

Суперкомпьютеры для администратора привычные и нет

Жуматий С.А.

serg@parallel.ru

в.н.с. НИВЦ МГУ им.М.В.Ломоносова

2017

Из чего состоят суперкомпьютеры

- Аппаратные компоненты
- Системное ПО

Компоненты

- Вычислитель
 - Сети
 - Узлы
- Инфраструктура
 - Помещение
 - Электричество
 - Охлаждение

Инфраструктура: помещение

- Площадь
- Шумность
- Доступ для техобслуживания
- Противопожарное оборудование
- Контроль доступа

Инфраструктура: электричество

- Источники питания, проводка, автоматы
- ИБП, место
- Расчётное время работы от батарей
- Фальшпол, коробка, проводка в стойках
- Заземление, статика

Инфраструктура: охлаждение

- Мощность: потребляемая и реальная
- Влажность
- Центральное отопление, солнце
- Зимнее время
- «Открытое окно»

Структура кластера

- Сети: управляющая, вычислительная, сервисная
- Управляющий узел
- Служебные узлы
- Вычислительные узлы
- Сетевая файловая система
- Система архивирования

Как устроен суперкомпьютер



Служебные узлы

- DHCP — управление адресами
- TFTP — загрузка, установка
- NFS — загрузка, root-fs
- DNS — управление именами
- NTP — синхронизация времени
- LDAP — управление учётными записями

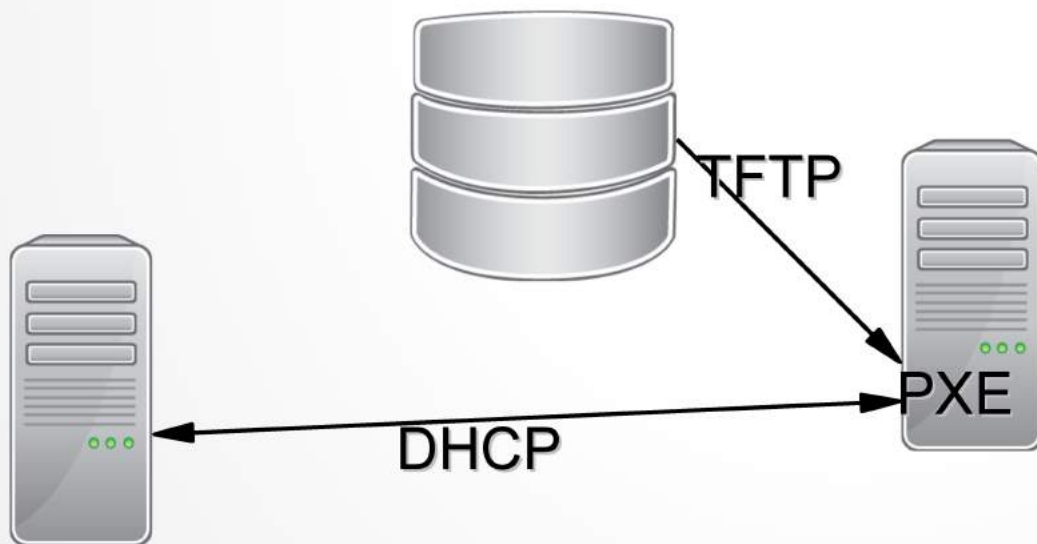
Служебные узлы

- Доступ (управляющий)
- Управление задачами
- Лицензии (flexlm)
- Мониторинг
- Статистика
- Визуализация
- Копирование данных



Операционная система

Установлена локально



Загружается по сети

Операционная система

Установлена локально:

Плюсы

- Быстрая загрузка
- Нет перегруза сети
- Не занимает лишнего места в памяти

Минусы

- Долгая установка и обновление
- Рассинхронизация настроек и версий

Операционная система

Загрузка по сети:

Плюсы

- Единый образ для всех узлов группы
- Возможность быстро обновить всё

Минусы

- Нагрузка на сеть
- Сложность обновления на лету
- Необходимо хранить файлы настроек в памяти

Установка системного ПО

- «по одному», клонирование
- Автоматизированная установка
- Готовые стеки
- ХСАТ

