CERN: ґрід та інформаційні технології

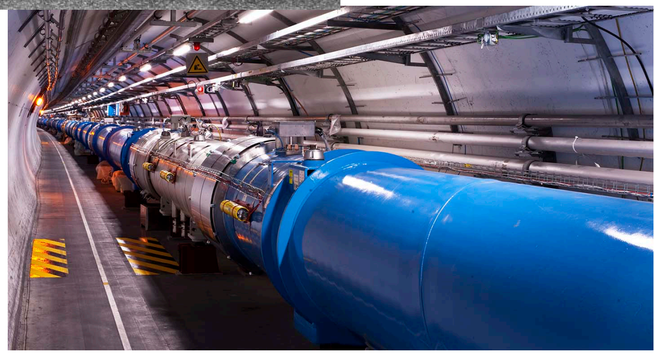
Свірін Павло
ATLAS, CERN
LAPP

Коротко про себе

- Закінчив НТУУ КПІ, факультет електроніки, магістр комп’ютерних наук
- 2008-2018: викладач в НТУУ КПІ
- Ph.D: НТУУ КПІ (2014 рік)
- 2014-2017 роки: CERN, експеримент ALICE, представництво Інституту теоретичної фізики ім. Боголюбова
- 2017-2019: Брукхевенська національна лабораторія (США)
- З 2019 року: CERN, експеримент ATLAS, представництво університету Техасу в Арлінґтоні
- З 2021 року: Барселонський центр суперкомп’ютингу (BSC)

CERN

- Зосновано 1954 року, 12 країн-зосновників
- Головна функція - надавання доступу до прискорювачів частинок, на базі яких побудовані експерименти
- 2009 року запущено в роботу Великий Адронний колайдер (LHC)



Країни-члени організації

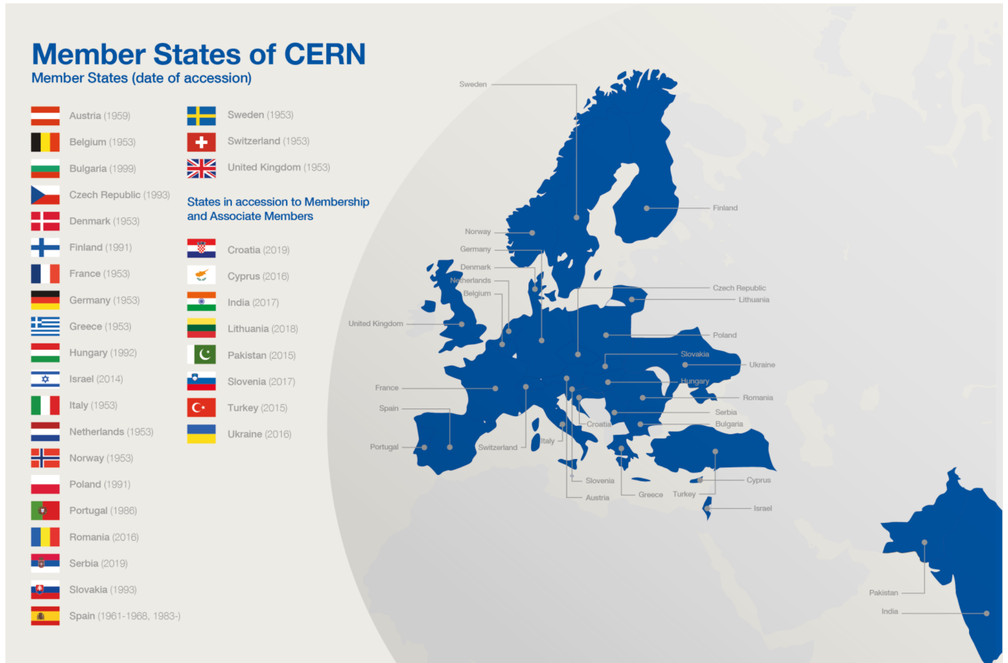
23 країни з повним членством, 6 країн з асоційованим членством, 2 країни з перехідним статусом до повного членства

Спостерігачі

- Індія, Японія, Росія, США
- Міжнародні організації: ЮНЕСКО, Європейська комісія, ОІЯД

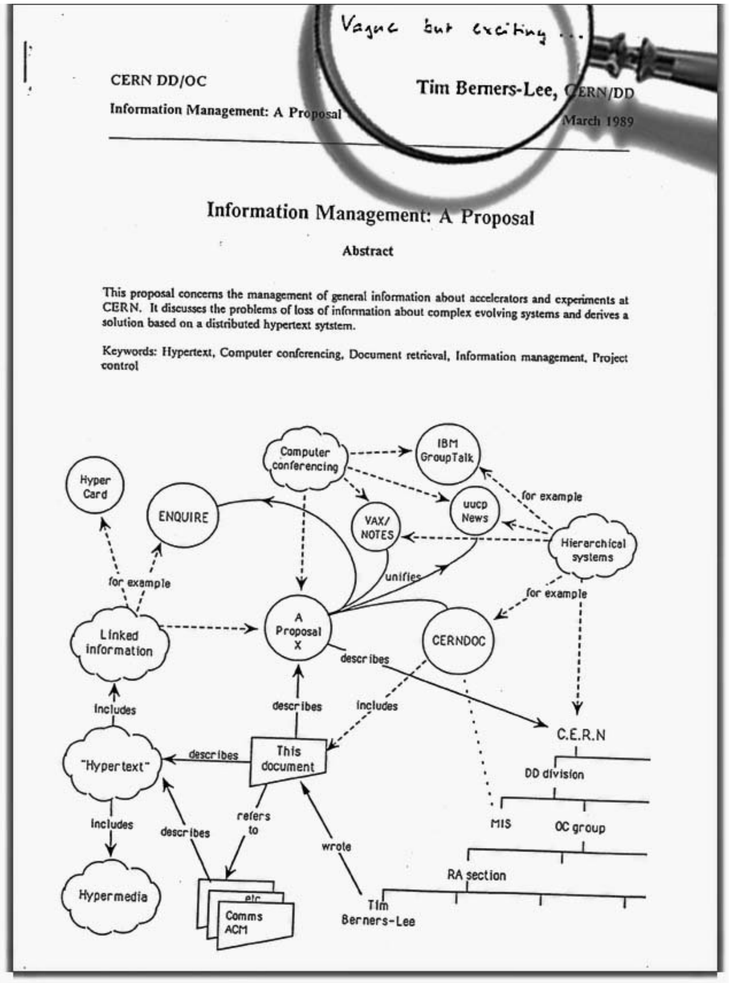
Наукові контакти

- 34 країни залучено до міжнародних проектів CERN
- інші контакти: 19 країн Европи, Азії, Африки, Південної та Північної Америки



Дослідження в CERN

- Фундаментальні дослідження в сфері фізики високих енергії
- Прикладні дослідження і результати:
  - World Wide Web
  - перше в світі впровадження тач-скрінів для панелі керування суперпротонним синхротроном (1975)
    - https://www.youtube.com/watch?v=tQe5dlzScwU
- Інші дослідження:
  - Медицина
  - Інженерія: https://home.cern/science/engineering
  - Аерокосмічні: https://www.nasa.gov/feature/nasa-cern-timepix-technology-advances-miniaturized-radiation-detection
  - https://kt.cern/cern-technologies-society



