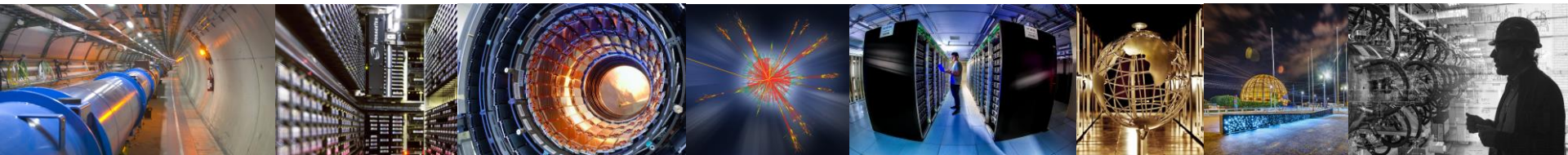




Разпределени изчисления в ATLAS

Част 2: !GRID

(суперкомпютри, облаци, доброволчески изчисления, графични карти)



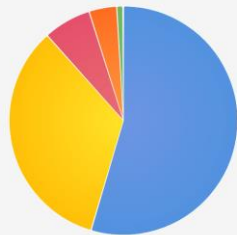
*Др. Иван Глушков
Тексаски университет / АТЛАС
Българска инженерна учителска програма
ЦЕРН, септември 2022*

Видове ресурси

384к

GRID

(HS06)
55% - GRID
35% - Суперкомпютри
7% - Облаци
1% - Доброволчески
0% - GPU



213к

Суперкомпютри

4

GPU

63к

Облаци

32к

Доброволчески

Разликата

HTC: High-Throughput Computing (GRID)

- Какво е?
 - Голяма изчислителна мощност за дълго време.
- За какво служи?
 - Ефективното изпълнение на много и слабо свързани задачи.

HPC: High-Performance Computing (Суперкомпютри)

- Какво е?
 - Голяма изчислителна мощност за ограничено време.
- За какво служи?
 - Ефективно изпълнение на много, тясно свързани задачи.

Облачни изчисления

- Какво е?
 - Виртуални машини - колкото и каквито са нужни
- За какво служи?
 - Гарантирани ресурси в сигурна и изолирана среда без нужда от поддръжка на хардуер. API достъп

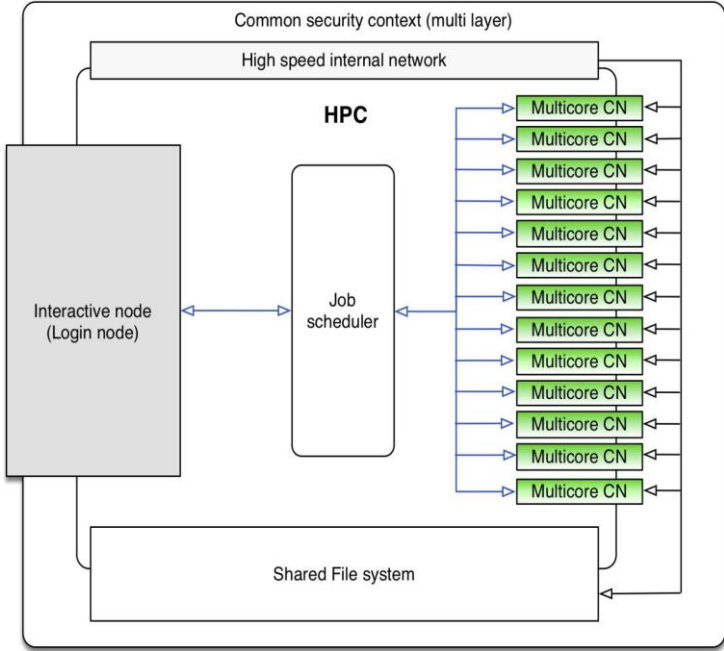
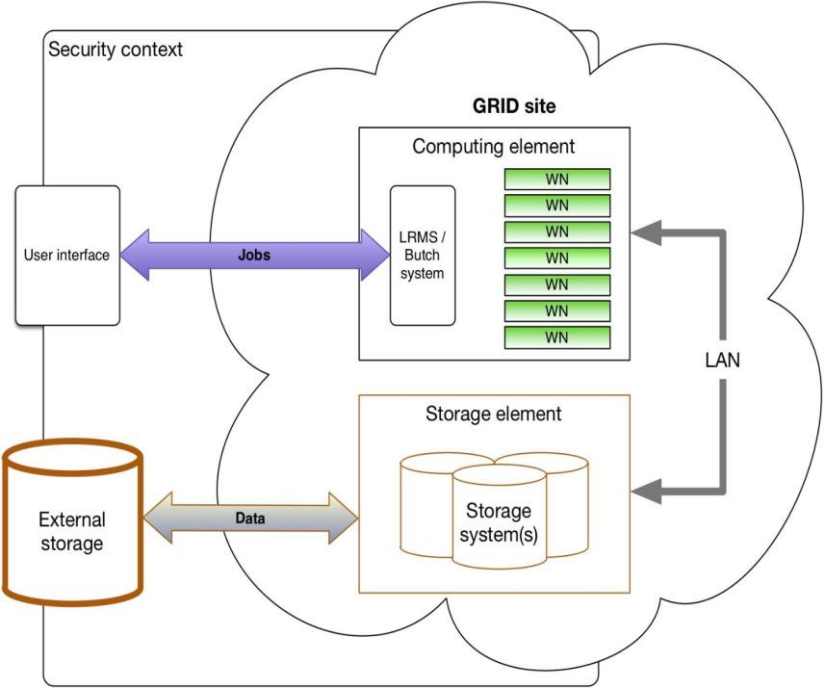
Доброволчески изчисления

- Какво е?
 - Централизиран изчисления на ресурси предоставени от доброволци
- За какво служи?
 - Оползотворяване на ресурси които иначе биха били прахосани