Установка системного ПО: Автоматизированная установка

RH: anaconda-ks.cfg,

SuSE: autoyast.xml,

Debian: preseed.file

linux **ks=http://путь-к-файлу.cfg**

linux **autoyast=http://путь-к-файлу.xml**

linux **auto preseed=http://путь-к-файлу**

Установка системного ПО: ГОТОВЫЕ СТЕКИ

- ROCKS
- PelicanHPC / Parallel Knoppics

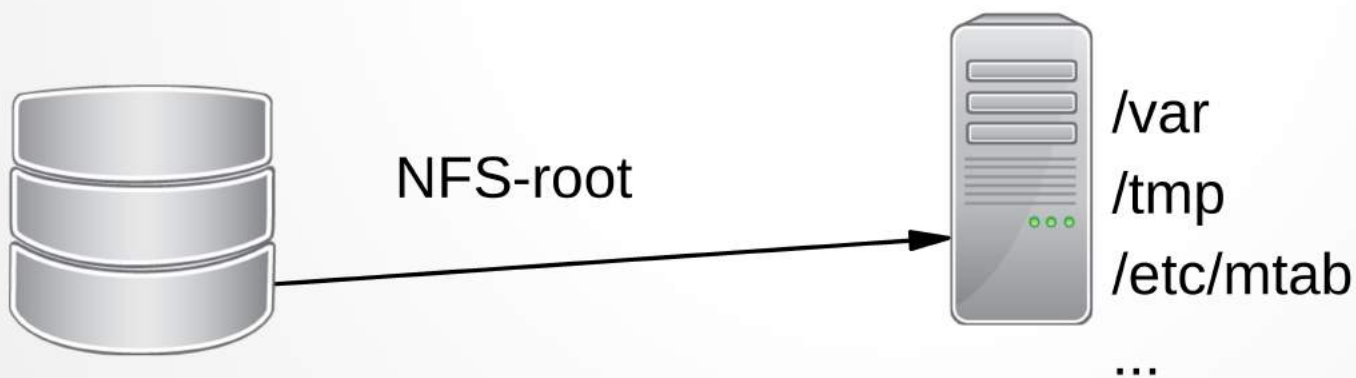
Установка системного ПО: xCAT

- Использует anaconda/autoyast
- Управляет DNS, DHCP, TFTP
- Поддерживает удалённое управление питанием (через IPMI, ILO, ...)
- Имеет встроенные «массовые» команды

xCAT

Виды установки ОС:

- Локально (statfull)
- По сети в память (stateless)
- По сети, в памяти только изменения (statelite)



