

Московский государственный университет имени М.В.Ломоносова

Факультет Вычислительной Математики и Кибернетики

**СУПЕРКОМПЬЮТЕРЫ И
ПАРАЛЛЕЛЬНАЯ ОБРАБОТКА ДАННЫХ**

Жуматий С.А.

в.н.с. НИВЦ МГУ, к.ф.-м.н.

*ВМК МГУ, 2020
serg@parallel.ru*

Компоненты и системное программное обеспечение суперкомпьютеров

Компоненты

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

Инфраструктура: помещение

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

Инфраструктура: электричество

- Источники питания, проводка, автоматы
- ИБП, PDU, место
- Резервирование
- Расчётное время работы от батарей
- Проводка: фальшпол, коробка, организаторы
- Заземление, статика

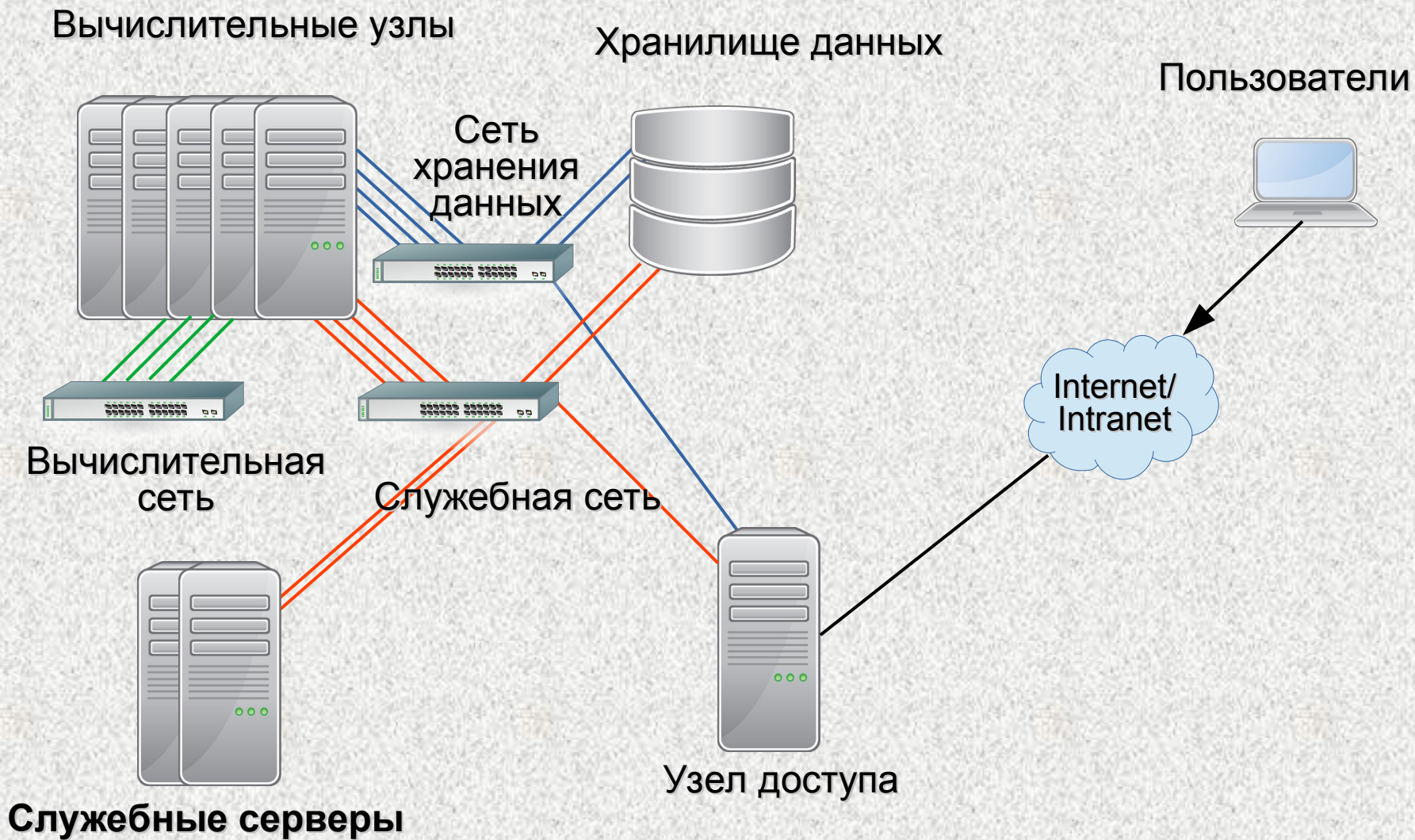
Инфраструктура: охлаждение

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

Структура кластера

- Сети: вычислительная, управляющая, хранения данных, сервисная, другие
- Вычислительные узлы
- Узлы доступа
- Сетевая файловая система
- Служебные узлы
- Система архивирования