Суперкомпютри

GRID и Суперкомпьютри



GRID и Суперкомпютри



- GRID се състои от компютърни клъстери
- По-голяма част от изчисленията в другите области на науката стават в суперкомпютри.
- Разлики: суперкомпютри <> GRID:
 - Нишки: много (~100 000 нишки) <> една
 - Работа на отделния компютър: част от задача <> една задача
 - Входно - изходни операции: малко <> много
 - Брой на файлове: малко <> много
 - Потребителска идентификация: логин/парола (еднократна!) <> сертификат
 - Процесори / операционни системи: много <> няколко
 - Директна връзка с интернет: не <> да

Мащаб на задачите на един суперкомпютър

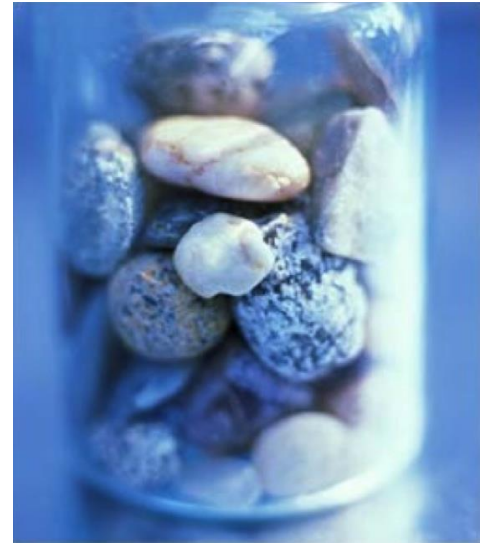
(на днешните машини)

Вид задача	Минимален брой компютри	Максимален брой Компютри	Максимална дължина на задачата	Приоритет
1	11,250	—	24.0	15
2	3,750	11,249	24.0	5
3	313	3,749	12.0	0
4	126	312	6.0	0
5	1	125	2.0	0

Два начина на ползване на суперкомпютър: квота и backfill

HPC: Backfill

HPC: Backfill



Источник: David Cameron

Суперкомпютри: Backfill



Източник: David Cameron

Суперкомпютри: Backfill

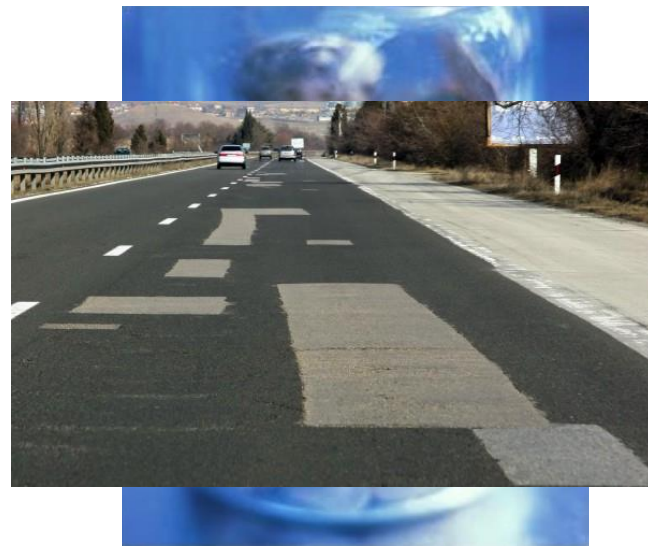
(пълним дупки всякакви)



Източник: David Cameron

Суперкомпютри: Backfill

(пълним дупки всякакви)



Източник: David Cameron

Суперкомпютри: Backfill

(пълним дупки всякакви)

Размер на дупките

- 400M CPU*часа на година (Titan)
- Обичайна ефективност на използване на суперкомпютър във времето - 90%.
- 10% == 400M CPU*часа на година (Titan)
- Всяка неефективност е нежелана!

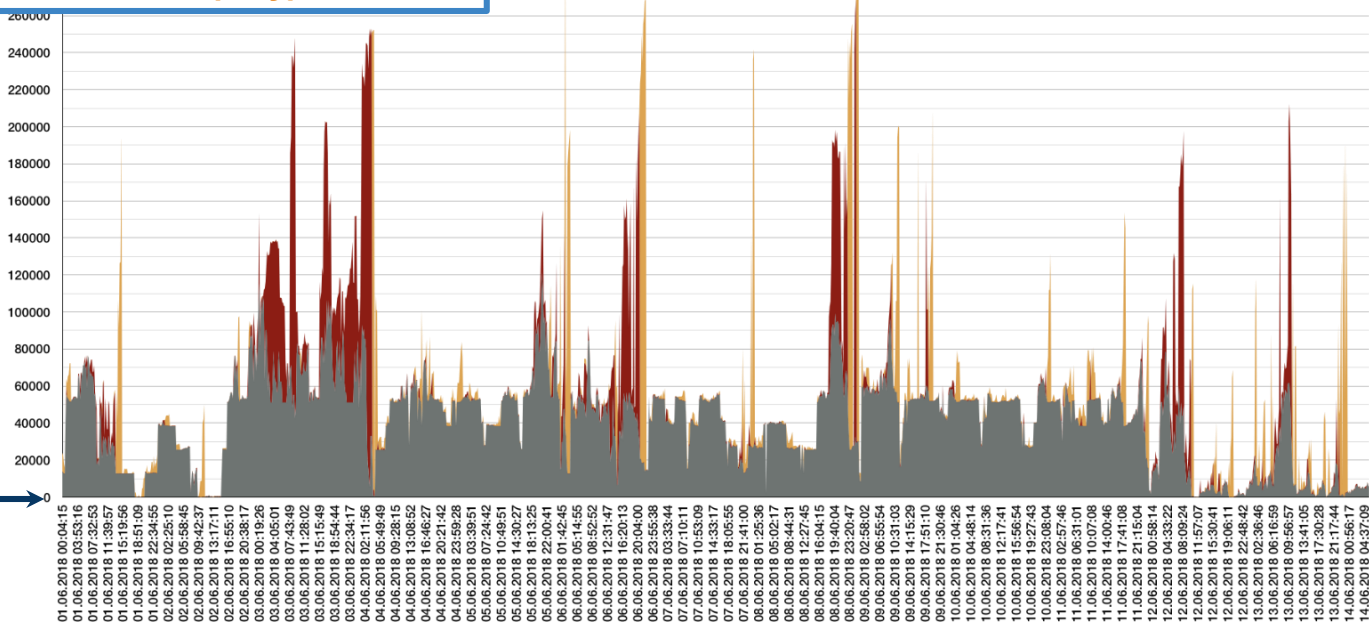
Запълване

- С какво?
 - Много, с различен размер, но малки задачи (т.е. - ние)
- На каква цена?
 - Безплатно



