

Система за разпределени изчисления на експеримента ATLAS

Катедра “Атомна физика” на 80 години

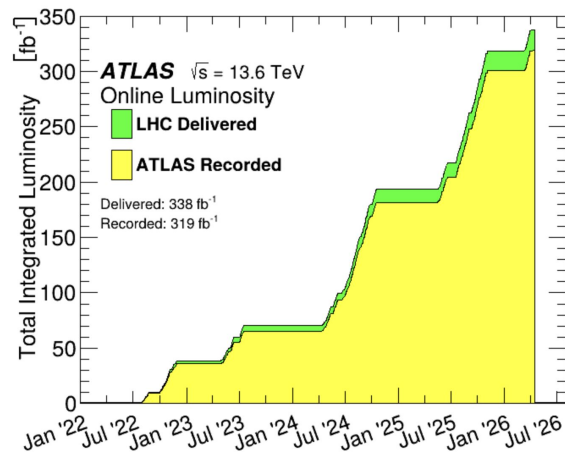
17 април 2026

Иван Глушков (BNL)

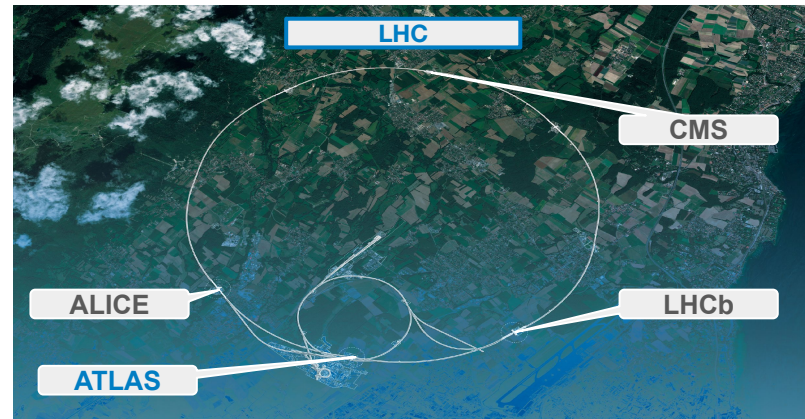
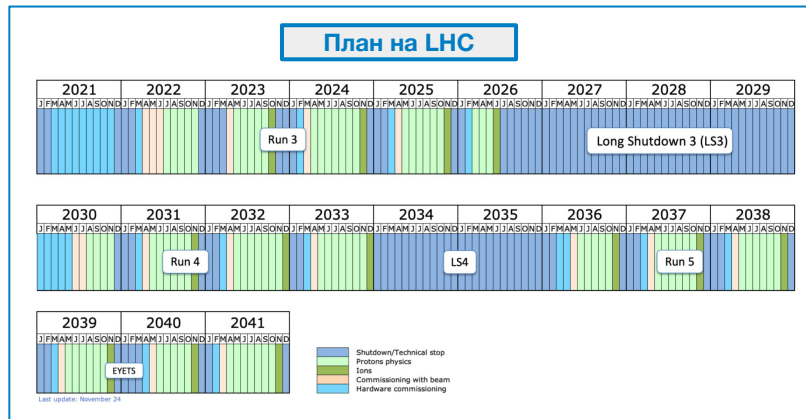


Адеронен колайдер

- p-p, ion-p, ion-ion
- E = 13 TeV
- L = 338 fb⁻¹
- 2021 - 2041
 - Текущо състояние:
 - Край на Run3: 29 юни 2026
 - Следващи етапи:
 - LS3: 2026 - 2030
 - Run4(HL-LHC): (L: x5-x7) 2031 - 2033



- P1: ATLAS
- P2: ALICE
- P5: CMS
- P8: LHCb



- Колаборация

- 3000 члена
 - 1200 докторанта
 - 170 института
 - 40 страни

- Детектор

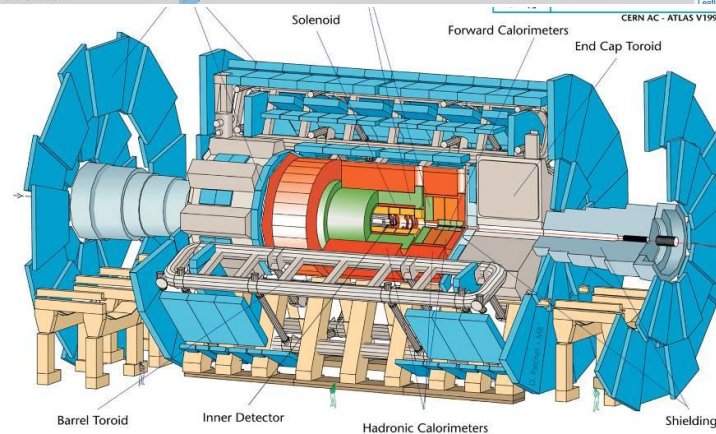
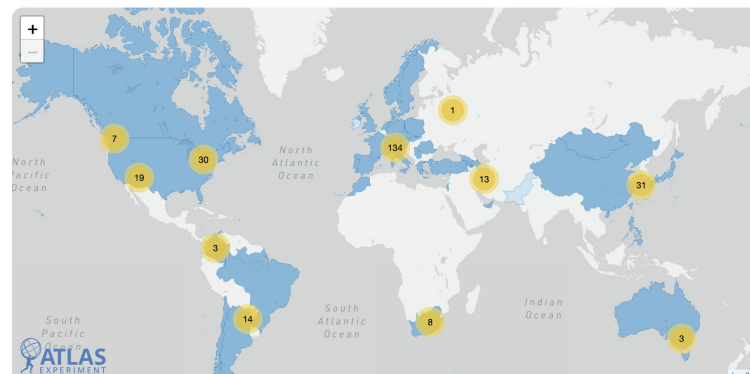
- “Общо предназначение”
- 44 x 22 метра
- 7000 тона