Как устроен суперкомпьютер



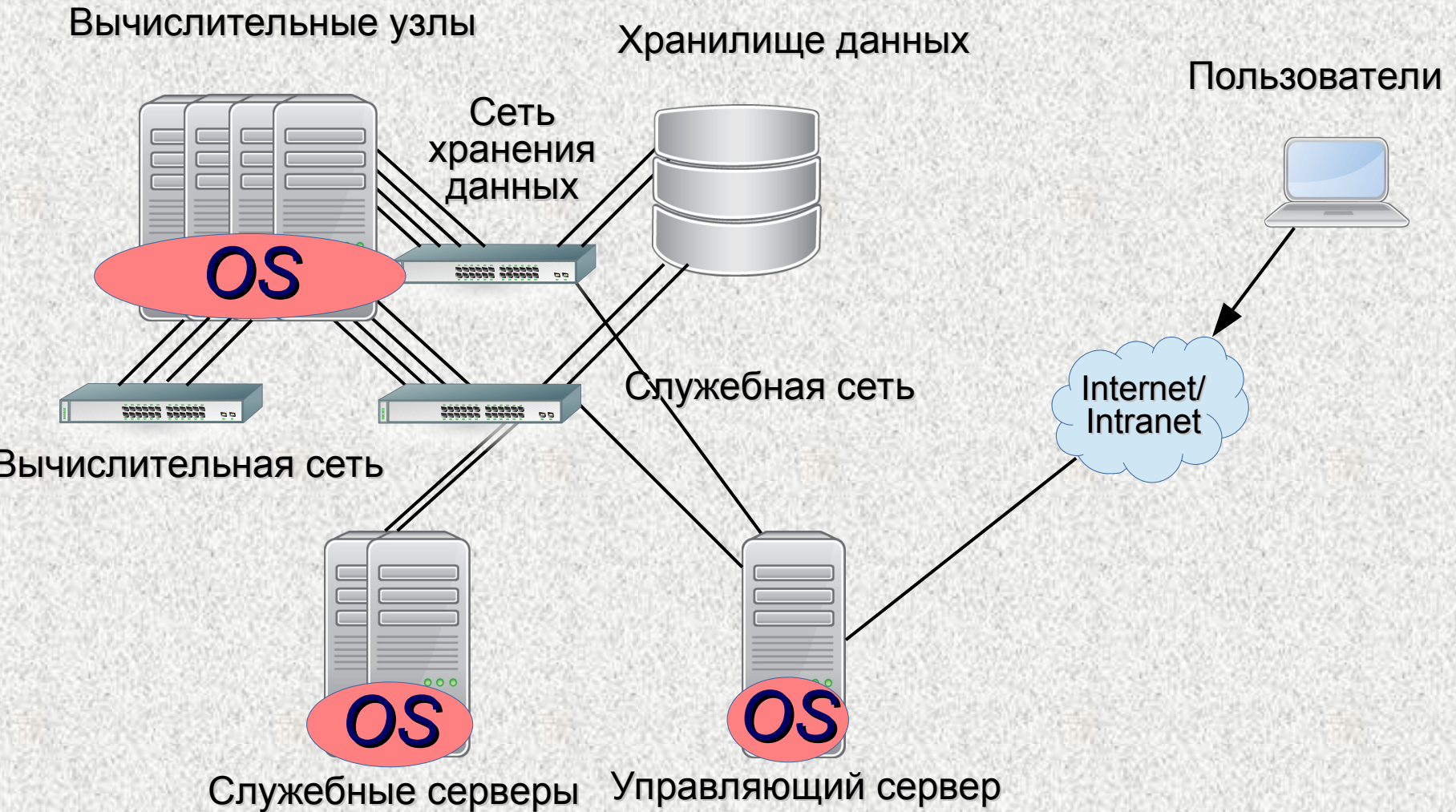
Служебные узлы (1)

- xCAT — управление образами т.п.
- DHCP — управление ip-адресами
- TFTP — загрузка, установка
- NFS — загрузка, root-fs
- DNS — управление dns-именами
- NTP — синхронизация времени
- LDAP — управление учётными записями