Използване на Titan от ATLAS

- Изчисления по квота
- Backfill
- Свободни ресурси

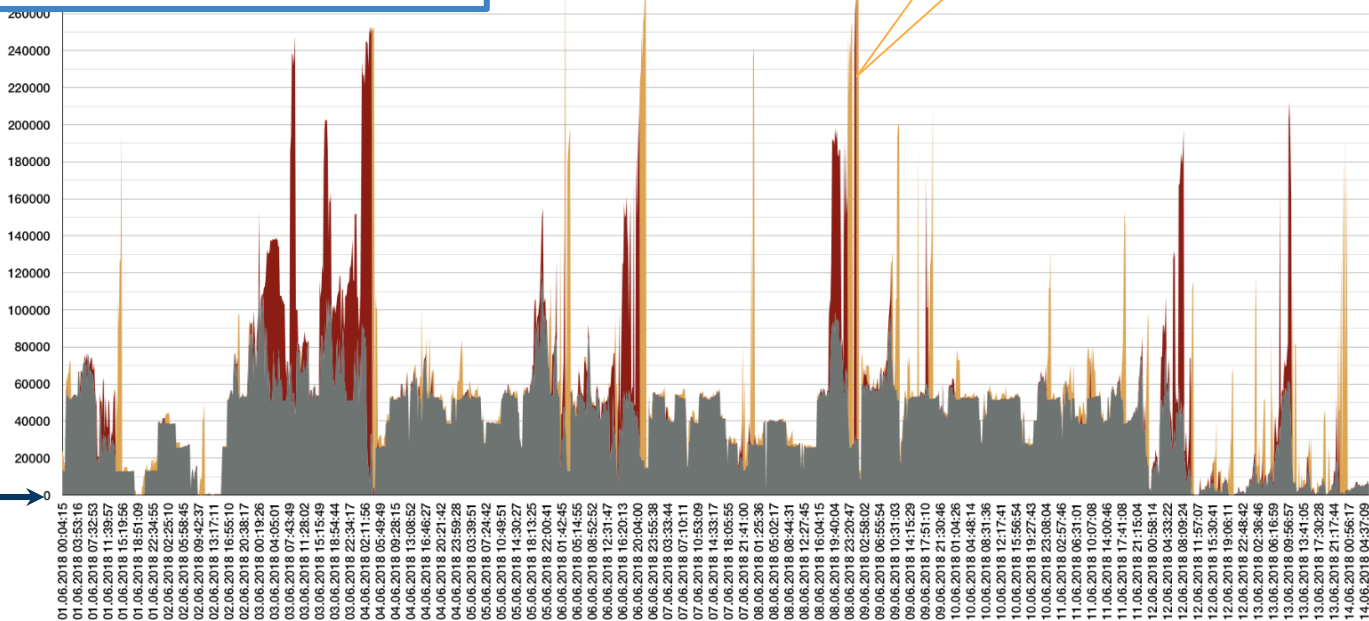


две седмици

Използване на Titan от ATLAS

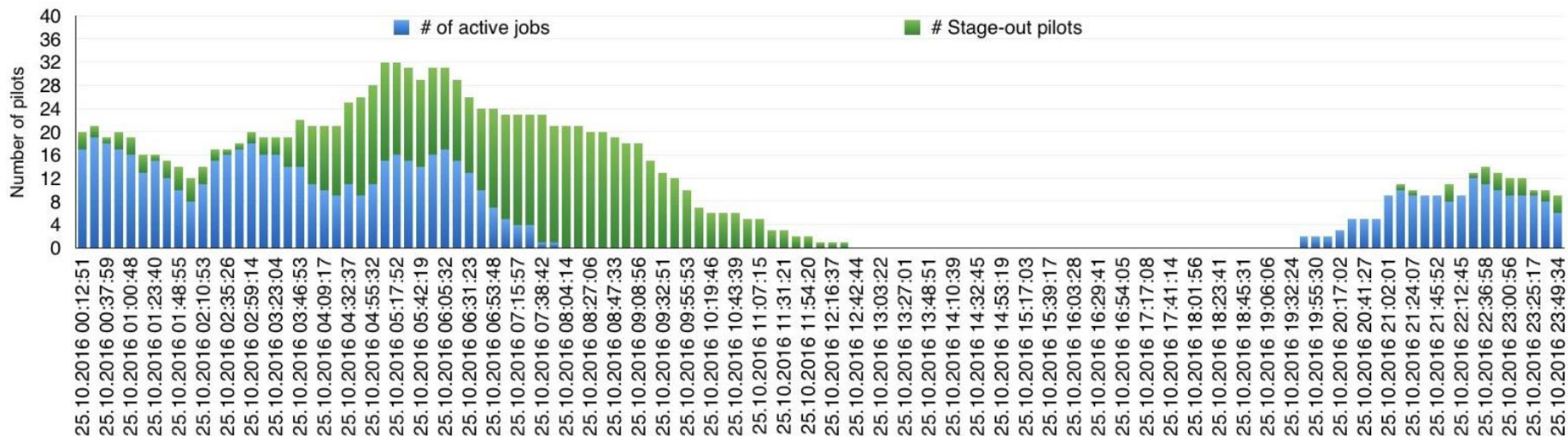
- Изчисления по квота
- Backfill
- Свободни ресурси

ЗАЩО?!



