

Adatelemzés és számítástechnika

Krasznahorkay Attila



ATLAS
EXPERIMENT

Mit is csinálunk...

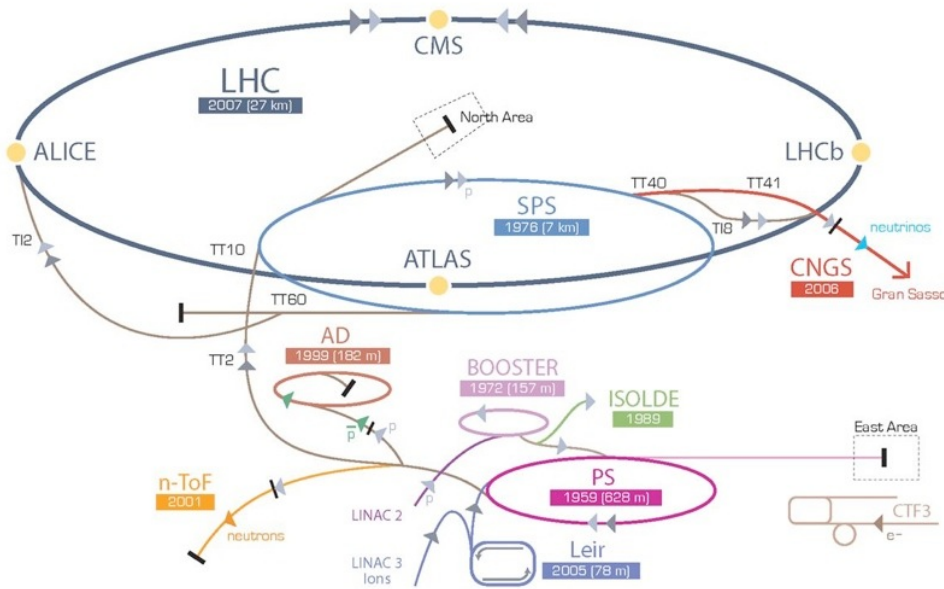


- Muszály pár dologgal előreszaladnom, hogy a mondandómnak értelme lehessen
- A legtöbb dolog amit csinálunk kesze-kuszáan egymásra épül, nehéz olyan sorrendben bemutatni őket amiben csak már ismert dolgokra hagyatkozhatunk.
- Vagyis lássuk először is, hogy nagy vonalakban hogyan is “csinálunk fizikát” az LHC-nál

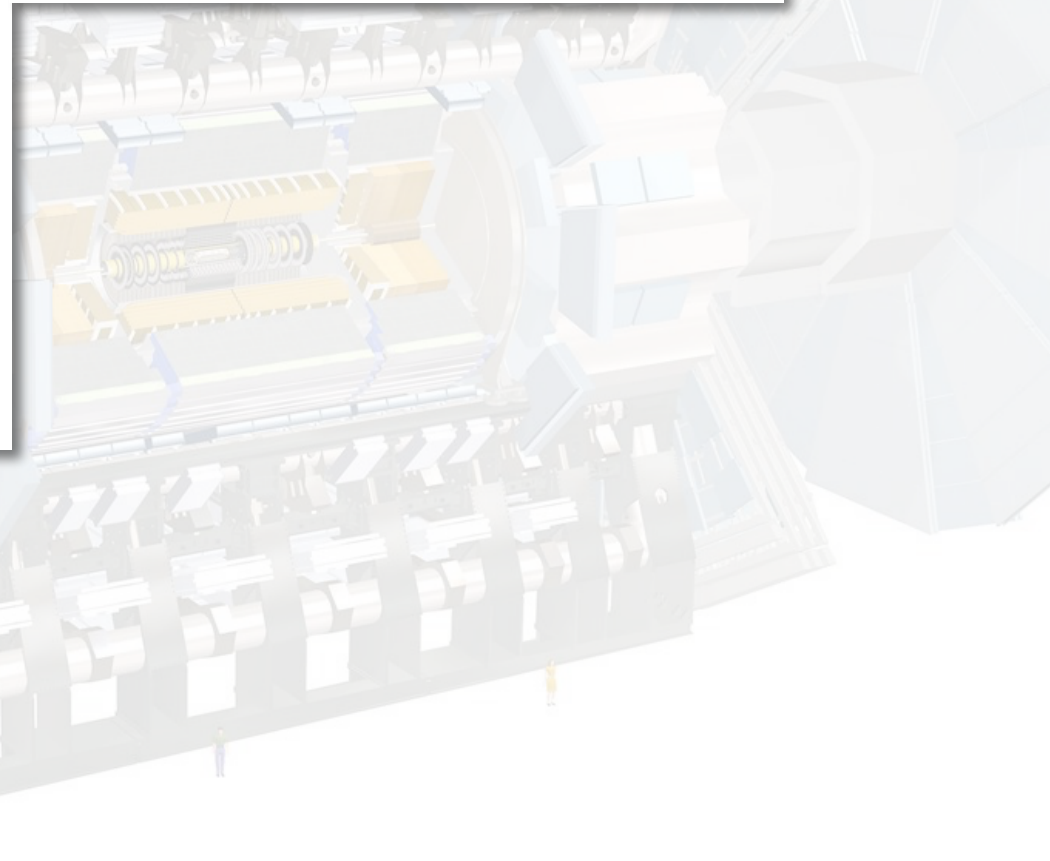
A madártásvlat



CERN Accelerator Complex



Létrehozzuk az “érdekes”
reakciókat
(Barna Dániel előadása)