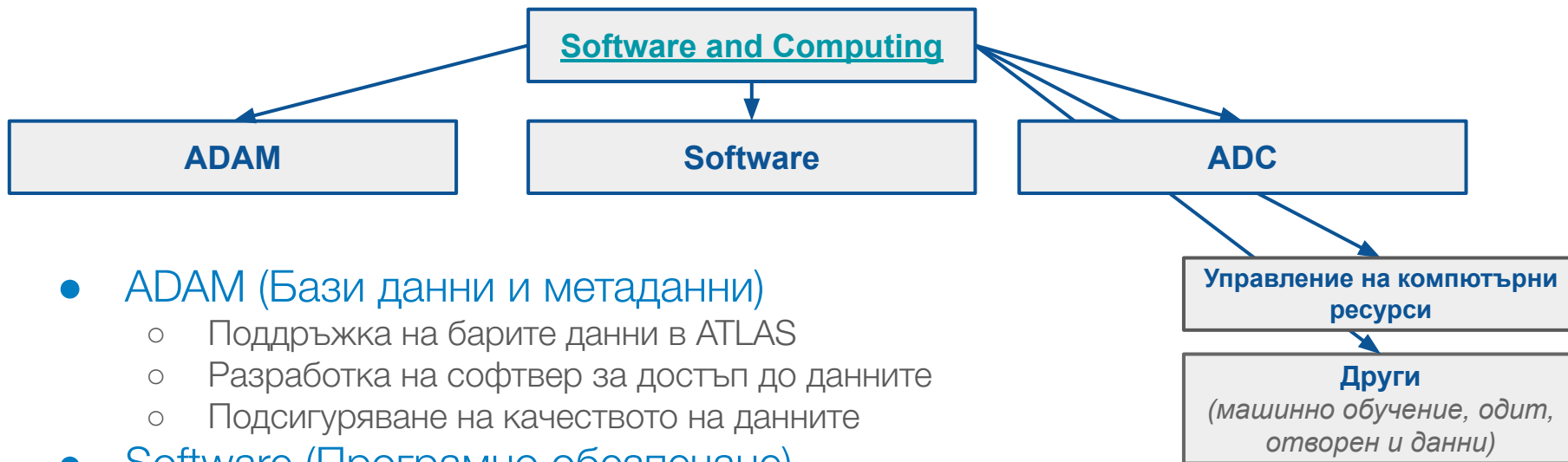
- Данни

- 10 PB/година

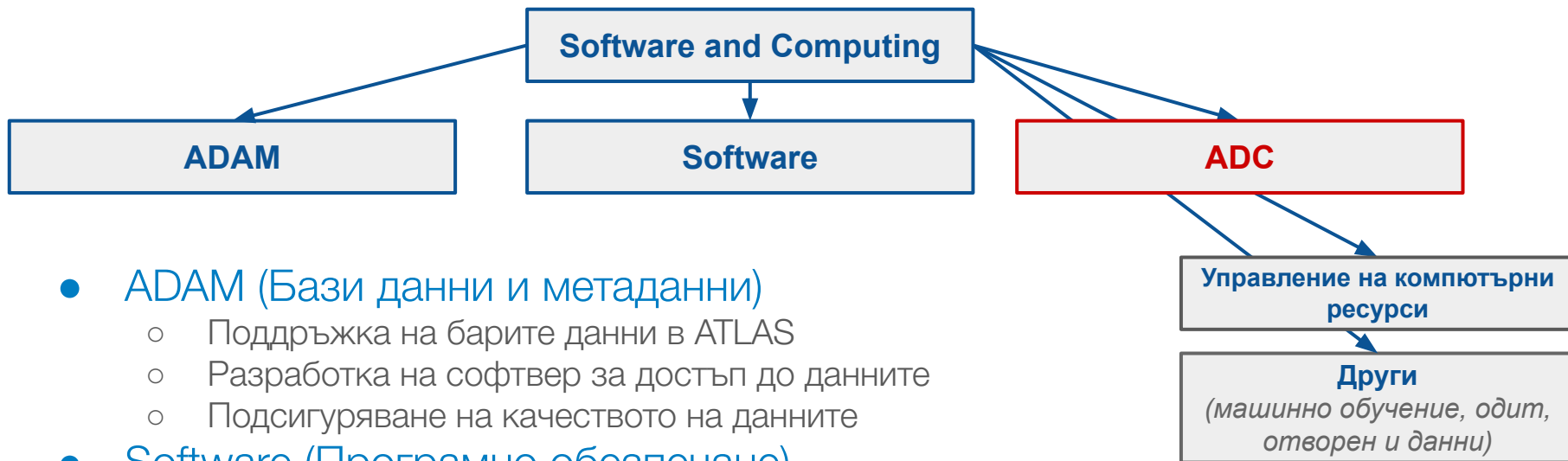
- Организация

- Организация
 - ...
- Комитети
 - ...
- Детекторни проекти
 - ...
- Дейности
 - **Software and Computing**
 - ...





- **ADAM (Бази данни и метаданни)**
 - Поддръжка на базите данни в ATLAS
 - Разработка на софтвер за достъп до данните
 - Подсигуряване на качеството на данните
- **Software (Програмно обезпечане)**
 - Разработка на софтвер за симулция, реконструкция и анализ на данни
 - Поддръжка и развитие на инфраструктура за разработване и регулярно тестване на софтуера на ATLAS
- **ADC (Система за разпределени изчисления)**
 - Хардуерно, софтуерно и оперативното осигуряване необходимо за подсигуряването на разпределени изчисления в ATLAS



- **ADAM (Бази данни и метаданни)**

- Поддръжка на барите данни в ATLAS
- Разработка на софтвер за достъп до данните
- Подсигуряване на качеството на данните

- **Software (Програмно обезпечане)**

- Разработка на софтвер за симулция, реконструкция и анализ на данни
- Поддръжка и развитие на инфраструктура за разработване и регулярно тестване на софтуера на ATLAS

- **ADC (Система за разпределени изчисления)**

- Хардуерно, софтуерно и оперативното осигуряване необходимо за подсигуряването на разпределени изчисления в ATLAS

Физика

Централизирана обработка на данни

Симулация/Генериране
Филтриране на данни
Подновяване на данни
Валидиране на данни

Потребителски анализ

Пограмно обезпечение
24/7 поддръжка

Инфраструктурно осигуряване

Инфраструктура

Tier0
GRID

Суперкомпютри
Доброволчески изчисления
Ферми за анализ

Управление

Потребителски анализ
Централизирани изчисления
Тостова инфраструктура
Мониторинг
Структура за 24/7 поддръжка

ADC

Управление на данни

Системно развитие

Rucio

Управление

Системно внедряване
Управление на системата

Мониторинг

Изследвания

Мрежи

Кеширане

Състеми за запазване на данни
Облачни услуги

Управление на процеси и задачи

Системно развитие

Дефиниране на процеси
Управление на процеси
Изпълнение на процеси

Управление

Системно внедряване
Мониторинг

Изследвания

Анализиране на данни
Ферми за анализ
Облачни услуги
Суперкомпютри



Доставчици на инфраструктура

CERN
WLCG
OSG
Nordugrid
Суперкомпютри
Изчислителни центрове

Интеграция на нововъведения

Нови архитектури (ARM, GPU)
Комерсиални облаци
Еластичност на ресурсите
Нови процеси (ML/AI)
Нов комерсиален софтуер (Varnish)

Приоритети на колаборацията

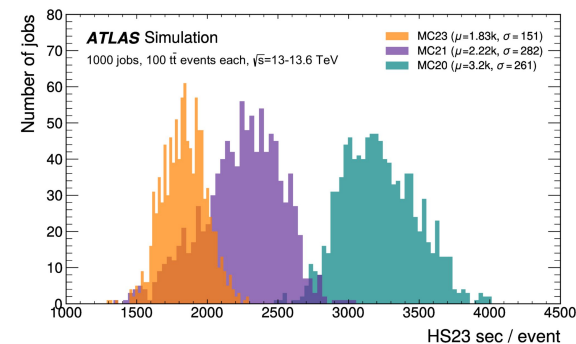
Важни анализи
Предстоящи конференции
Статистика vs Подновяване vs
Пресяване на данни