две седмици

Проблемът



Американски суперкомпютри използвани от АТЛАС



[Cori](#) (САЩ, [top500: 41](#))
Cray/HPE
Intel Xeon Phi 7250 68C
1.4GHz
622336 ядра
Производителност: 14
PFLOP/сек



[Theta](#) (САЩ)
Cray XC40
Intel Phi 7230 64C 1.3GHz
280320 ядра / 914 TB памет / 11 PB диск
Производителност: 11.66 PFLOP/сек



[Edison](#) (САЩ)
Cray XC30
Intel Xeon E5-2695v2 121C 2.4GHz
133824 ядра / 357 TB памет / 7.56 PB диск
Производителност: 2.57 PFLOP/сек

.....

[Titan](#) (САЩ)
Cray XK7
Opteron 6274 16C 2.2 GHz
560640 ядра / 694 TB памет / 40 PB
диск
Производителност: 27.11
PFLOP/сек

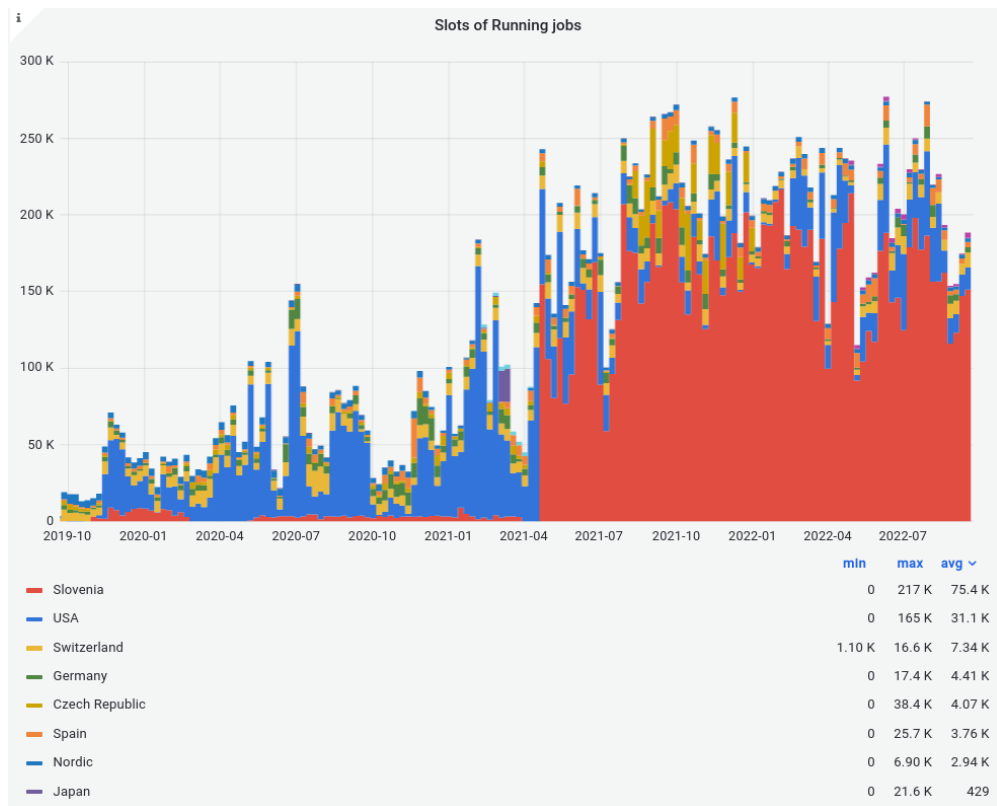


Еднаквото между всички..

Няма такава..

И така до май 2021..

Vega



Vega е:

- Суперкомпютър / HPC
- Открит на 20.04.2021
- В Словения
- Финансиран от EuroHPC и Словения
- На [131-то](#) място в [top500](#)
- Atos/BullSequana XH2000
- AMD EPYC 7H12 2.6GHz
- 122800 ядра
- Производителност: 375 PFLOP/сек
- Не се използва напълно от потребителите
- ATLAS използва ресурсите в режим “backfill”

EuroHPC JU

(European High-Performance Computing Joint Undertaking / Европейска HPC инициатива)

- Инициатива между държавни и частни институции за консолидиране на ресурси на европейско ниво за развитие на HPC
- Членове: ЕС, Австрия, Белгия, България, Хърватска, Кипър, Чехия, Дания, Естония, Финландия, Франция, Германия, Гърция, Унгария, Исландия, Ирландия, Италия, Латвия, Литва, Люксембург, Малта, Холандия, Северна Македония, Норвегия, Полша, Португалия, Румъния, Сърбия, Словакия, Словения, Испания, Швеция, Турция, представители на 3 частни партньори
- Цел: Разработка на пан-европейска HPC инфраструктура
- Централна: Люксембург
- Начало: 2018
- Финансиране: 7 млрд. Евро за периода 2021-2027
 - 3 млрд. - дългосрочен бюджет на ЕС
 - 3 млрд. - членове на инициативата
 - 1 млрд. - частни донори



Суперкомпютрите на EuroHPC JU



[LUMI](#) (Финландия, [top500: 3](#))
Hewlett Packard/Cray EX
AMD EPYC 64C 2GHz
1110144 ядра
Производителност: 375 PFLOP/сек



[MELUXINA](#) (Люксембург, [top500: 48](#))
Atos/BullSequana XH2000
AMD EPYC 7452 32C 2.35GHz
172544 ядра
Производителност: 13 PFLOP/сек



[LEONARDO](#) (Италия, в строеж)
Atos/BullSequana XH2000
Производителност: 250 PFLOP/сек



[KAROLINA](#) (Чехия, [top500: 79](#))
Hewlett Packard
AMD EPYC 7452 32C 2.45GHz
102272 ядра
Производителност: 10 PFLOP/сек



[MARENOSTRUM 5](#) (ще бъде в Испания)
Bull SAS/Bull Sequana XH3000
Intel Sapphire Rapid
Производителност: 205 PFLOP/сек



[DISCOVERER](#) (България, [top500: 113](#))
Atos/BullSequana XH2000
AMD EPYC 7H12 64C 2.6GHz
144384 ядра
Производителност: 5 PFLOP/сек



[VEGA](#) (Словения, [top500: 131](#))
Atos/BullSequana XH2000
AMD EPYC 7H12 2.6GHz
122800 ядра
Производителност: 7 PFLOP/сек