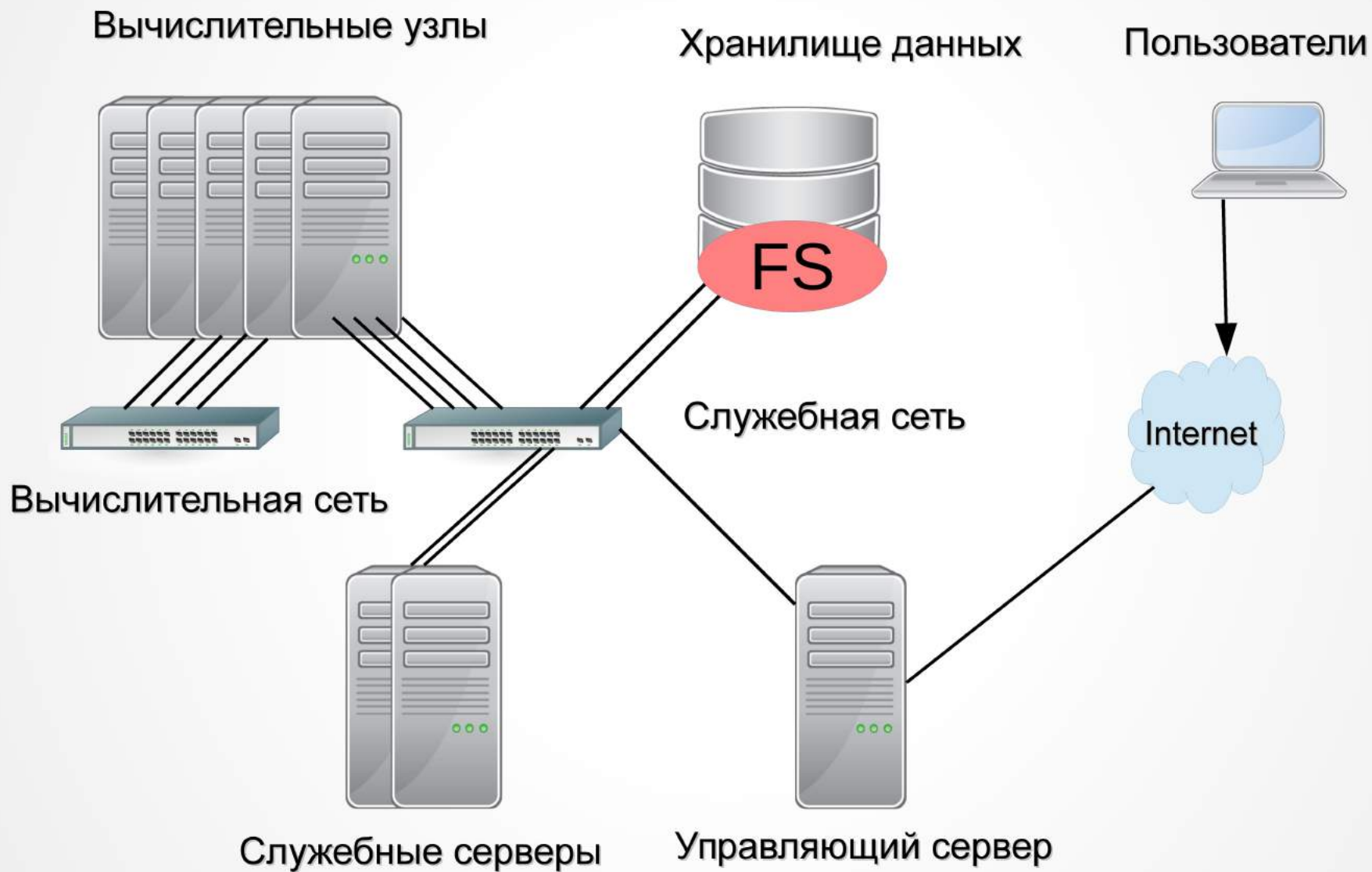
xCAT

Объекты: site, node, group, network, osdistro, osimage, route...

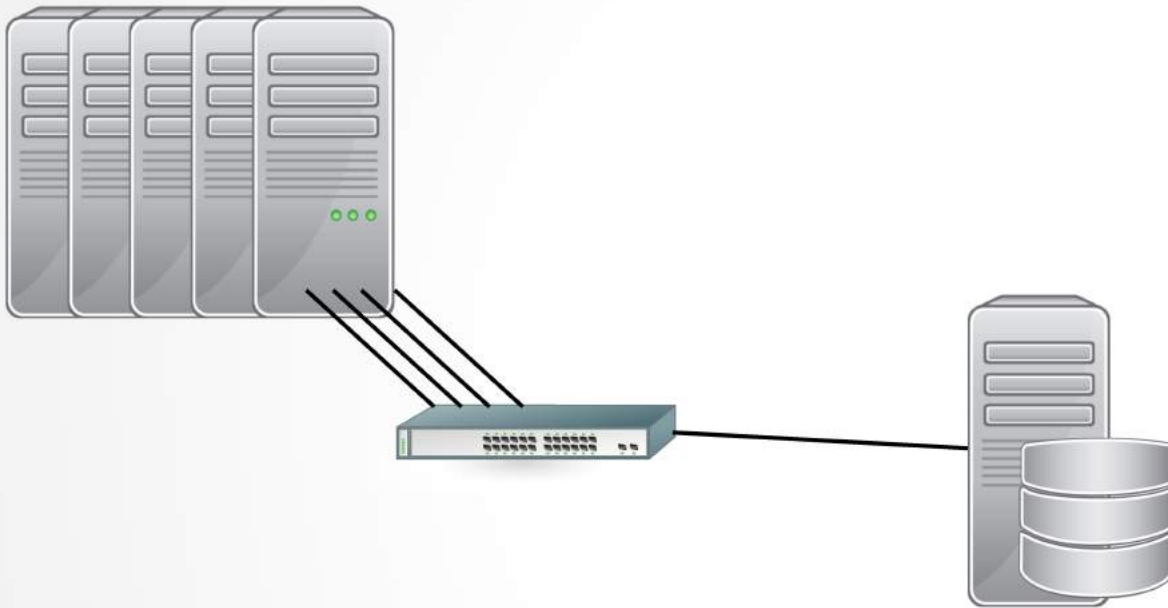
Таблицы: domain, hosts, ipmi, litetree, mac, nodegroup, nodelist, nodetype, postscripts,...

```
mkdef -t node node-1 groups=all,compute arch=x86_64 \  
    bmc=node-1-ipmi bmcusername=ADMIN bmcpassword=admin \  
    mac=xx:xx:xx:xx:xx:xx mgt=ipmi netboot=pxe \  
    provmethod=centos67-x86_64-statelite-compute
```

