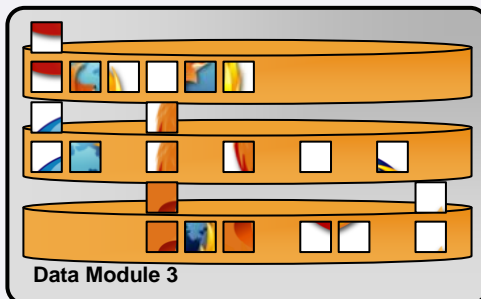
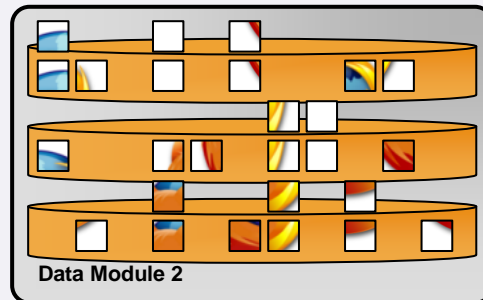
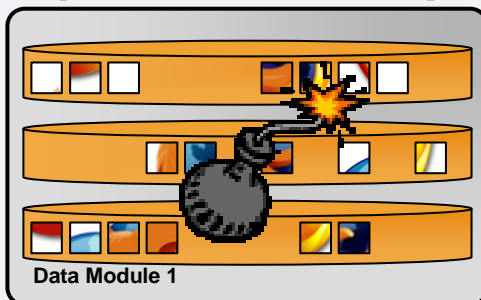


Создание нового раздела



IBM XIV

[hardware failure]



- Данные восстанавливаются из распределенных копий
- Производительность восстанавливается
- Самовосстановление



Тестирование отказоустойчивости

Test series I: Single + Double Disk (36 TB Date)

1. Test: Single Disk Failure

17:14 – Removed one Disk. XIV starts redistribution

17:16 - Redistribution finished. System full redundant.

2. Test: Double Disk Failures

17:25:53 and 17:25:57 : Disk 14:7 and 14:6 removed

17:27:17 Data Redistribution completed

Test series II: Total Module Failure (52 TB Date)

3. Test: Total Module Failure

14:20:12 Remove Power from Data Module

-> immediate down

14:45:45 Data Rebuild finished



- Отказ 17 дисков в течении часа, но XIV восстанавливается в течении нескольких минут






| | Date ▼ | Event Code | User | Description |
|---|---------------------|---------------------------------|------|--|
| i | 2009-09-02 14:45:45 | DATA_REBUILD_COMPLETED | | Rebuild process completed. System data is now prot... |
| i | 2009-09-02 14:29:43 | USER_LOGIN_SUCCEEDED | | User 'admin' successfully logged into the system. |
| i | 2009-09-02 14:20:36 | USER_LOGIN_SUCCEEDED | | User 'admin' successfully logged into the system. |
| i | 2009-09-02 14:20:36 | HOST_MULTIPATH_OK | | Host 'win_linux_02' has redundant connections to th... |
| i | 2009-09-02 14:20:36 | HOST_MULTIPATH_OK | | Host 'solaris' has redundant connections to the syste... |
| i | 2009-09-02 14:20:36 | USER_LOGIN_SUCCEEDED | | User 'admin' successfully logged into the system. |
| i | 2009-09-02 14:20:35 | DATA_REBUILD_STARTED | | Rebuild process started because system data is not ... |
| x | 2009-09-02 14:20:33 | TARGET_CONNECTION_DISCONNECTED | | Target named 'XIVloopTarget' is no longer accessibl... |
| x | 2009-09-02 14:20:27 | MODULE_FAILED | | 1:Module:7 failed. |
| i | 2009-09-02 14:19:43 | USER_LOGIN_SUCCEEDED | | User 'admin' successfully logged into the system. |
| i | 2009-09-02 14:07:30 | HOST_MULTIPATH_OK | | Host 'win_linux_02' has redundant connections to th... |
| i | 2009-09-02 13:56:53 | HOST_NO_MULTIPATH_ONLY_ONE_PORT | | Host 'win_linux_02' is connected to the system throu... |
| i | 2009-09-02 13:56:41 | HOST_MULTIPATH_OK | | Host 'win_linux_02' has redundant connections to th... |



- Протестирован отказ 3 модулей в течении 2 часов
- Одновременный отказ
 - Модуль данных
 - коммутатор
 - UPS

У XIV нет единой точки отказа!