[DEUCALION](#) (Португалия, в строеж)
Fujitsu/PRIMEHPC+BullSequana
A64FX, AMD EPYC
Производителност: 7 PFLOP/сек

Надпреварата

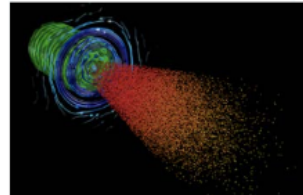
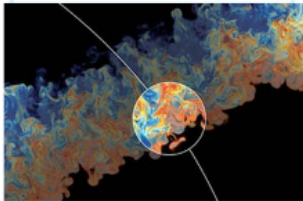
- **Целта:** Пускане в експлоатация на компютър който може да направи 10^{18} изчисления за секунда (exaFLOP)
 - Целта е постигната през юни 2022 в САЩ!
- **Статус:** [top500.org](https://www.top500.org/)
- **Тест:** [Linpack Benchmark](https://www.netlib.org/linpack/)
 - Задача: Решаване на система от линейни уравнения
 - Класиране: Само за линейни уравнения
 - Резултати:
 - R_{peak} : Теоретична максимална производителност
 - R_{max} : Измерена производителност
- **Резултати:**
 - Top10: Само 2 машини в Европа
 - Top500:
 - Всички машини в таблицата са с производителност > 1 petaFLOP
 - Повечето машини са в САЩ
 - Нито една машина не е в България
- **Защо е необходимо?**

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	Frontier - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 26Hz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	LUMI - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 26Hz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	1,110,144	151.90	214.35	2,942
4	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	200.79	10,096
5	Sierra - IBM Power System AC922, IBM POWER9 22C 3.16GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438
6	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93.01	125.44	15,371
7	Perlmutter - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70.87	93.75	2,589
8	Setene - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63.46	79.22	2,646
9	Tianhe-2A - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, NUDT National Super Computer Center in Guangzhou China	4,981,760	61.44	100.68	18,482
10	Adastra - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 26Hz, AMD Instinct MI250X, Slingshot-11, HPE Grand Equipement National de Calcul Intensif - Centre Informatique National de l'Enseignement Supieur (GENCI-CINES) France	319,072	46.10	61.61	921

ECP applications target national problems in six areas

National security

- Next-generation, stockpile stewardship codes
- Reentry-vehicle-environment simulation
- Multi-physics science simulations of high-energy density physics conditions



Energy security

- Turbine wind plant efficiency
- Design and commercialization of SMRs
- Nuclear fission and fusion reactor materials design
- Subsurface use for carbon capture, petroleum extraction, waste disposal
- High-efficiency, low-emission combustion engine and gas turbine design
- Scale up of clean fossil fuel combustion
- Biofuel catalyst design

Economic security

- Additive manufacturing of qualifiable metal parts
- Urban planning
- Reliable and efficient planning of the power grid
- Seismic hazard risk assessment



Scientific discovery

- Cosmological probe of the standard model of particle physics
- Validate fundamental laws of nature
- Plasma wakefield accelerator design
- Light source-enabled analysis of protein and molecular structure and design
- Find, predict, and control materials and properties
- Predict and control stable ITER operational performance
- Demystify origin of chemical elements

Earth system

- Accurate regional impact assessments in Earth system models
- Stress-resistant crop analysis and catalytic conversion of biomass-derived alcohols
- Metagenomics for analysis of biogeochemical cycles, climate change, environmental remediation

Health care

- Accelerate and translate cancer research



Взето от [презентацията](#) на Lori Diachin, зам-директор на “Exascale Computing Project”, от конференцията “Supercomputing 18”, 14.11.2018

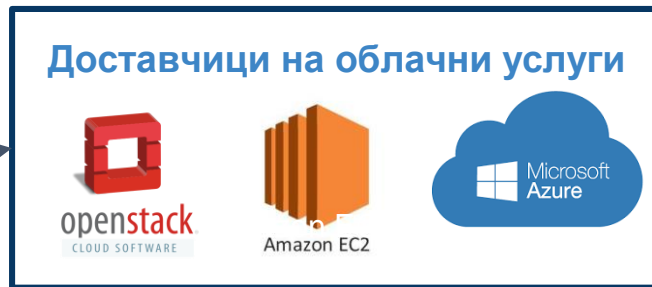
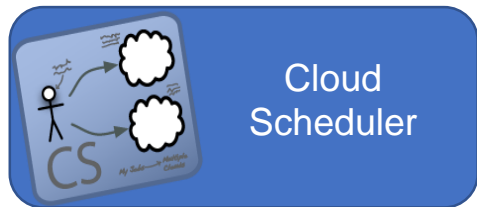


Облаци

На кратко

- ATLAS използва собствени, научни и комерсиални.
- Предимства
 - Комерсиални: При нужда от много ресурси за кратко време
 - Научни: Споделяне на ресурси и цена между различни институти и различни области на изследване
 - Собствени: Бързо и контролирано усвояване на ресурси
- Недостатъци
 - Комерсиални: Скъпи