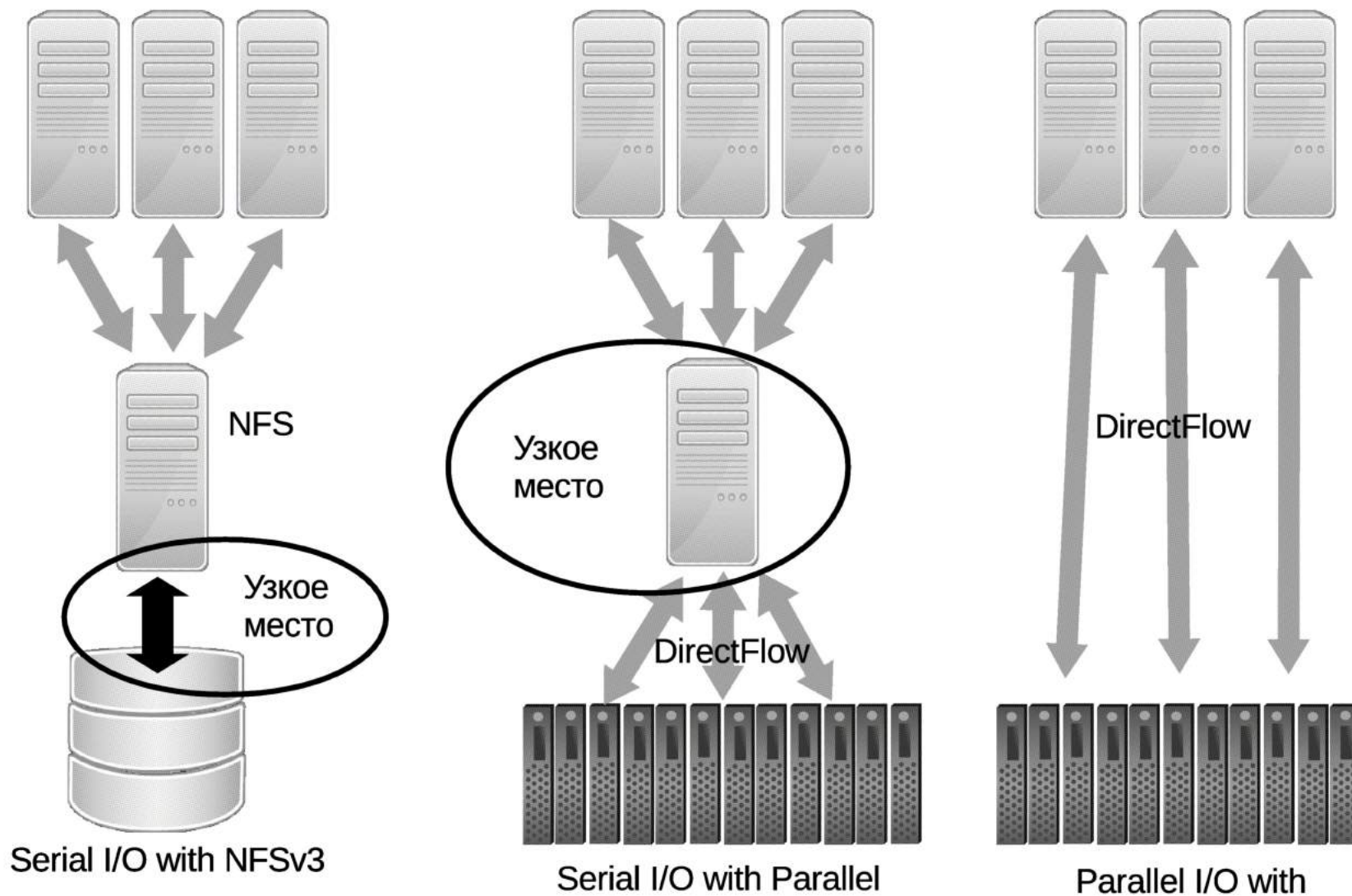
```
chtab node=compute hosts.ip='|\D+(\d+)|10.0.0.(10+$1)|'
```