SnapShots

- Мгновенное создание и удаление логических копий
- Без потерь производительности
- Неограниченное количество копий
- Дифференциальные копии
- Копии

Высокая производительность логических копий дает преимущества :

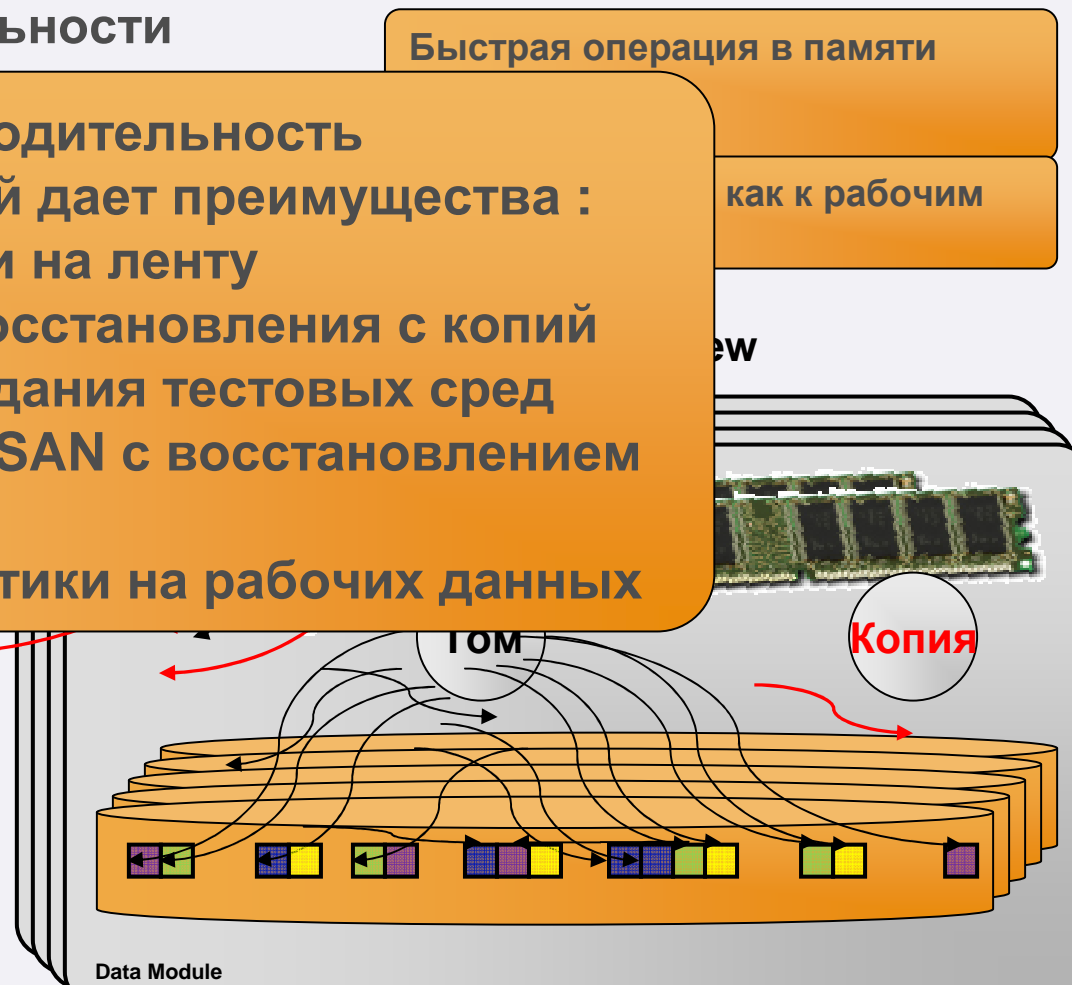
- При сохранении на ленту
- Мгновенного восстановления с копий
- Упрощение создания тестовых сред
- Загрузка через SAN с восстановлением данных
- Легкость аналитики на рабочих данных