Поддръжка и управление на ресурси

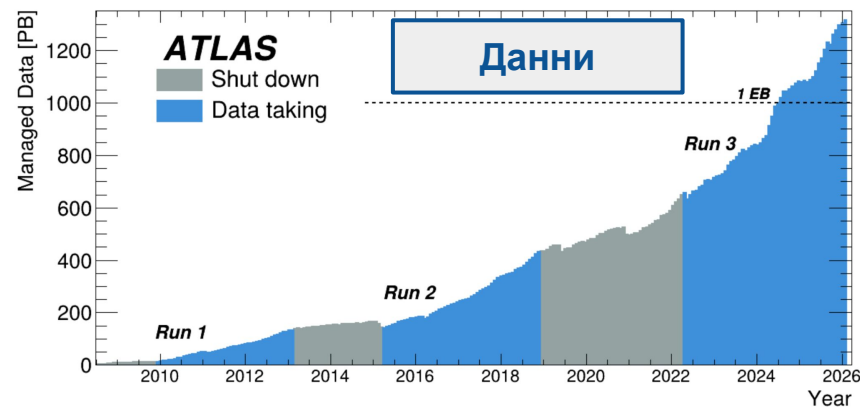
Изчислителни центрове
Конфигуриране
Оптимизация
Ъпгрейди
Потребители
Ресурси
Помощ

ATLAS

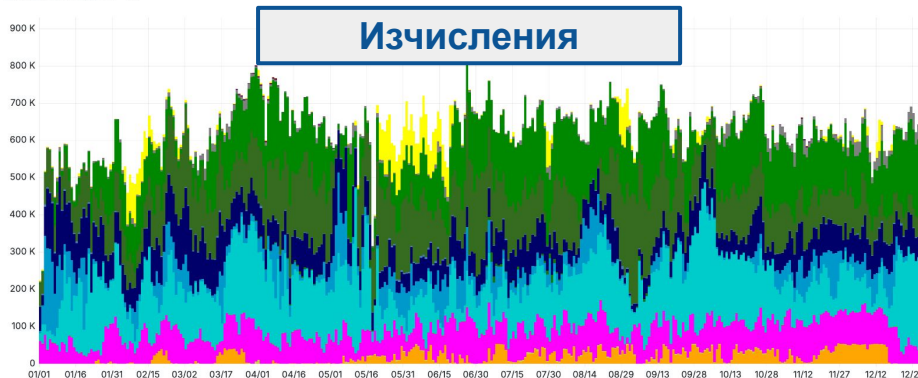
Подобрения в софтуер
Оптимизация в достъпа до мета-данни
Координация с екипа за набиране на данни в реално време



- **Оперативни параметри**
 - Паралелни изчисления: 600 хил.
 - Данни: 1.1 EB
 - Трансфер на данни: 500 Gb/s
 - Уникални потребители x100
 - Брой задачи: x10 хил.
- **Статични параметри**
 - 150 изчислителни центъра
 - 10 суперкомпютъра
 - 10 кеш-сайта
 - Разширения чрез:
 - Облаци (собствени и/или комерсиални)