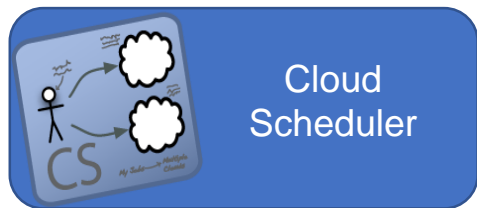
Принцип на работа



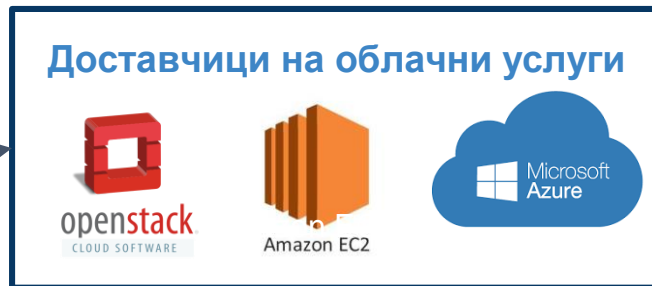
Изпраща
задачи



Принцип на работа



Направи ми виртуални
машини

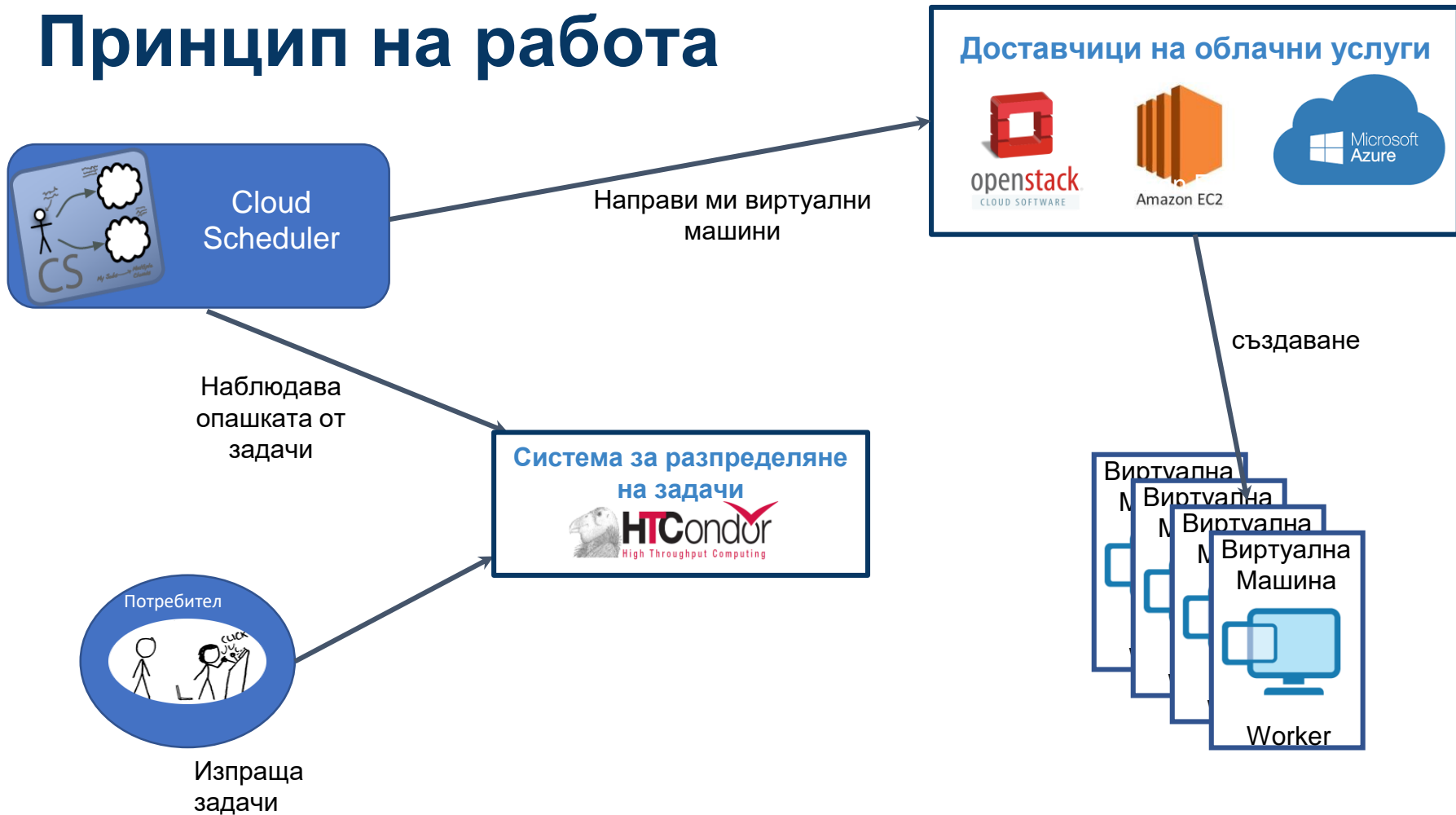


Наблюдава
опашката от
задачи

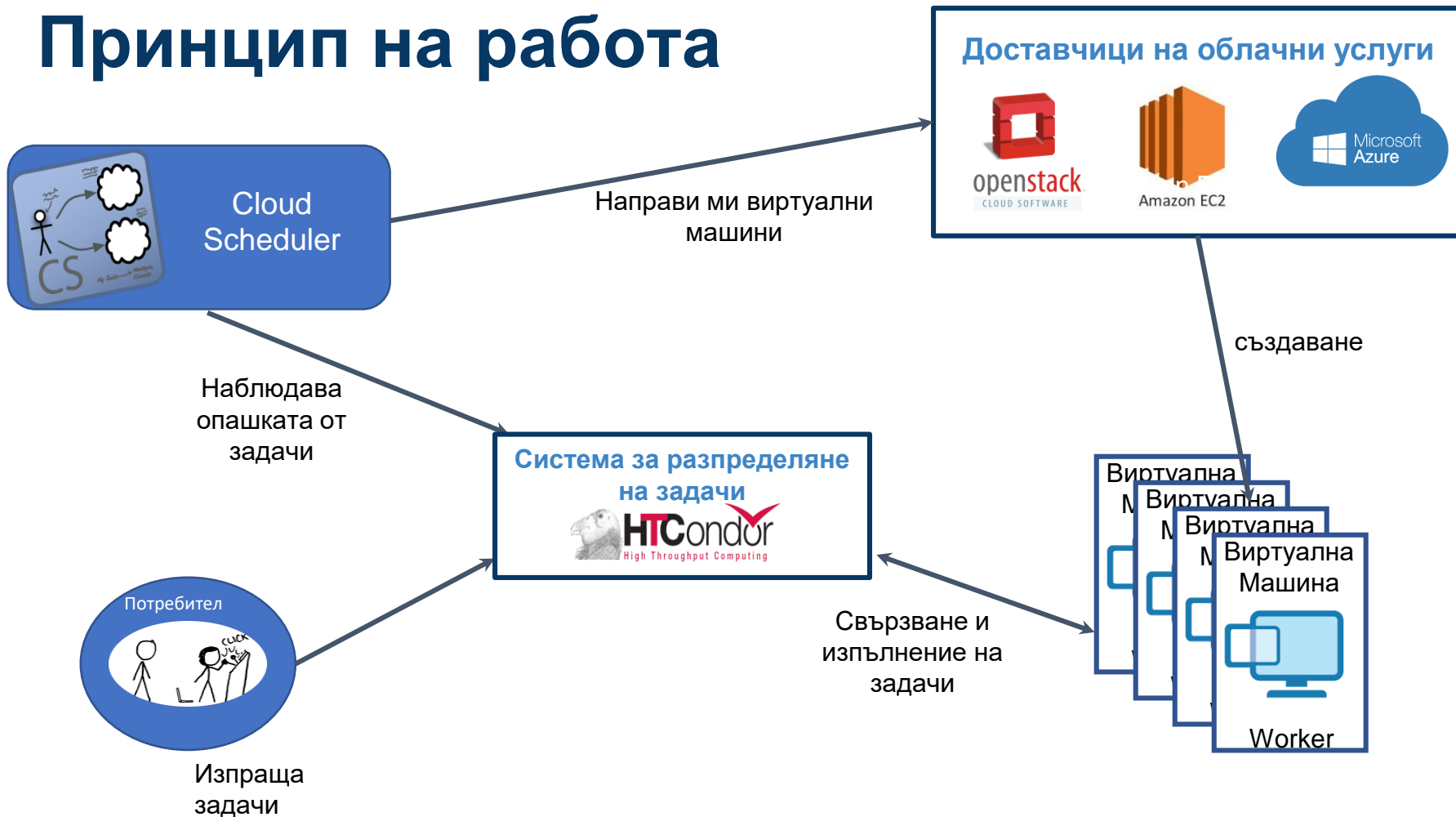


Изпраца
задачи

Принцип на работа

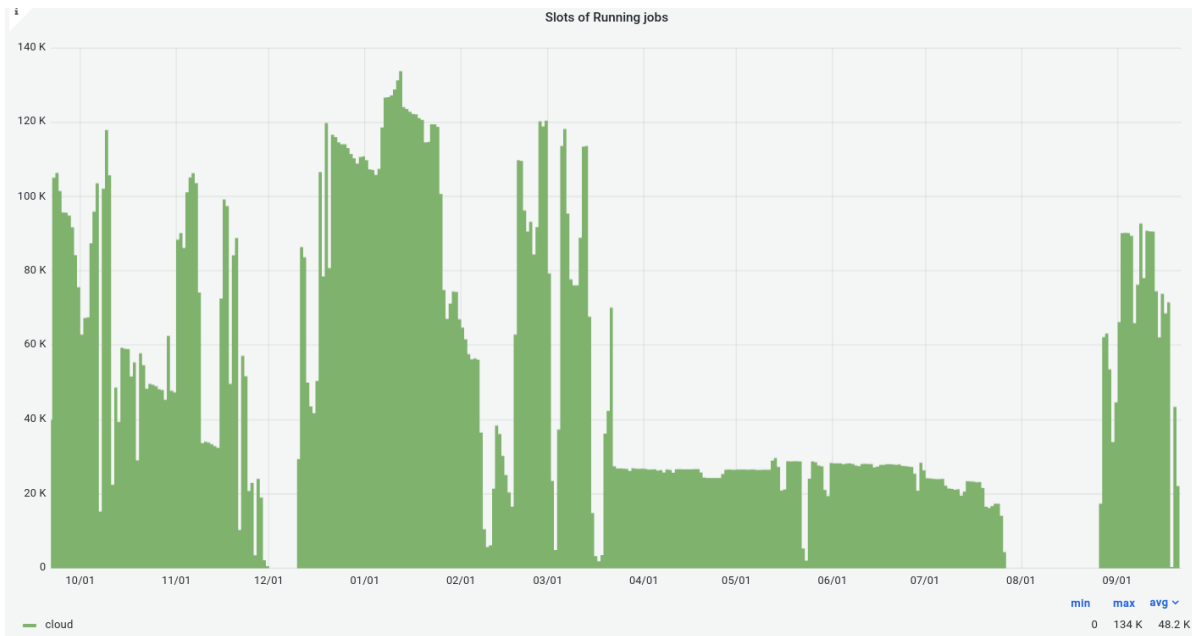