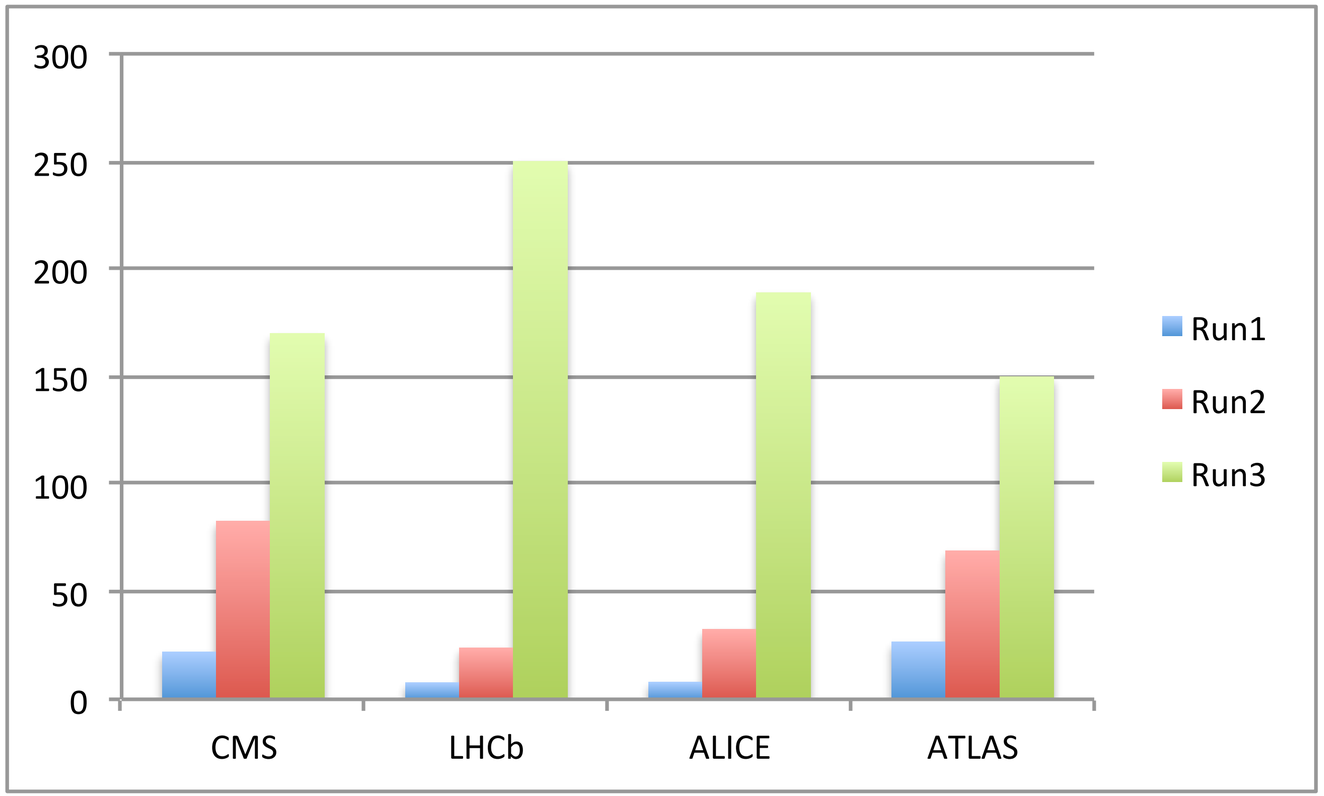
Великі експерименти CERN

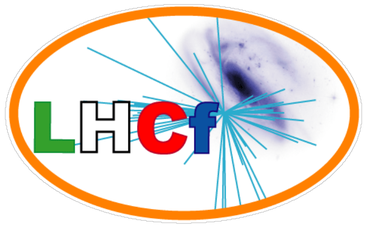
Найбільші експерименти CERN, які генерують переважну кількість даних.

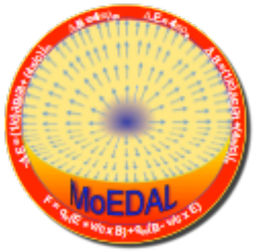
В колаборацiях - близько 10 тисяч чоловік

Об’єм даних на експериментах

- Об’єм даних, що отримано лише 2015 року дорівнює об’єму даних за перший сезон роботи LHC
- Протягом 2016 року було отримано 88 Пб даних (20 млн. DVD)
- На наступний запуск (2020-2022 pp) прогнозується 25-кратне зростання об’єму отриманих даних
- на 2021 ATLAS оперував приблизно 390 Пб даних (отриманих з детекторів та симульованих)





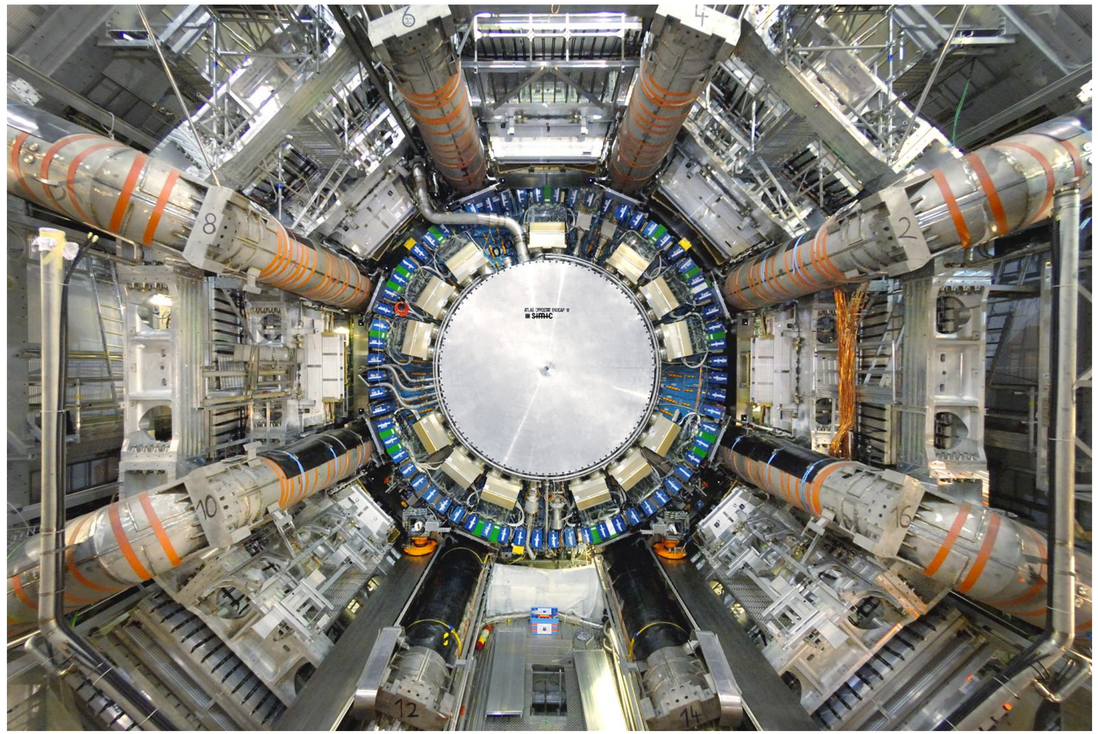


CERN: малі експерименти

- 21 малий експеримент
- Невелика кількість вчених у колабораціях (по декілька сотень вчених)
- Не всі знаходяться на LHC, тому деякі з них генерують дані щороку
- Генерують відносно незначні об’єми даних

Експеримент ATLAS

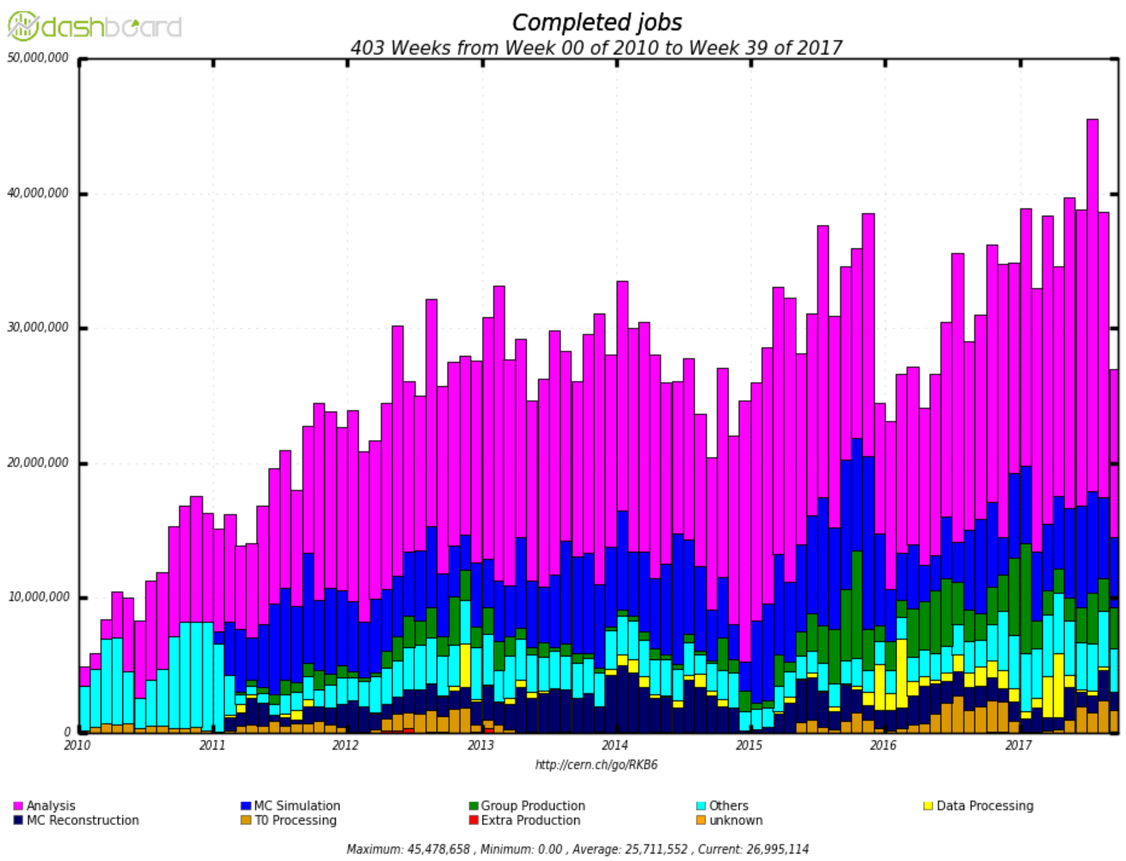
- найбільший експеримент фізики високих енергій, коллаборація - бл. 2000 науковців і інженерів з 165 інститутів
- розроблений для вимірювання найширшого діапазону сигналів, а отже - продукує велику кількість даних