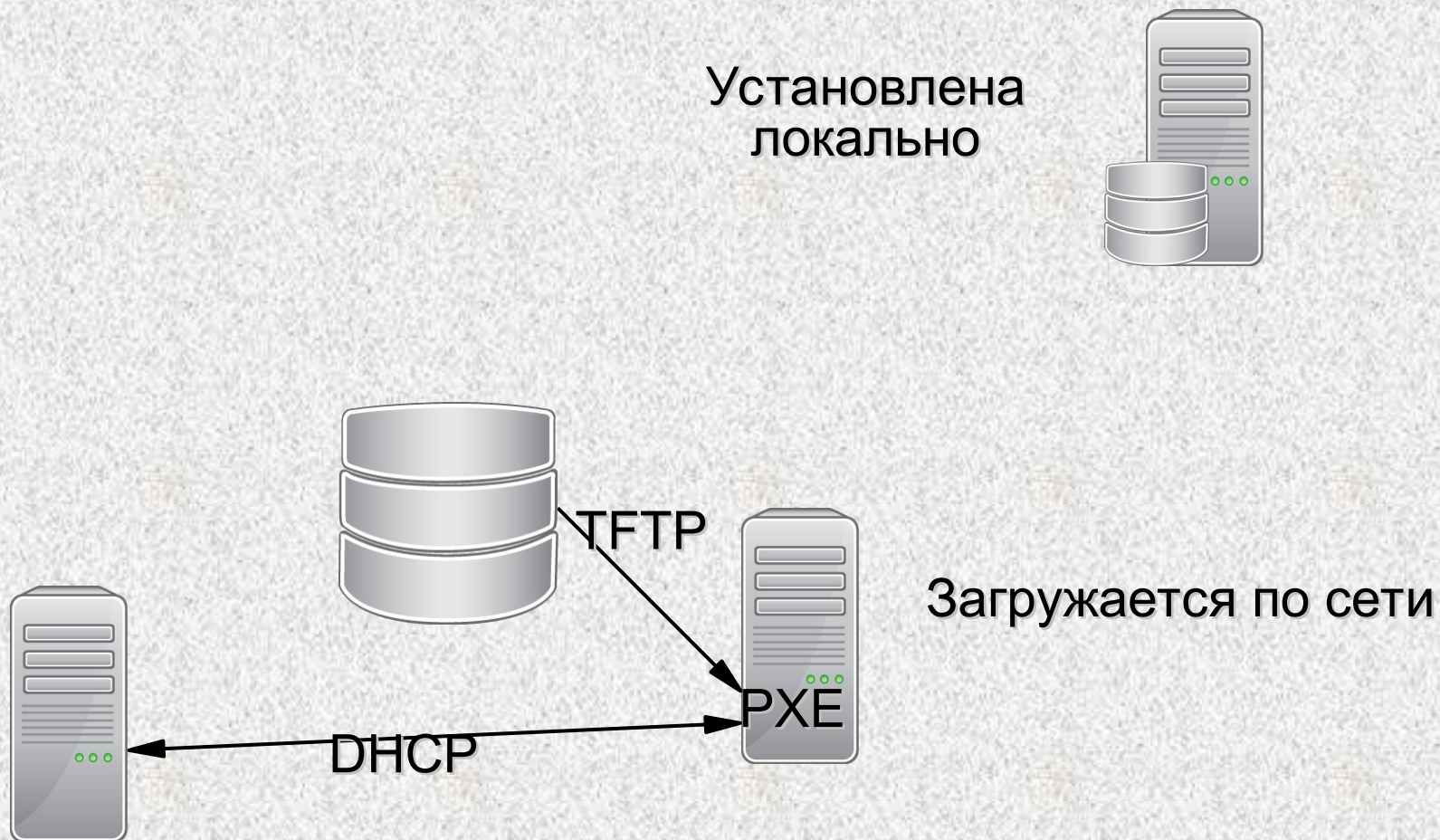
Служебные узлы (2)

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

Загрузка ОС



Операционная система



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
- Много места в памяти
- Нет возможности сохранить изменения

Установка ОС на диск

- «По одному», клонирование образа
- Автоматизированная установка
- Готовые стеки
- xCAT

Установка ОС на диск: автоматизированная

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 ОС на диск: готовые стеки

- ROCKS
- PelicanHPC / Parallel Knoppics

Установка ОС на диск: xCAT

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

XCAT

Виды установки ОС:

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



XCAT

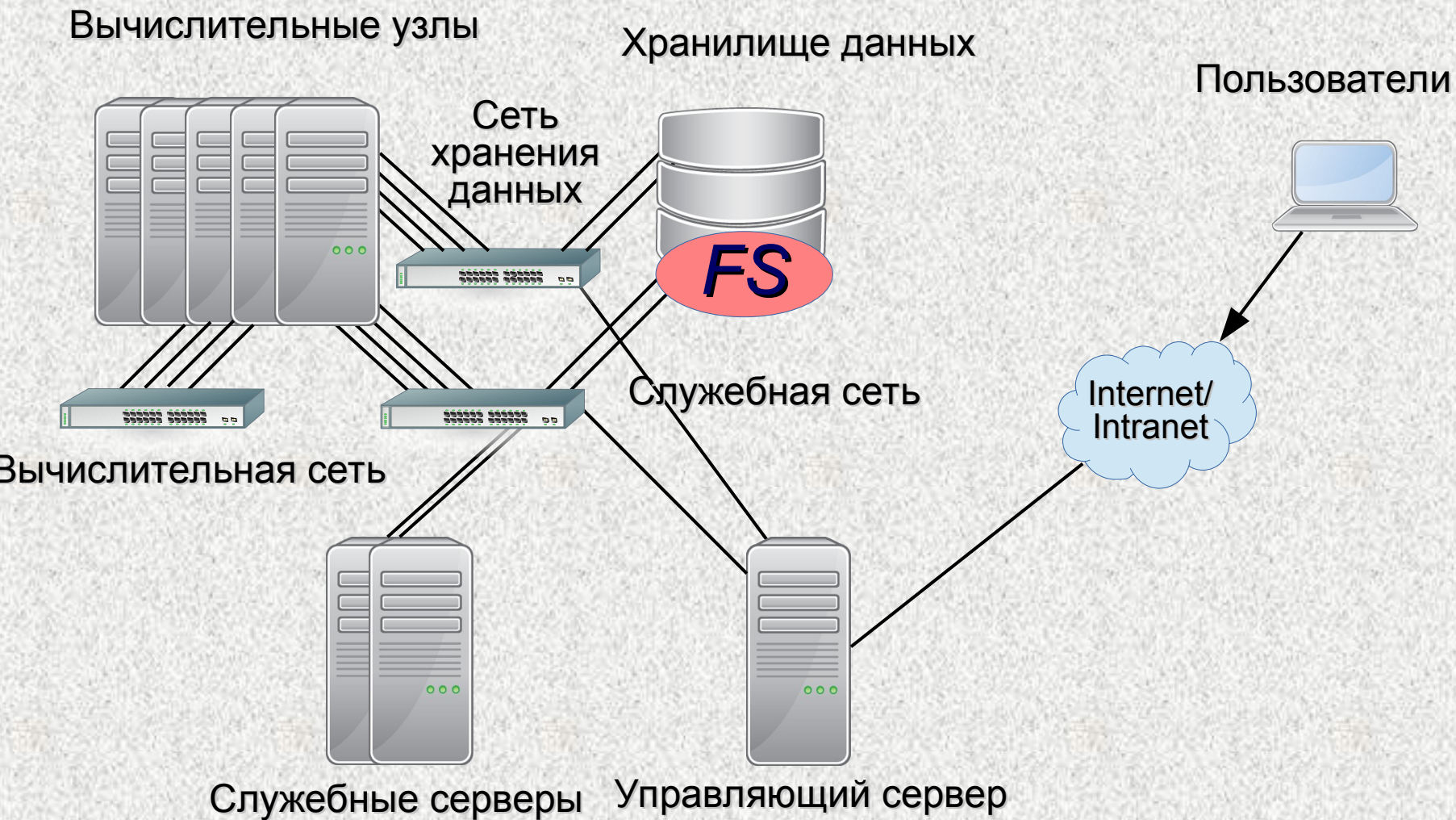
Объекты: site, **node**, group, **network**, osdistro, osimage, route...

