

# A Detektortól a Végső Ábráig

*Az adatok feldolgozása...*

Ifj. Krasznahorkay Attila

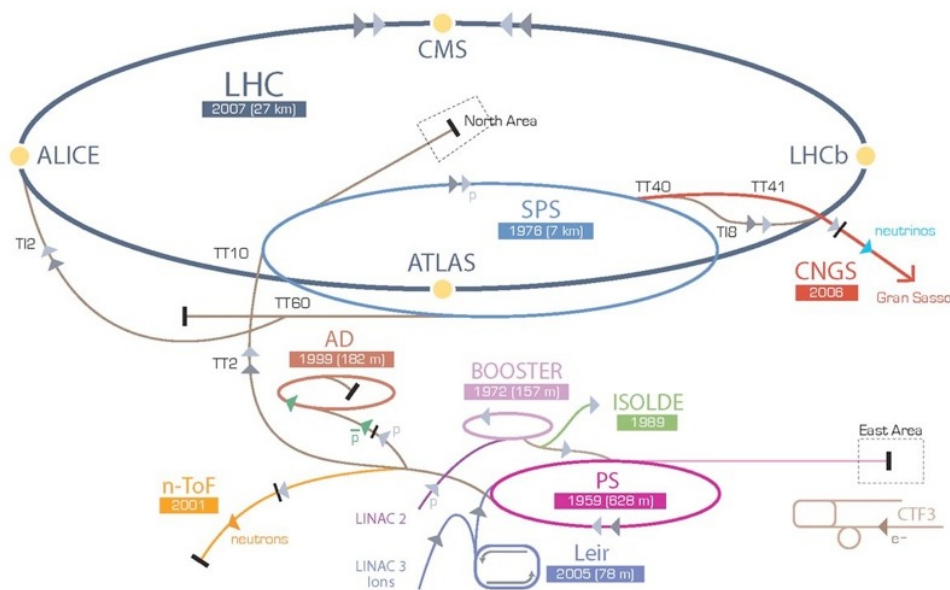


# ATLAS EXPERIMENT

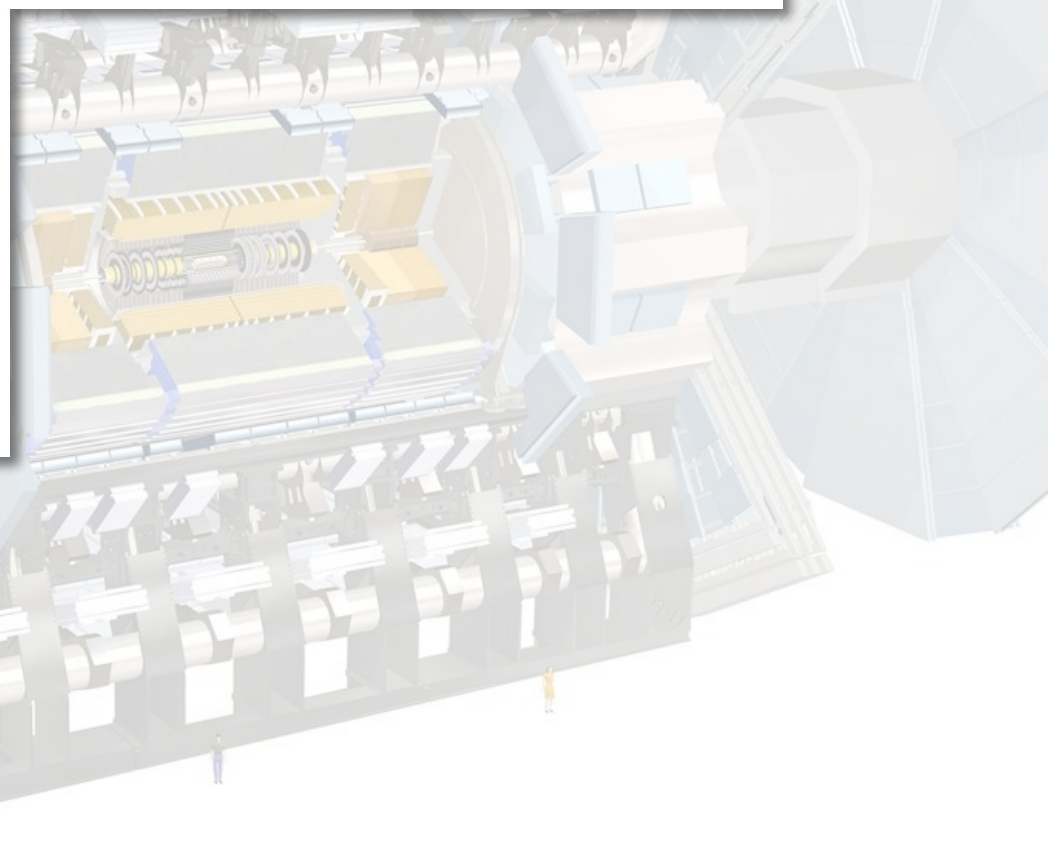
# A Madártásvlat



CERN Accelerator Complex



Létrehozzuk az “érdekes”  
reakciókat  
(Varga Dezső előadása)

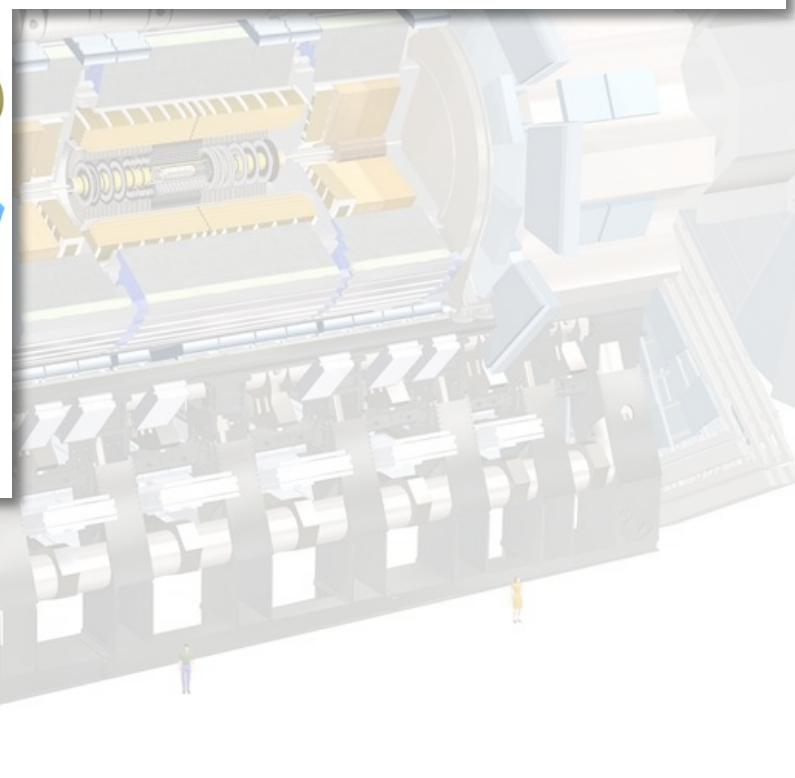
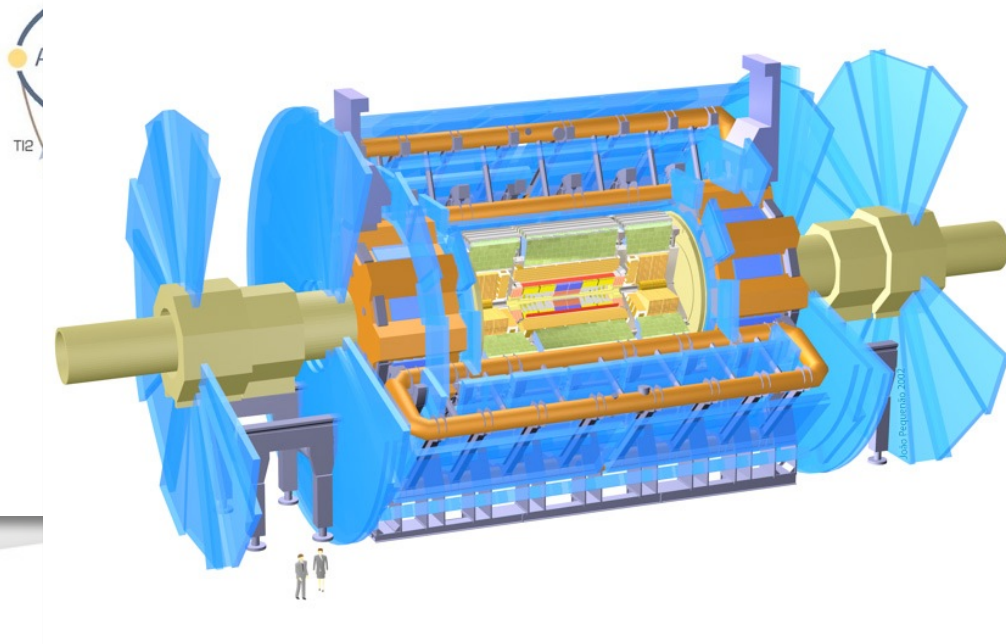


# A Madártávlat



CERN Accelerator Complex

Létrehozzuk az “érdekes”  
Érzékeljük a kijövő  
részecskéket  
(Varga Dezső előadása)