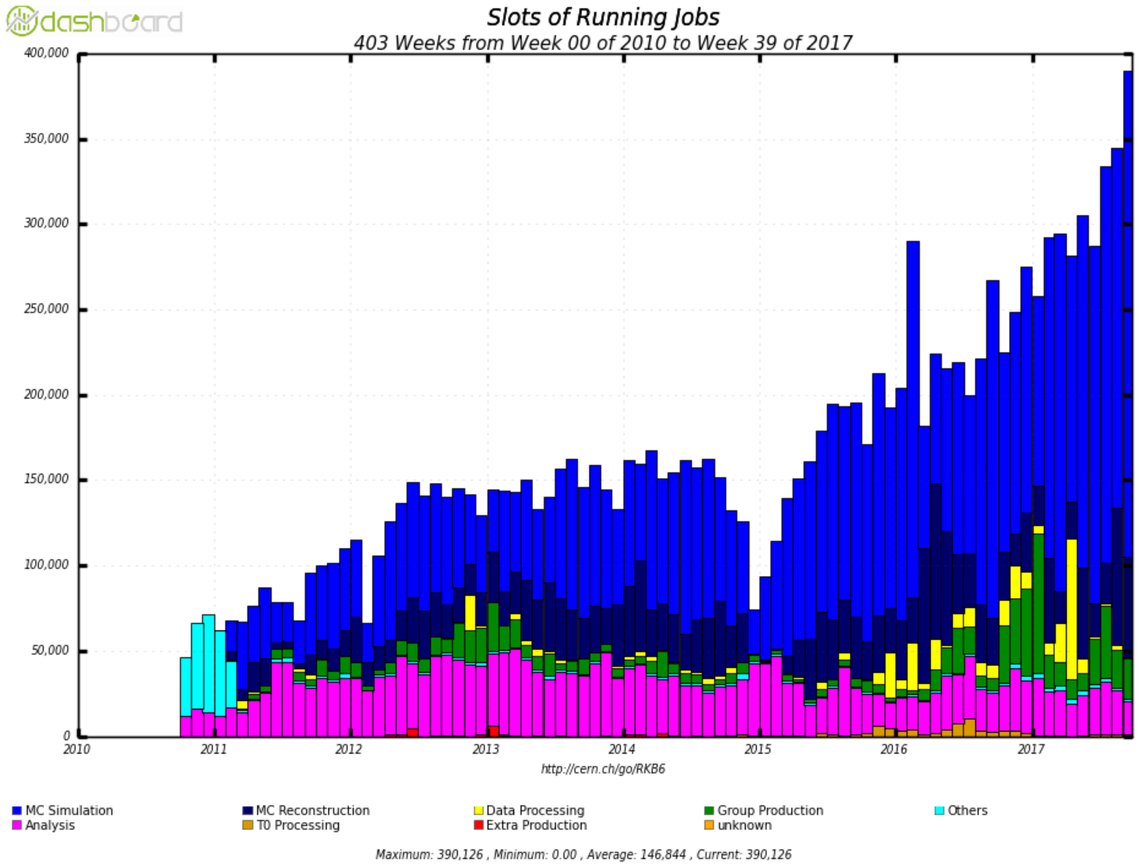


ATLAS EXPERIMENT

Статистика ATLAS

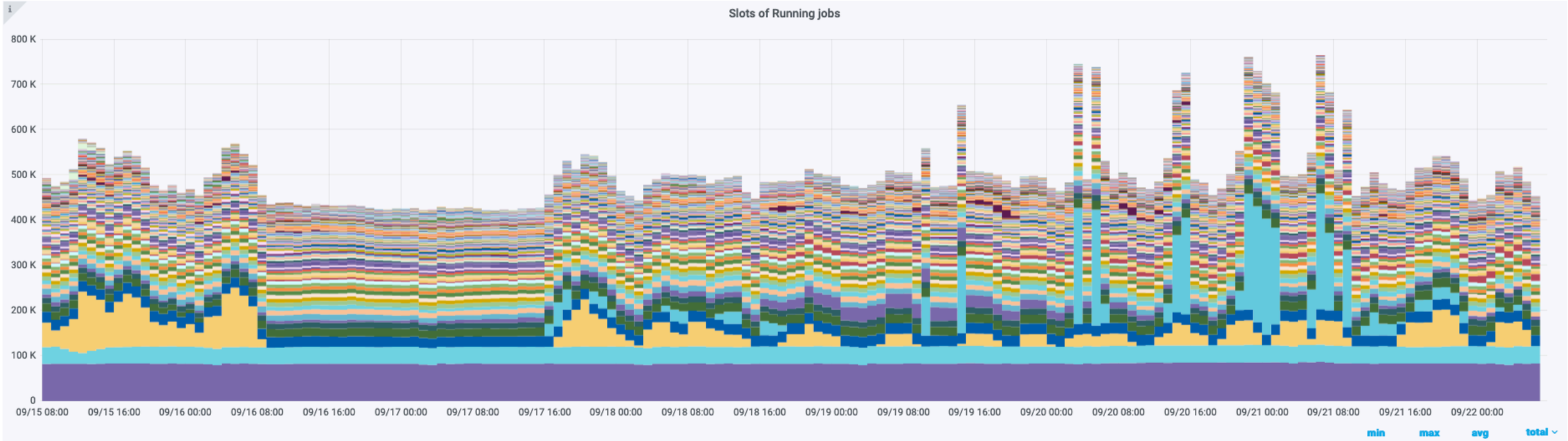
Статистика експерименту ATLAS





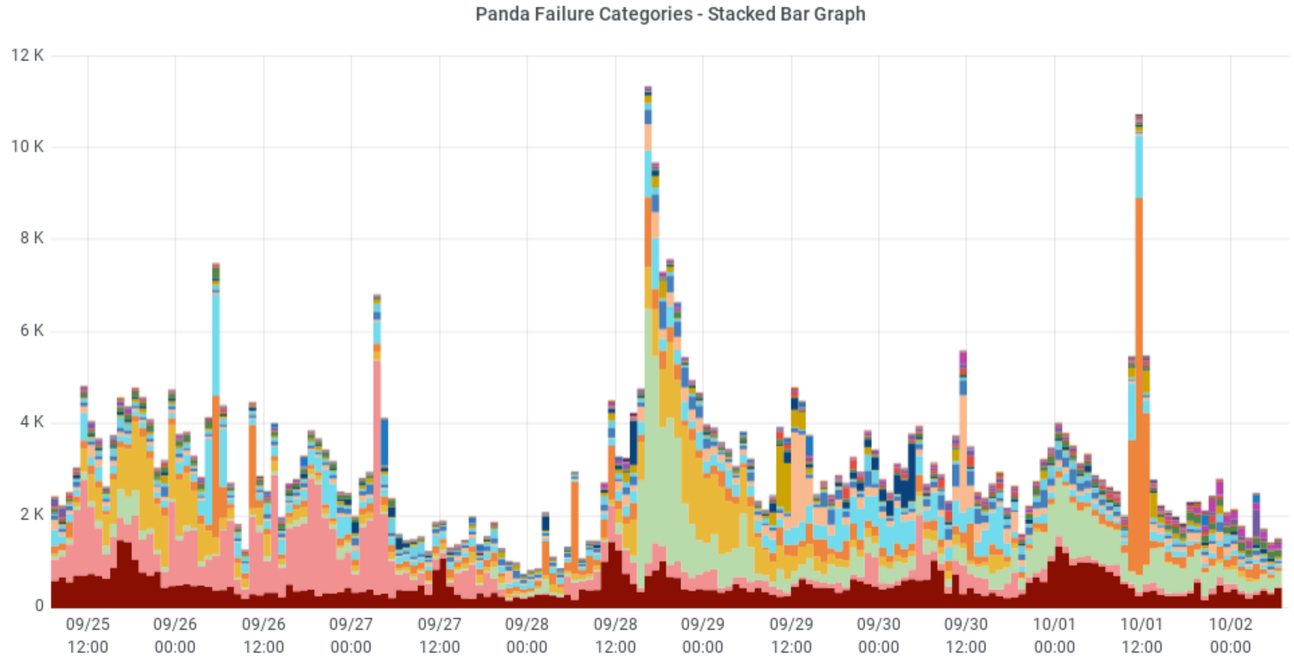
На 2020 рік - 140-150 тисяч задач, запущених паралельно, приблизно стільки ж - в стані очікування

Слоти задач ATLAS



	min	max	avg	total
CERN-P1	0	86.9 K	82.5 K	13.9507 Mil
CERN-T0	0	38.7 K	37.1 K	6.2683 Mil
NERSC_Cori_p2_mcore	0	136.0 K	29.6 K	4.9958 Mil
BNL	0	24.8 K	23.4 K	3.9462 Mil
NERSC_Cori_p2_ES_Test	0	256.0 K	21.6 K	3.6489 Mil
MWT2	0	21.4 K	18.6 K	3.1397 Mil
FZK-LCG2	0	14.7 K	12.9 K	2.1806 Mil
SiGNET	0	32.6 K	11.8 K	1.9890 Mil
IN2P3-CC	0	13.1 K	11.0 K	1.8599 Mil