NFS

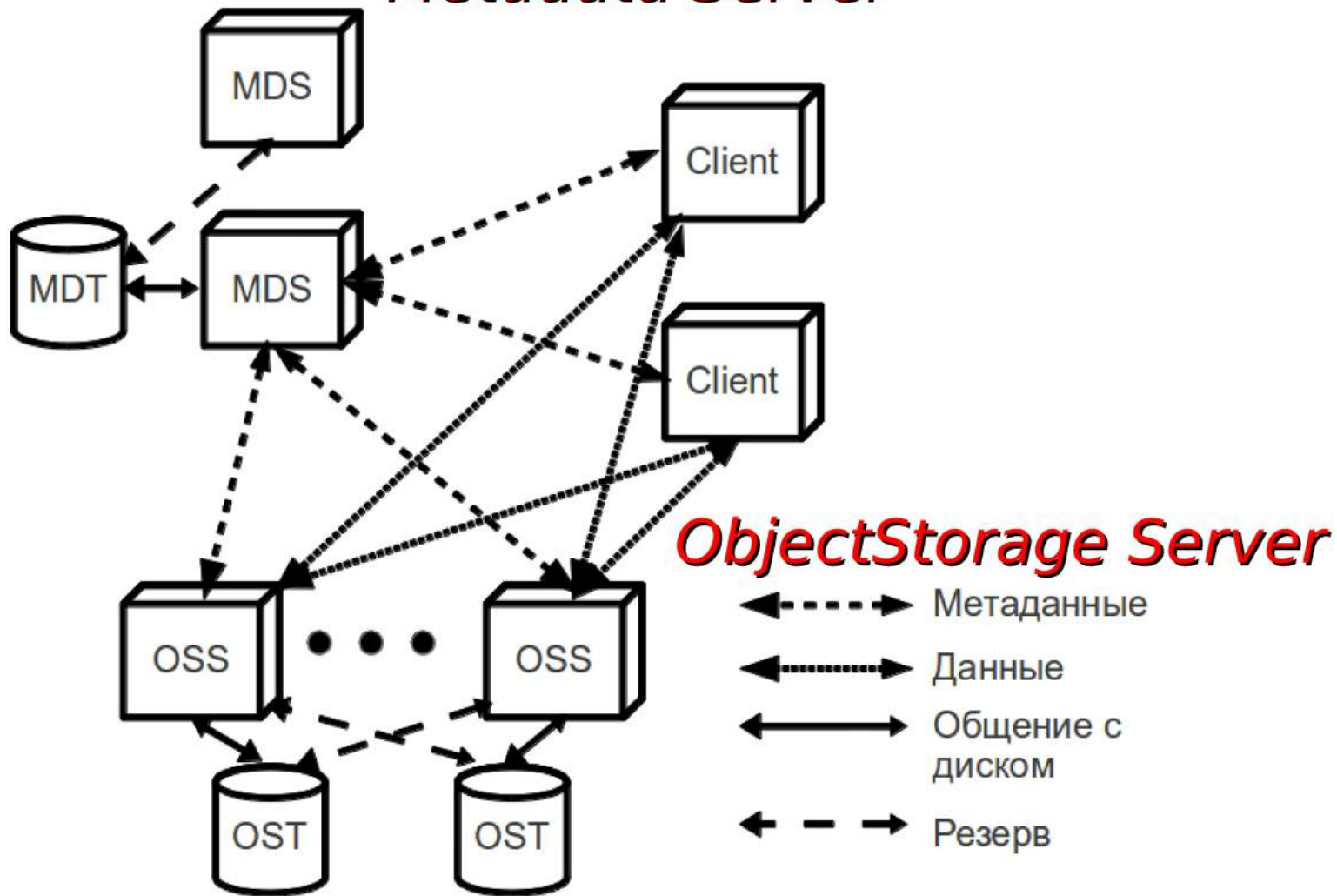


PanFS



Lustre

Metadata Server

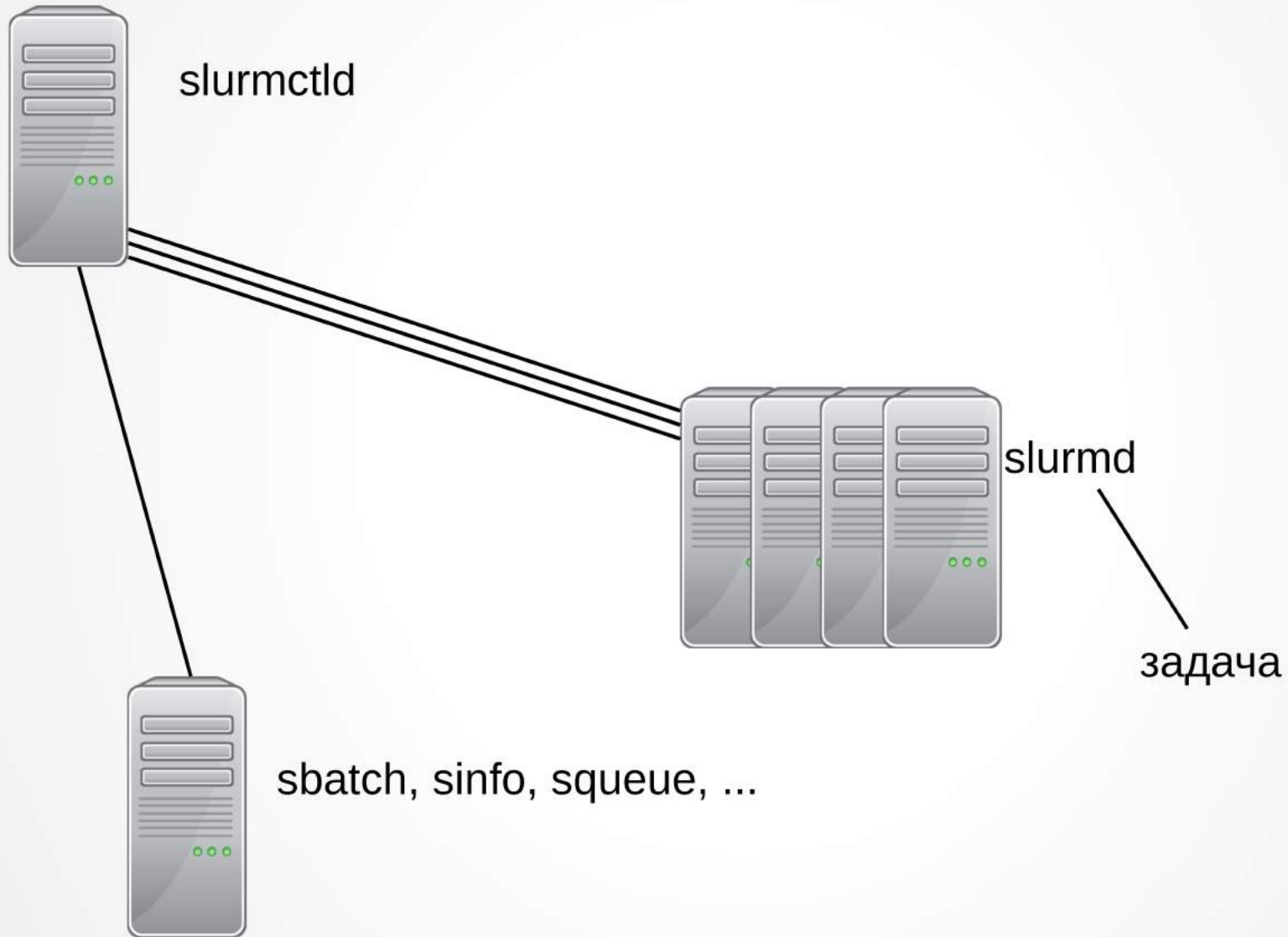




Управление задачами

- Slurm
- OpenPBS
- Torque
- LSF, BrightManager, Moab, ...