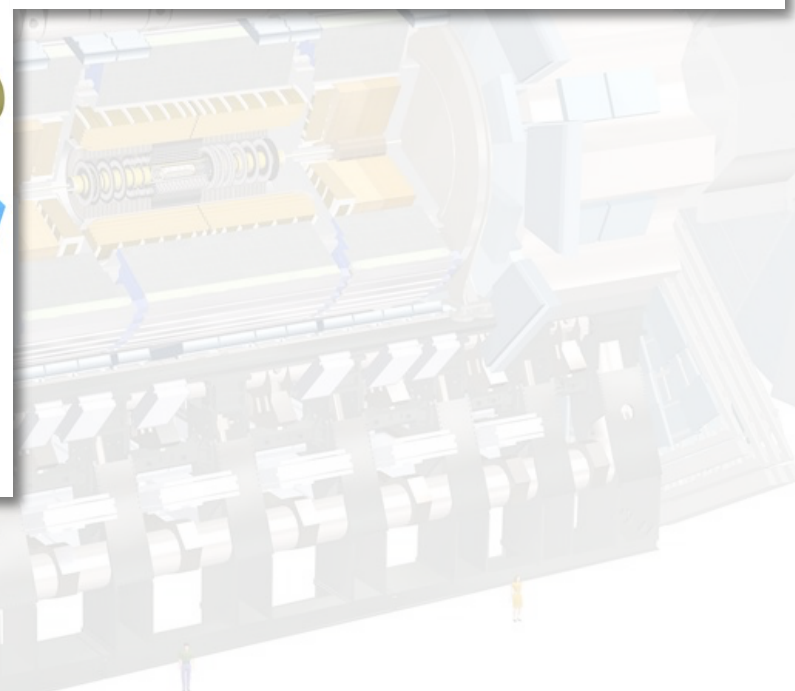
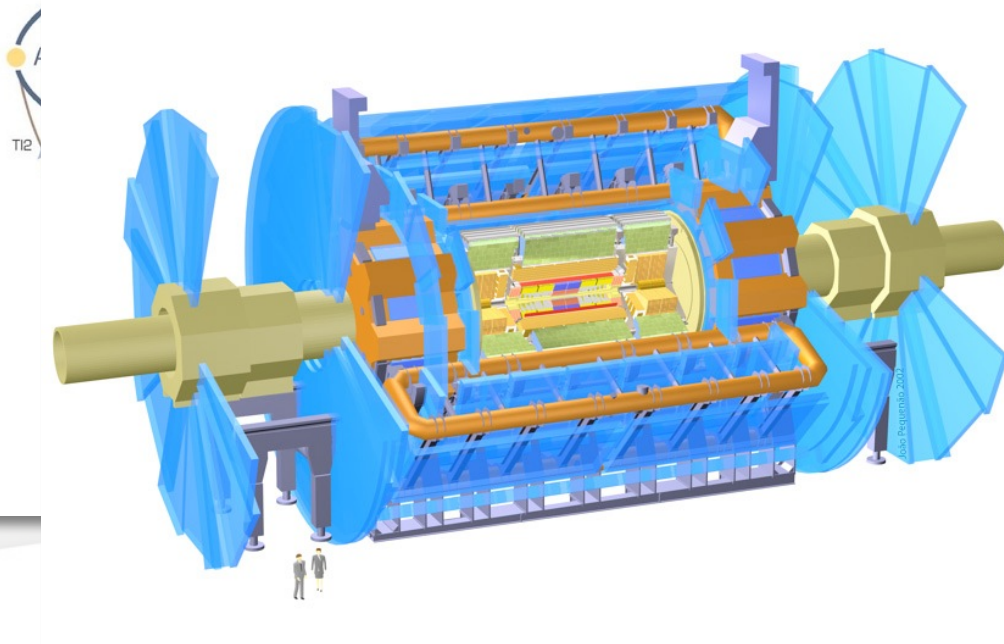