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

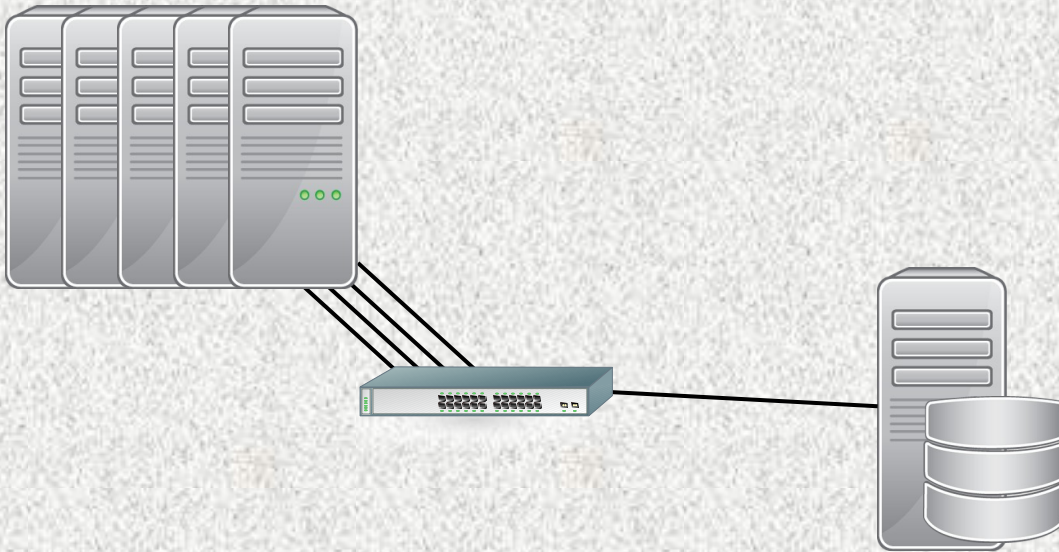
```
mkdef -t node node-1 groups=all,compute arch=x86_64 \  
    bmc=node-1-ipmi bmcusername=ADMIN bmcpassword=admin \  
    mac=xx:xx:xx:xx:xx:xx mgt=ipmi netboot=pxe \  
    provmethod=centos67-x86_64-statelite-compute
```