Slurm



Slurm



slurmctld

/etc/slurm/slurm.conf

scontrol

- show node|job|partition|...
- create ...
- delete ...
- modify ...

Хранение учётных записей

- passwd
- nis+
- LDAP
- Гибридный метод

Хранение учётных записей

- passwd

root:x:0:0:root:/root:/bin/bash

логин UID GID инфо дом шелл

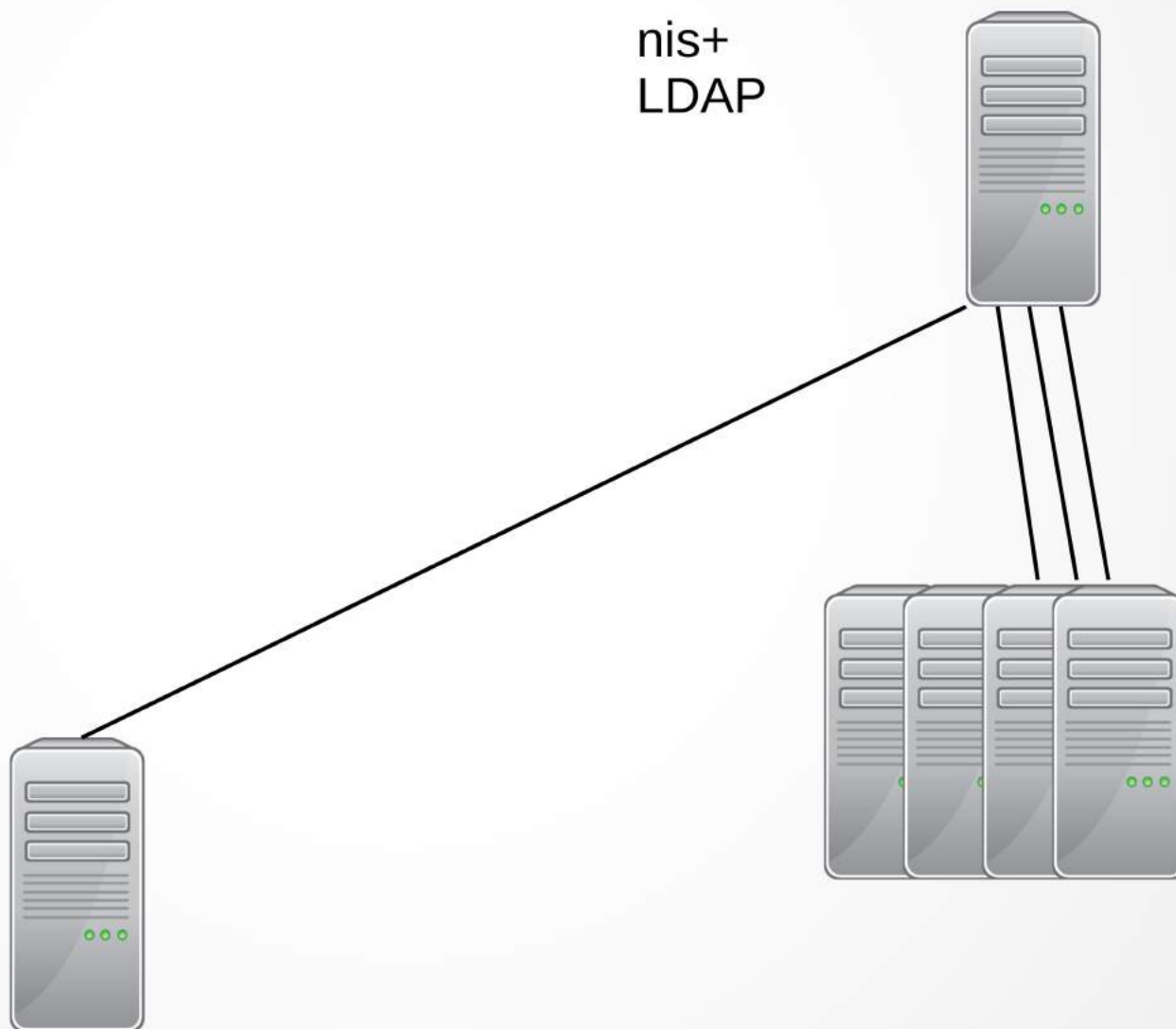
- shadow

root*:16219:0:99999:7:::