Статистика експерименту ATLAS: помилки



	min	max	avg	total
Pilot Error	165	1.481 K	500	84.582 K
Athena/Exec Error: Athena	16	5.079 K	485	81.936 K
Execution Error 65	18	5.524 K	427	72.100 K
Transformation Error: not installed in CE	5	2.039 K	343	57.946 K
TaskBuffer Error 300	45	8.176 K	321	54.332 K
DDM Error	25	2.186 K	294	49.634 K
Athena/Exec Error: Proot	11	2.540 K	162	27.459 K
Pilot/DDM Error: get error	22	595	124	20.875 K
Transformation Error	7	361	80	13.573 K

Приблизно 3–8 тисяч задач на годину закінчується з помилками

Розслідуваннями помилок, оптимізацією процесу займається окрема команда (DPA: Distributed Physics and Analysis)

Зберігання і доступ до даних

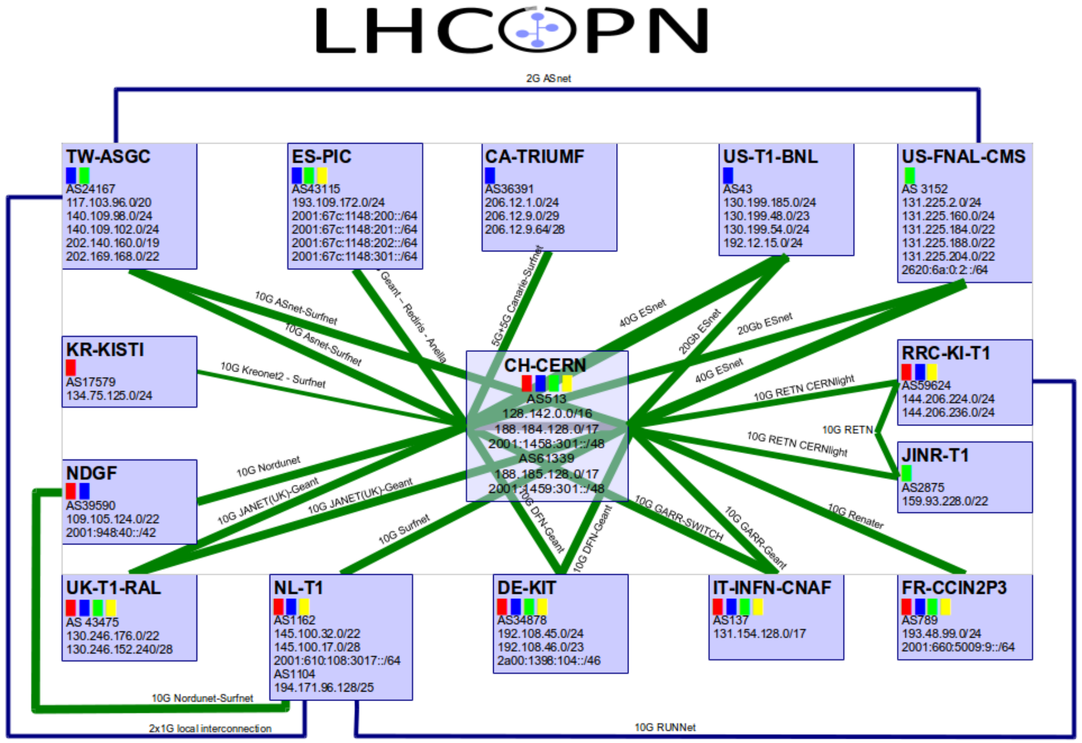
Системи зберігання даних

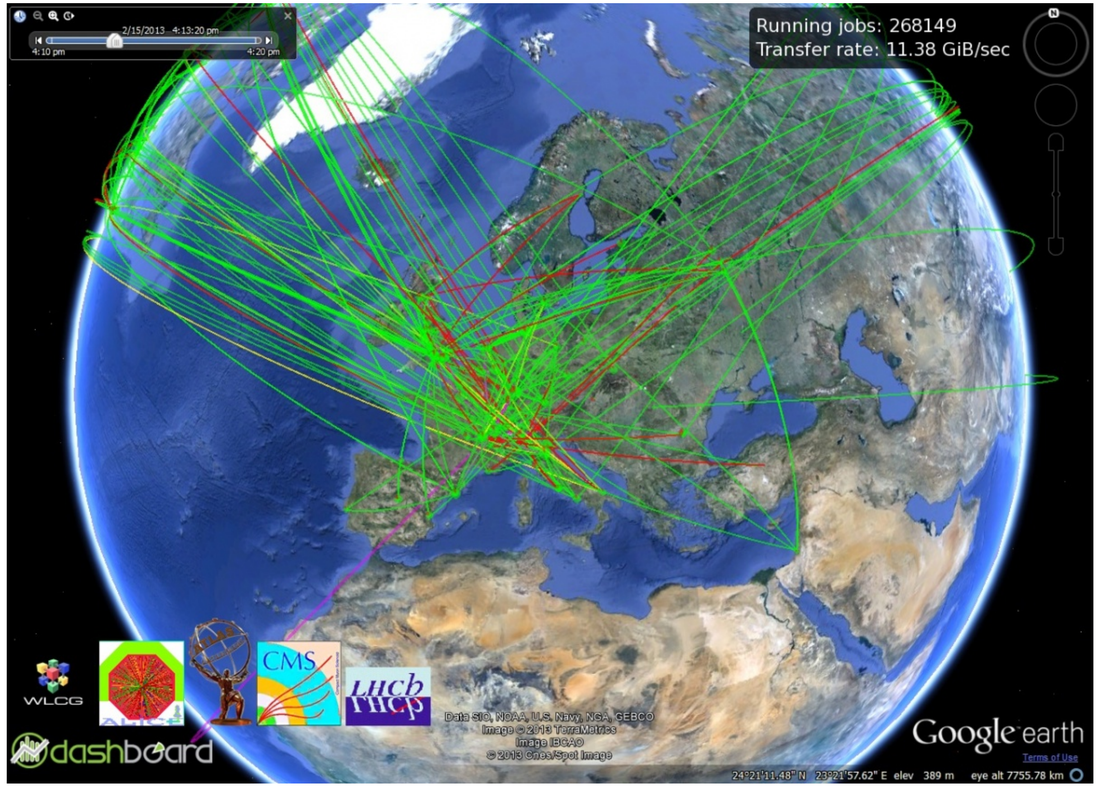
Для зберігання даних використовуються наступні системи:

- EOS - на жорстких дисках, працює за протоколом XROOTD
- CASTOR - використовуються плівкові носії. Переваги: економія електроенергії, дешевизна. Недоліки: швидкість доступу.

Загальний об’єм плівкових накопичувачів у CERN - близько 100 Пб (на січень 2013 року)

Обмін даними між Tier0-Tier1-Tier2





OUTDATED!!!