Быстрая операция в памяти

как к рабочим



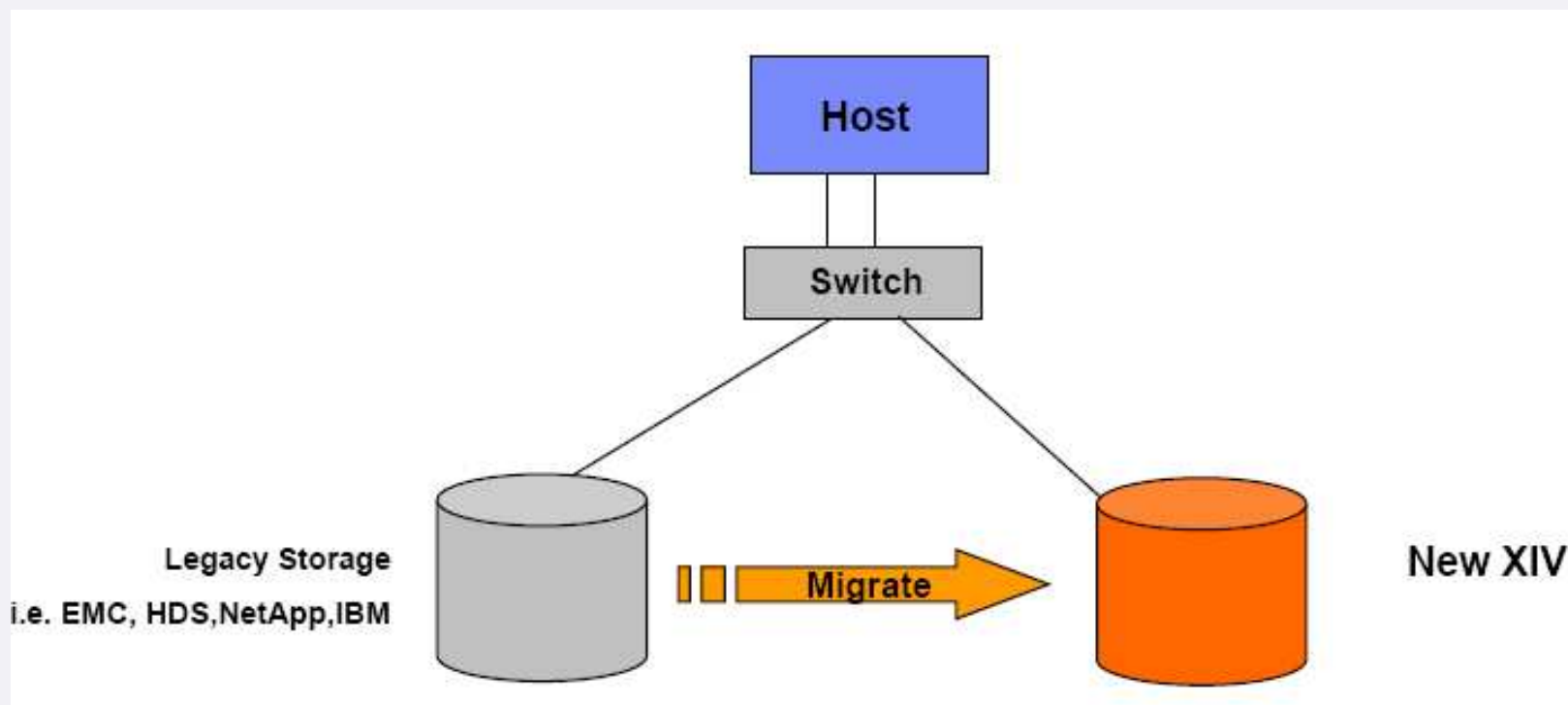
Восстановление тома с копии



Data Module

Data Migration

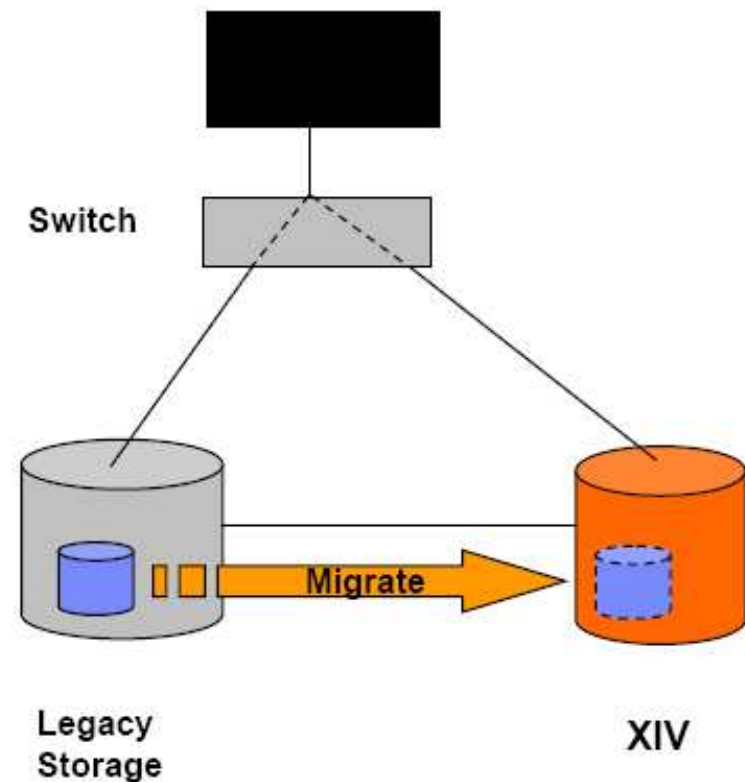
- XIV позволяет перенести данные с другого устройства
- Процесс занимает несколько минут