A madártávlat



CERN Accelerator Complex

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

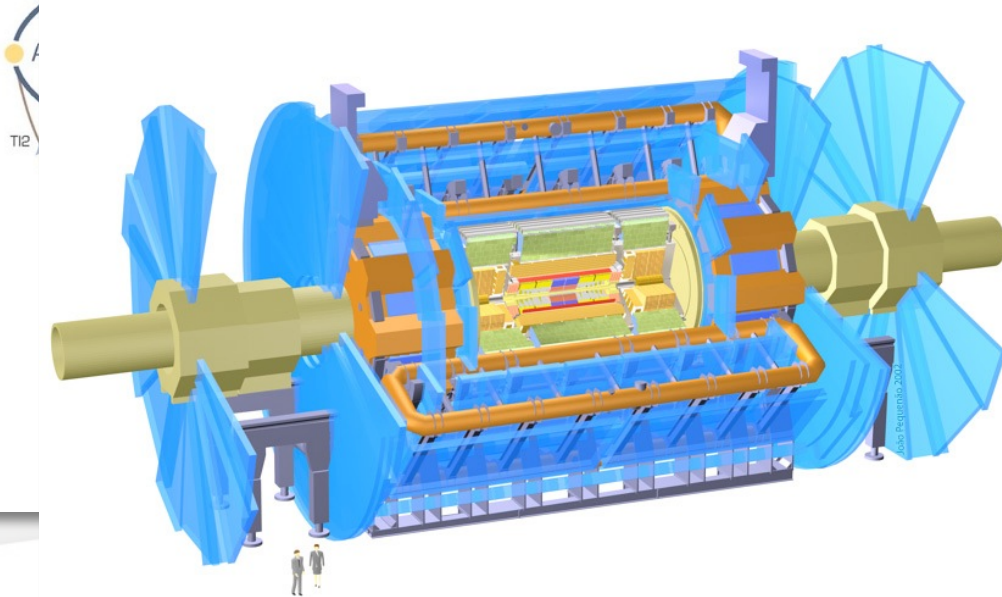


A madártávlat

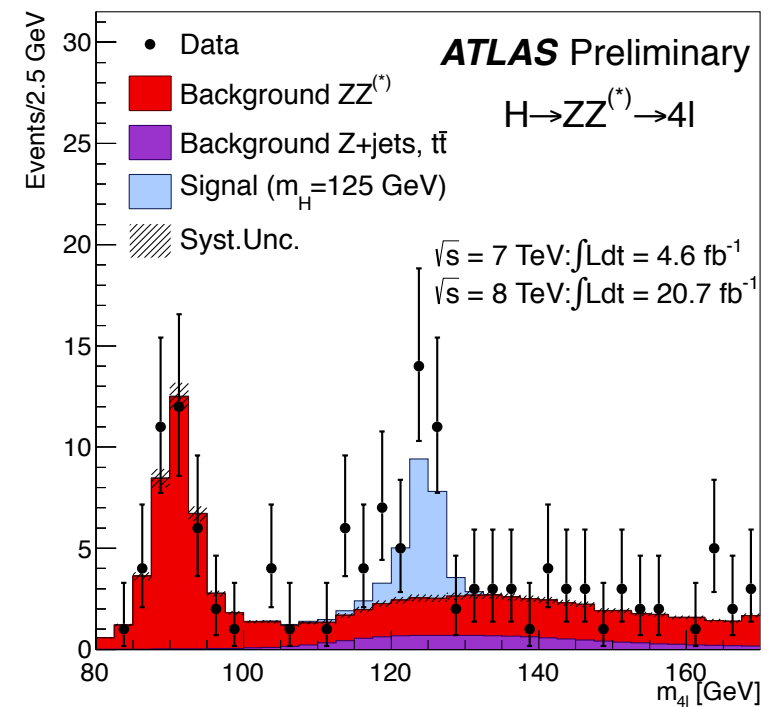


CERN Accelerator Complex

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



Levonjuk a fizikára vonatkozó
következtetéseket
(Újvári Balázs és Fodor
Zoltán előadásai)

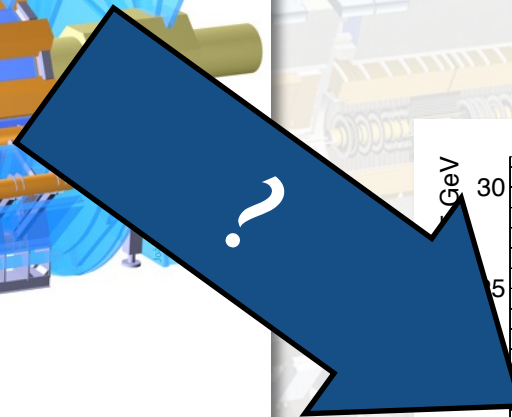
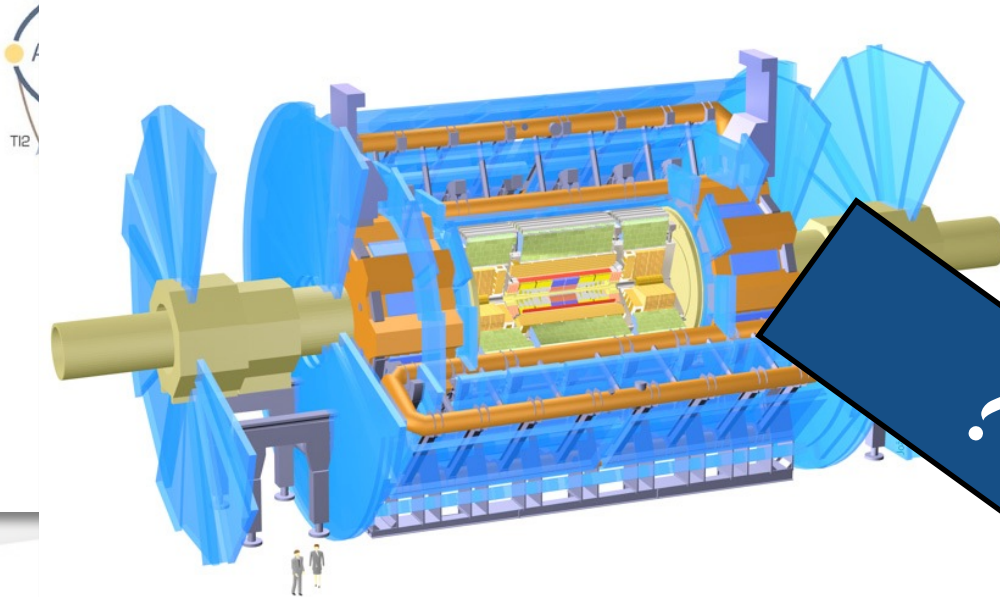