Трансфер даних

<u>Avg throughputs</u>

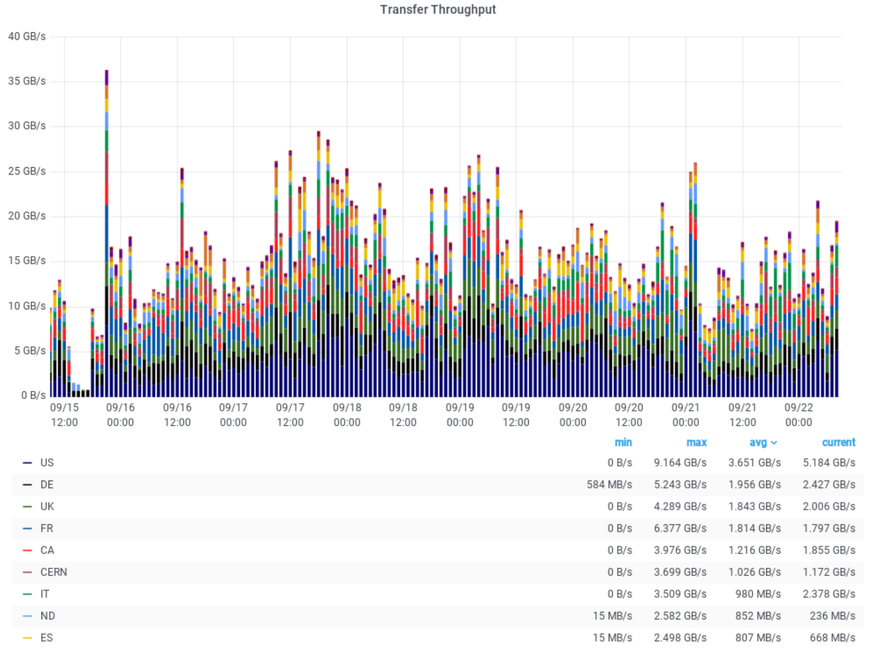
transfer: 18 GB/sec

staging: 2 GB/sec

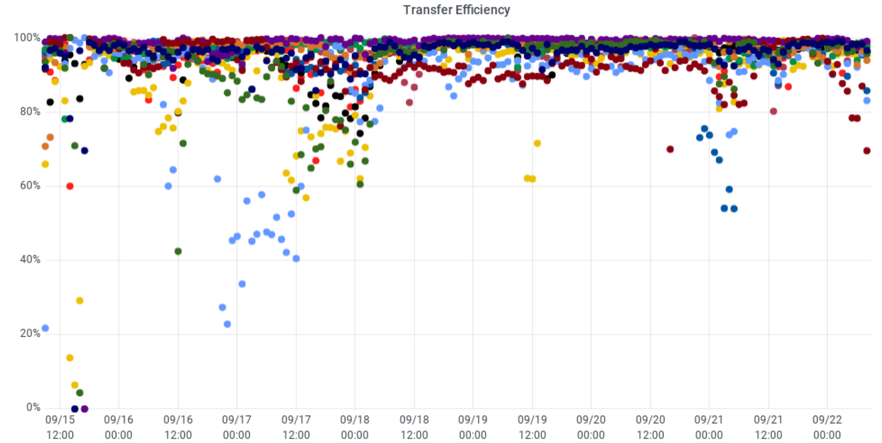
deletions: 20 GB/sec

– – – – – – – –

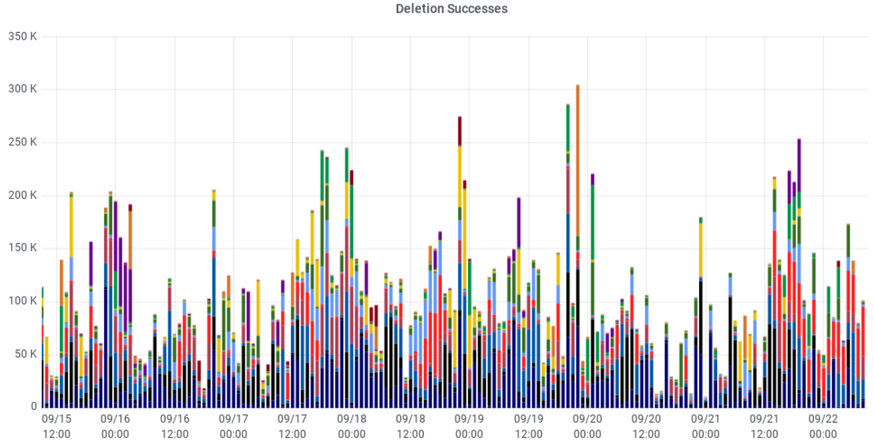
- Rucio 1.23.6 validated and deployed today on production nodes



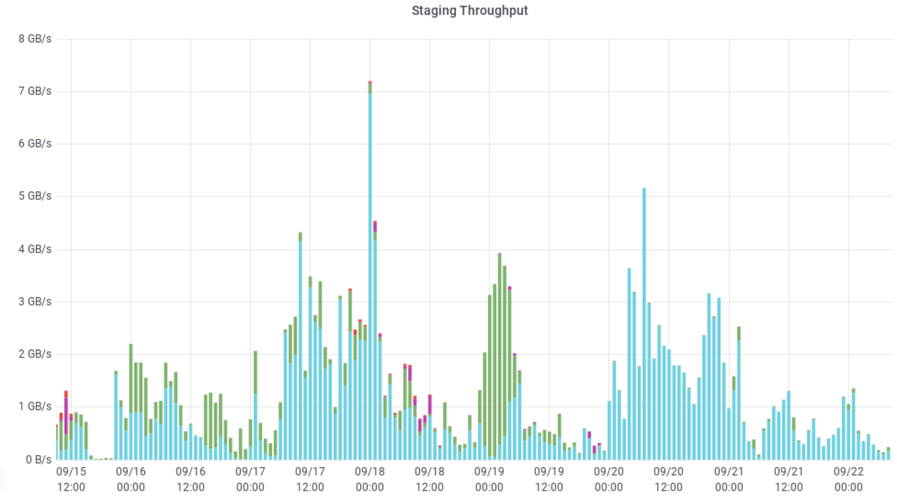
	min	max	avg	current
US	0 B/s	9.164 GB/s	3.651 GB/s	5.184 GB/s
DE	584 MB/s	5.243 GB/s	1.956 GB/s	2.427 GB/s
UK	0 B/s	4.289 GB/s	1.843 GB/s	2.006 GB/s
FR	0 B/s	6.377 GB/s	1.814 GB/s	1.797 GB/s
CA	0 B/s	3.976 GB/s	1.216 GB/s	1.855 GB/s
CERN	0 B/s	3.699 GB/s	1.026 GB/s	1.172 GB/s
IT	0 B/s	3.509 GB/s	980 MB/s	2.378 GB/s
ND	15 MB/s	2.582 GB/s	852 MB/s	236 MB/s
ES	15 MB/s	2.498 GB/s	807 MB/s	668 MB/s



	min	max	avg	current
ND	22%	100%	87%	83%
ES	6%	99%	89%	95%
UK	4%	100%	92%	96%
TW	70%	100%	93%	70%
FR	54%	100%	95%	86%
DE	74%	99%	95%	96%
US	0%	99%	95%	97%
CA	60%	100%	96%	96%
CERN	80%	100%	97%	98%

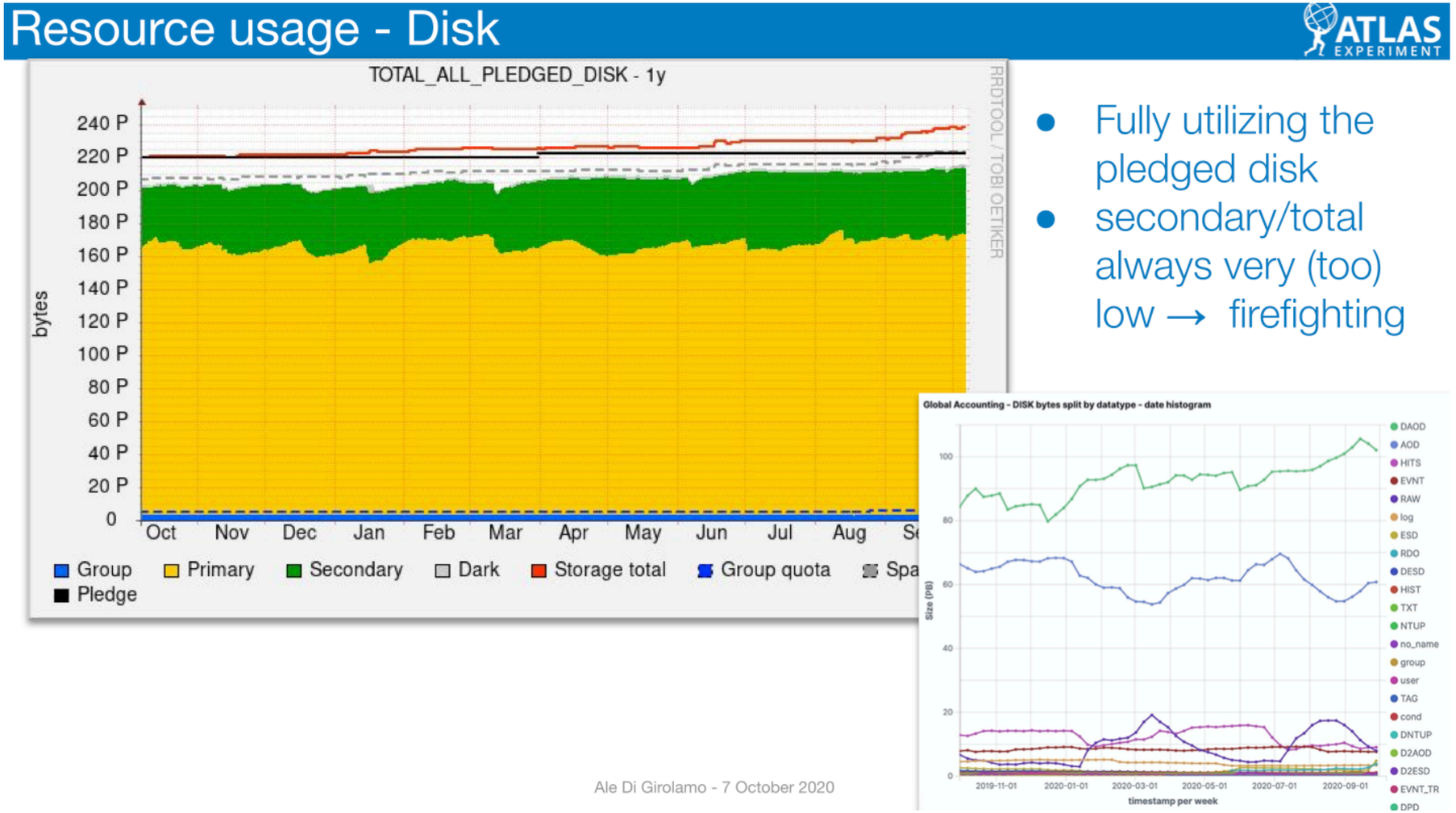


	avg	total
US	19.9 K	3.3390 Mil
DE	15.9 K	2.6647 Mil
FR	13.0 K	2.1838 Mil
CERN	12.0 K	2.0087 Mil
CA	10.4 K	1.7548 Mil
ND	8.4 K	1.4066 Mil
UK	7.8 K	1.3033 Mil
ES	7.2 K	1.2160 Mil
IT	5.2 K	865.5 K



	min	max	avg	current
Production Input	0 B/s	6.956 GB/s	1.012 GB/s	166 MB/s
Analysis Input	0 B/s	3.644 GB/s	310 MB/s	74 MB/s
User Subscriptions	0 B/s	699 MB/s	17 MB/s	0 B/s
Staging	0 B/s	132 MB/s	6 MB/s	0 B/s
Data Consolidation	0 B/s	2 MB/s	12 kB/s	0 B/s
T0 Export	0 B/s	0 B/s	0 B/s	0 B/s
Recovery	0 B/s	0 B/s	0 B/s	0 B/s