Шаги миграции

1. Prepare new SAN Zones
2. Establish migration connection
3. Setup XIV as host in legacy storage system
4. Shutdown application
5. Change zoning (from old storage to XIV)
6. Map source LUN to XIV (on host)
7. Start migration (XIV GUI)
8. Map new volumes to host
9. Startup host
-Migration is performed in background
10. When migration ready: remove old storage

Can be done in some minutes ...





Простота XIV начинается уже при создании конфигурации

Не более 30 возможных параметров!!!

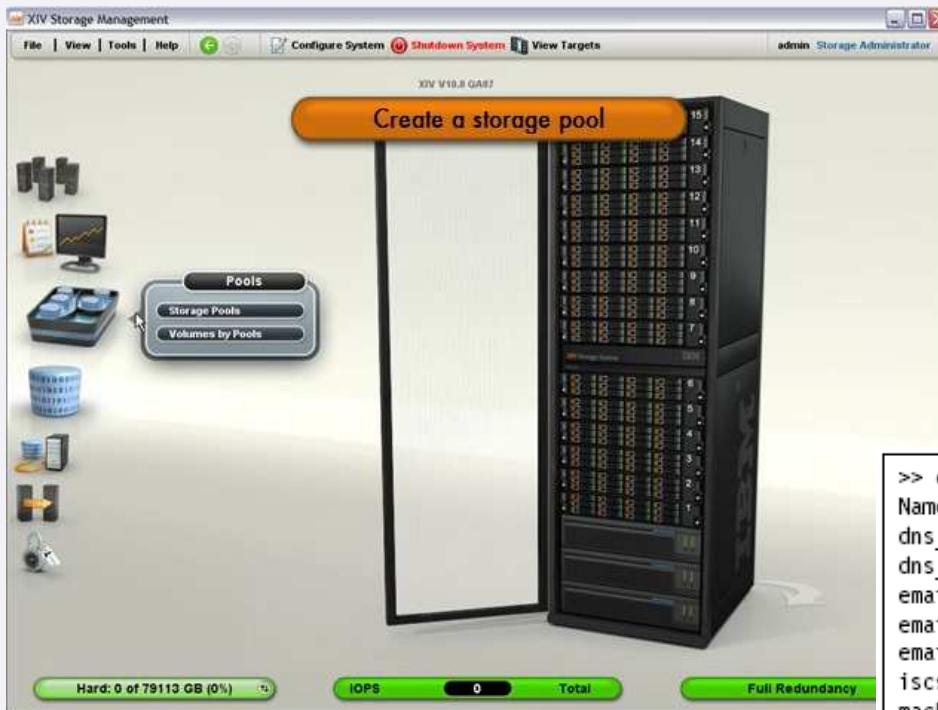
The screenshot shows a configuration window titled "2810-A14 IBM XIV Storage System (1 yr Warranty): # 1: - Data Modules". It features several tabs: "Quantity", "Indicator", "Data Modules", "Power Cord", "Cable/Modem", and "Misc". The "Data Modules" tab is active, showing an "Order type" dropdown set to "Dual-CPU" and "Dual-CPU Capacity on Demand (CoD)". Below this is a table for "Interface and Data Modules":

| Description | Proposed |
|--------------------------------|----------|
| 1101 Dual-CPU Interface Module | 3 |
| 1106 LV Data Module | 3 |

At the bottom, there is a "Feature Code:" field, a "Search" button, and a "Results:" field. Navigation buttons include "< Previous", "Next >", "OK", "Cancel", and "Reset Page".