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

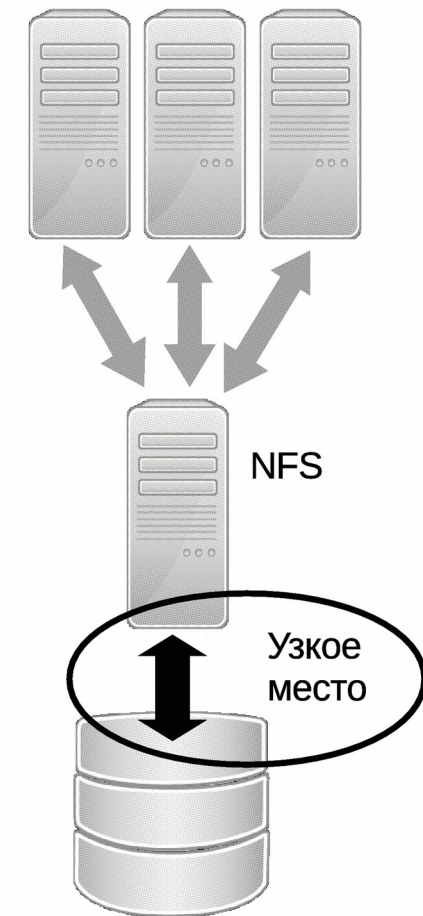
Файловые системы



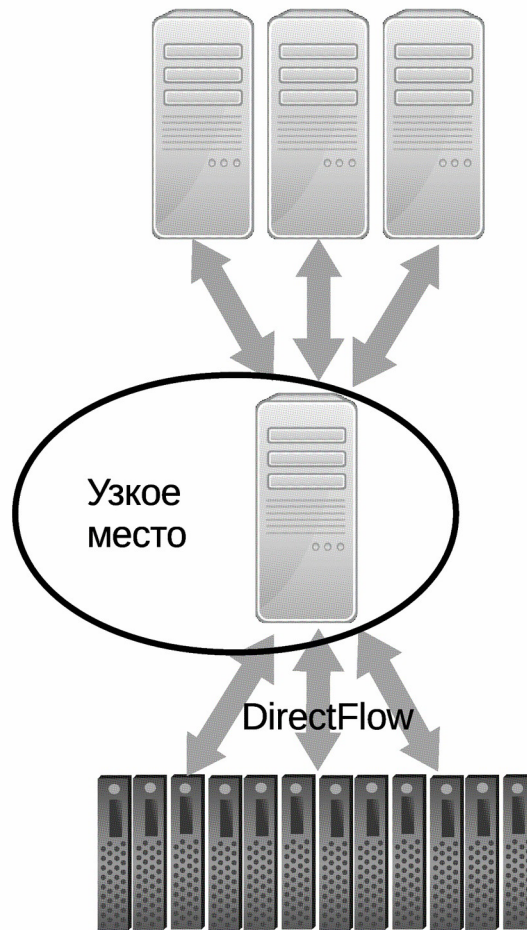
NFS



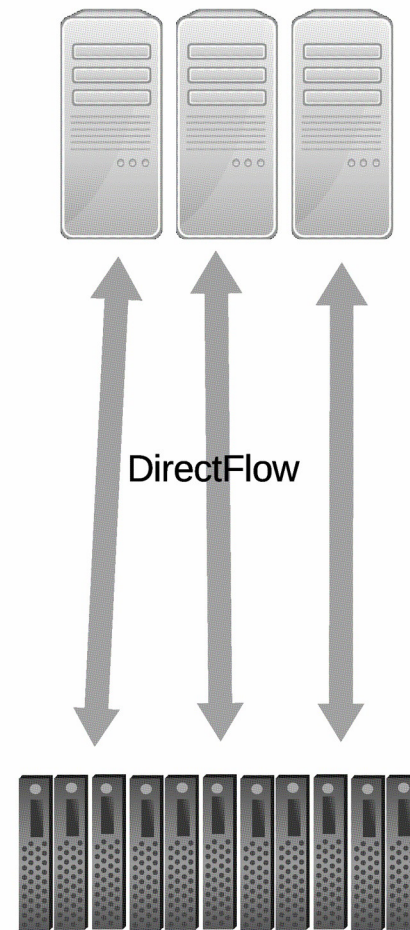