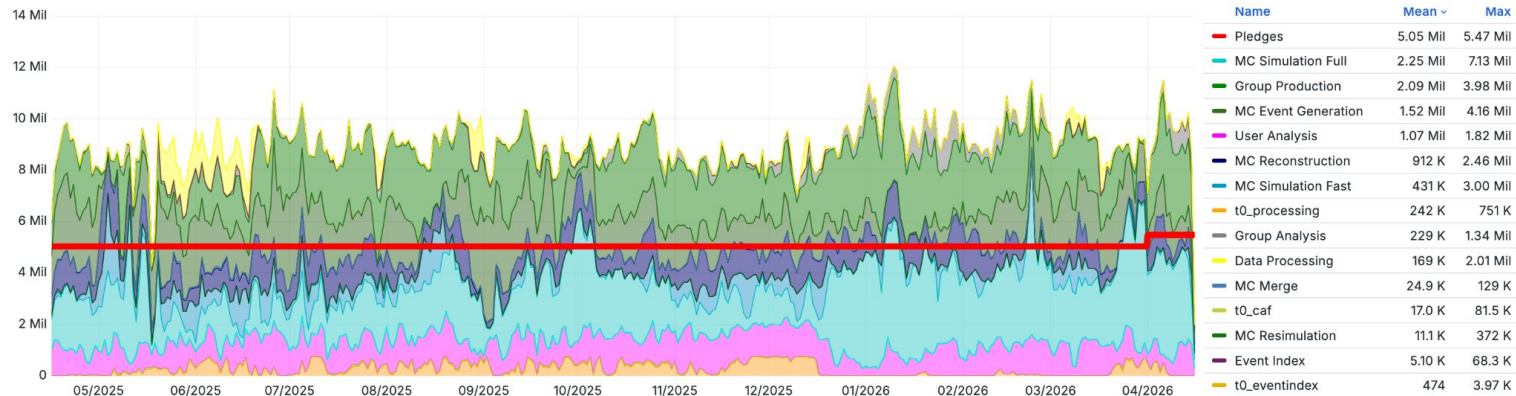
Slots of Running jobs



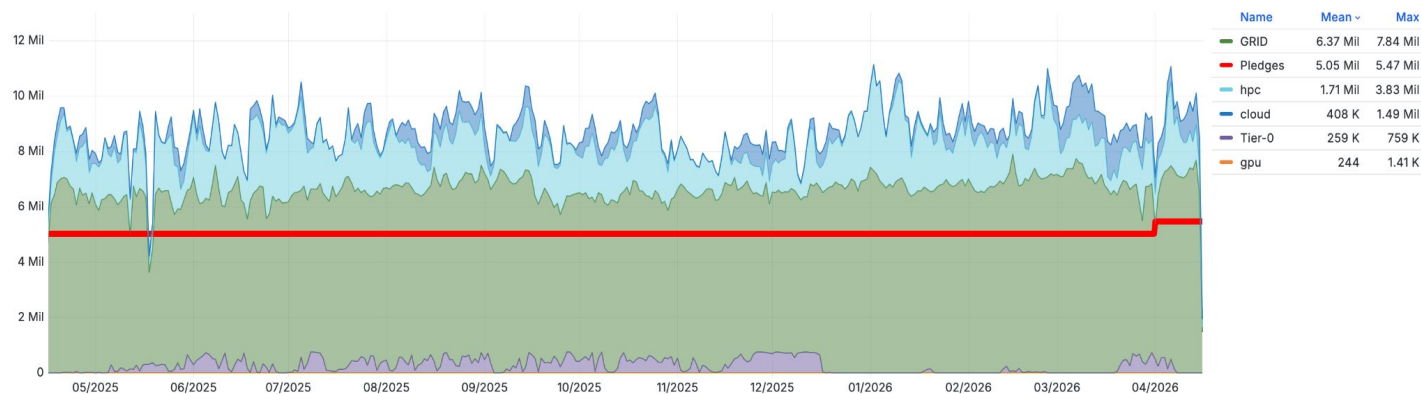
Name	Mean	Max	Min
MC Simulation Full	139 K	441 K	14.1 K
Group Production	129 K	237 K	1.34 K
MC Event Generation	122 K	300 K	13.7
MC Reconstruction	76.6 K	210 K	5.85 K
User Analysis	73.1 K	130 K	17.4 K
MC Simulation Fast	41.9 K	239 K	25.9
t0_processing	17.4 K	53.6 K	0
Data Processing	14.4 K	145 K	0
Group Analysis	9.48 K	44.6 K	76.0
MC Merge	1.97 K	9.29 K	456
t0_caf	1.21 K	9.90 K	0
MC Resimulation	866	25.4 K	0
Testing	621	1.02 K	266
Event Index	586	4.21 K	0
t0_eventindex	30.6	217	0
Others	0.0000190	0.00694	0



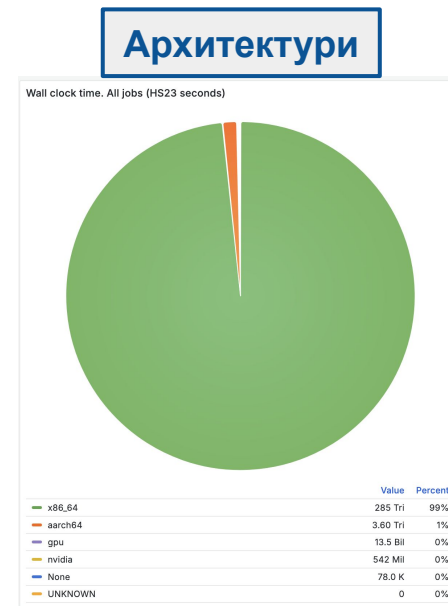
Slots of Running jobs (HS23) by ADC activity 🔍



Slots of Running jobs (HS23) by Resource type (Grouped) 🔍

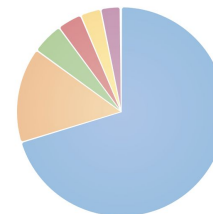


- x86
 - Стандартни ресурси (99%)
 - Всичко работи: софтуер, отчетност
- ARM
 - По-евтини (производителност, енергия, CO2)
 - Различни инструкции - всичко трябва да се прекомпилира. Направено!
 - ATLAS приема до 50% от предлаганите ресурси да са ARM
- GPU
 - Широко паралелизиране/Различна парадигма
 - Софтуер трябва да се пренапише
 - В ATLAS имаме първа версия на симулация с EM компонента на GPU
 - Отчетността не е тривиална ($\frac{1}{2}$ GPU, различно количество памет, отношение CPU/GPU)



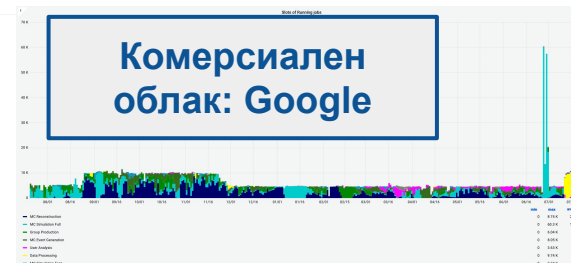
- **Стандартен изчислителен център**
 - Стандартни ресурси (99%)
 - Перфектна интеграция
- **Суперкомпютър (HPC)**
 - Много агенции на държавно ниво преминават към HPCs
 - Оптимизирани за високо-паралелни задачи (x100 хил нижки)
 - Лимитиран достъп - “National Security Infrastructure”
- **Облаци**
 - Основно предимство: Еластичност на ресурсите
 - Комерсиални облаци:
 - Много по-скъпи от стандартните
 - Много по-еластични
 - Собствени облаци
 - Позволяват бързо използване на тригерни ресурси
- **Доброволчески изчисления (BOINC)**
 - Използване на част от ресурсите на компютър
 - Приложение:
 - Компютри които временно не се използват
 - На заден план - запълване на ресурси които не се използват ефективно
 - До 60 хил процеса в ATLAS