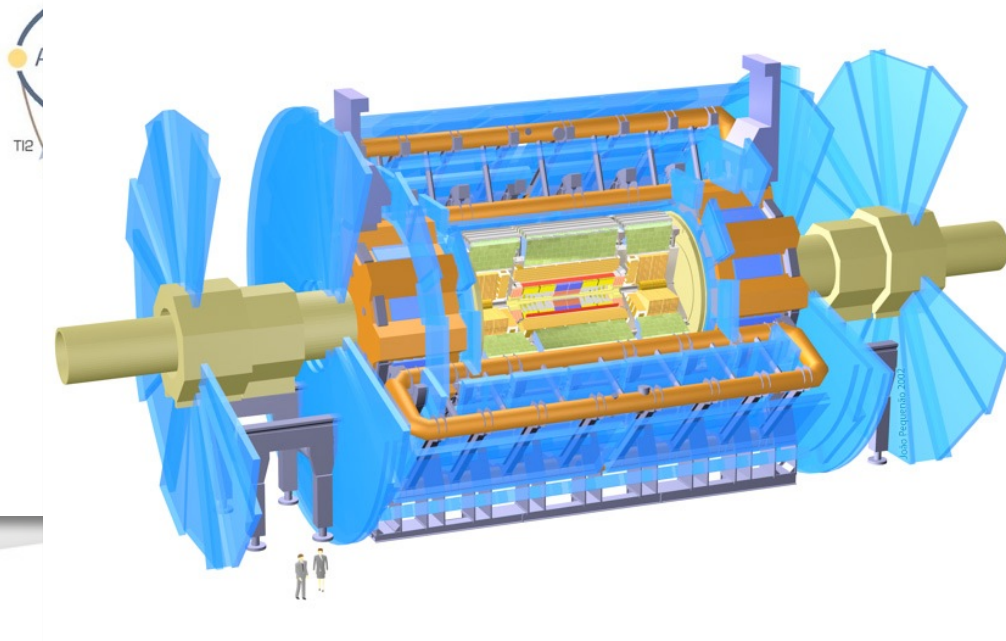


# A Madártávlat

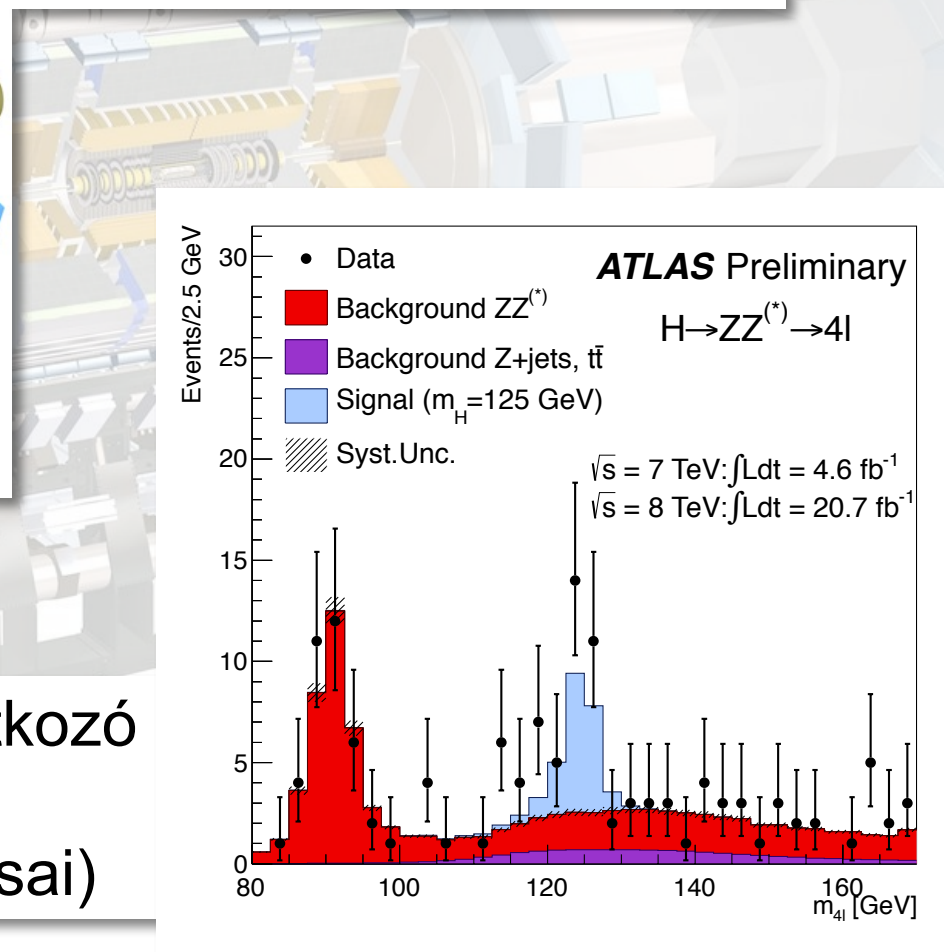


CERN Accelerator Complex

Létrehozzuk az “érdekes”  
Érzékeljük a kijövő  
részecskéket  
(Varga Dezső előadása)



Levonjuk a fizikára vonatkozó  
következtetéseket  
(Horváth Dezső előadásai)

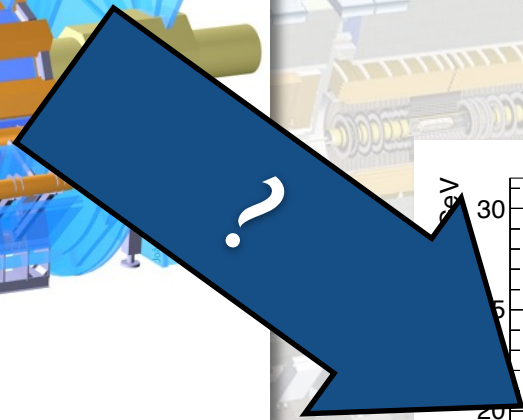
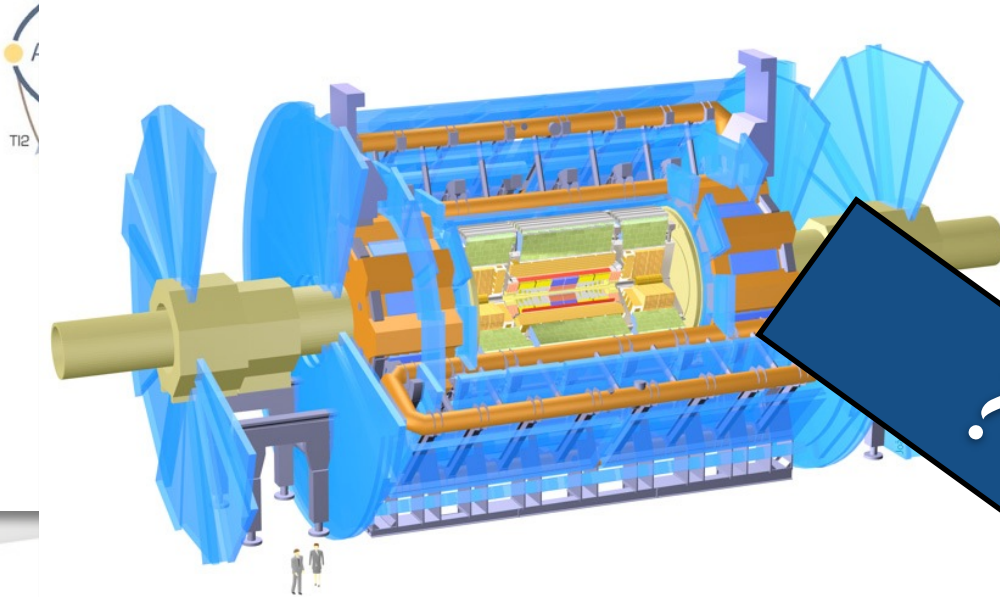


# A Madártávlat

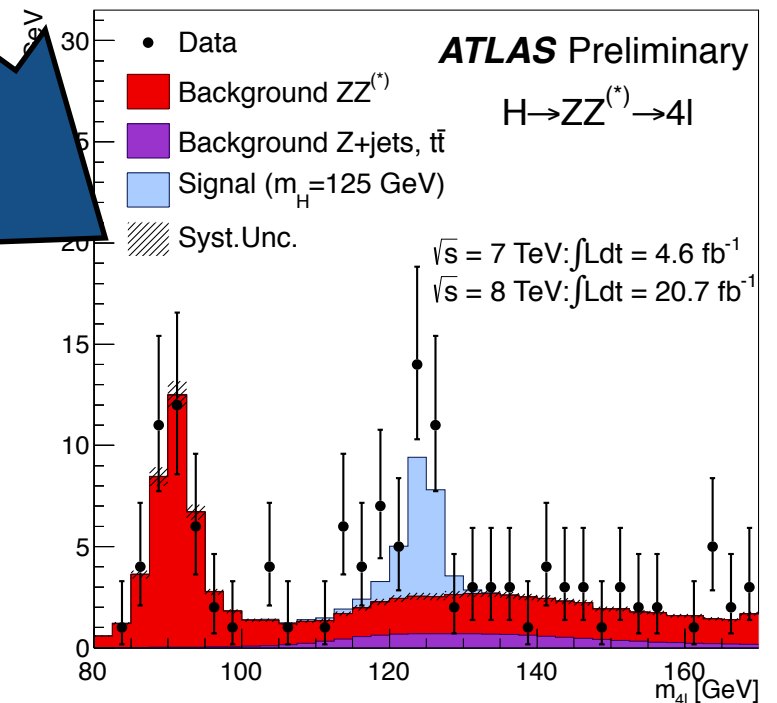


CERN Accelerator Complex

Létrehozzuk az “érdekes”  
Érzékeljük a kijövő  
részecskéket  
(Varga Dezső előadása)



Levonjuk a fizikára vonatkozó  
következtetéseket  
(Horváth Dezső előadásai)



# Adat-analízis Madártávlatból



- Minden adatot és szimulációt ugyanúgy feldolgozunk.
- Olyan mennyiségeket akarunk megmérni az adatokban amiknek a segítségével a fizikára tudunk következtetni.
- Mindezt sok alkalommal végezzük el egymás után, hogy a mérés bizonytalanságait meg tudjuk becsülni
- A kísérleti fizikus diákok/doktoranduszok ennek a menetét tanulják, sokszor éveken keresztül
- A technikai részletekkel pedig még ők is ritkán vannak tisztában...

# Számítástechnikai Alapok



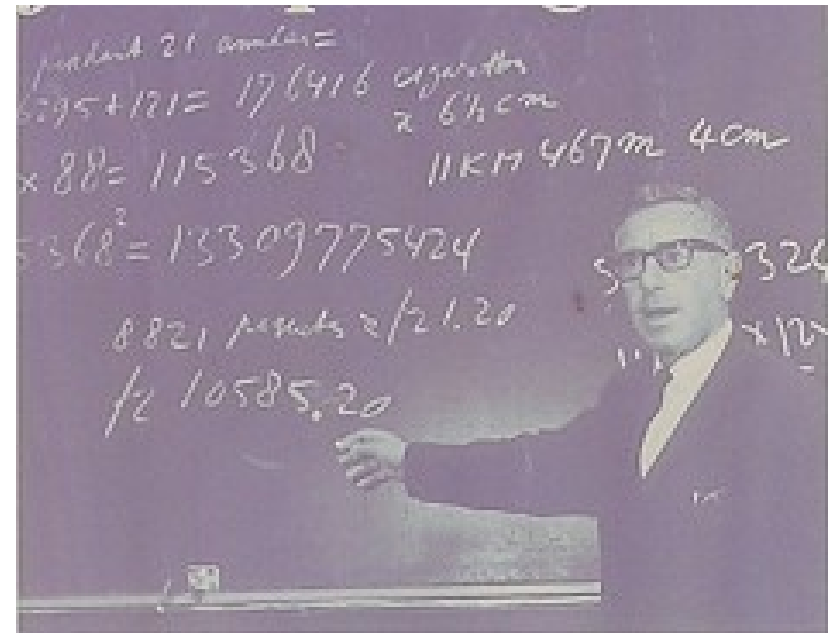
- 1 db ATLAS esemény mérete: kb. 1.6 MB
  - Teljes intenzitás mellett 40 millió eseményt “kapunk” másodpercenként (64 TB/s???)
- Az elsődleges eseményválogatás (trigger) kb. 500 eseményt választ ki másodpercenként
  - Ezzel még mindig >30 TB adatot rögzítünk naponta
- Évente kb. 1 milliárd eseményt rögzítünk, és kb. 3 milliárd eseményt szimulálunk
  - 1 esemény teljes szimulálása: kb. 5 perc
  - 1 esemény rekonstruálása: kb. 30 másodperc
- -> Sok-sok millió “processzor órát” és kb. 50 PB tárhelyet használunk
  - Pénzre inkább nem is fordítanám ezt le...

A detailed 3D cutaway illustration of a particle accelerator, likely the Large Hadron Collider (LHC). The diagram shows the complex internal structure, including the main tunnel, various magnets, and the central collision point. The structure is rendered in shades of blue, grey, and yellow. Several small human figures are placed throughout the structure to provide a sense of scale, highlighting the massive size of the machine. The text is overlaid in the center of the image.

# A CERN-i Számítástechnika (Rövid) Története



# A kezdetek...



Wim Klein

Képes volt egy 133 számjegyből álló szám 19. gyökét fejben meghatározni. Egyszerű matematikai műveletekben jóval gyorsabb volt kora számítógépeinél.



# Az első számítógépek



1958–61: AZ első valódi nagyszámítógép a CERN-ben, a MERCURY. Két 40 bites szám összeadása 180, szorzása 300 us-ig tartott. Lyukszallag programozás. 1966-ban elajándékozták a lengyel 'ásványtani és bányászati' minisztériumnak.



# Az első számítógépek



1961-63: Az IBM709 4x gyorsabb a MERCURY-nál. Mágnes szallagos egység 200 bpi sűrűséggel ír/olvas.

Támogatja a FORTRAN programozási nyelvet !



# Megjelent az Internet...

Mi az Internet ? – Az Internet összekapcsolt számítógépek millióinak rendszere. Neve az 'Interconnected networks' kifejezésre utal. A kapcsolat lényege nem a kábel vagy a rádiókapcsolat, hanem a közös nyelv (protokoll).