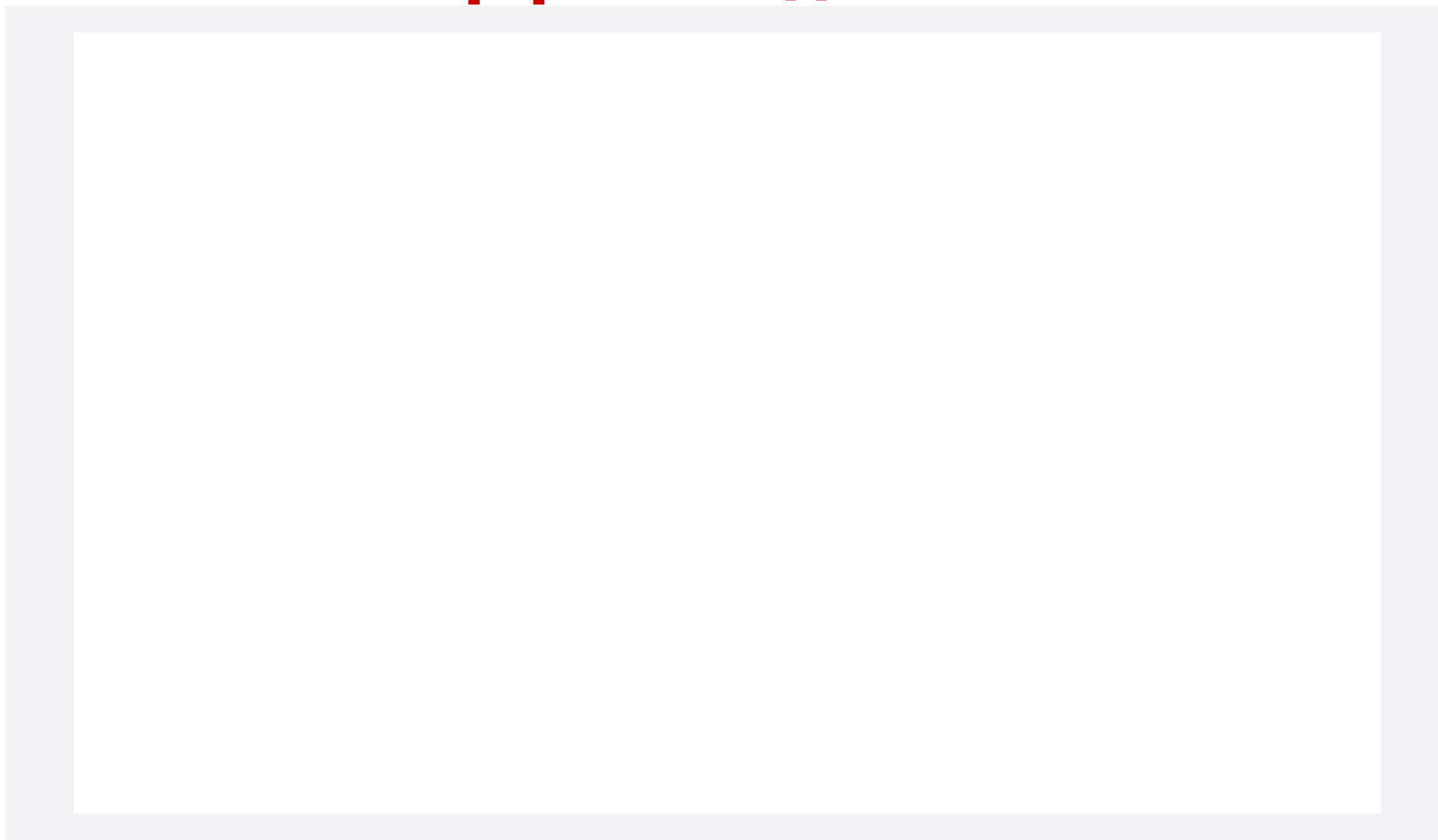
Интерфейс управления



```
>> config_get
Name                Value
dns_primary
dns_secondary
email_reply_to_address
email_sender_address
email_subject_format {severity}: {description}
iscsi_name           iqn.2005-10.com.xivstorage:000019
machine_model        A14
machine_serial_number MNO0019
machine_type         2810
ntp_server
snmp_community       XIV
snmp_contact         Unknown
snmp_location        Unknown
snmp_trap_community  XIV
support_center_port_type Management
system_id            19
system_name          XIV MNO0019
```



Интерфейс в действии



**Спасибо за
внимание!**