Типове ресурси

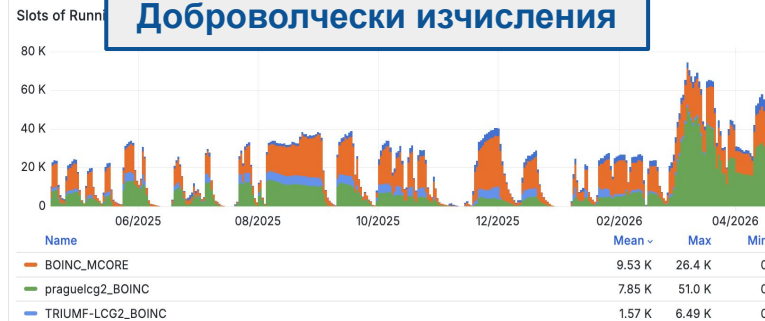


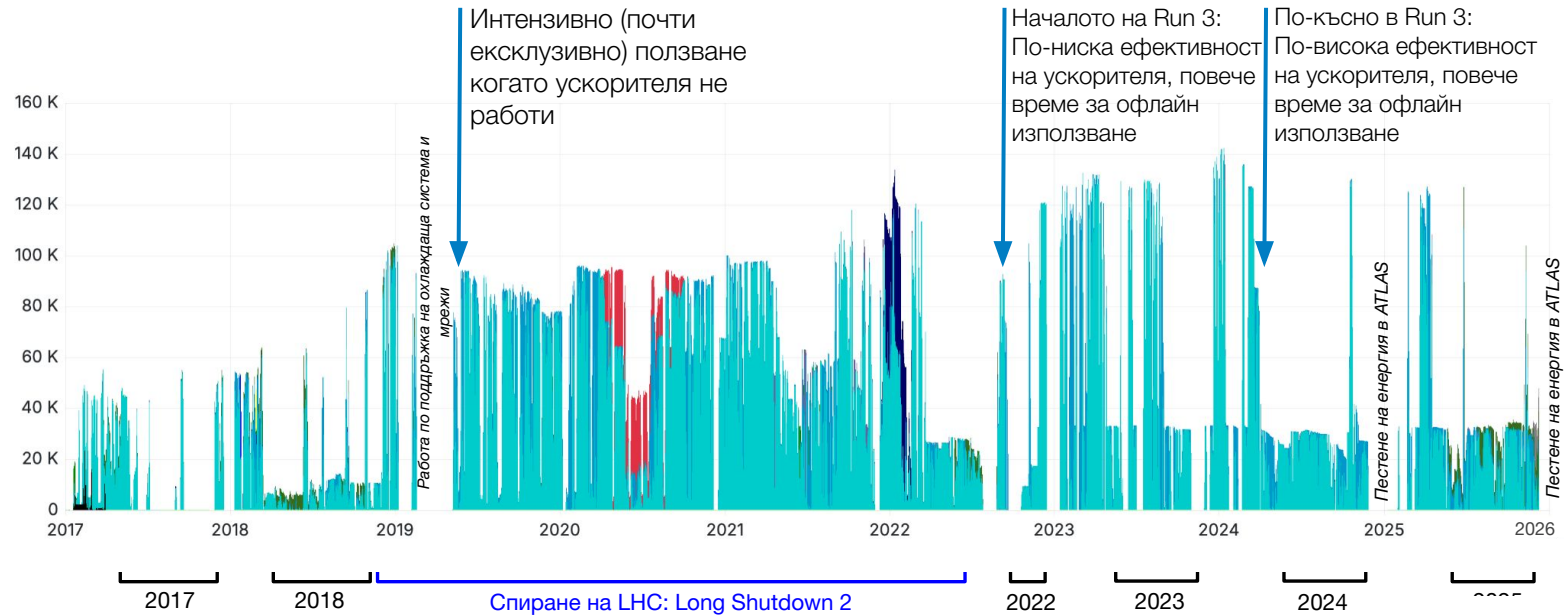
	Value	Percent
GRID	204 Tri	71%
hpc	42.4 Tri	15%
cloud	13.9 Tri	5%
cloud_special	10.7 Tri	4%
hpc_special	9.02 Tri	3%
Tier-0	8.50 Tri	3%
gpu	14.0 Bil	0%
UNKNOWN	0	0%

Комерсиален облак: Google



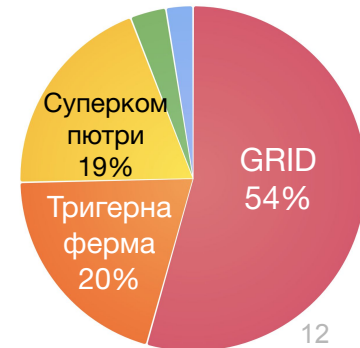
Доброволчески изчисления



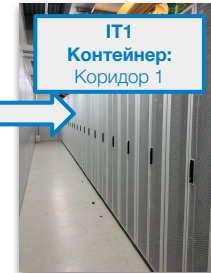
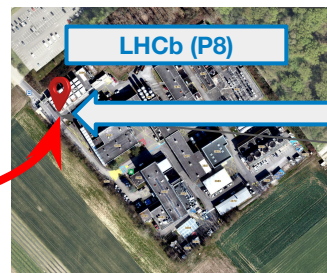
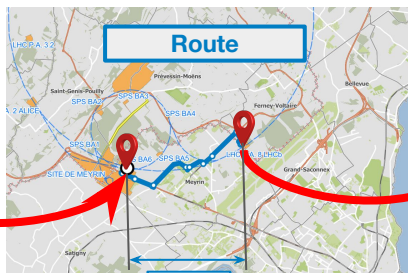


Тригерната ферма е от огромно значение за изчисленията в ATLAS

- Почти 400 милиарда симуирани събития в ATLAS от 2017 година насам, 20% от тях - на тригерната ферма
- Използваме когато можем - по-малко когато оперативния цикъл на LHC е по-ефективен
- През HL-LHC е планирано е да се използва между филовете на ускорителя



- LS3: 2026 - 2030
 - Преместване на 456 сървъра в контейнер на LHCb
 - Мрежов ъпгрейд който ще позволи използване за всички изчислителни нужди на ATLAS
 - 1.3 MHS23 (~ 1/8 от всички изчислителни ресурси на ATLAS)
- Run4: 2031 - 2033
 - Новата ферма ще се състои от CPU + GPU
 - Точни параметри все още не са известни
 - Използването на офлайн ще трябва да се направи от нула
 - Контейнери
 - Използване на CPU+GPU за офлайн обработка на данни



- Видове данни

- Детекторни данни
 - RAW > AOD > DOAD
- Монте Карло
 - EVNT > HITS > RDO > AOD > DAOD

- Инфраструктура

- Дисково пространство
 - Компютърните центрове предпочитат изчислителни сървъри
 - Идеално би било да имаме всички данни навсякъде, постоянно
 - Нямаме пространство дори и за едно копие на всички данни
- Лентови носители
 - По дизайн - записваме веднъж, четем - “никога”
 - “Data Carousel” - четем при всяка голяма изчислителна кампания
- Кеш
 - Дисков масив за временно, динамично съхранение на данни
 - Данните които не са използвани най-дълго се изтриват първи
 - Използват се най-вече за доставка на данни на сайтове без собствени дискови масиви
 - Само няколко такива системи в ADC

- Данни

- Правила за записване
 - Определят се от приоритетите на колаборацията
 - Например:
 - RAW: две копия на лента (T0 + T1)
 - AOD: едно копие на лента и две временни копия на диск
 - DAOD: едно копие на T2 диск
- Пресяване (премахване на ненужни..)
 - ..събития (Skimming)
 - ..обекти от събития (Thinning)
 - ..параметри на събития (Slimming)
- Разчистване

- **Време на живот (Lifetime model)**

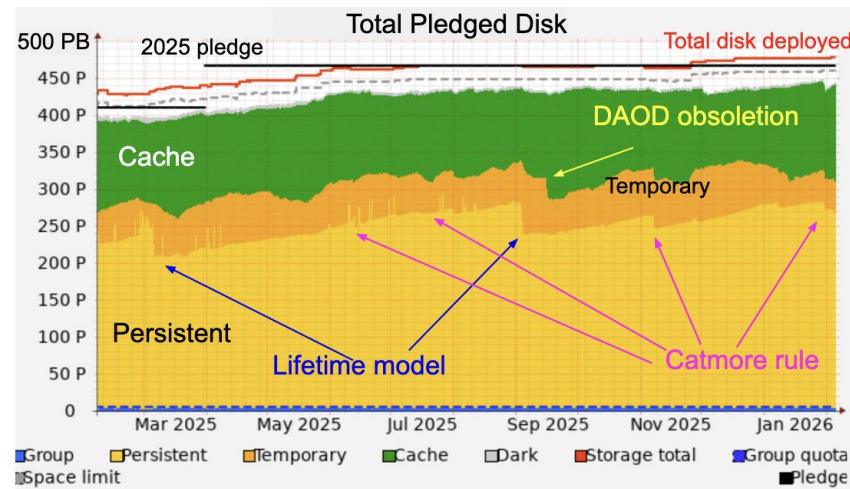
- Данните които не са използвани повече от X дни се предлагат за изтриване
- X зависи от типа данни (напр. 6m за DAOD)
- Всеки член на колаборацията може да поиска изключване на данните
- Честота на изпълнение: 3-4 пъти годишно