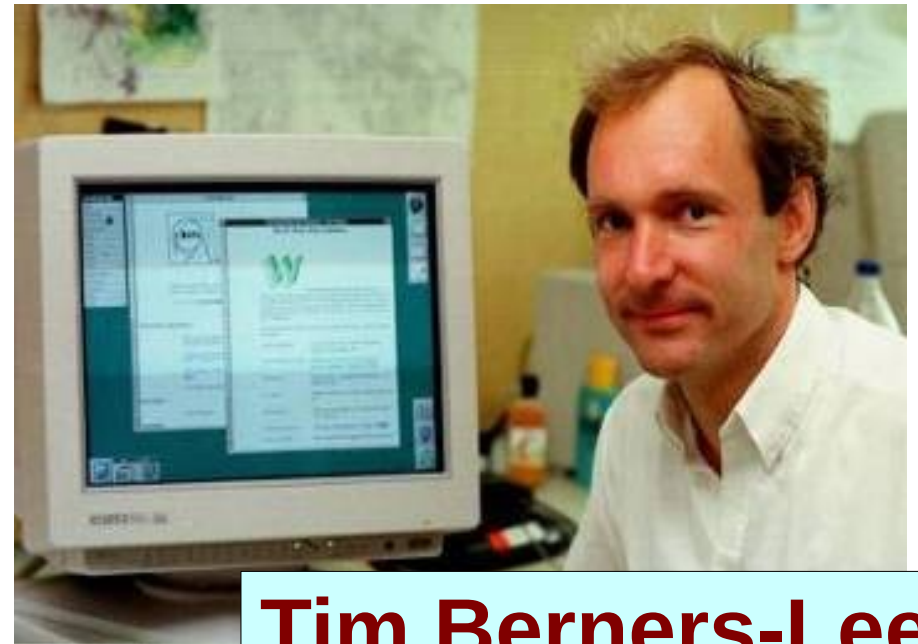
A protokoll – A protokoll az egymással való kapcsolattartás és információcsere szabályainak összessége. Az Interneten való kommunikációhoz a Transmission Control Protokol/Internet Protokolt használjuk.

A TCP/IP – A TCP/IP óriási előnye, hogy rengeteg, teljesen különböző hálózaton használható. Ez a tulajdonsága kulcsfontosságú volt az Internet gyors elterjedésében.

# ...és a WEB.

A World Wide Web (WWW) – 'csak' egy, az Internetet használó szolgáltatás. (WEB  $\neq$  Internet !!!). Segítségével információt oszthatunk meg, tehetünk nyilvánossá. A felhasználó a dokumentumokat feltölti egy webszerverre, amelynek egyedi címe (web address, vagy Uniform Resource Locator, URL) segítségével utalhatunk dokumentumokra.

Kezdetben – az információmegosztás statikus és egyirányú volt. Manapság a többirányú információmegosztást is támogató interaktív weboldalak, portálok, blogok uralják az internetet.



**Tim Berners-Lee**

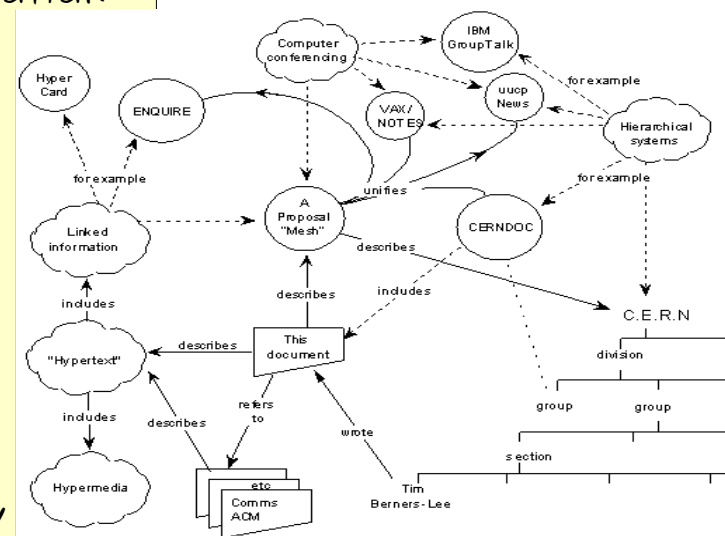
# Miért is a CERN-ben fejlesztették ki a Web-et ?

- A tudományos kutatásban rendkívül fontos az információ és az ötletek szabad áramlása.
- A CERN 80 országában dolgozó 6500 kutató nagy közös barkácsműhelye.
- Az LHC tervezésekor kritikus fontosságúvá vált az információ gyors, könnyű és globális megosztásának megvalósítása.

- 1989: Tim Berners-Lee javaslata az LHC információs rendszerére

- 1991: első www rendszerek

- 1993: első és sokáig népszerű Mosaic browser, ekkor már 500 webszerver, 1%-os forgalom ! :-)

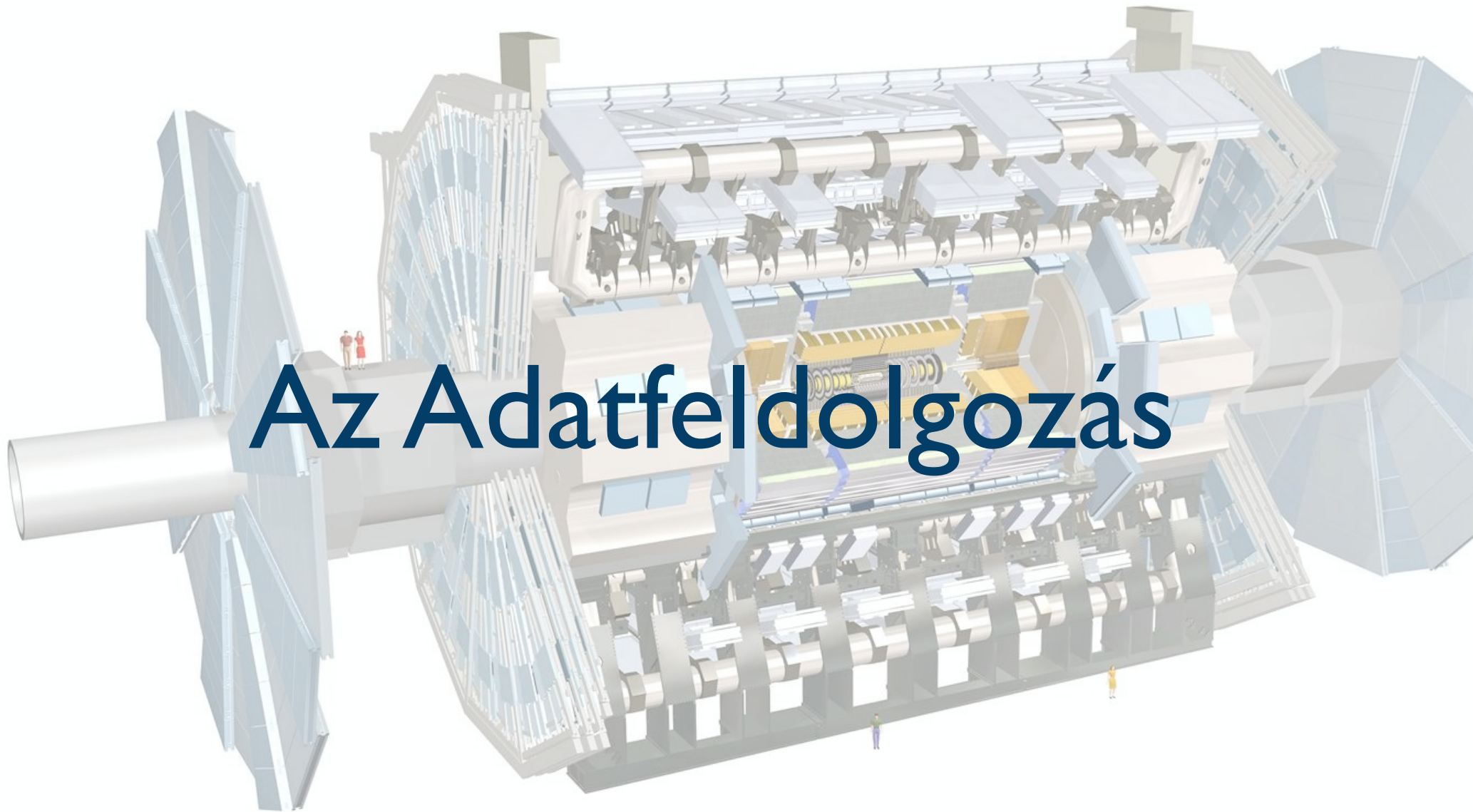


# És mi a Grid ?

A Grid - egy, az Internetre épülő szolgáltatás, csakúgy mint a Web. A Gridhez kapcsolt eszközök nem csak információt, osztanak meg, hanem tárterületet, számítási kapacitást, adatszázis információt, alkalmazásokat, hálózati forgalmat is !

- **Neve** az elektromos hálózatok analógiájára lett kitalálva.
- **Hasonlat:** Fogasztók és szolgáltatók → kenyérpirító és erőmű
- **Ötlet:** Évtizedek óta létezik, de globális méretekben csak most valósult meg.





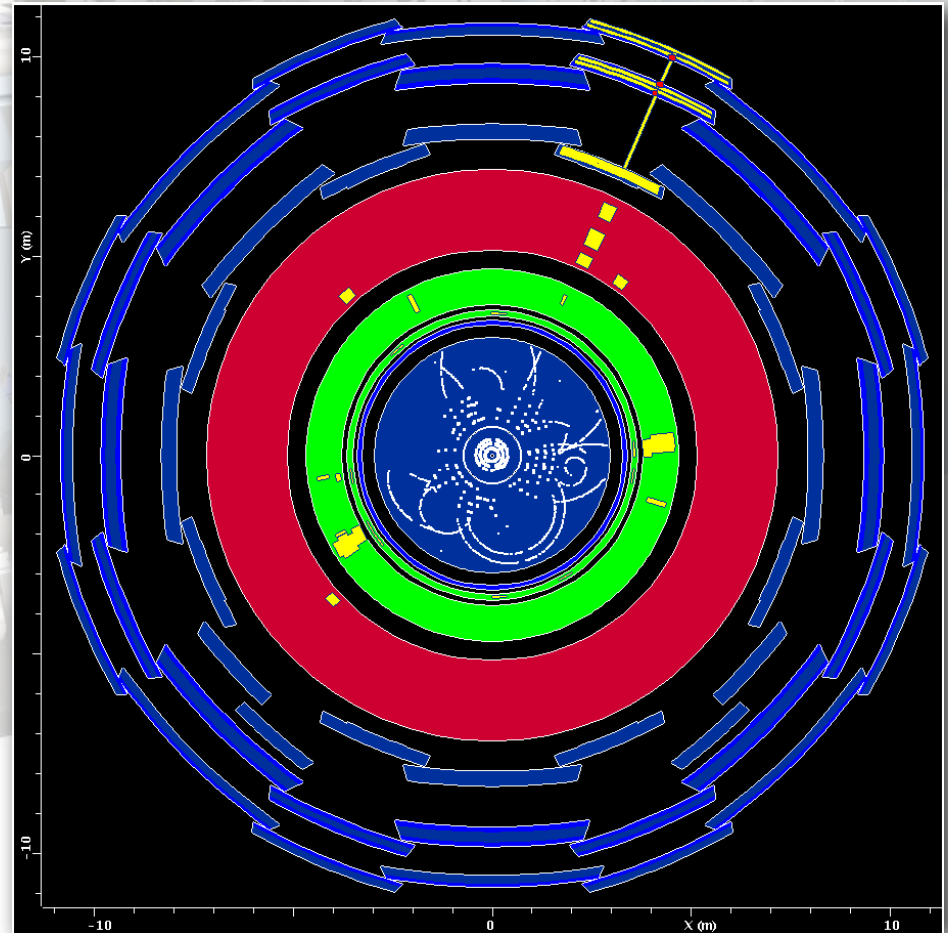
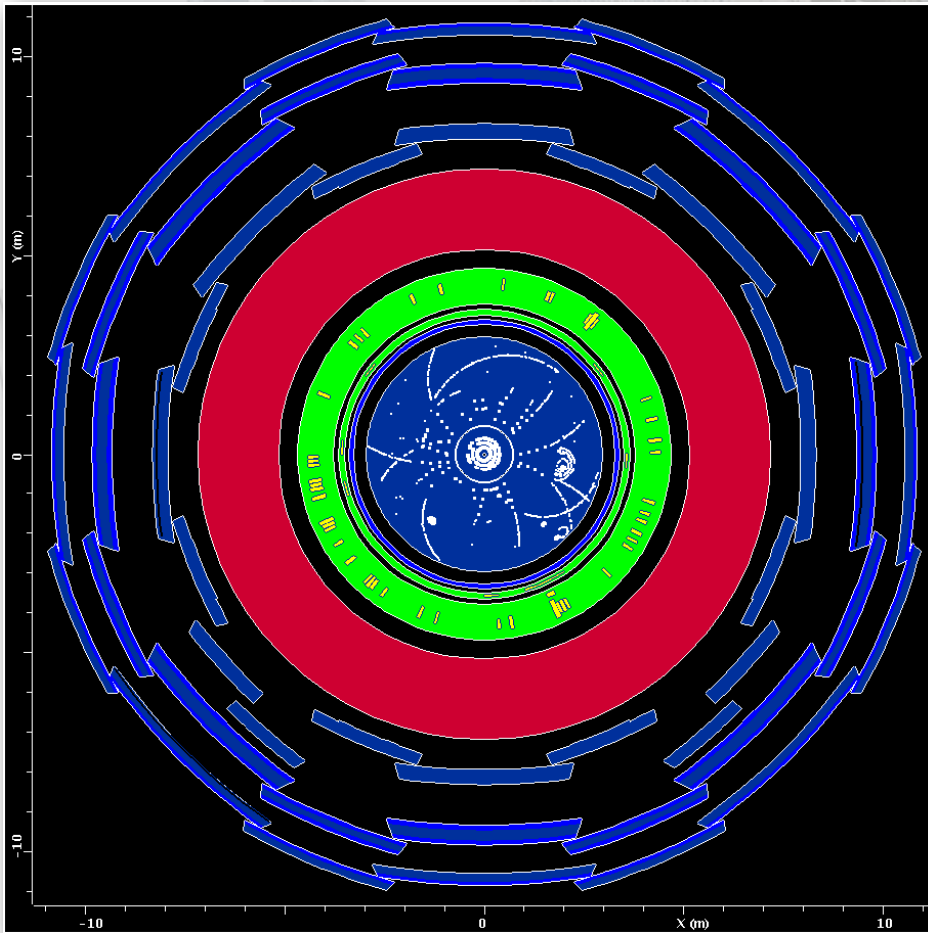
# Az Adatfeldolgozás