Принцип на работа



Собствени облаци - тригерна ферма

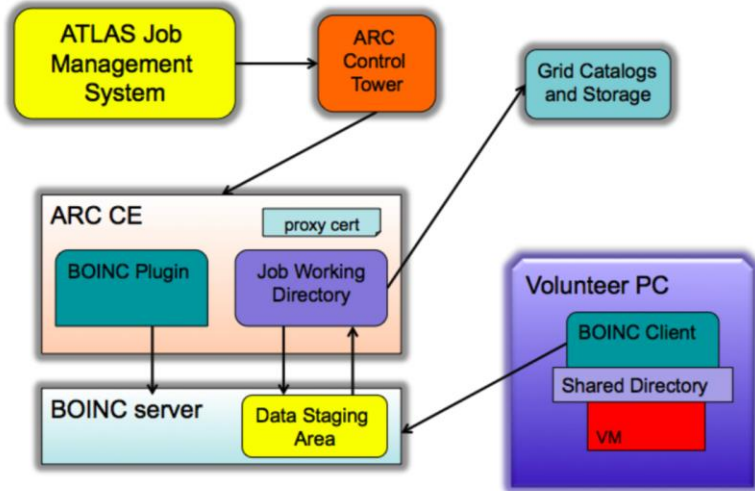
- Ресурс - 120 000 ядра
- Използване - само когато не се набират данни
- Облачно решение:
 - Бързо усвояване на ресурси
 - Няма нужда от инсталиране
 - Светкавично връщане на ресурси при нужда (kill)
- Все още не се използва максимално



