

# Учебный кластер из одноплатных компьютеров Raspberry Pi

Никольский И.М., Фурманов К.К.

GRID2016, Дубна

# Raspberry Pi 2 model B



<https://www.raspberrypi.org/>

# Raspberry Pi 2 model B

- A 900MHz quad-core ARM Cortex-A7 CPU
- 1GB RAM
- 4 USB ports
- 40 GPIO pins
- Full HDMI port
- Ethernet port
- Combined 3.5mm audio jack and composite video
- Camera interface (CSI)
- Display interface (DSI)
- Micro SD card slot
- VideoCore IV 3D graphics core

# Raspberry Pi - модели

- Raspberry Pi (model A, model B)
- Raspberry Pi 2 (model A, model B)
- Raspberry Pi 3(model A, model B)
- Raspberry Pi Zero

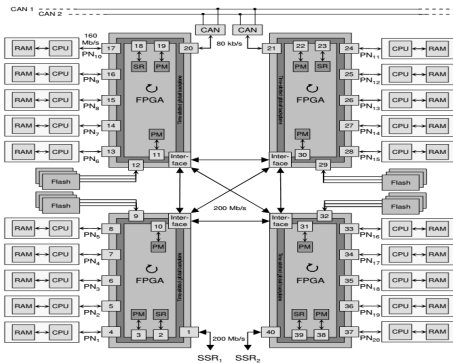
## Другие одноплатные компьютеры

- Banana Pi, Orange Pi, NanoPi
- Beaglebone, Odroid
- Samsung Exynos, Samsung Arktik
- Intel Edison, Intel Galileo

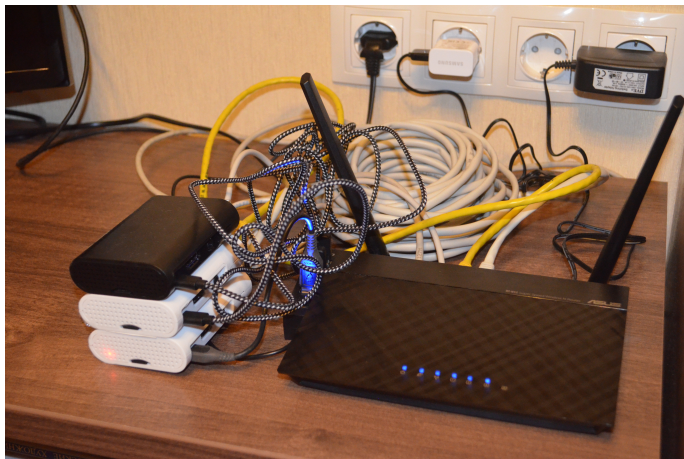
# Кластеры из Raspberry Pi

- Iridis-Pi (университет Саутгемптона) - 64 узла
- Resin.io - 128 узлов
- любительские кластеры

# Встраиваемые кластеры (embedded clusters)



# Учебный кластер





# Компоненты кластера

- Raspberry Pi 2 model B × 3
- Корпуса
- WiFi-роутер,
- USB-хаб,
- Блок питания - 5В, 2А,
- Патч-корды, MiniUSB(питание)

# Простейшее ПО для распр вычислений

- xargs - стандартная утилита Линукс, позволяет запускать параллельно несколько процессов
- GNU Parallel - аналог, процессы можно запускать на нескольких узлах

## GNU Parallel, сложение векторов

```
q=$(seq 0 190 | tr "\n" " ")
r=$(seq 310 500 | tr "\n" " ")
parallel --xapply "echo {1}+{2}|bc -l" ::: $q ::: $r
```

```
parallel --sshloginfile nodes.txt ...
```