# Adatgyűjtés



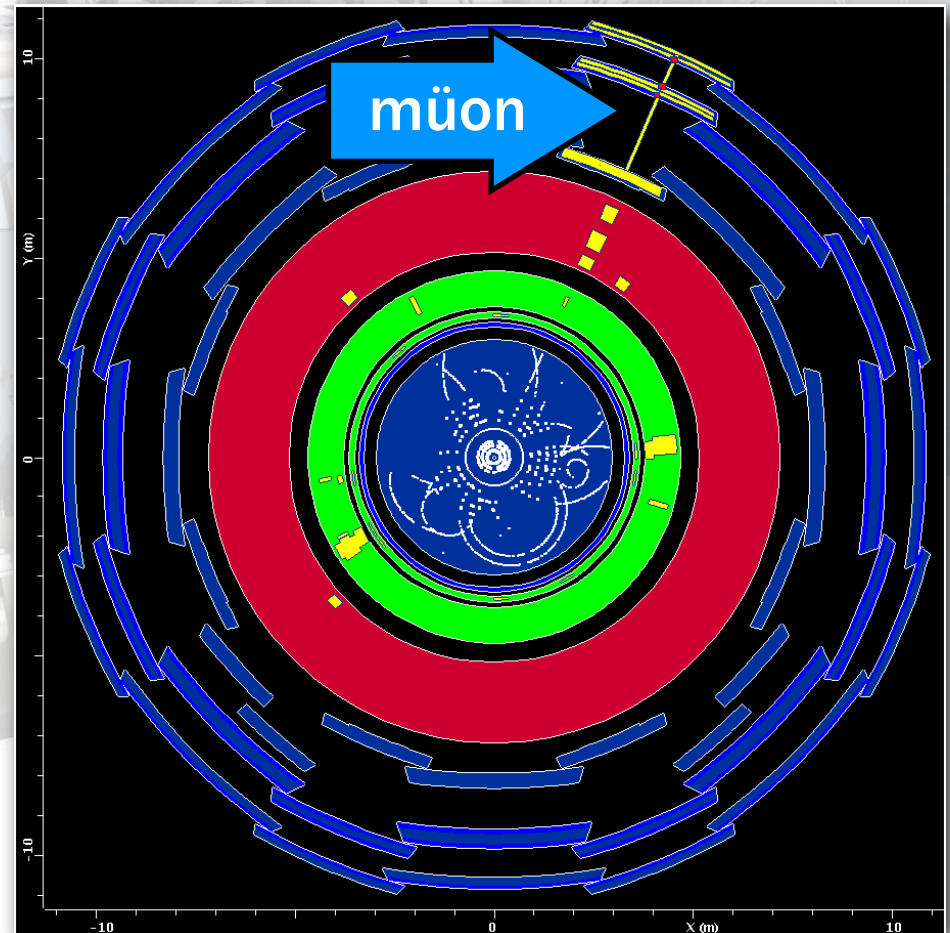
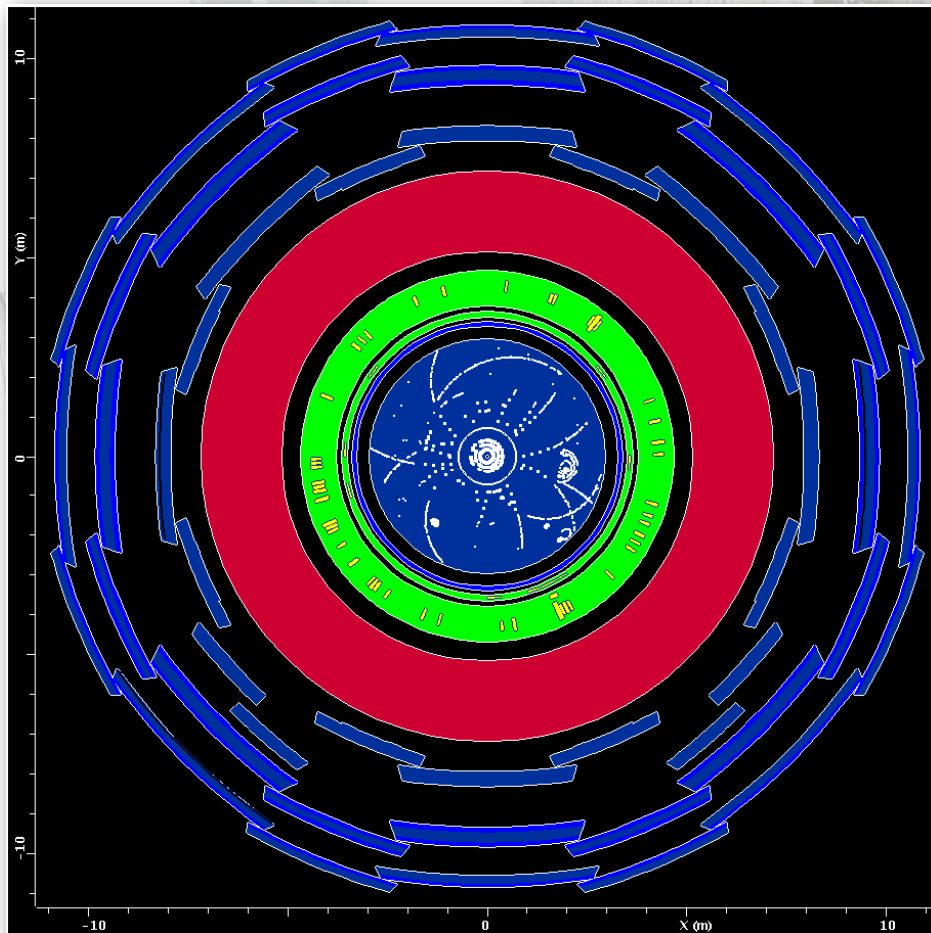
- Nem tudunk minden érzékelt eseményt rögzíteni
- Azokat, amiktől nem várunk új fizikát, azonnal el kell dobjunk
  - Hogyan?



# Adatgyűjtés



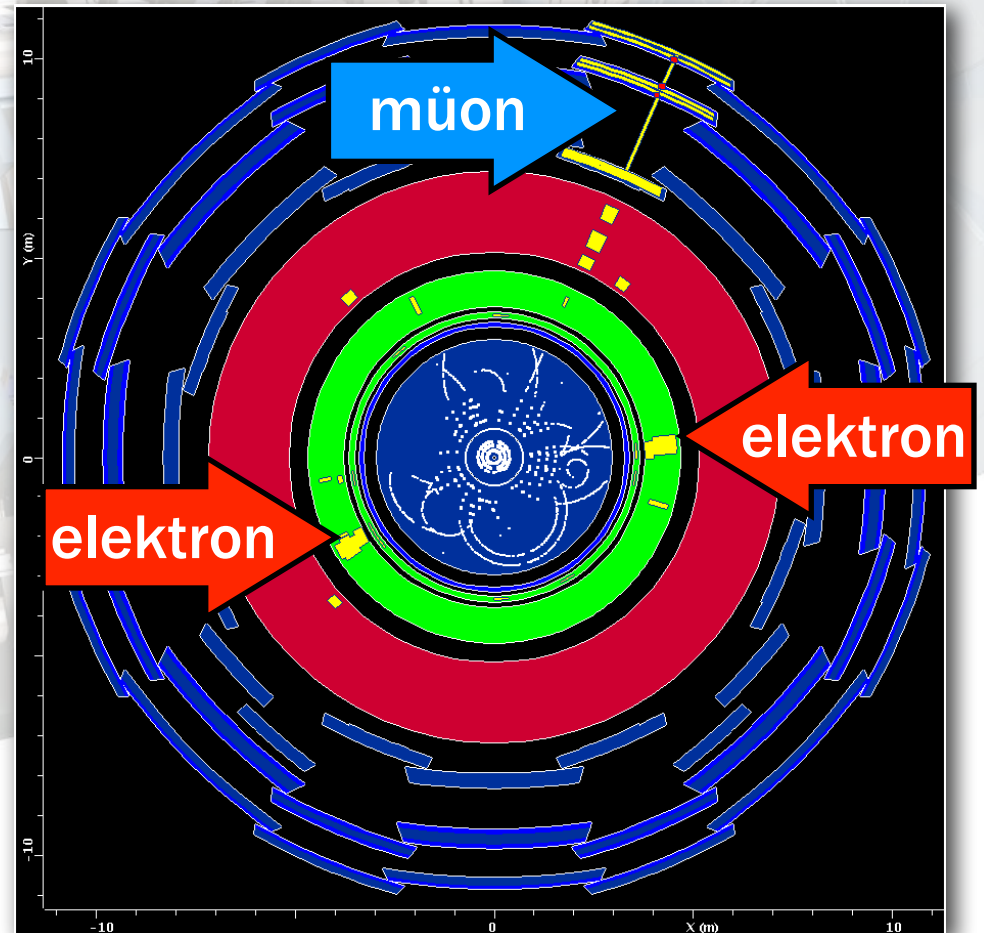
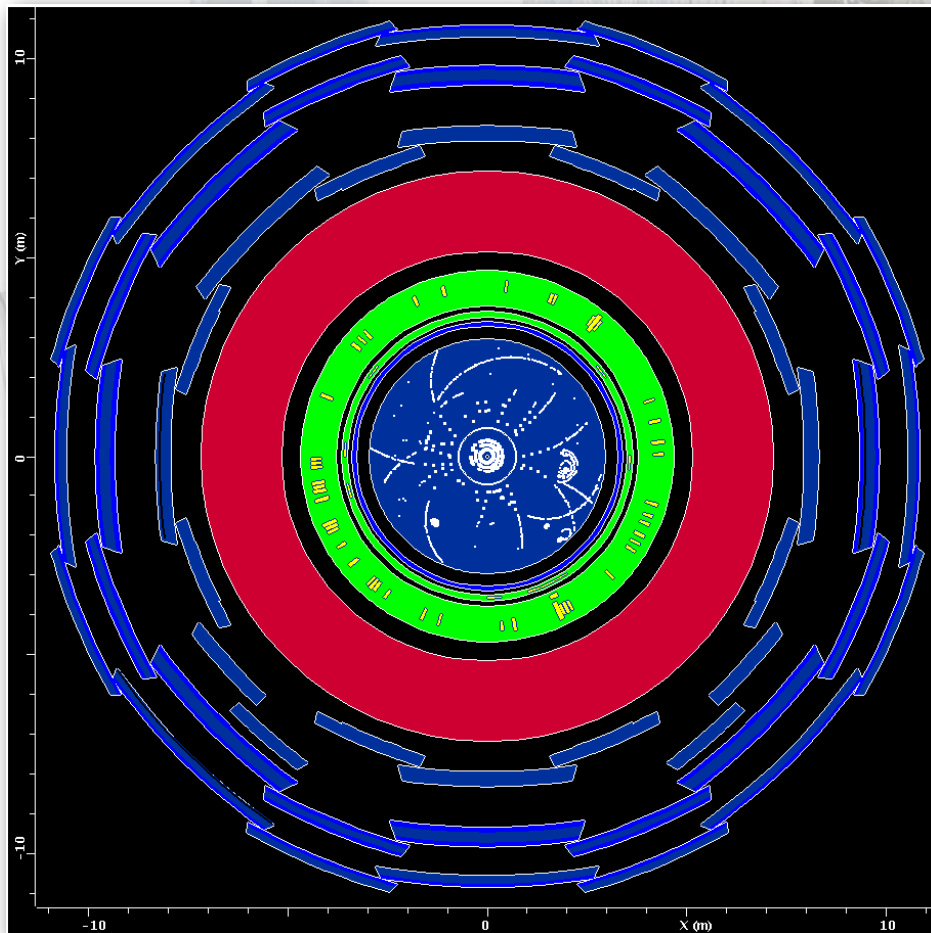
- Nem tudunk minden érzékelt eseményt rögzíteni
- Azokat, amiktől nem várunk új fizikát, azonnal el kell dobjunk
  - Hogyan?