A madártávlat

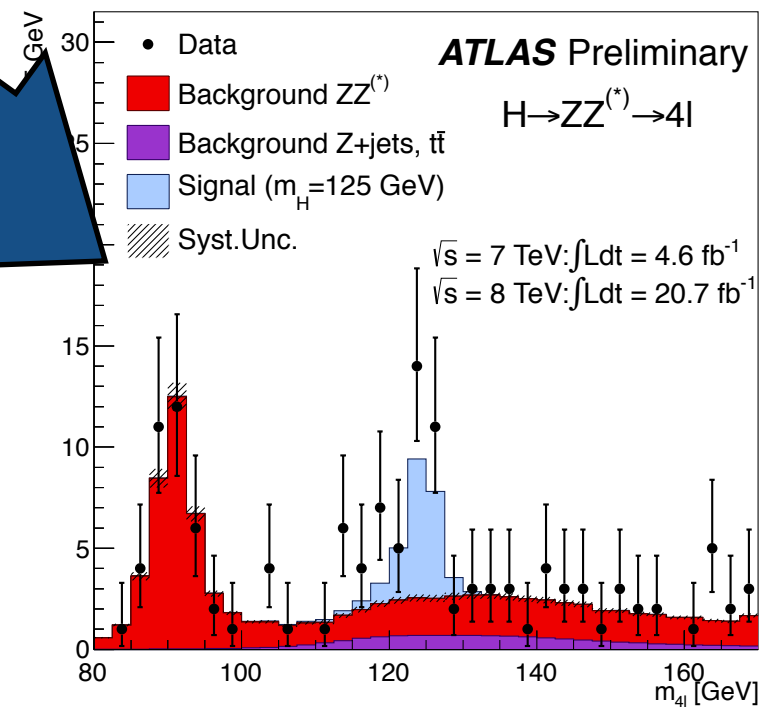


CERN Accelerator Complex

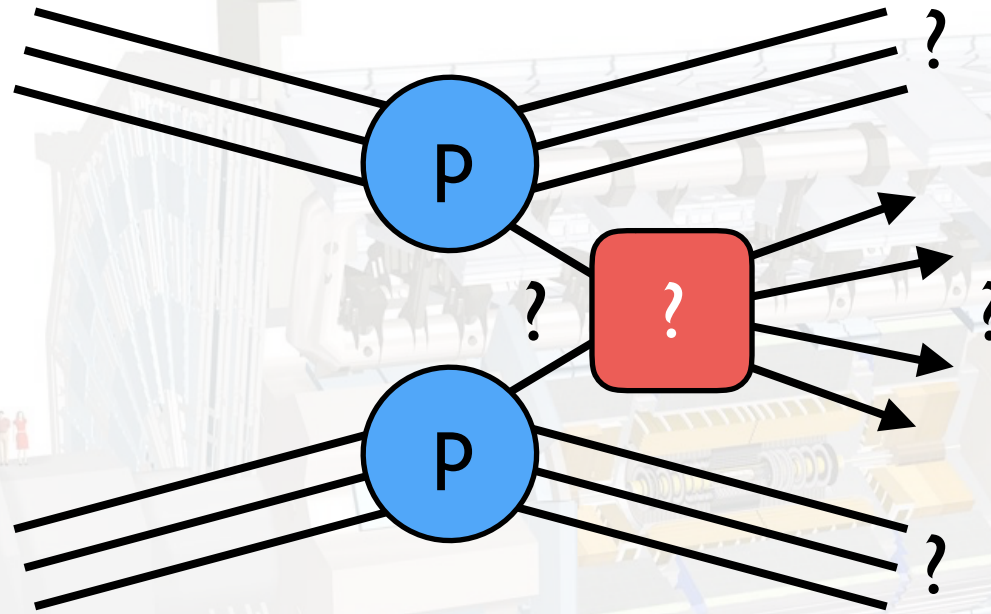
Létrehozzuk az “érdekes”
Érzékeljük a kijövő
részecskéket
(Varga Dezső előadása)



Levonjuk a fizikára vonatkozó
következtetéseket
(Újvári Balázs és Fodor
Zoltán előadásai)



A kísérlet



- Amire igazán kíváncsiak vagyunk az LHC-nál, az a piros dobozban végbemenő folyamat
 - Fizika szempontjából (többek között) az is nagyon érdekes ugyanakkor, hogy a protonból “mi is jön ki”, és mekkora valószínűséggel
- Az alapján, amit látunk kijönni a reakcióból, próbálunk visszakövetkeztetni, hogy:
 - Valóban egy proton-proton ütközést láttunk-e éppen
 - Milyen reakció zajlott le pontosan