логин пароль

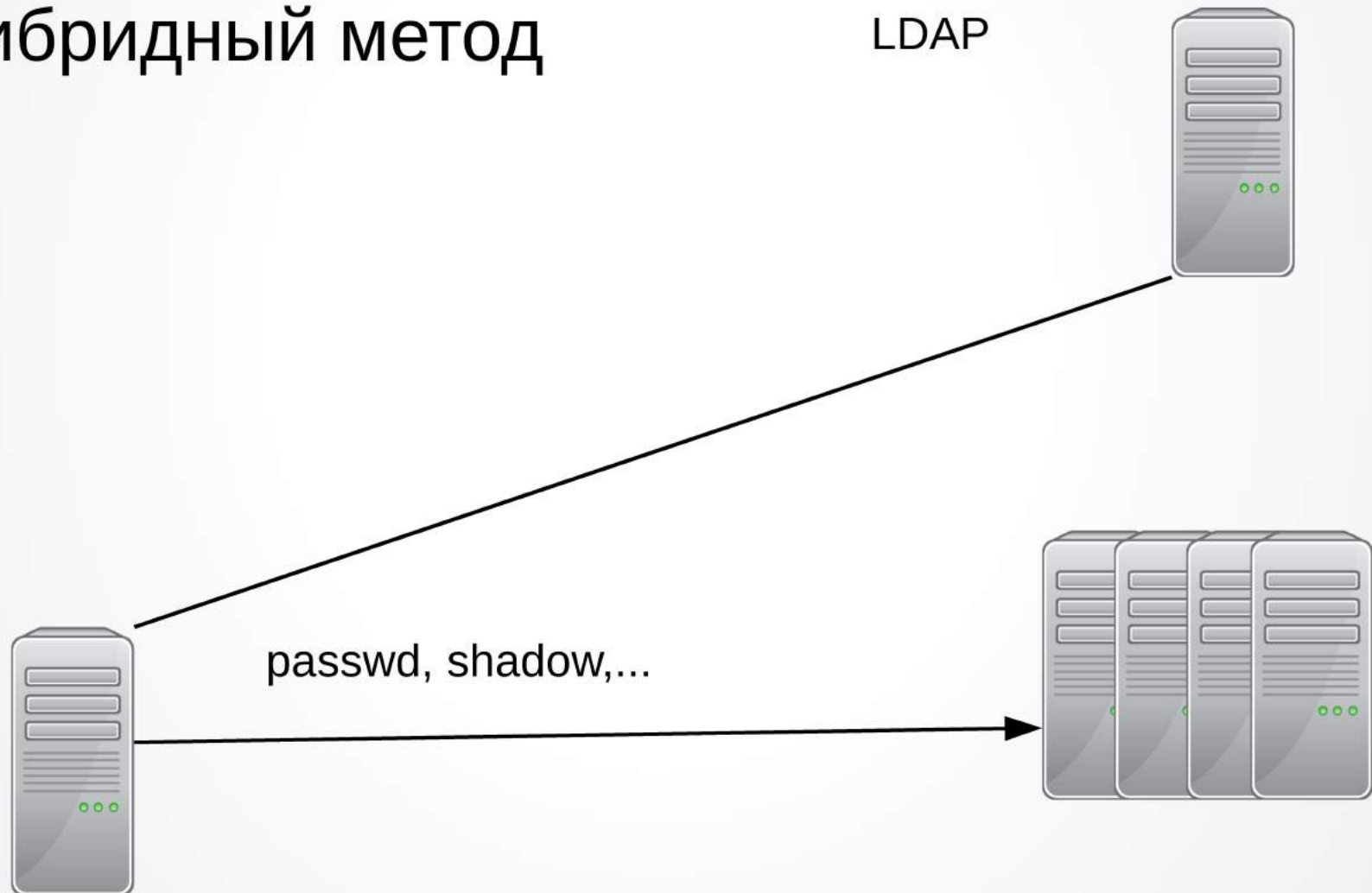
Хранение учётных записей

- LDAP



Хранение учётных записей

- Гибридный метод



За бортом...

- лимиты, ограничения доступа,
- мониторинг, статистика,
- общение с пользователями,
- учёт сбоев, лицензий, ПО.

Спасибо!

Вопросы?...

serg@parallel.ru