# Adatgyűjtés



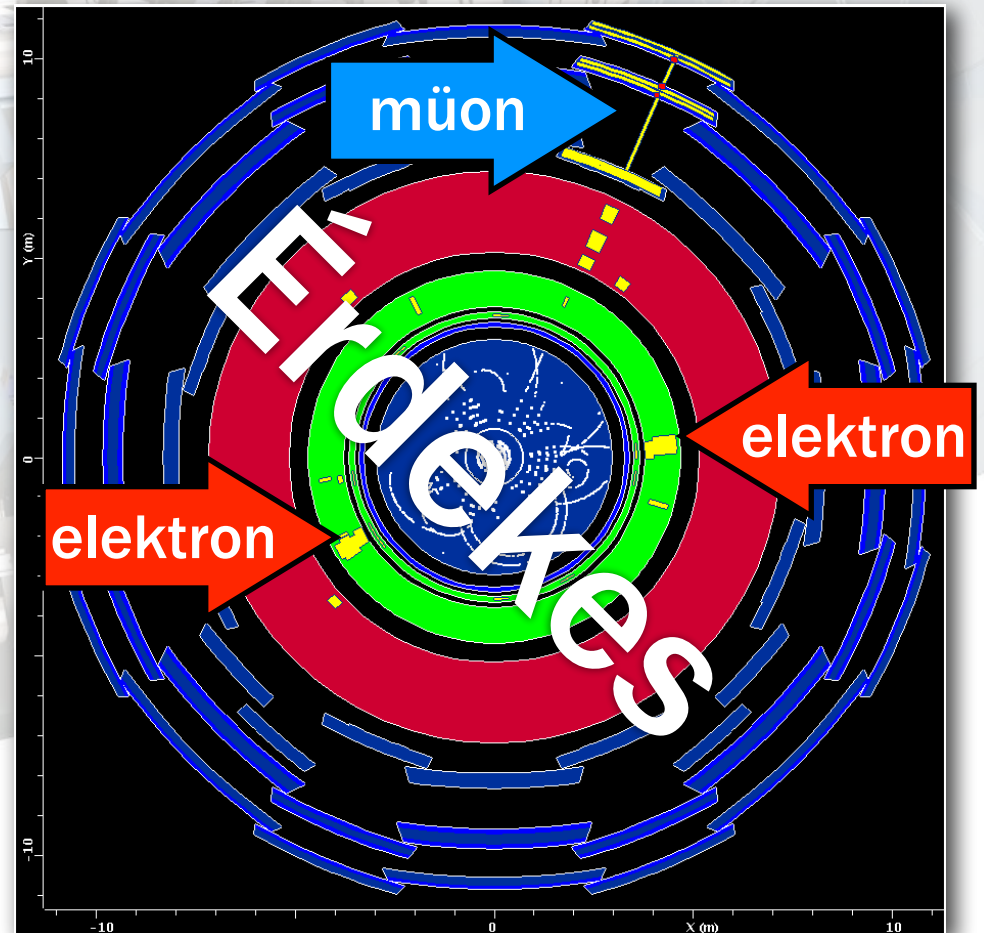
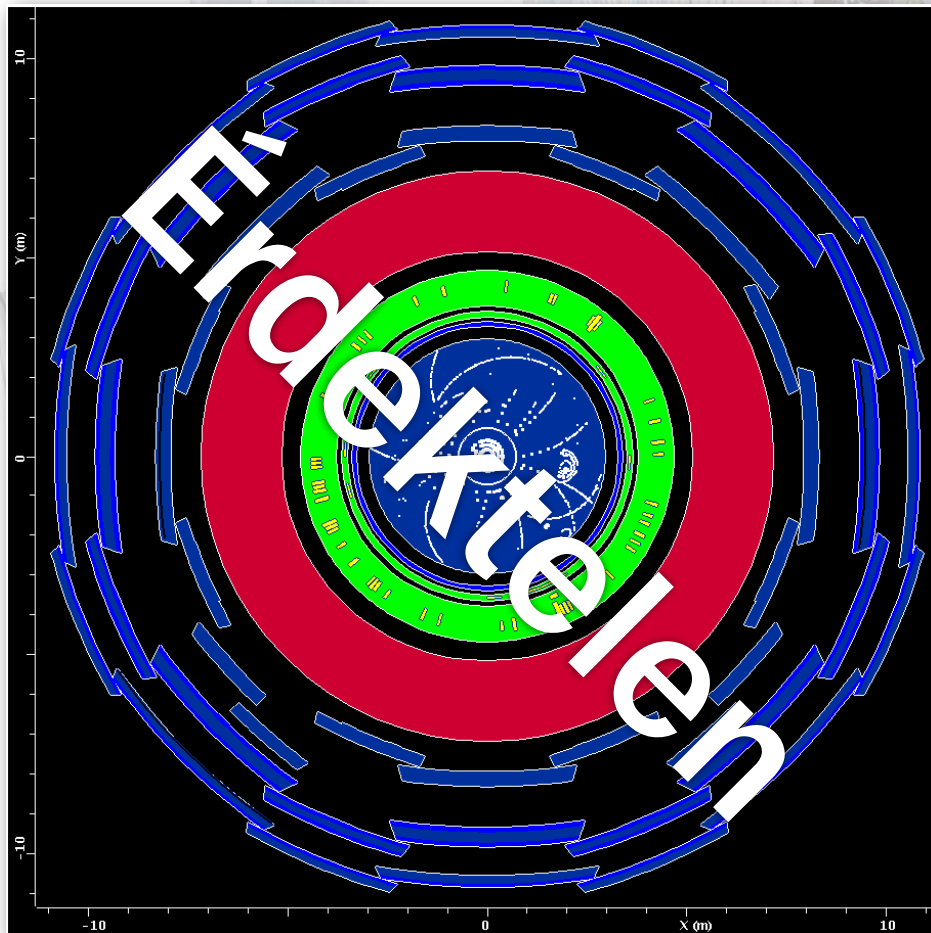
- Nem tudunk minden érzékelt eseményt rögzíteni
- Azokat, amiktől nem várunk új fizikát, azonnal el kell dobjunk
  - Hogyan?



# Adatgyűjtés



- Nem tudunk minden érzékelt eseményt rögzíteni
- Azokat, amiktől nem várunk új fizikát, azonnal el kell dobjunk
  - Hogyan?

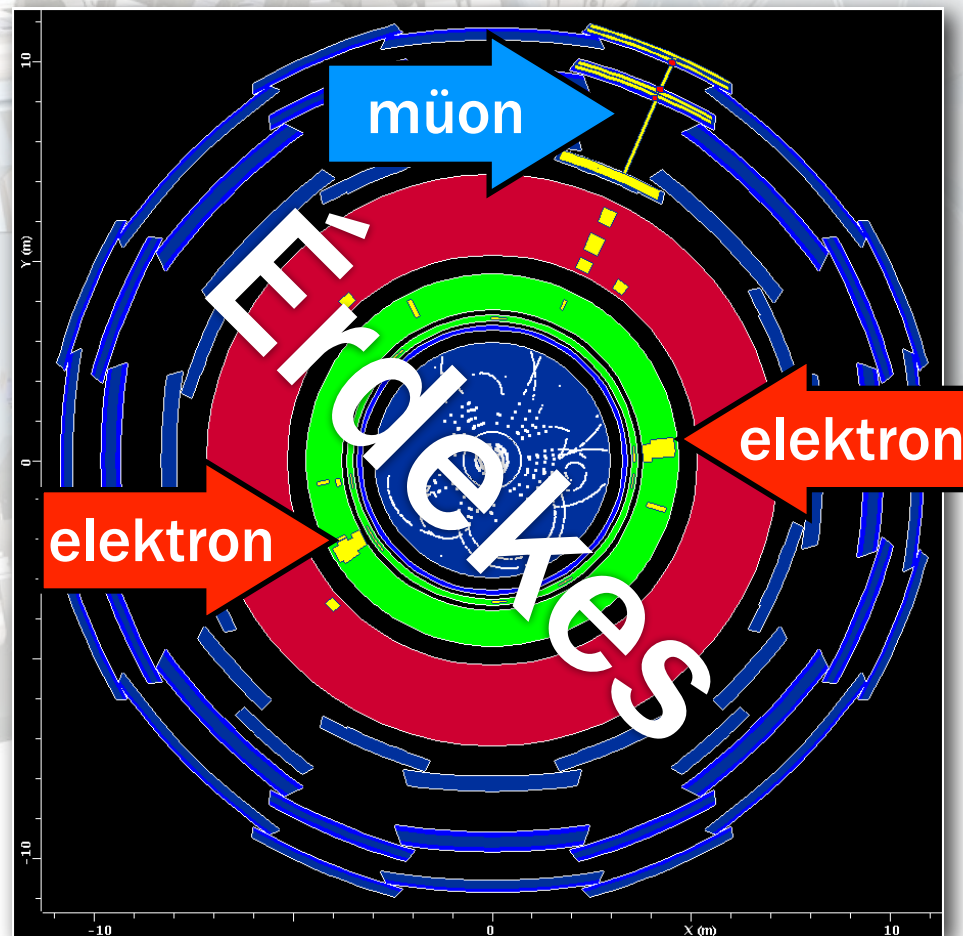
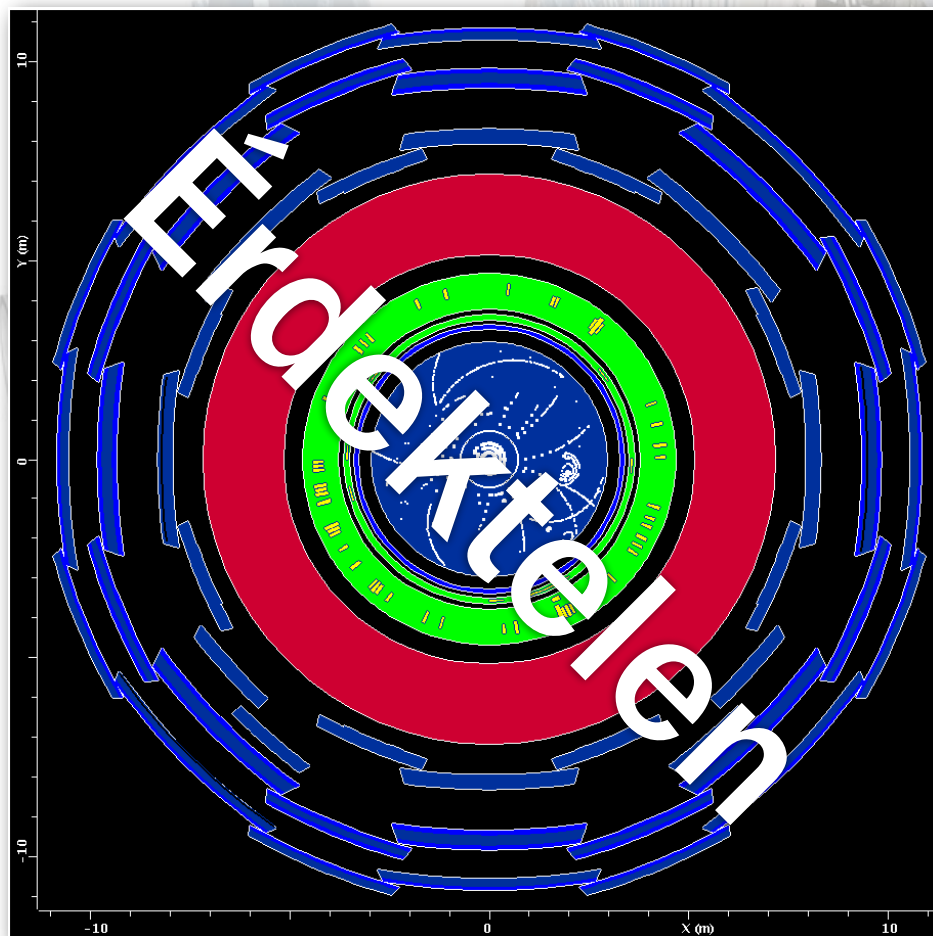