Az adatgyűjtés

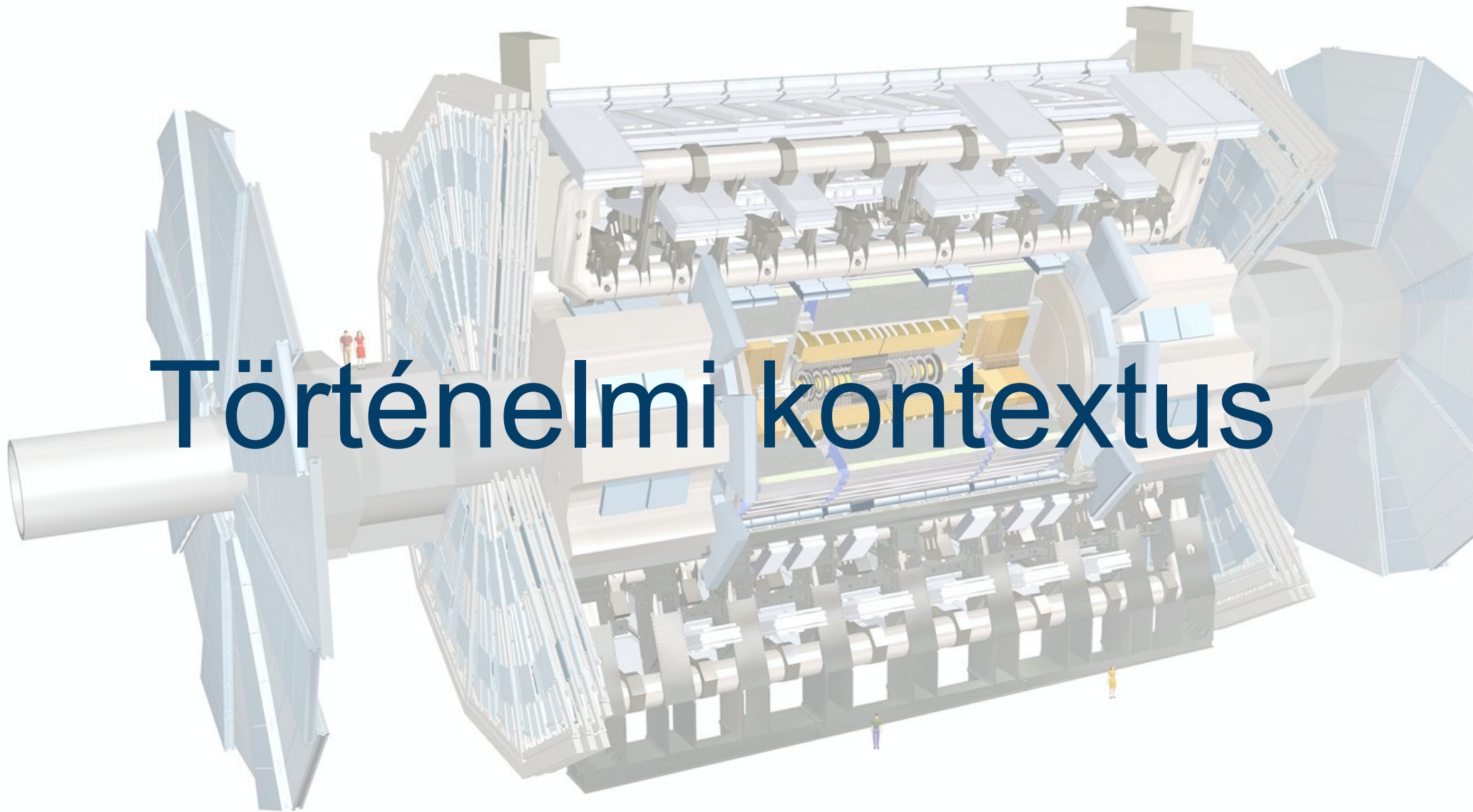


- A detektorokkal egymástól független “eseményeket” rögzítünk
 - Minden detektor típus a rá jellemző módon tudja érzékelni azt ha különböző részecskék kölcsönhatásba lépnek vele
 - A rögzített információt, mint egy fényképezőgép CCD kameráját, kiolvassuk
- Mivel az LHC másodpercenként több mint 1 milliárd proton-proton ütközést hoz létre (kb. 40 millió proton csomag ütközést keltünk másodpercenként, és egy ilyen ütközésben kb. 40 proton-proton ütközés jön létre) minden eseményt nem rögzíthetünk
 - Egy válogató (trigger) rendszer feladata kiválasztani a legérdekesebb eseményeket, hogy másodpercenként csupán 1-2 GB adatot kelljen rögzítenünk
 - Az ATLAS esetében ez ~1.6 MB-os események 1 kHz-es rögzítését jelenti
- A nyers adatokat a CERN számítóközpontjába küldjük, ahonnan biztonsági másolatok készülnek róla a világ sok pontjára
 - Évente “néhány” milliárd eseményt szoktunk rögzíteni

Az adatfeldolgozás



- A nyers adatokat erre fejlesztett célprogramokkal “rekonstruáljuk”. Így állapítva meg, hogy milyen részecskék milyen tulajdonságokkal repültek át (és a legtöbb esetben semmisültek meg) a detektoron
 - Megkapjuk, hogy mennyi elektron, müon, hadronzápor, stb. keletkezett az egyes eseményekben, és milyen tulajdonságokkal
- Az egységesen rekonstruált adatokat, a legtöbbször több lépésen keresztül analizáljuk, ekkor már egy adott analízisnek megfelelően
 - Például kereshetünk két leptonos eseményeket kevés “hiányzó energiával”, hogy Z bozonos eseményeket akarunk vizsgálni
 - A kiválasztott eseményekben egyenként kiszámíthatunk különböző komplex paramétereket (pldával az eseményben feltételezett Higgs bozon tömegét), amiknek az eloszlásából statisztikai analízissel a háttérben zajló fizikára tudunk visszakövetkeztetni.