PanFS



Serial I/O with NFSv3



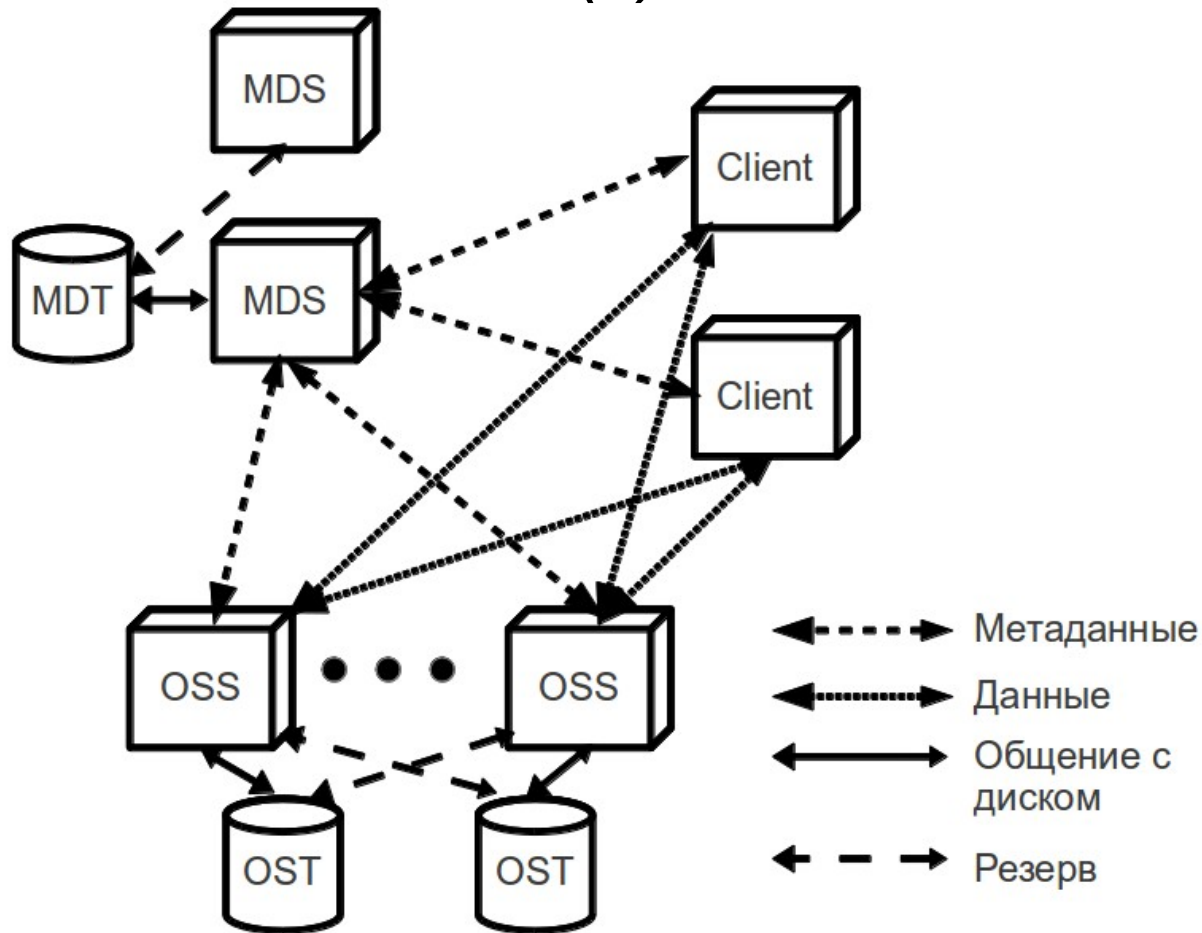
Serial I/O with Parallel



Parallel I/O with

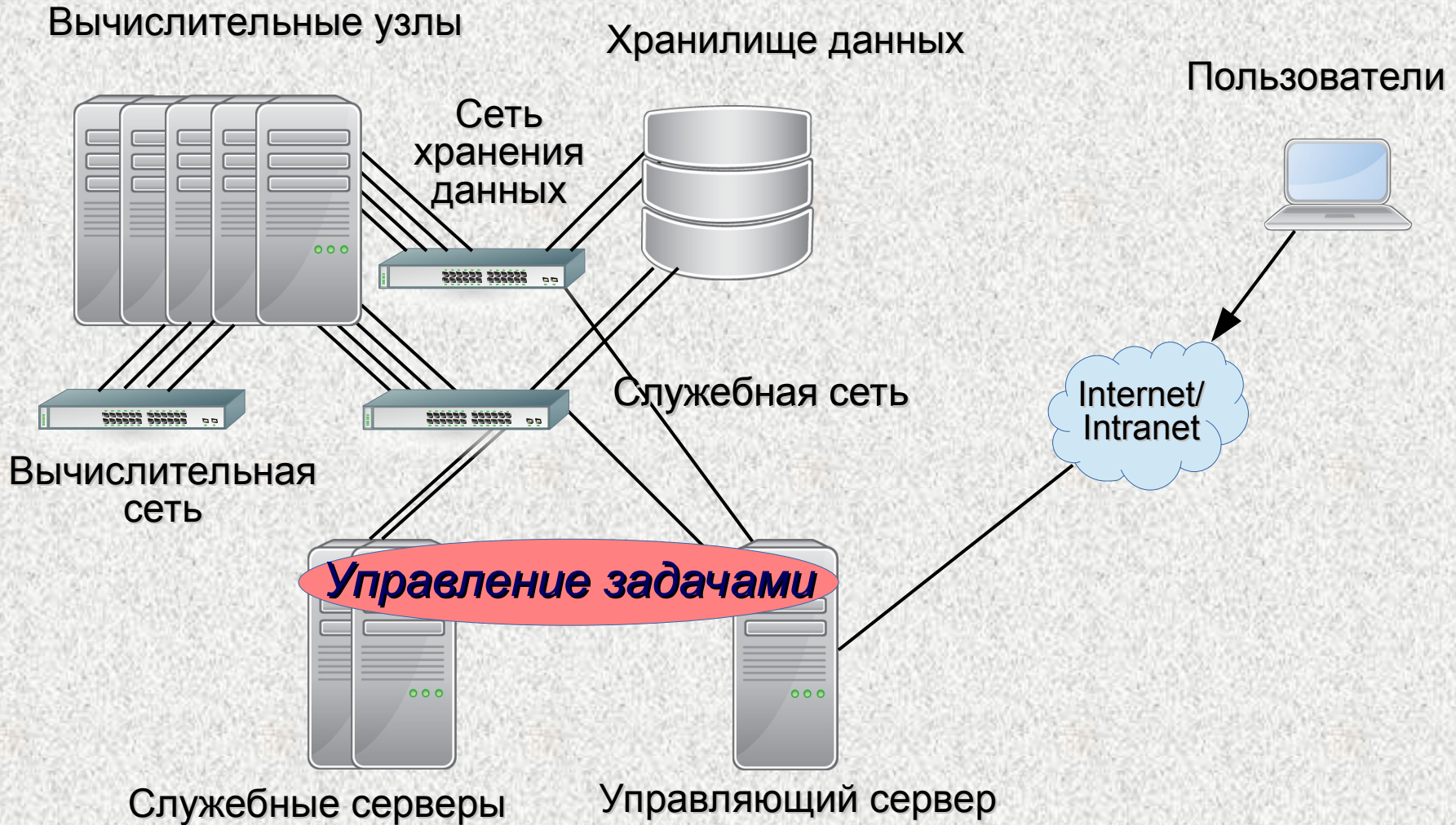
Lustre

Metadata Server(s)