nodes.txt:

```
2/ pi@192.168.100.1
```

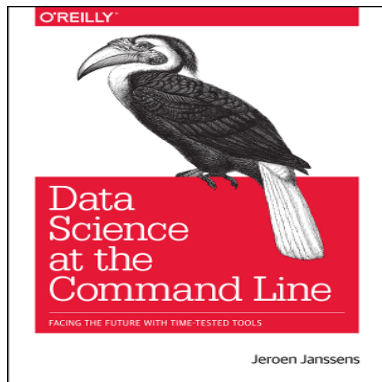
```
2/ pi@192.168.100.2
```

# GNU Parallel, вычисление $\pi$

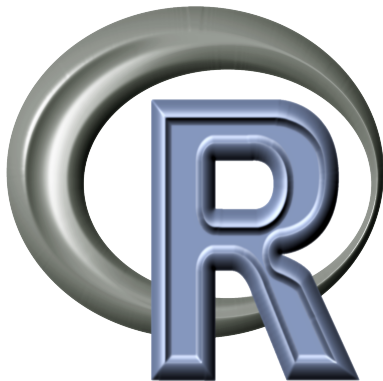
```
n_proc=2
num_per_proc=8
num_rnd=$((2*n_proc*num_per_proc))

od -N$num_rnd -An -d /dev/hwrnd | \
./parallel --jobs $n_proc ./dist.sh | \
tr "\\n" " " | xargs ./pi_script.sh
```

# Data Science at the Command Line (O'Reilly, September 2014)



# Среда стат вычислений R



<https://www.r-project.org/>

# Среда стат вычислений R

- Свободно распространяемый
- Добавление новых возможностей с помощью пакетов
- Элементы функционального программирования (apply, ..)

# Возможности R в области распр вычислений

Пакеты snow, parallel, foreach

Основная схема

```
cl = makeCluster(список хостов, мастер)
parApply(cl, набор данных, имя функции)
stopCluster(cl)
```



# Тестовая задача

Определение доверительных интервалов прогнозов регрессии методом бутстрап

*Bradley Efron Bootstrap Methods: Another Look at the Jackknife // Annals of Statistics. — 1979. — Vol. 7, no. 1., pp 1–26.*

# Тестовый набор данных

- Набор данных abalone из UCI Machine Learning Repository.
- данные о моллюсках галиотис (4177 экземпляров)
- 9 параметров: пол, размеры панциря, масса и т.д.

# Регрессионная модель

$$\log(N_r) \sim s + \log(S_m) + w_s + w_{sh},$$

где

- $N_r$  - число колец на раковине,
- $s$  - пол,
- $S_m = (l * d * h)^{\frac{1}{3}}$ ,
- $l, d, h$  - размеры раковины,
- $w_s$  - вес тела моллюска,
- $w_{sh}$  - вес раковины.

# Метод бутстрап - основная идея

## Задача

- Пусть надо определить доверит интервал для некоторого параметра распределения.
- У нас нет генеральной совокупности, только одна выборка.

## Решение

- Мы имитируем выбор из генеральной совокупности, делая выборки из имеющейся выборки (повторные выборки, resamples).
- Производится  $n_{boot}$  повторных выборок с возвращением.
- По каждой получаем оценку нужного параметра - всего  $n_{boot}$  оценок параметра.
- По этому набору оценок строим доверительный интервал.

Параметры программы:

- `smp_len` - количество наблюдений
- `nboot` - количество перевыборок
- `alpha` - уровень доверия

Код:

```
host_ar = c('localhost', '192.168.1.101', '192.168.1.102')  
  
cl=makeCluster(host_ar)  
  
parLapply(cl, 1:nod_num,  
          function(i) return(Boot(nboot, alpha, smp_len)))  
  
stopCluster(cl)
```

## Результаты вычислительных экспериментов

nboot	sample.len	параллЗузла	послед
1000	50	42.099	27.527
1000	80	41.708	27.926
1000	200	45.225	29.534
1500	200	43.884	67.292
2000	200	89.659	59.166
1500	200		67.605
1300	200	58.901	

# Кластеры из одноплатных компьютеров

- + Низкая цена, низкое энергопотребление, компактность
- Невысокая производительность

Хороши для образовательных целей

Но! Parallela Board (фирма Adapteva, [www.adapteva.com](http://www.adapteva.com)) 16 ядер - 32 Гфлопс, 64 ядра - 102 ГФлопс

# Результаты

- Построен кластер из Raspberry Pi
- Предложены варианты начального обучения распределенным вычислениям



Спасибо за внимание!

- *I. V. McLoughlin, T. R. Bretschneider, Chen Zheming* Virtualized Development and Testing for Embedded Cluster Computing *International Journal of Networking and Computing* Volume 2, Number 2, pages 160–187, July 20A
- *Simon J. Cox, James T. Cox, Richard P. Boardman, Steven J. Johnston, Mark Scott, Neil S. O'Brien* Iridis-pi: a low-cost, compact demonstration cluster *Cluster Computing*, June 2013, Volume 17, Issue 2, pp 349-358