Történelmi kontextus

Adat-analízis a múltban



- Érdeemes látni, hogy hogyan is jutottunk idáig
 - A jelenlegi komplex adatfeldolgozás természetesen nem egyik percről a másikra lett kitalálva
- Milyenek voltak a kísérletek a múltban?
 - Legkorábban: Fényképek és besugárzott minták szemmel tanulmányozása
 - Gáz- és buborékkamra képeket ellenőriztek egyenként, betanított munkával
 - Később: Kevés paraméter (energialeadás, pozíció) mérése, és közvetlenül analízálása
 - Képesek voltak a különböző detektorokból kijövő elektromos jeleket olyan formára hozni, hogy azokból a detektor által érzékelt részecskék milyenségére tudjanak következtetni.
 - De a jeleket lehetetlen volt egyesével rögzíteni, a kor számítástechnikája ezt nem engedte meg
 - Sokcsatornás analízátorokkal elektronikusan készítették eloszlásokat a mért paramétereiből
- Végül: Amint a mért elektronikus paramétereket rögzíteni lehetett, az adatok analízise a kísérlet elvégzésétől későbbre tolódhatott
 - Lehetővé téve, hogy kifinomultabb adatfeldolgozással pontosabb mérést lehessen később elvégezni

Számítástechnika a múltban



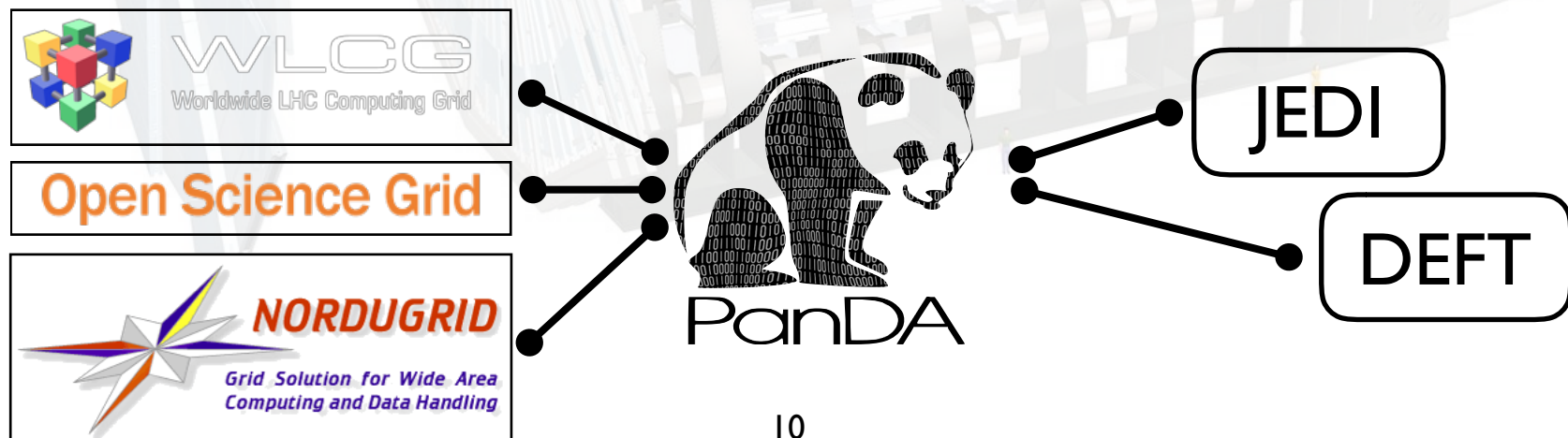
- A CERN első “valódi” számítógépe 1958-ban állt üzembe
 - Programokat és adatokat lyukszalagon tudott kapni, és természetesen mai szemmel nagyon lassú volt
- Az évek folyamán egyre nagyobb és nagyobb teljesítményű szinguláris gépeket helyeztek üzembe
- A nagy váltást a személyi számítógépek megjelenése hozta
 - Ezeket lehetett használni az adatgyűjtés irányításától kezdődően egészen az adatfeldolgozás végső lépéseiig
 - Azóta is a személyi számítógépek uralják a fizika kísérleteket