“Online” szoftver

# Adatgyűjtés



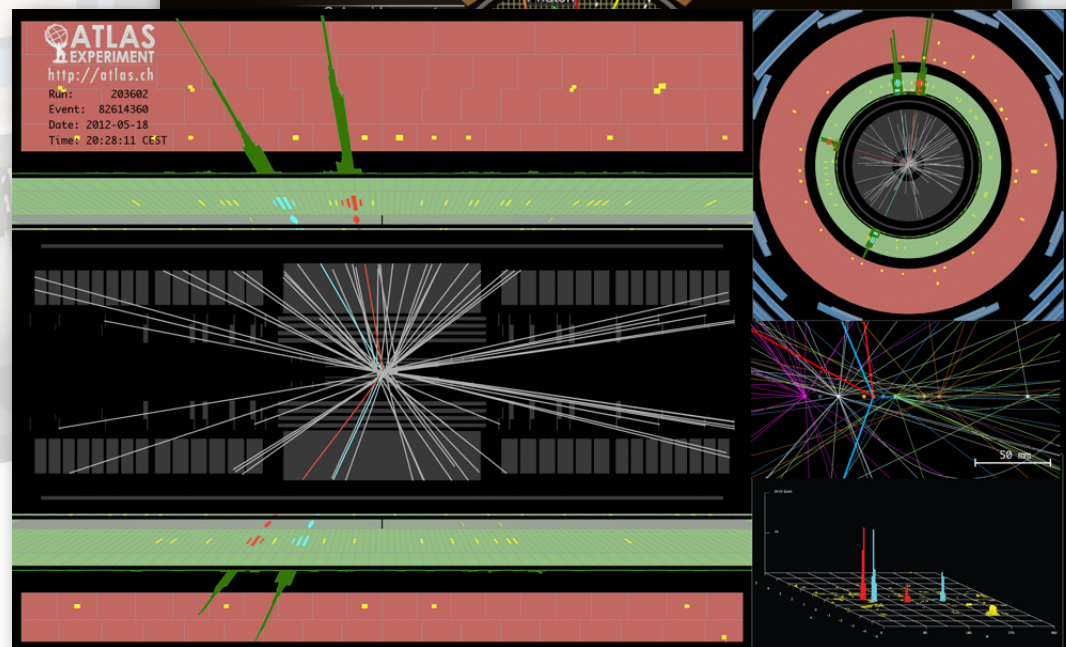
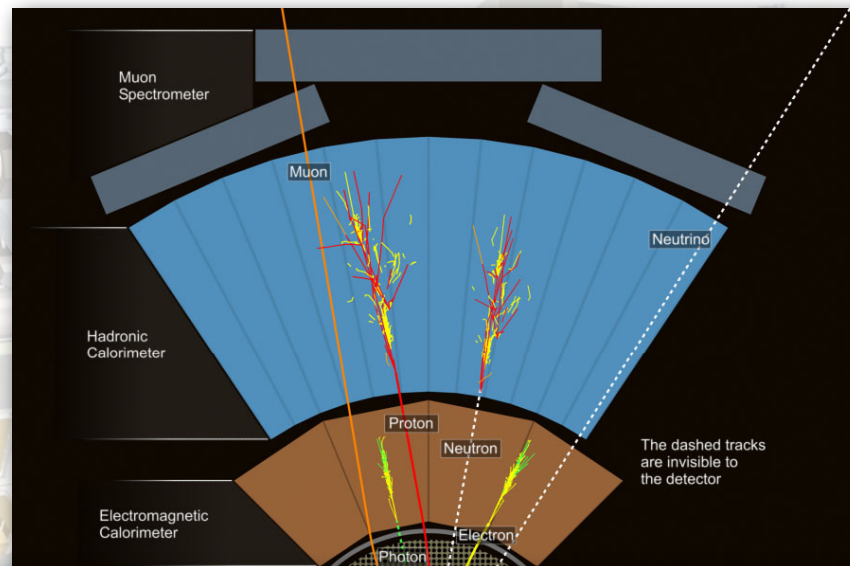
- Nem tudunk minden érzékelt eseményt rögzíteni
- Azokat, amiktől nem várunk új fizikát, azonnal el kell dobjunk
- Hogyan?



# Rekonstrukció



- Varga Dezső beszélt ennek az alapjairól...
- A kísérletek saját szoftvert írnak/írtak a bejövő adataik első feldolgozásához
  - Hasonló ahhoz, ahogyan egy digitális fényképezőgép létrehozza a képet az érzékelőjéből jövő jelek alapján
  - A világ legösszetettebb szoftverei...
    - Az ATLAS-ban kb. 8 millió sor programkód (>\$200M)

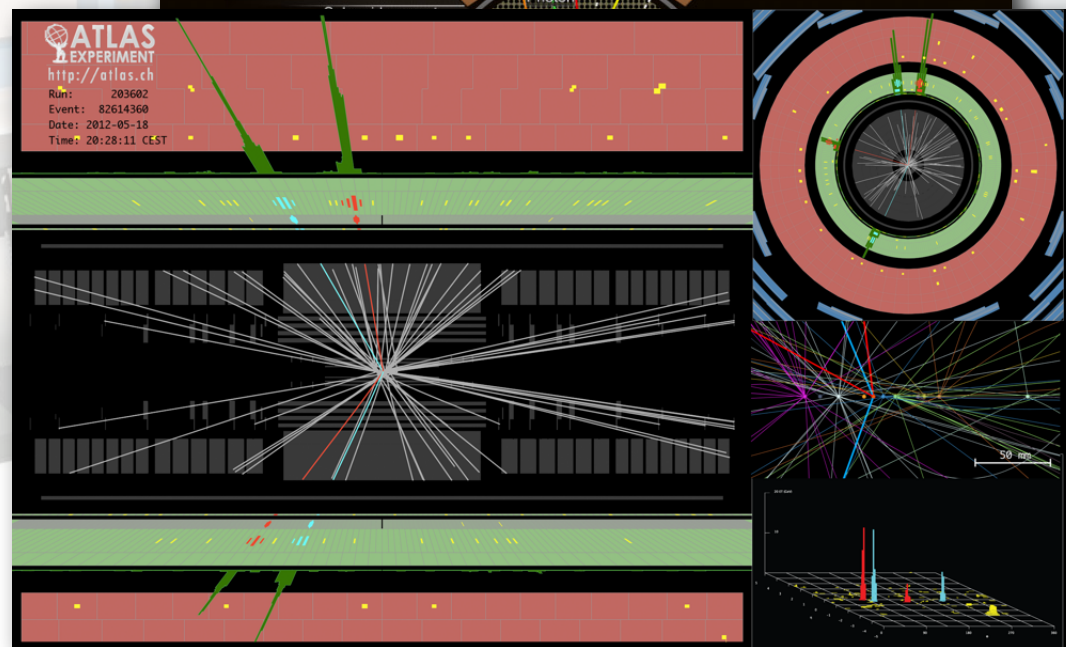
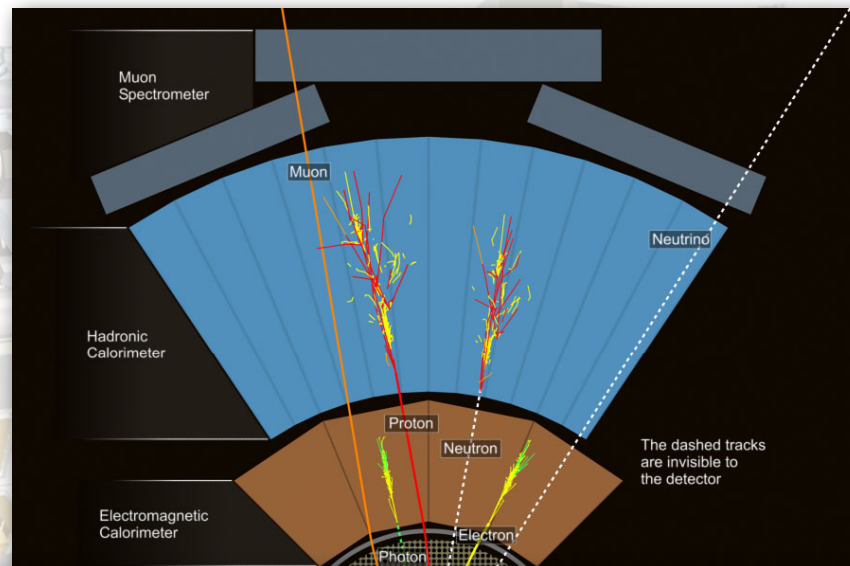