Заповнення дискових накопичувачів



- Fully utilizing the pledged disk
- secondary/total always very (too) low → firefighting

Керування даними

Даними керує система Rucio

Являє собою каталог файлів, вказує на знаходження файлів і метод доступу

Проводить автоматичне резервне копіювання і знаходження оптимальних копій файлів

Інфраструктура обробки даних

PanDA WMS

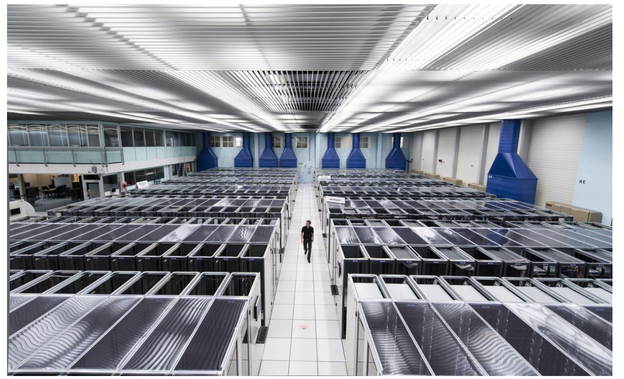
Система PanDA WMS (Production and Distributed Analysis) дозволяє будувати обчислювальні системи з різних типів обчислювальних елементів

Дозволяє ефективно розподіляти задачі між обчислювальними кластерами, а також проводить менеджмент даних до початку виконання завдання

Наразі використовується в експериментах ATLAS і COMPASS, планується використання в CMS

Датацентр CERN

- Забезпечує функціонування основних сервісів (електронна пошта, відеоконференції, керування даними)
- 10 тис. серверів, загалом 110 тис. процесорів
- Щодня обробляє близько 1 Пб інформації (бл. 210,000 DVD)
- Для обміну даними використовується швидкісний оптичний кабель загальною довжиною 35 тис. км.



Ґрід-обчислення

Ґрід є формою розподілених обчислень, в якому багато комп'ютерів об'єднані в один потужний віртуальний комп'ютер, і які працюють разом для виконання трудомістких завдань. Для певних додатків, «ґрід» обчислення можна розлядати як спеціальний тип паралельних обчислень які покладаються на цілі комп'ютери(обладнані процесорами, пам'ятю, живленням, мережевим інтерфейсом і тд.), під'єднані до комп'ютерної мережі(приватної або публічної) звичайним мережевим інтерфейсом, таким як Ethernet.

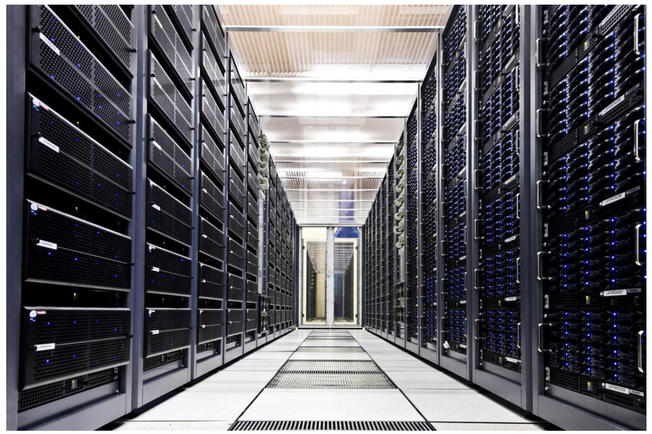
Термін з'явився на початку 1990-х років, як метафора, що демонструє можливість простого доступу до обчислювальних ресурсів як і до електричної мережі (англ. Power grid)

WLCG: The Worldwide LHC Computing Grid

До 2006 року - LCG (LHC Computing Grid).
Міжнародна колаборація, яка підтримує ґрід-інфраструктуру зі 170 обчислювальних центрів у 36 країнах (дані 2012 року). Інфраструктура розроблена спеціально для обробки даних з Великого адронного колайдера.

Найбільший в світі обчислювальний ґрід.
Складається з інфраструктур:

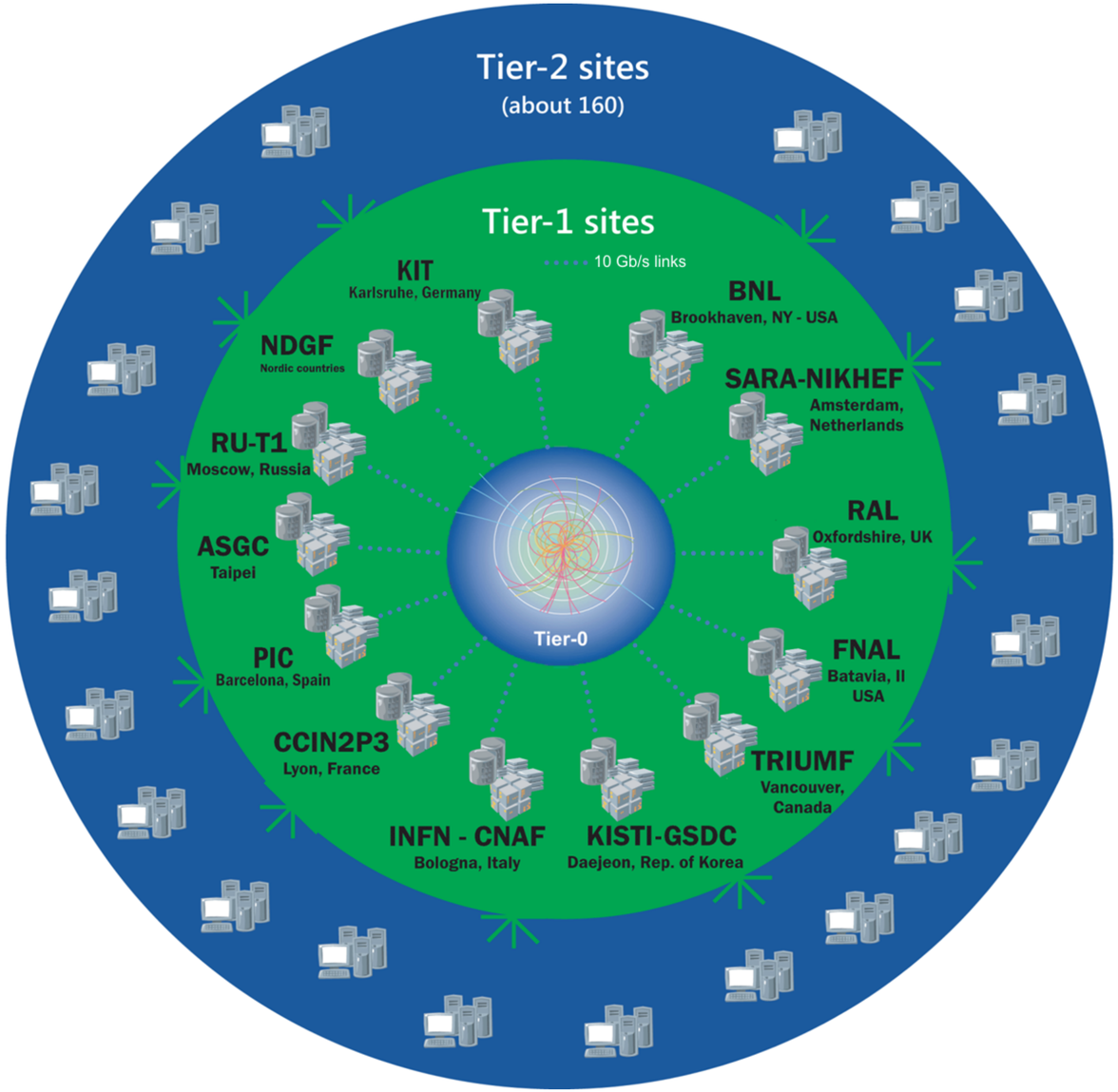
- European Grid Infrastructure
- Open Science Grid (США)
- TWGrid (Тайвань)
- EU-IndiaGrid (ґрід-інфраструктури Європи та Азії)
- NorduGrid (скандинавські країни)



Надає єдиний доступ до обчислювальних та дискових ресурсів, інструментарію візуалізації, тощо. Розробляє вимоги до роботи ресурсів, аутентифікації користувачів.