ObjectStorage Servers

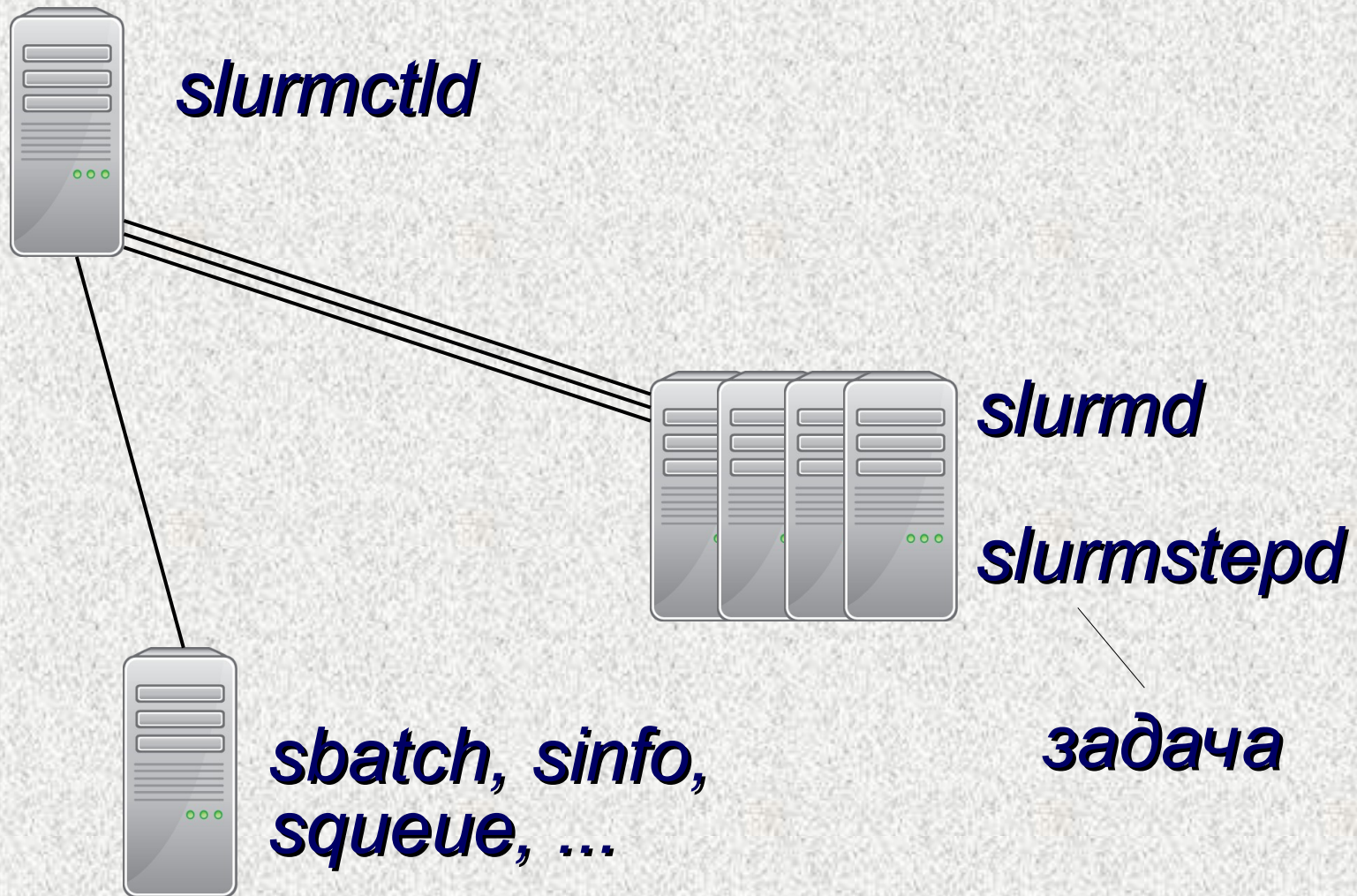
Контроль ресурсов



Управление задачами

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

Slurm



Slurm



slurmctld

/etc/slurm/slurm.conf

scontrol

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

Хранение учётных записей

- passwd
- nis+
- LDAP
- Гибридный метод

Хранение учётных записей

- **passwd**

root:x:0:0:root of evil:/root:/bin/bash

логин UID GID инфо дом шелл



- **shadow**

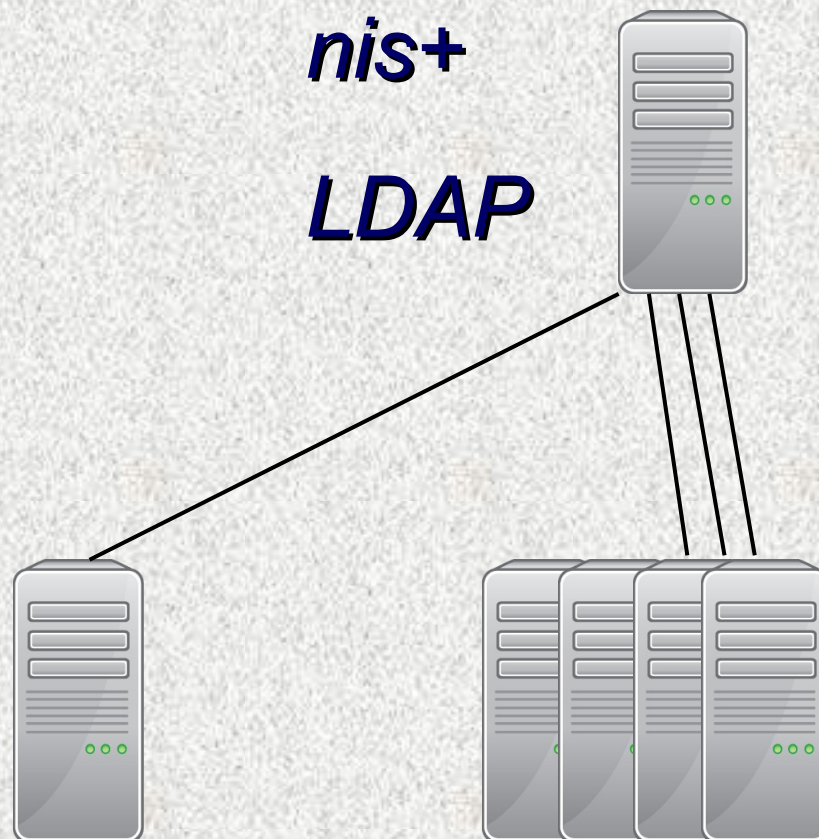
root:*:16219:0:99999:7::