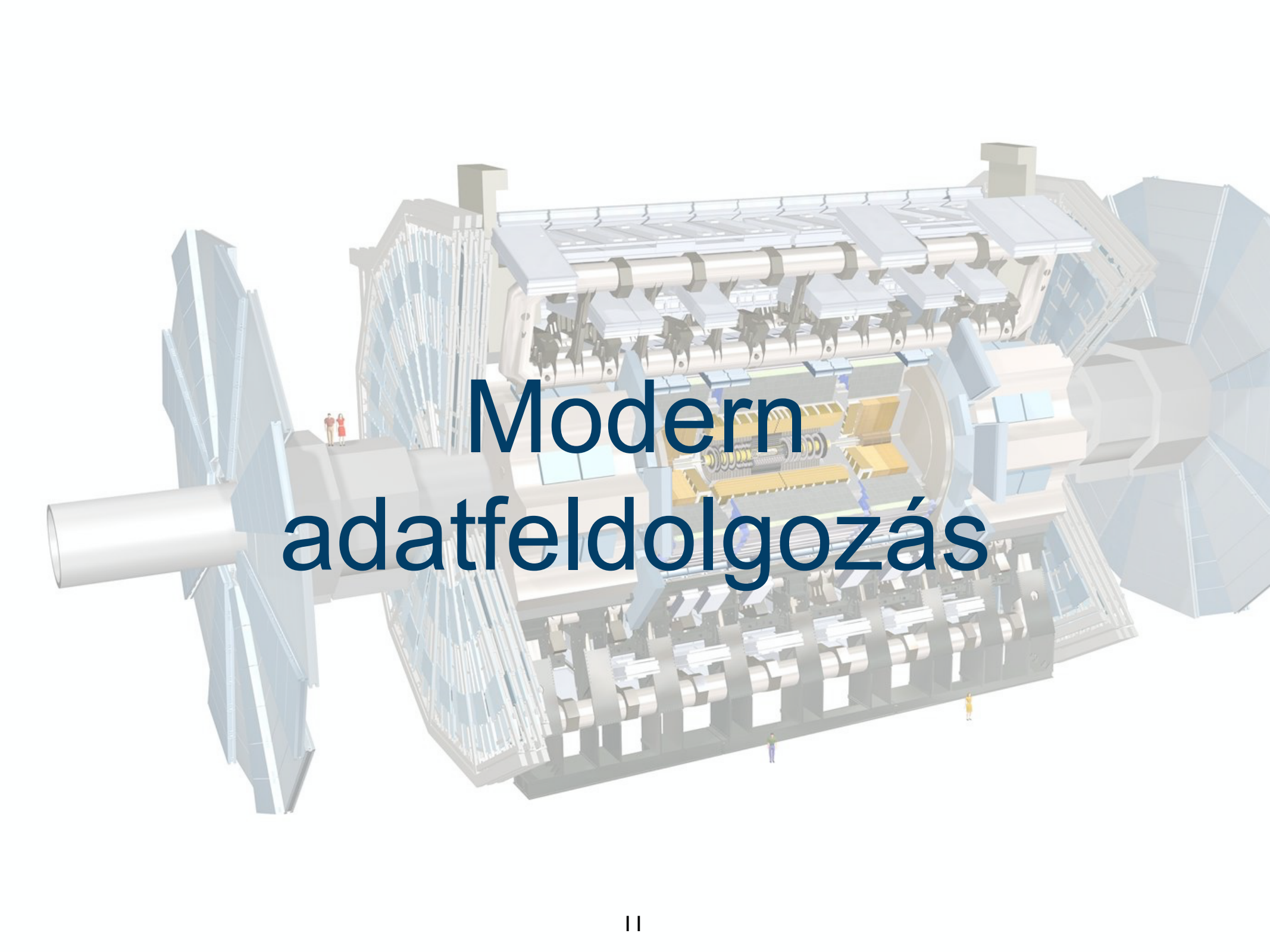


A grid



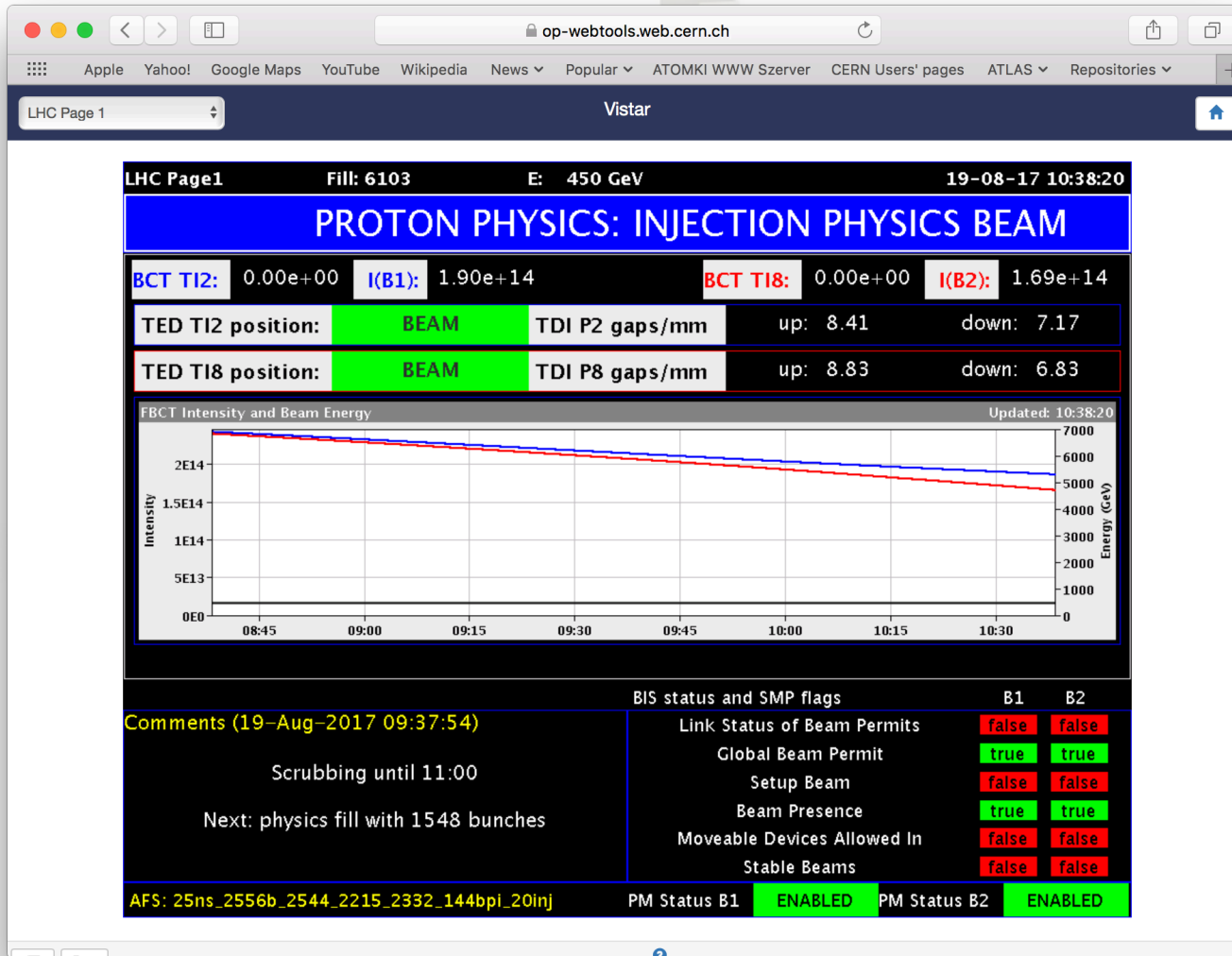
- A személyi számítógépek előretörésével hamar megjelentek az ezekből épített klaszterek
 - Az ilyen “batch cluster”-ek arra lettek kitalálva, hogy ugyanazt a feladatot, egyenként kis változtatásokkal, futtatni tudják egyszerre sok gépen/processzoron
 - A LEP adatai is ilyen klasztereken lettek feldolgozva annak idején
- Az LHC tervezésénél hamar világossá vált, hogy anyagilag nem célszerű egy nagy klasztert építeni az adatok feldolgozására → A világ sok egyéni klaszterét kell inkább összekötni egy egységes rendszerré
 - Amire az ember úgy tud feladatokat küldeni, hogy nem kell tudnia az adatai pontosan hol vannak, vagy hogy a programja pontosan hol fog lefutni
- A rendszer természetesen kellően bonyolult
 - Többféle alsóbb szintű rendszeren futtatunk sokféle “middleware” szoftvert, azon felül a kísérletek saját ütemező szoftverét, ami az egyes kísérletek offline szoftverét futtatja végül