“Offline” szoftver

# Rekonstrukció

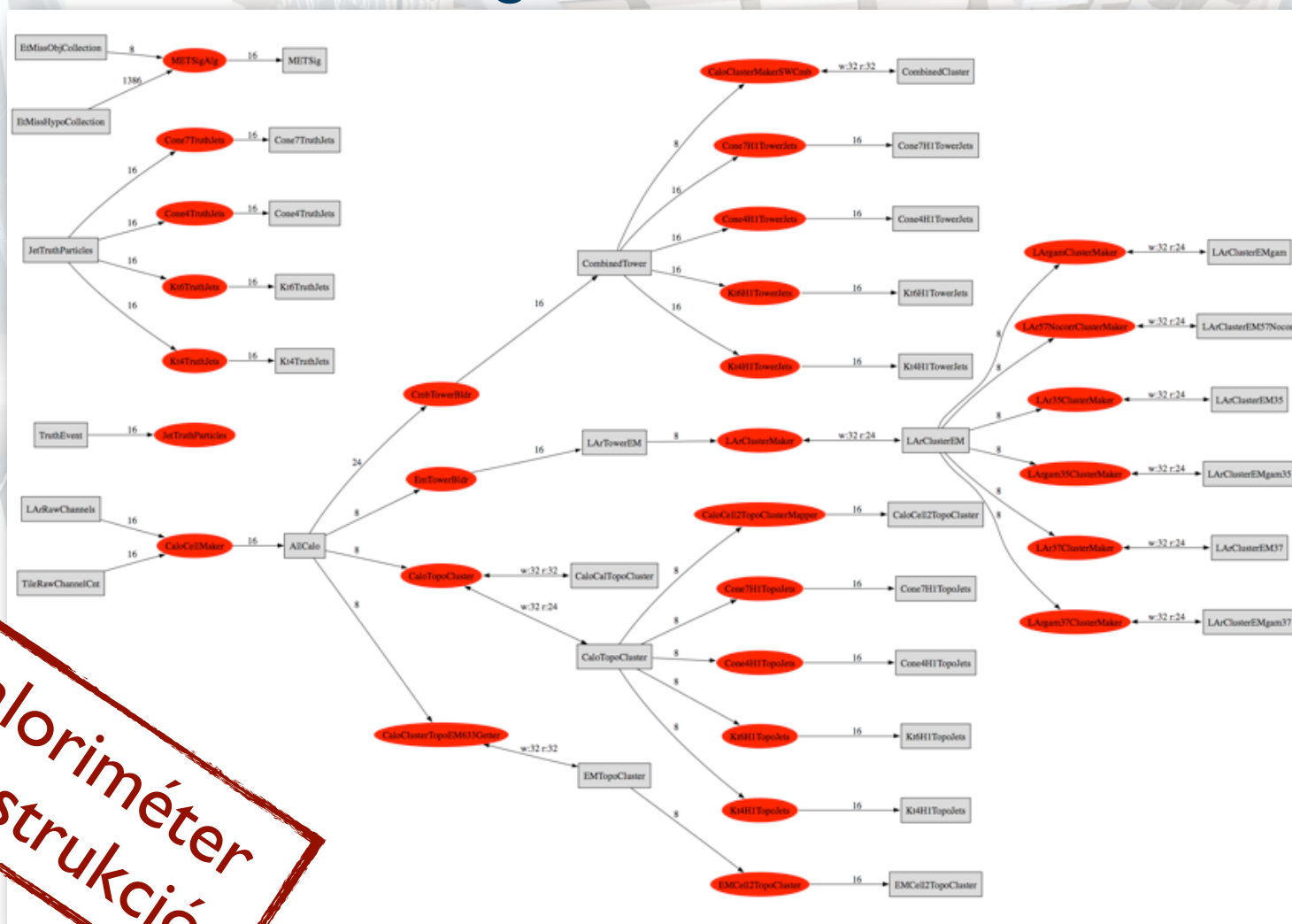


- Varga Dezső beszélt ennek az alapjairól...
- A kísérletek saját szoftvert írnak/írtak a bejövő adataik első feldolgozásához
- Hasonló ahhoz, ahogyan egy digitális fényképező-gép létrehozza a képet az érzékelőjéből jövő jelek alapján
- A világ legösszetettebb szoftverei...
  - Az ATLAS-ban kb. 8 millió sor programkód (>\$200M)



# A Szoftver

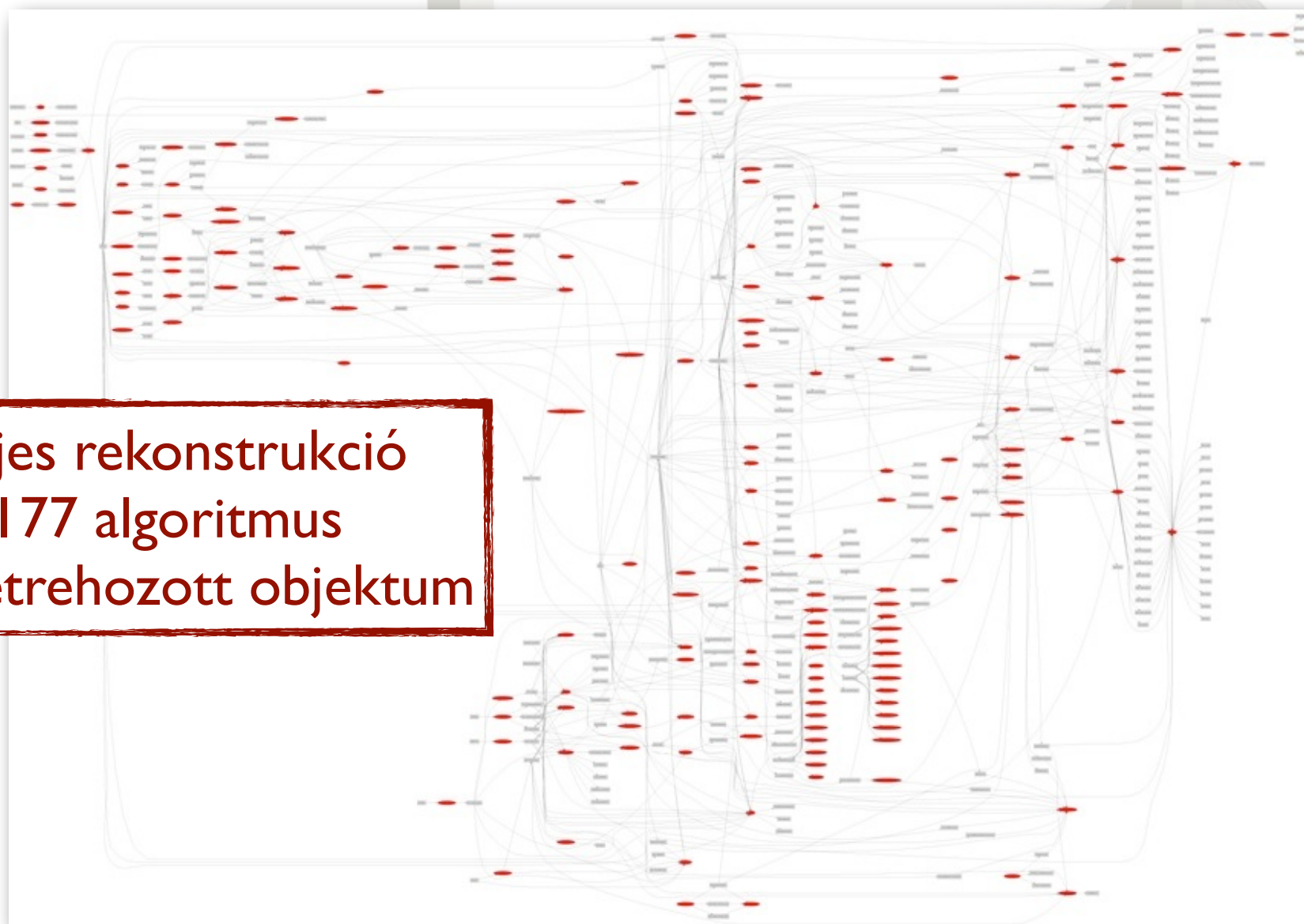
- Egymás után futtatott, egymástól “nagyban független” algoritmusok összessége



**A kaloriméter  
rekonstrukció**



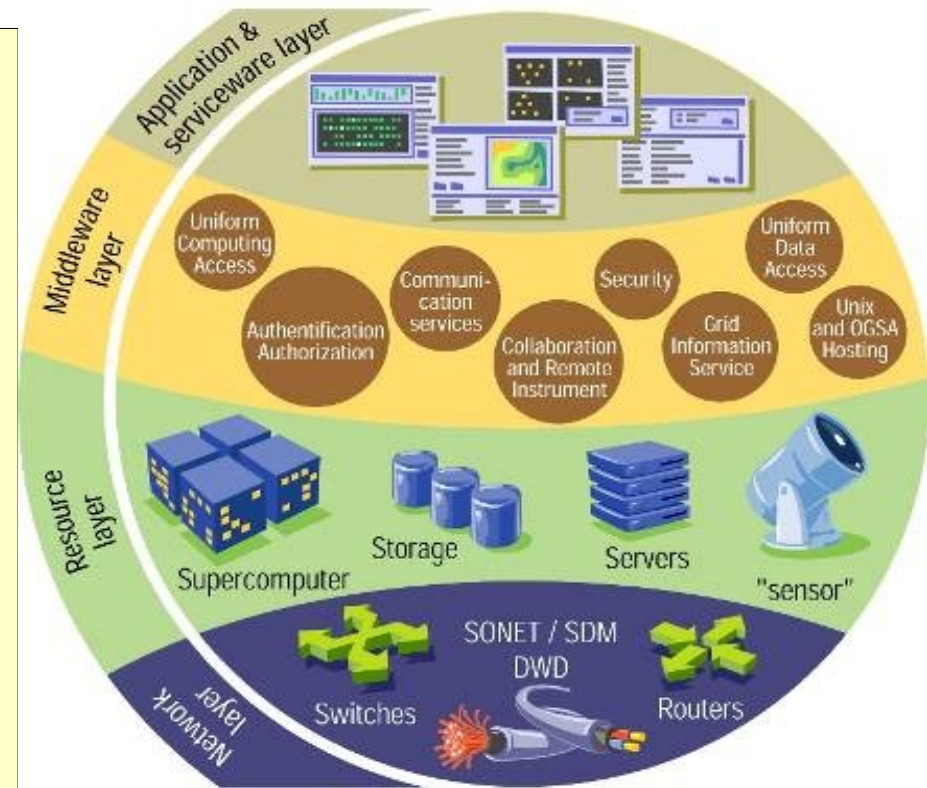
# A Szoftver



Teljes rekonstrukció  
177 algoritmus  
372 létrehozott objektum

# Hogyan működik?

- A Gridet egy 'speciális' szoftver az ún. middleware (köztesréteg) tartja életben.
- A middleware 'automatikusan megtalálja' a felhasználó számára szükséges adatsomagokat és legmegfelelőbb végrehajtási helyet (számítógépet).



- A middleware feladata továbbá a számítási terhelések egyenlő elosztása, a hálózati biztonság megteremtése, az erőforrások felügyelete, monitorozása, naplózás, számlázás és sok minden más.