- **Правило на Катмор (Catmore rule)**

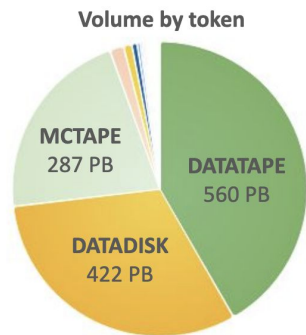
- Ако група от данни на диск и с копие на лента не са ползвани в последните X месеца, биват освобождавани за изтриване

- **Премахване на стари филтрирани данни (DAOD Obsolescence)**

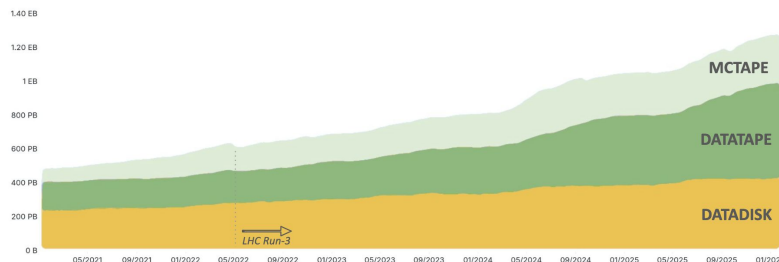
- При произвеждане на ново филтриране на данни
- Старите версии се изтриват след като всички анализи са преминали към новата версия
- Честота на изпълнение: 3-4 пъти годишно



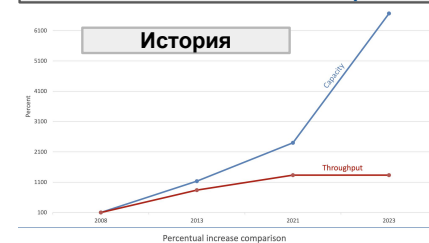
- **Параметри:**
 - Евтини и бавни
 - Пишем веднъж - четем веднъж
- **Бъдеще:**
 - Увеличаване на обема но не и на скоростта
 - Цената се увеличава двойно тази година
- **ADC:**
 - Четем данни при всяка по-голяма кампания
 - За увеличаване на скоростта на четене - метаданни за всеки файл



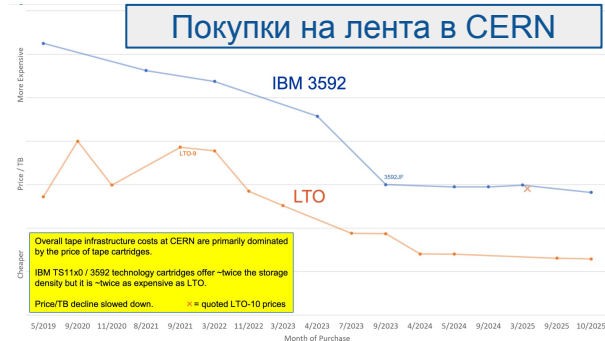
Данни в ATLAS



Касета за данни: Обем и Скорост

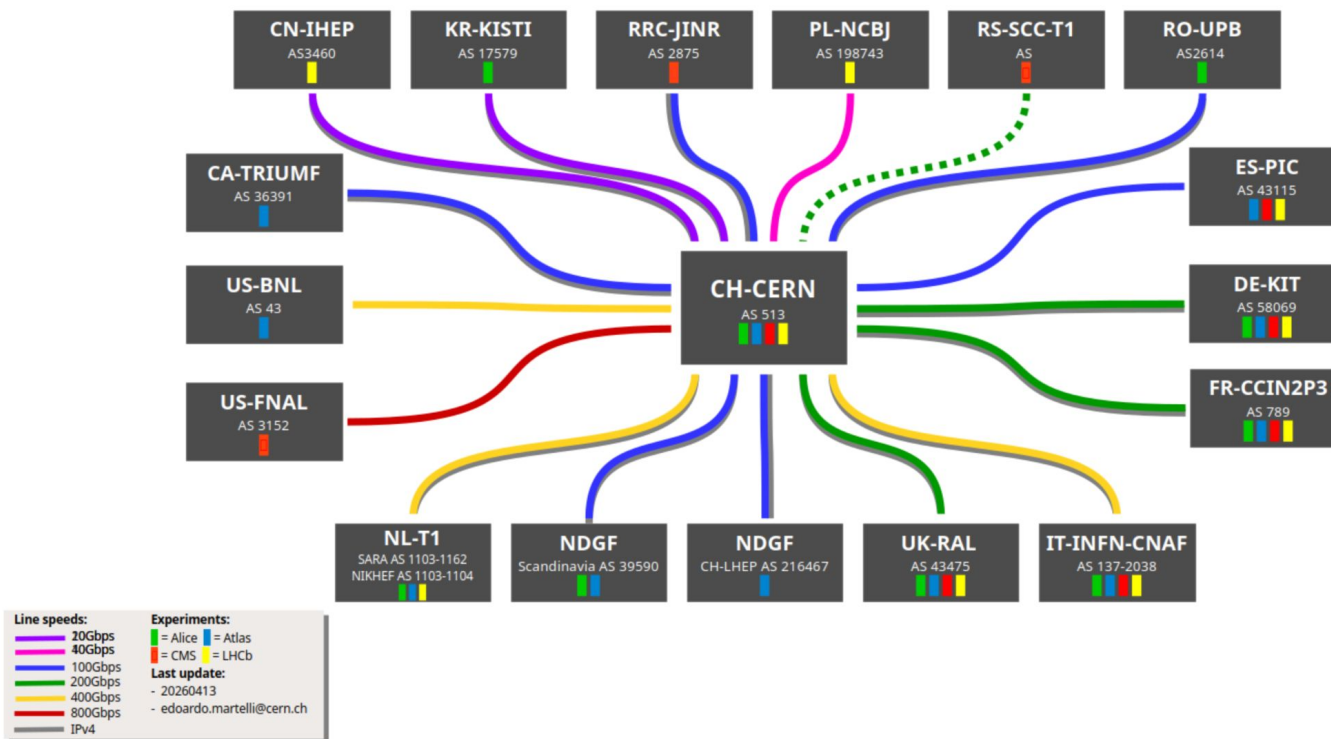


Покупки на лента в CERN



* P.S.: Източник на графиките: "Technology and Market Evolution of Tape", Vladimir Bahyl (CERN), WLCG Open Technical Forum (OTF) #8