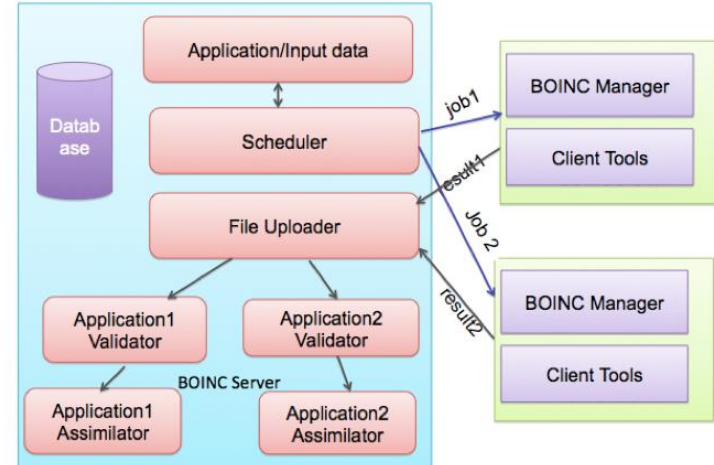


Доброволчески споделени изчисления (Volunteer Computing)

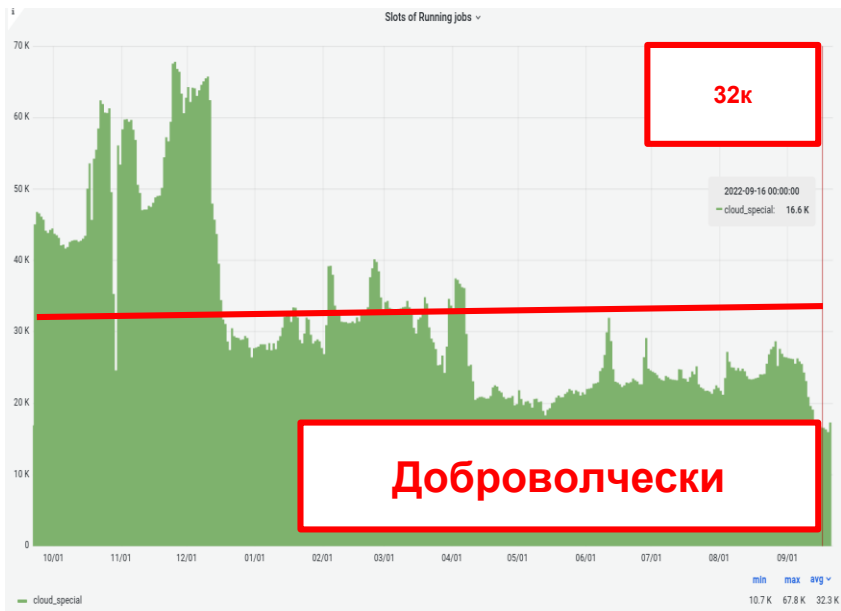
BOINC



BOINC Architecture



ATLAS@Home

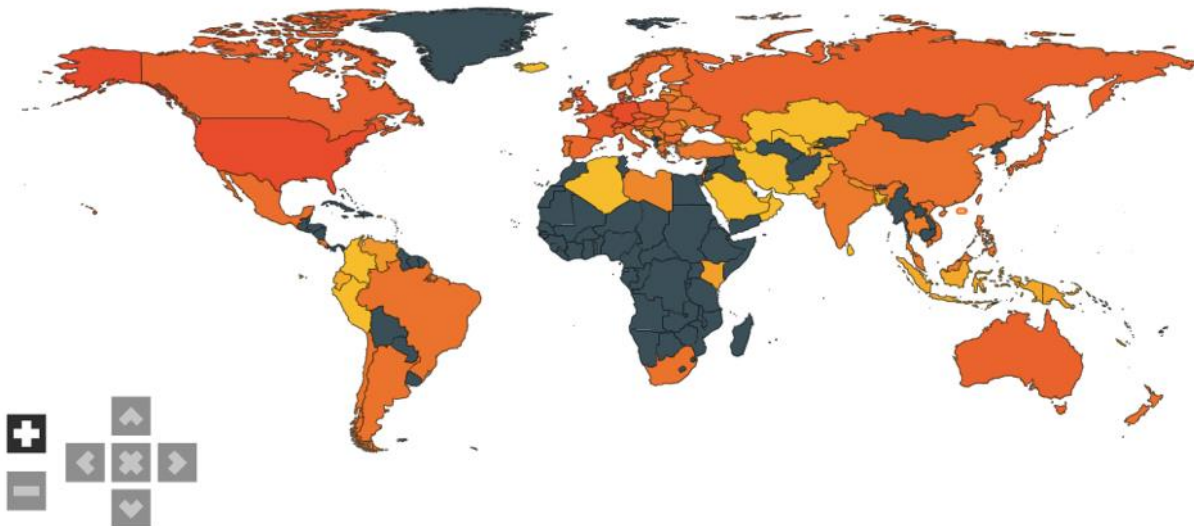
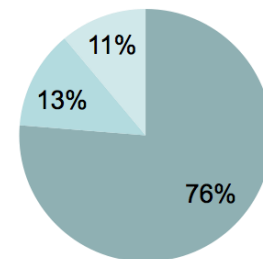


- Обработка ~ 1% от симулираните събития в ATLAS
- Сравним с голям безплатен компютърен център
- Не точно.. доброволците очакват все пак някаква поддръжка
- Голям потенциал в настолните компютри по учреждения които не се изключват нощно време

Кой ни помага

- VOINC е огромно общество от отдадени хора
- Много от тях участват в няколко проекта едновременно

■ Windows ■ Linux ■ Mac



И ние инсталираме BOINC

- На новопристигналите машини в ЦЕРН
- На машини с “неефективни” услуги
- На малки GRID сайтове
- На сайтове с ограничения

Обобщение

- GRID е все още тук и не се очаква скоро да си тръгне
- Все повече организации и спонсори избират алтернативи на конвенционалните компютърни центрове - суперкомпютри, облаци..
- Колаборацията ATLAS успешно ги интегрира и използва



Въпроси?
Ivan.Glushkov@cern.ch



Технологии в ЦЕРН