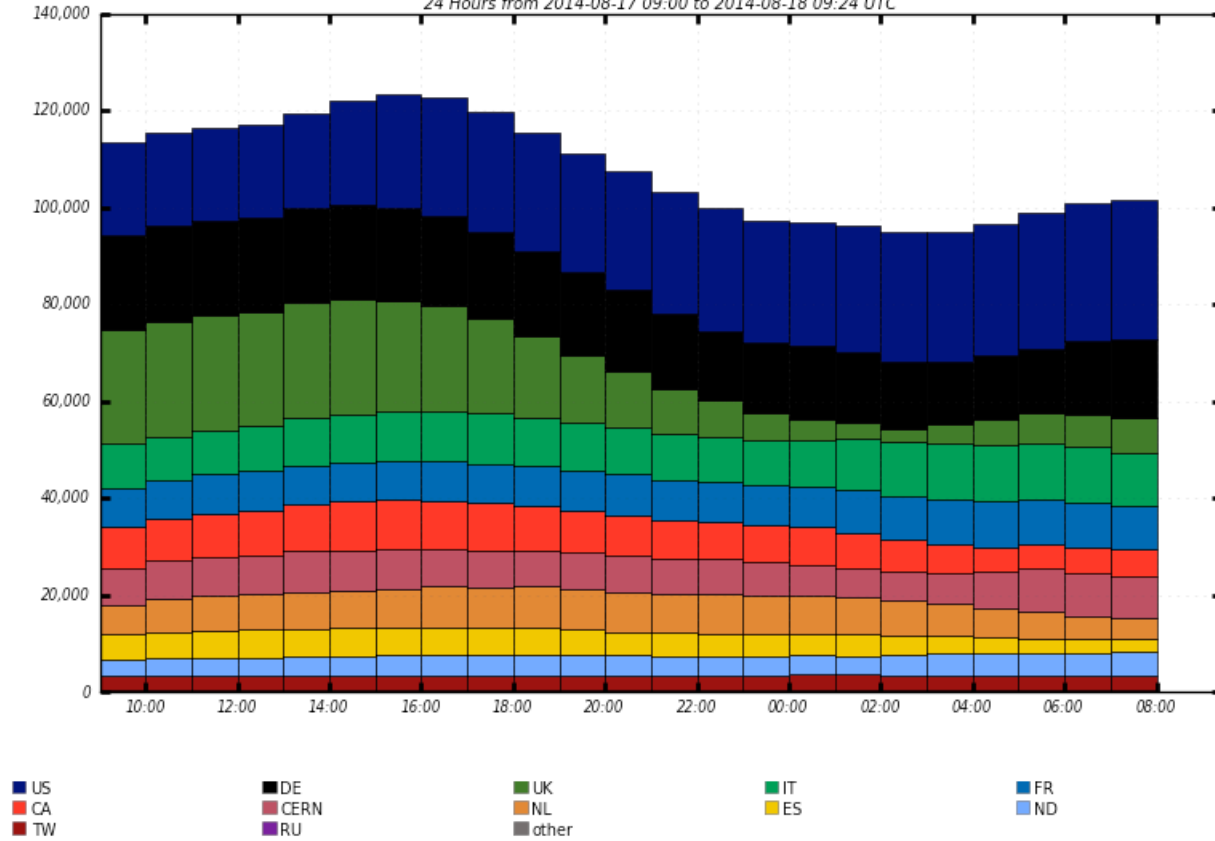
# A Feladatok a Griden



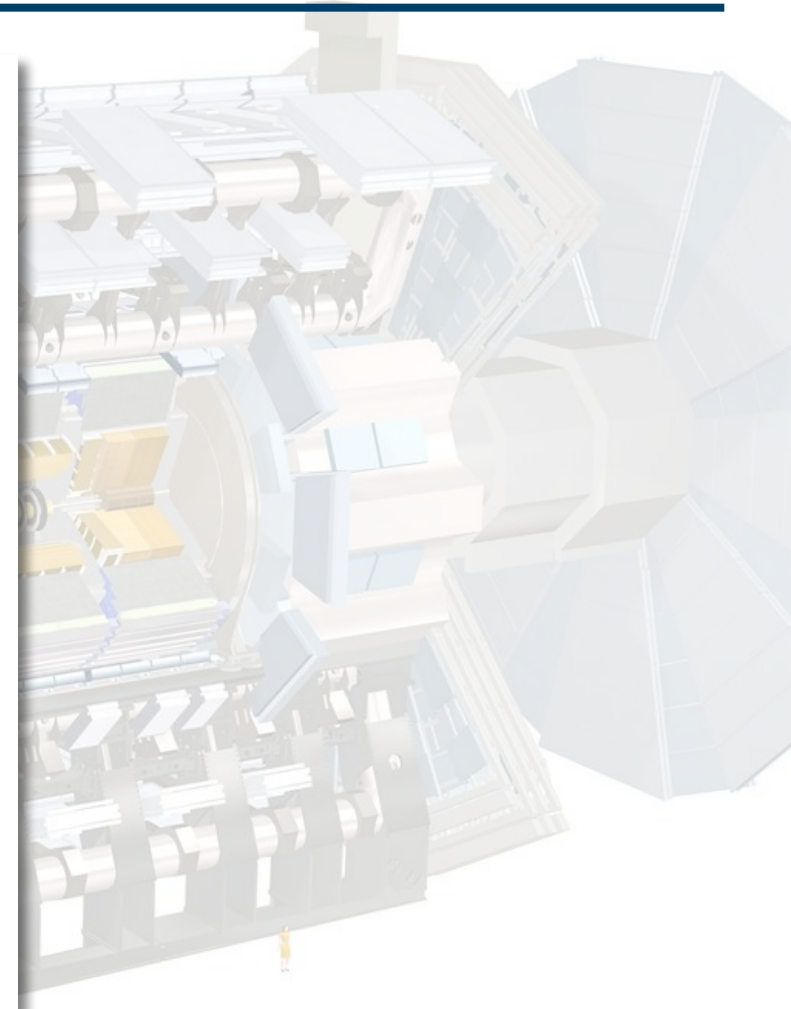
dashboard

## Running jobs

24 Hours from 2014-08-17 09:00 to 2014-08-18 09:24 UTC



Maximum: 123,515 , Minimum: 94,958 , Average: 108,124 , Current: 101,505

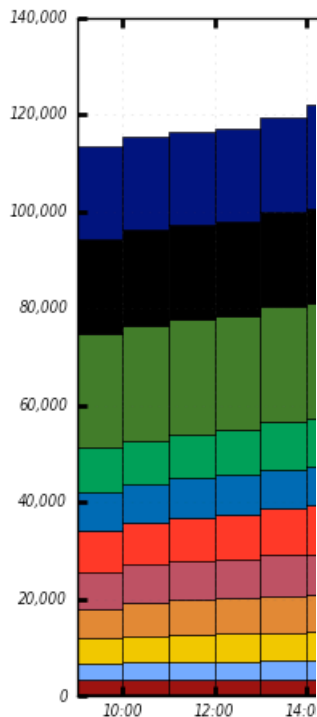


# A Feladatok a Griden

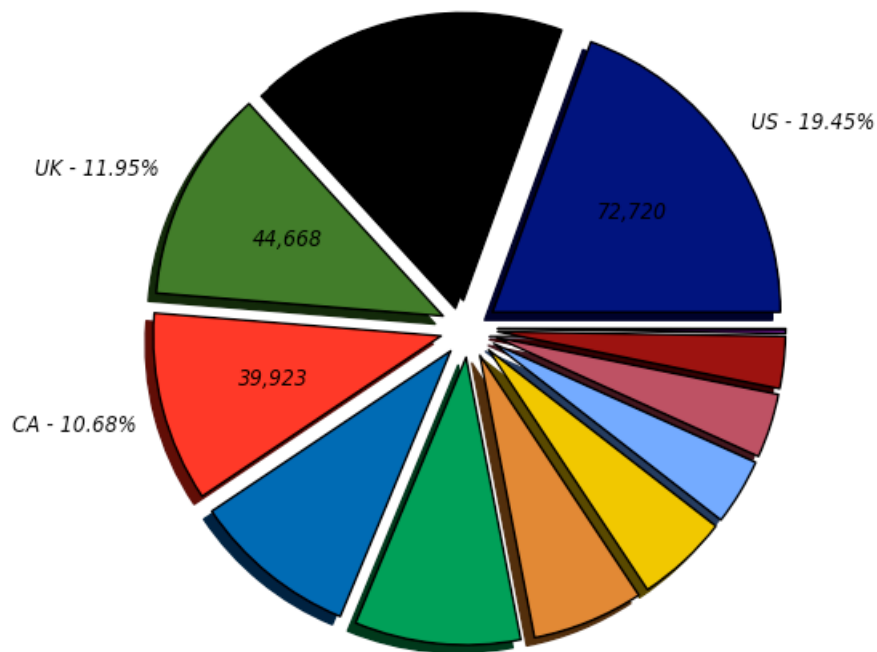


## Running jobs

24 Hours from 2014-08-17 09:00 to 2014-08-18 09:24 UTC



## Completed jobs (Sum: 373,861) DE - 17.33%



- US
- CA
- TW
- DE
- CERN
- RU

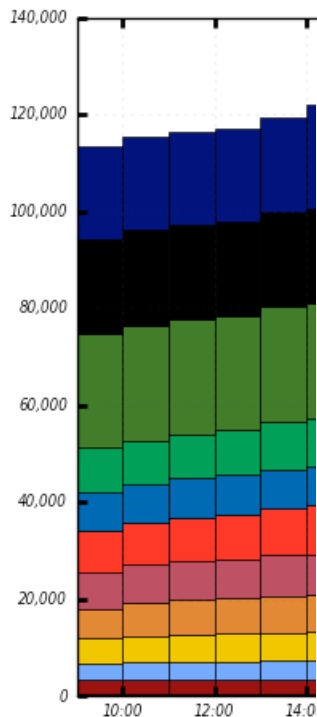
- |                        |                        |                         |                        |                         |
|------------------------|------------------------|-------------------------|------------------------|-------------------------|
| ■ US - 19.45% (72,720) | ■ DE - 17.33% (64,795) | ■ UK - 11.95% (44,668)  | ■ CA - 10.68% (39,923) | ■ FR - 9.32% (34,847)   |
| ■ IT - 9.32% (34,834)  | ■ NL - 6.26% (23,411)  | ■ ES - 5.18% (19,377)   | ■ ND - 3.73% (13,927)  | ■ CERN - 3.61% (13,495) |
| ■ TW - 2.92% (10,903)  | ■ RU - 0.24% (902.00)  | ■ other - 0.02% (59.00) |                        |                         |

# A Feladatok a Griden



Running jobs

24 Hours from 2014-08-17 09:00 to 2014-08-18 09:24 UTC

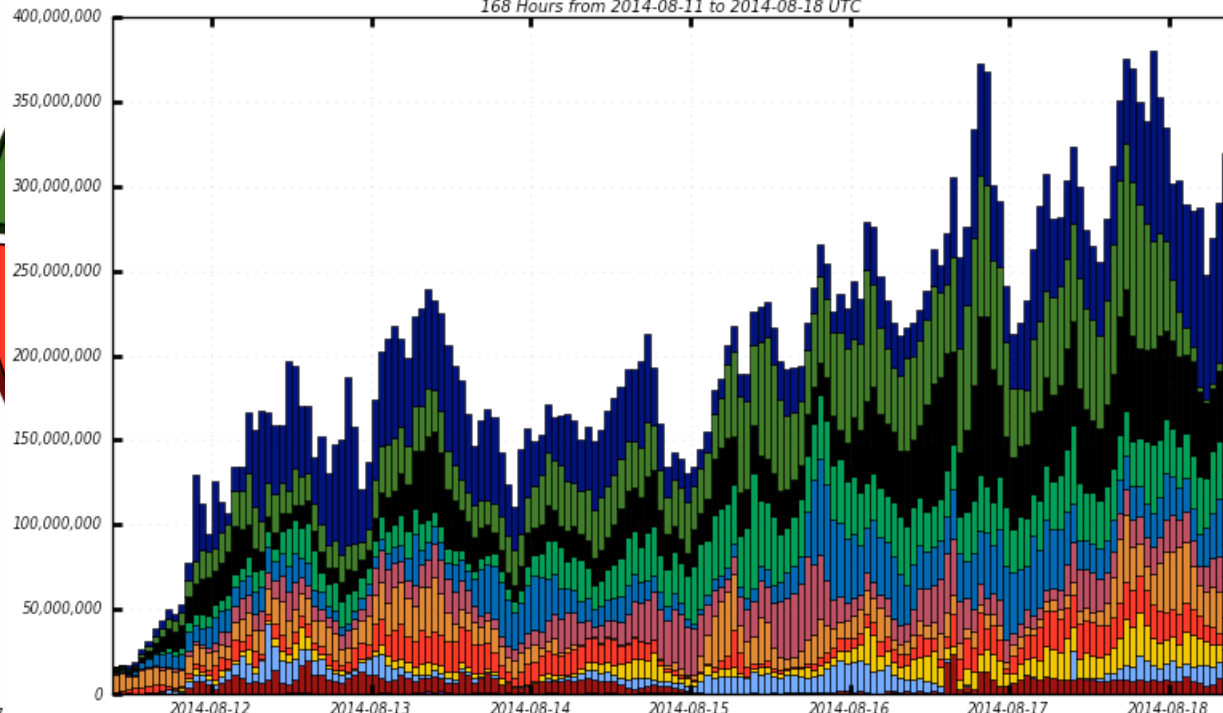


Completed jobs (Sum: 373,861)  
DE - 17.33%



CPU consumption All Jobs in seconds

168 Hours from 2014-08-11 to 2014-08-18 UTC



US - 19.45% (72,720)  
IT - 9.32% (34,834)  
TW - 2.92% (10,903)  
DE - 17  
NL - 6.2  
RU - 0.2

US UK DE IT FR  
CERN NL CA other ES ND  
TW RU

Maximum: 380,546,784 , Minimum: 16,434,253 , Average: 202,260,220 , Current: 319,381,883

# Összefoglalás



- Az LHC adatainak feldolgozása hatalmas számítástechnikai kapacitásokat igényel
- Kizárólag világméretű összefogással teremthető ez elő -> Ez a GRID
- Sok százezer processzor, néhányszor 100 PB tárolókapacitás, ...
- A kollaborációkban sok ember csak az adatok feldolgozásának szervezésével foglalkozik
- Ugyancsak sok ember szükséges magának a feldolgozó-szoftvernek a fejlesztéséhez
- Csak a legnagyobb szoftvercégek végeznek hasonló méretű fejlesztéseket az egész világon!