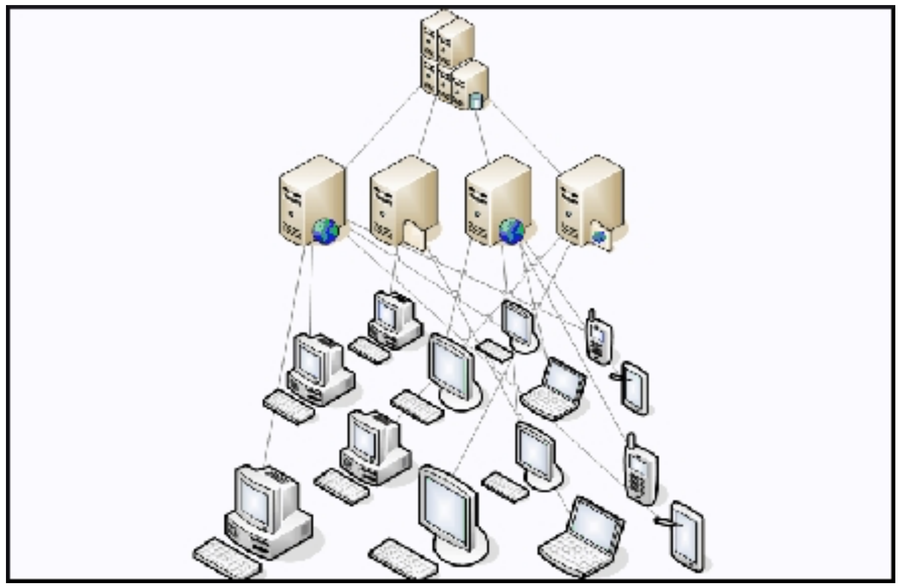
TIER-центри

- Tier-0: знаходяться в CERN та в датацентрі Wigner (до 2019 року). Відповідають за збереження "сирих даних" (перша копія даних), перший прохід реконструкції та взаємодію з Tier-1. В періоди простою ВАК беруть участь у загальній обработці даних. (~15% обчислювальних потужностей)
- Tier 1: великі комп’ютерні центри з відповідними обчислювальними можливостями, також зберігають великі об’єми даних. Відповідають за взаємодію з обчислювальними ресурсами Tier-2 (~40% обчислювальних потужностей).
- Tier 2: університети чи наукові інститути, що збрерігають достатньо інформації та надають обчислювальні потужності для виконання необхідних задач з аналізу даних. (близько 160, ~45% обчислювальних потужностей)
- Tier 3: окремі комп’ютери чи локальні кластери.



Волонтерські обчислення

- Комп’ютерні потужності надають користувачі персональних комп’ютерів, тощо.
- Використовуються для математичних розрахунків (наприклад, пошук простих чисел), біологічні розрахунку
- В CERN використовує експеримент ATLAS, хоча на даний момент ефективність такого типу обчислень не є великою



Кластер TDAQ

- бл. 100 тис. ядер (надає 20-25% всіх обчислювальних потужностей ATLAS)
- Встановлений в шахті ATLAS, звичайно використовується для первісної обробки і реєстрації даних, отриманих з детектора

Суперкомп’ютери

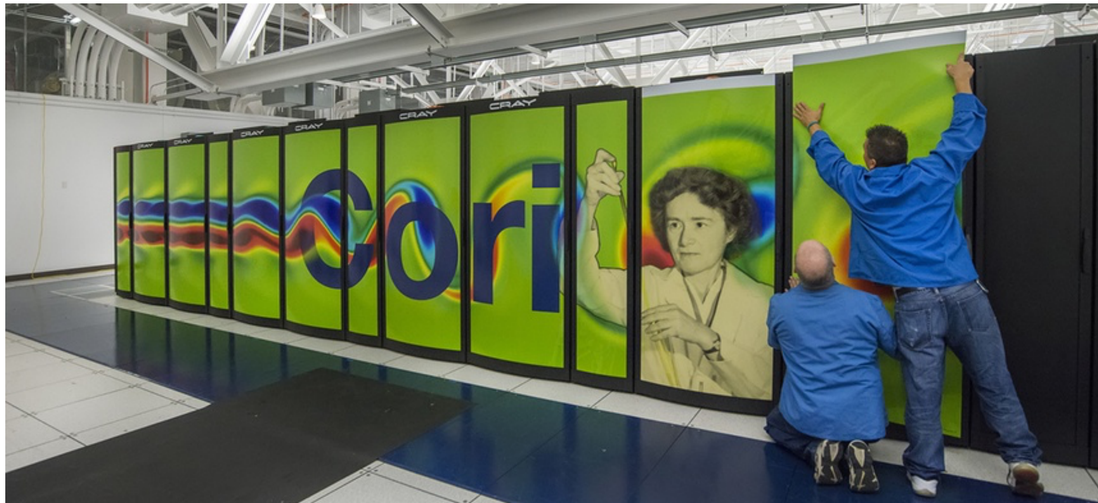
Комп’ютери з великими обчислювальними потужностями порівняно зі звичайними. Швидкодія розраховується у кількості операцій над числами з плаваючою точкою на секунду (FLOPS).

Вперше побудовані у 1960-х роках компанією Cray Research.

З 1990-х років суперкомп’ютери використовують тисячі процесорів, в 2000-х - вже десятки тисяч.

Використовуються для розрахунків прогнозів погоди, молекулярної динаміки, ядерних реакцій, комп’ютерної безпеки.

Суперкомп’ютери в ATLAS



ORNL Titan

Суперкомп’ютер в лабораторії Oak Ridge National Lab, США. Відкрито 2012 року. На момент пуску займав першу сходинку в світі.

6 “критичних кодів”:

- Молекулярна динаміка (LAMMPS)
- Молекулярна фізика (S3D)
- Моделювання ядерних реакцій (Denovo)
- Глобальні атмофсерні моделювання (CAM-SE)
- Астрофізика (NRDF)
- Термодинаміка (WL-LSMS)

ORNL Summit

Запущено в червні 2018 року, позиція #5 світовому рейтингу

Нова архітектура процесорів: IBM Power9

Пікова продуктивність - ~200 ПФлопс

Енергоспоживання - 13 МВт (Titan - 7 МВт)

MareNostrum

MareNostrum4 запущено в червні 2017 року, знаходиться в Барселонському центрі суперкомп’ютингу (BSC)

Пікова продуктивність - ~14 ПФлопс, на момент запуску: #13 в світі

На 2018 рік - #1 серед європейських суперкомп’ютерів за енергоефективністю

Архітектури: IBM Power9, AMD x86_64, ARM, NVIDIA Volta GPUs

В даний момент монтується MareNostrum5, очікувана продуктивність: ~200 ПФлопс

Українська інфрастуруктура

Український Грід

Розпочав роботу 2006 року.

На поточний момент грід інфраструктура України об’єднує 38 кластерів з загальною кількістю ядер більше за 2900 и доступним дисковим об’ємом 250TB.

Кластери працюють під управлінням Nordugrid ARC.

Використовується для досліджень у сферах фізики, біології, медицини та ін.

Україна	BITP ARC Training	11	0+0	0+0
	BITP Cluster	88	0+34	0+0
	CHIMERA	120	0+0	116+0
	CSTU ARC CE	4	0+0	0+0
	DFTI Cluster	112	0+44	1+0
	HPC and FOSS Center	13	0+0	0+0
	IAP Cluster	16	0+8	0+0
	IAPMM Cluster	16	4+8	0+0
	ICMP Cluster	192	0+89	0+0
	ICYB SCIT-3	1036	32+504	0+0
	IEP Cluster	48	0+0 (queue inactive)	0+0
	IFBG Cluster	72	16+41	0+0
	ILTPE ARC UA	88	0+0	1+0
	ILTPE Cluster	88	0+62	0+0
	IMAG cluster	44	0+0	0+0
	IMATH Cluster	16	0+2	0+0
	IMBG ARC	100	96+0	2+0
	IMMSP Cluster	40	0+0	6+0
	IMP ARC CE	84	0+0	0+0
	INPARCOM Cluster	8	0+0	0+0
	INPARCOM GPU Cluster	8	0+0	0+0
	IOP Cluster	104	0+0	0+0
	IPM Cluster	44	0+5 (queue inactive)	0+0
	IPMS Cluster	20	0+0	0+0
	IRE Cluster	64	0+0	0+0
	ISMA cluster	332	0+305	0+39
	ISOFTS Cluster	8	0+0	0+7605
	KIPT IPP	2	0+0	0+0
	KMA Grid Cluster	0		0+0
	KNU ARC	40	0+37	13+11
	KPI training cluster	24	0+0	0+0
	LNU Training Cluster	28	0+28	0+0
	MAO Cluster	104	0+48	0+0
	MHI Cluster	120	0+0	0+0
	PIMEE ARC	24	16+0	6+0
	RIAN	1	0+0	0+0
	SRI cluster	4	0+0	0+0