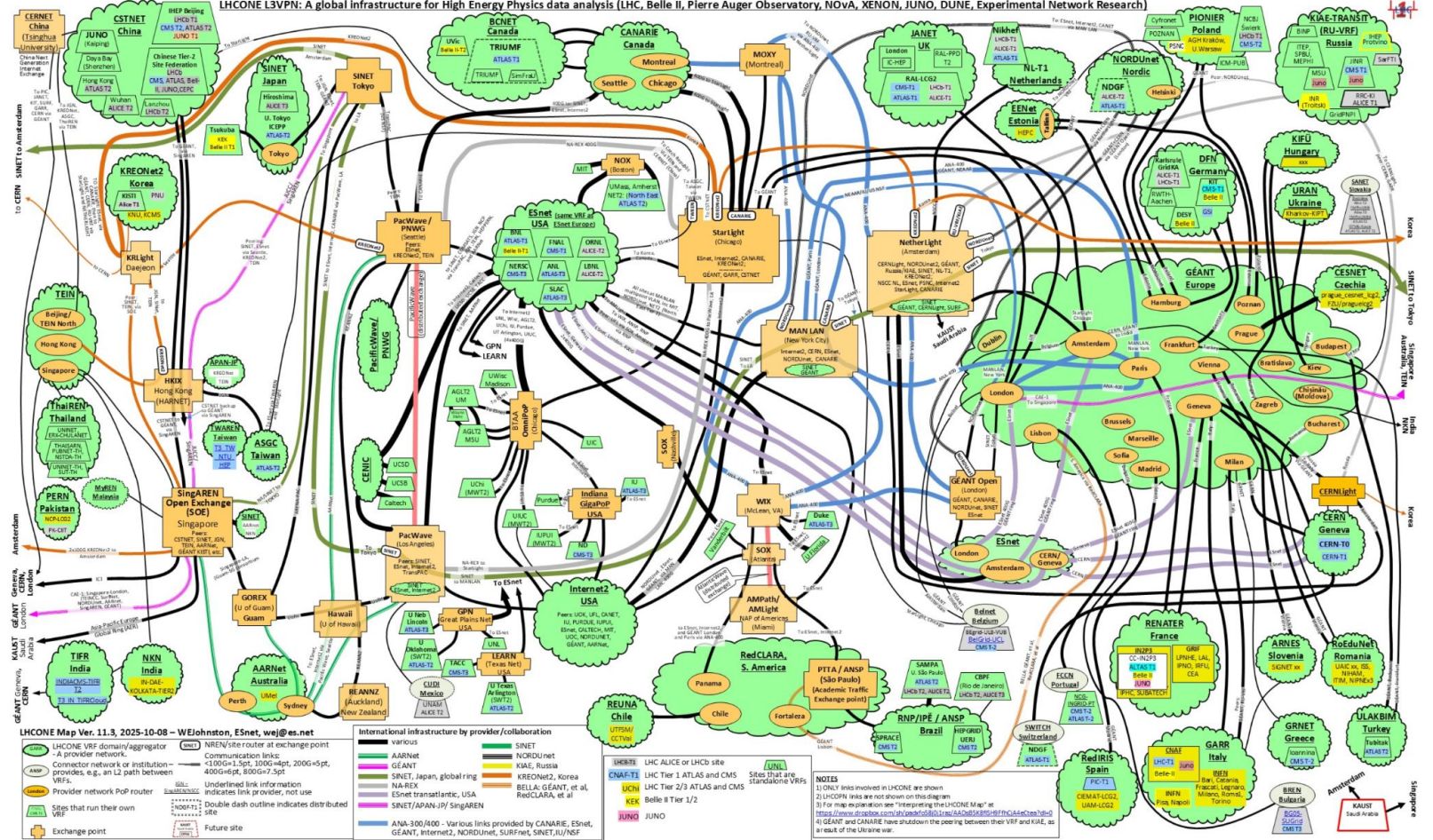
** P.P.S: Всичко за текущото състояние на лентите в CERN/WLCG можете да научите от [мазу](#) конференция



Мрежи: LHCCONE



LHCCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON, JUNO, DUNE, Experimental Network Research)



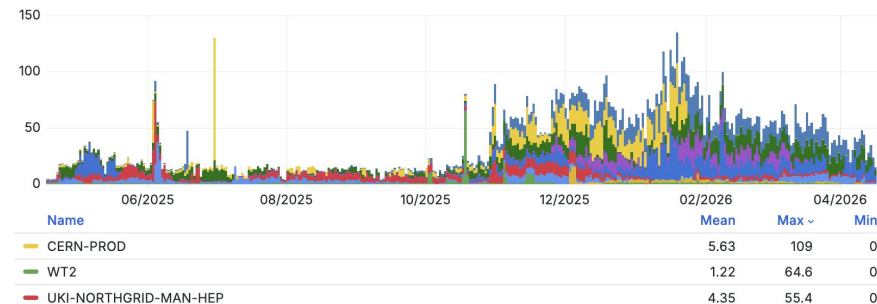
- Области на използване:

- Анализ на данни
- Реконструкция
- Детектиране на аномалии
- ...

- ADC:

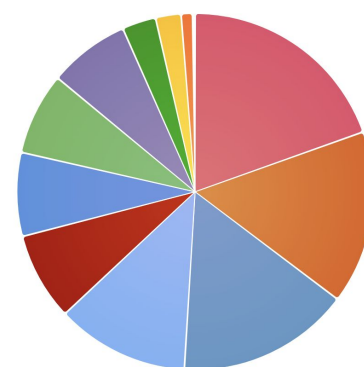
- Inference as a Service (IaaS)
 - Предоставяне на отдалечен достъп до обучени модели
 - В процес на развитие
- Разпределен достъп до графични процесори (13 изчислителни центъра)

Slots of Running jobs ⓘ



Разпределение на графични задачи по изчислителни центрове

Completed jobs



	Value	Percent
UKI-NORTHGRID-MAN-HEP	68.8 K	20%
CERN-PROD	55.3 K	16%
FZK-LCG2	55.1 K	16%
UKI-SOUTHGRID-RALPP	42.3 K	12%
BNL-ATLAS-OPP	27.8 K	8%
UKI-LT2-QMUL	26.5 K	8%
WT2	26.0 K	7%
OU-OSCAR-ATLAS	25.7 K	7%
LBNL_DSD_ITB	10.6 K	3%
Vega	7.88 K	2%
UKI-SCOTGRID-GLASGOW	3.35 K	1%
RAL-LCG2	252	0%
INFN-T1	1	0%