логин пароль



Хранение учётных записей

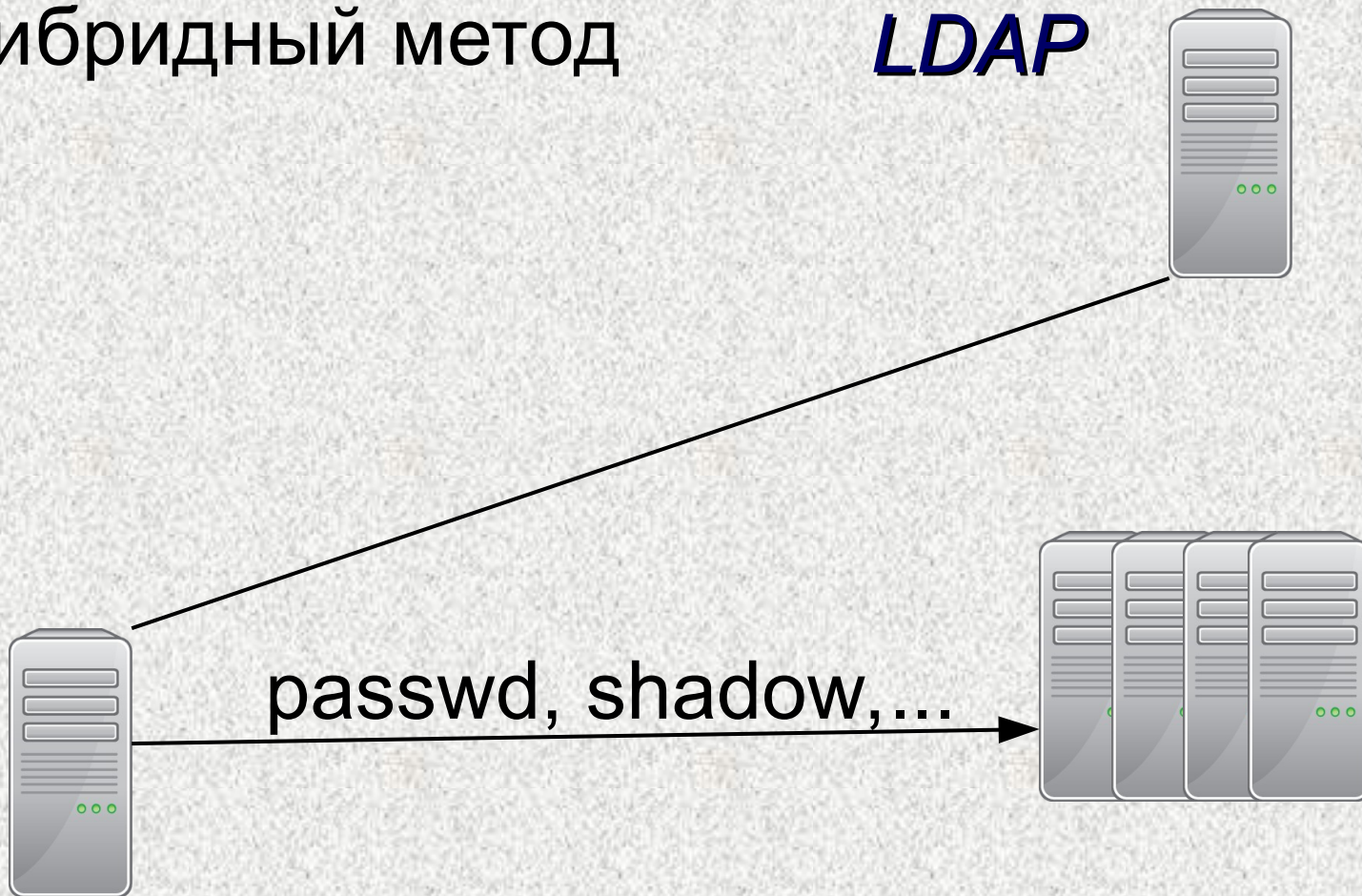
- LDAP



Хранение учётных записей

- Гибридный метод

LDAP



Лицензии

- Локальные и сетевые
- Привязка:
 - MAC-адрес
 - Серийный номер материнской платы
 - Файл
- Тип:
 - Список пользователей/серверов
 - Число пользователей (network)
 - Число одновременных запусков (floating)

Московский государственный университет имени М.В.Ломоносова

Факультет Вычислительной математики и кибернетики

**СУПЕРКОМПЬЮТЕРЫ И
ПАРАЛЛЕЛЬНАЯ ОБРАБОТКА ДАННЫХ**

Жуматий С.А.

в.н.с. НИВЦ МГУ, к.ф.-м.н.

*ВМК МГУ, 2020
serg@parallel.ru*