Українські суперкомп’ютери

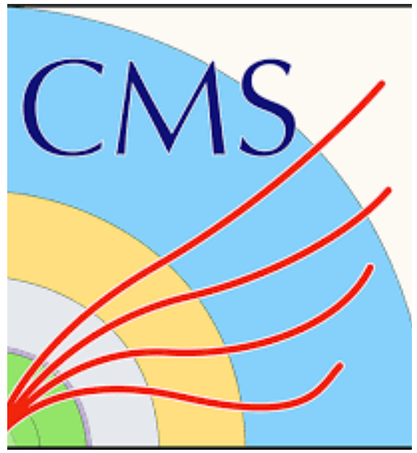
- “СКІТ-3” та “СКІТ-4”, встановлені в Інституті кібернетики НАН України.
- На момент запуску (2012 рік) входили до рейтингу “Top-50” на пострадянському просторі (43 TFlops, 170 Тб файлового сховища, пам’ять: 2.5 Тб, 716+448 ядер)
- Залучені до розрахунків у експерименті ALICE
- Серед інших проектів: медична кібернетика, комп’ютерний моніторинг грунтів та підземних вод, математичне моделювання літальних апаратів

Участь в експериментах CERN

BITP: Інститут теоретичної фізики ім. Боголюбова НАН України

KNU: Київський національний університет

ICYB: Інститут кібернетики ім. Глушкова НАН України (CKIT-3/CKIT-4)



UA-ISMA: Інститут сцинтиляційних матеріалів НАН України

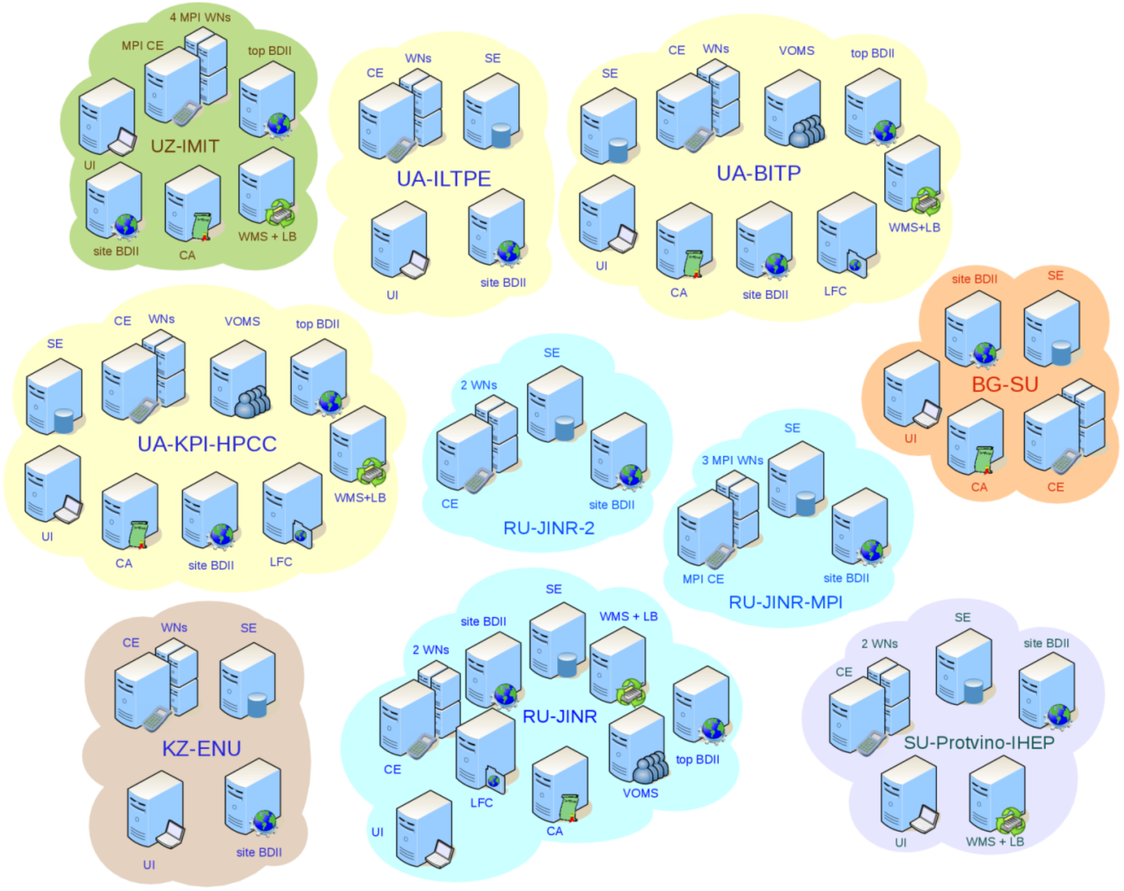
UA-KIPT-LCG2: Харківський фізико-технічний інститут

Навчальна інфраструктура

Однією з основних проблем Українской грід-спільности є нестача спеціалістів, що вміють використовувати Грід-технології для наукових розрахунків.

На базі НТУУ КПІ організовано дісплейний клас для підготовки грід-адміністраторів.

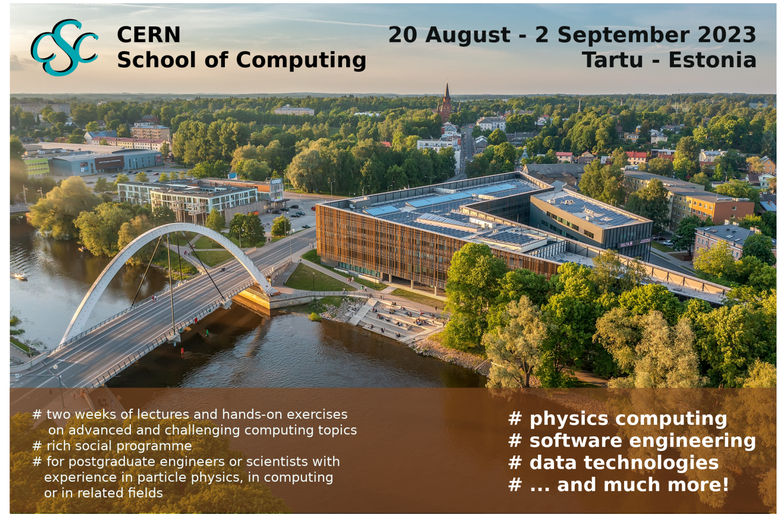
Інститут теоретичної фізики ім. Боголюбова проводить відеоконференції для грід-користувачів.



Освітні проекти CERN

CERN School of Computing (CSC)

- Кожного року проходить в новому університеті
- Кількість учасників: бл. 60 чол.
- Тематика:
  - паралельне програмування
  - технології зберігання даних
  - ефективне програмування і оптимізація програмного забезпечення
  - безпека
  - машинне навчання
  - аналіз фізичних даних
- Остання: в Кракові (2022), 2023 - університет Тарту
- Також існує iCSC (Inverted Cern School of Computing)

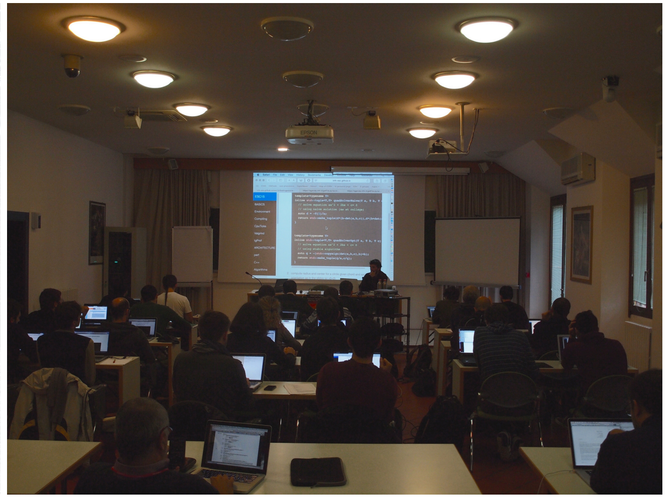


Thematic CSC (tCSC)

- Проходить кожного року у Спліті (Хорватія), організується факультетом електротехніки Сплітського університету
- Тематика:
  - ефективне програмування і оптимізація програмного забезпечення
  - паралельне програмування
  - C++ і його особливості

ESC INFN

- Проходить кожного року в Бертіноро (Італія)
- Тематика:
  - ефективне програмування і оптимізація програмного забезпечення, ефективне керування пам’яттю
  - паралельне програмування: OpenMPI та OpenMP
  - паралельне програмування на графічних прискорювачах
  - C++ і його особливості



Заключення

- З кожним запуском LHC кількість зібраних даних збільшується, програмне забезпечення стає все більш складним
- В майбутньому очікуються нові типи комп’ютерних ресурсів (наприклад, квантові комп’ютери) чи суперкомп’ютери нових зразків
- Очікуємо збільшення рівня участі українських студентів і аспірантів в освітніх програмах CERN, а також в програмах Technical Students/ Summer Students

ДЯКУЮ ЗА УВАГУ!

Свірін Павло
ATLAS, CERN
LAPP