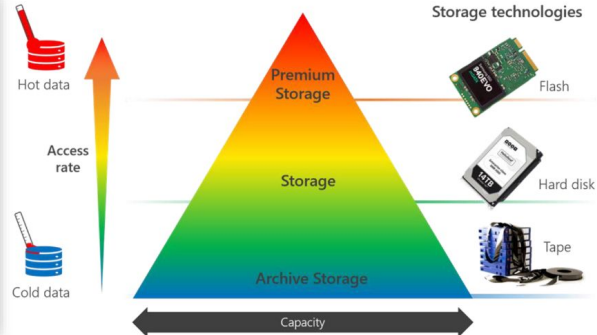
- LLMs
 - Всички използват LLMs
 - За генериране на код
 - За агенти / програми
 - За презентации, доклади, картинки...
 - Предимства (?): Бързо
 - Недостатъци:
 - Споделяне на конфиденциални данни с моделите
 - Преодоверяване на резултатите
 - Необходимост:
 - Подсигуряване на изолация на данните (договори с доставчиците)
 - Правила в рамките на колаборацията за използването на AI в софтуера на ATLAS
 - Prompt/agent engineering обучение
- Огромно повишаване на цените на хардуер
 - На всяко ниво - от памет през дискови носители до лентови носители
 - Пример: Сървър на Dell
 - Цена 2025: 13к USD
 - Цена 2026: 40к USD

"Thanks" to AI, storage is the new gold

TECH
AI memory is sold out, causing an unprecedented surge in prices
PUBLISHED SAT, JAN 10 2026:7:00 AM EST | UPDATED SAT, JAN 10 2026:10:55 AM EST
 KIF Leewing @KIFLEEWING WATCH LIVE

KEY POINTS

- This year, there won't be enough memory to meet worldwide demand because powerful AI chips made by the likes of Nvidia, AMD and Google need so much of it.
- Prices for computer memory, or RAM, are expected to rise more than 50% this quarter compared to the last quarter of 2025.
- Wall Street has been asking consumer electronics companies like Apple and Dell Technologies how they will handle the memory shortage, and if they might be forced to raise prices or cut margins.



PC Components > Storage

Many high-capacity NVMe SSDs are now as expensive as gold by weight as shortage intensifies — we ran the numbers, here's what we found
News | By Bruno Ferreira published January 16, 2026

Just like RAM, expect them to be made of unobtainium in short order

Image credit: Getty (imged)

Share | Comment

A keen-eyed Reddit user caught a brainwave while browsing for high-capacity SSDs and was struck with the following thought: with the AI-infused silicon shortages, we've reached the point where NVMe gumstick-style SSDs are more expensive than gold by weight. The thread generated discussion aplenty, so we figured we'd dig into this and look up pricing and weight across a range of models. Spoiler alert: It's very much true for 8 TB drives and quickly heading there for 4 TB models.

PC Components > Storage > HDDs

HDD prices spike as AI infrastructure and China's PC push collide — hard drives record biggest price increase in eight quarters, suppliers warn pressure will continue
News | By Zak Kilian published December 15, 2025

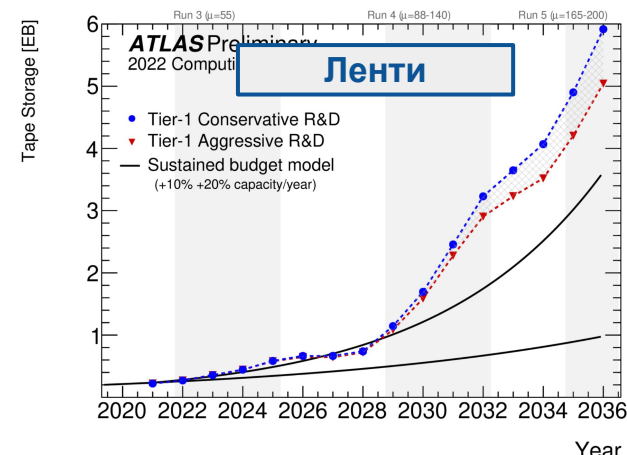
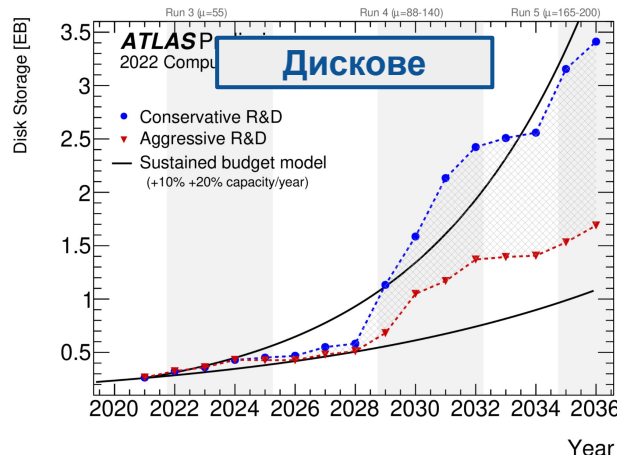
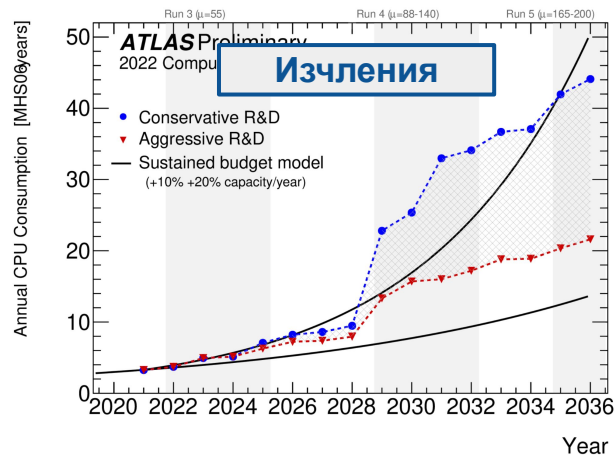
Hard drives, often treated as a legacy technology, are once again being pulled into the spotlight.

February 2026

Market Evolution of Tape

27

Источник: ["Technology and Market Evolution of Tape"](#), Vladimir Bahyl (CERN), WLCG Open Technical Forum (OTF) #8



● Екстраполация

- От 2022
- Базирана на
 - Физика: очаквани стойност на ускорителя
 - ИТ: размер на събития, нови алгоритми
- Текущо финансиране:
 - 10% увеличаване на ресурсите на година



Не взема предвид увеличаването на цените на процесори, памет (x4), дискови (x1.5) и лентови носители (x2) предизвикани от лавинообразното строене на ИИ изчислителни центрове

- Текущо финансиране:
 - 10% увеличаване на ресурсите на година

- **Организационно**
 - ADC е ключовия компонент позволяващ статистическата прецизност необходима за резултатите на ATLAS
 - Синхронизацията на всички заинтересовани страни е ключов елемент за успеха на ADC
- **Оперативно**
 - 24/7, 5 континента, 30 FTEs (Full-Time Equivalent)
 - До 1 мил. компютърни ядра / > 1 EB данни
 - Оперирание във хетерогенна среда
- **Предизвикателства**
 - Дисково и лентово пространство
 - Цени
 - Изкуствен интелект
 - Специалисти

Въпроси:

Ivan.Glushkov@cern.ch