A detailed 3D cutaway illustration of a particle detector, likely a proton synchrotron. The central feature is a long, horizontal cylindrical beam pipe. Surrounding it are multiple layers of complex, multi-faceted structures, possibly representing the detector's internal components like calorimeters or tracking chambers. The structure is highly symmetrical and intricate. Several small human figures are placed at various points within the structure to provide a sense of scale, highlighting the massive size of the equipment. The overall color palette is dominated by light blues, greys, and metallic tones, with some yellow and red highlights on specific internal components.

Modern adatfeldolgozás

Az ATLAS adatgyűjtés



Az ATLAS adatgyűjtés



op-webtools.web.cern.ch

LHC Page 1

LHC Page1 Fill: 6103 E: 450 GeV

PROTON PHYSICS: II

BCT T12: 0.00e+00 I(B1): 1.90e+14

TED T12 position: **BEAM** TDI P2 gap

TED T18 position: **BEAM** TDI P8 gap

FBCI Intensity and Beam Energy

Comments (19-Aug-2017 09:37:54)

Scrubbing until 11:00

Next: physics fill with 1548 bunches

AFS: 25ns_2556b_2544_2215_2332_144bpl_20inj

atlasop.cern.ch

ATLAS Operation Overview Systems Services Infrastructure

Peak Luminosity by Fill

Luminosity by Day

Total Luminosity

Efficiency by Day

Day

Summary

Peak Stable Lumi

Peak <Events>/BX

Avg <Events>/BX

	Lumi (mb ⁻¹)	Percent
Physics Beams Del.	0.0	-
ATLAS Ready Del.	0.0	-
ATLAS Ready Rec.	0.0	-
Del. after Warmstop	0.0	-

Luminosity

LHC Status

Az ATLAS adatgyűjtés



op-webtools.web.cern.ch

LHC Page 1 Fill: 6103 E: 450 GeV

atlasop.cern.ch

ATLAS Detector Systems ATLAS Operation Systems

Inner Detector

- PIX Pixel Detector 7
- SCT Semiconductor Tracker
- TRT Transition Radiation Tracker
- IDE Inner Detector 7

Calorimeters

- LAR Liquid Argon Calorimeter
- TIL Tile Calorimeter

Muon Spectrometer

- MDT

FSM ATLAS DETECTOR CONTROL 19-08-17 10:42:36

IDE	OK	BARREL B LAYER	S	OK	RPO	OK	FWD
EVC	OK	DISKS	S	OK	LCD	OK	Inner Detector
ENV	OK	IBL	S	OK	ZDC	D	
TEH	OK	DBM	S	OK			
BLM	OK	INF	S	OK			
BCM	OK	BARREL A	S	OK			
RAD	OK	ENDCAP A	S	OK			
		ENDCAP C	R	OK			
		INF	R	OK			
MUO	OK	BARREL A	R	OK			
BIS	OK	BARREL C	R	OK			
CAEN	OK	ENDCAP A	R	OK			
		ENDCAP C	R	OK			
		INF	R	OK			
TDG	OK	EMBA	R	OK			
LI	OK	EMBC	R	OK			
		EMECA	R	OK			
CIC	OK	EMECC	R	OK			
GOL	OK	HEFCAL A	R	OK			
ENV	OK	HEFCAL C	R	OK			
USA1	OK	INF	R	OK			
USA2	OK	LBA	R	OK			
US	OK	LBC	R	OK			
SDX	OK	EBA	R	OK			
LUX	OK	EBC	R	OK			
		INF	R	OK			
SAF	OK	MDT	BARREL A	S	OK		
DSS	OK		BARREL C	S	OK		
SNF	OK		ENDCAP A	S	OK		
EXT	OK		ENDCAP C	S	OK		

Infrastructure

Total Luminosity

Efficiency by Day

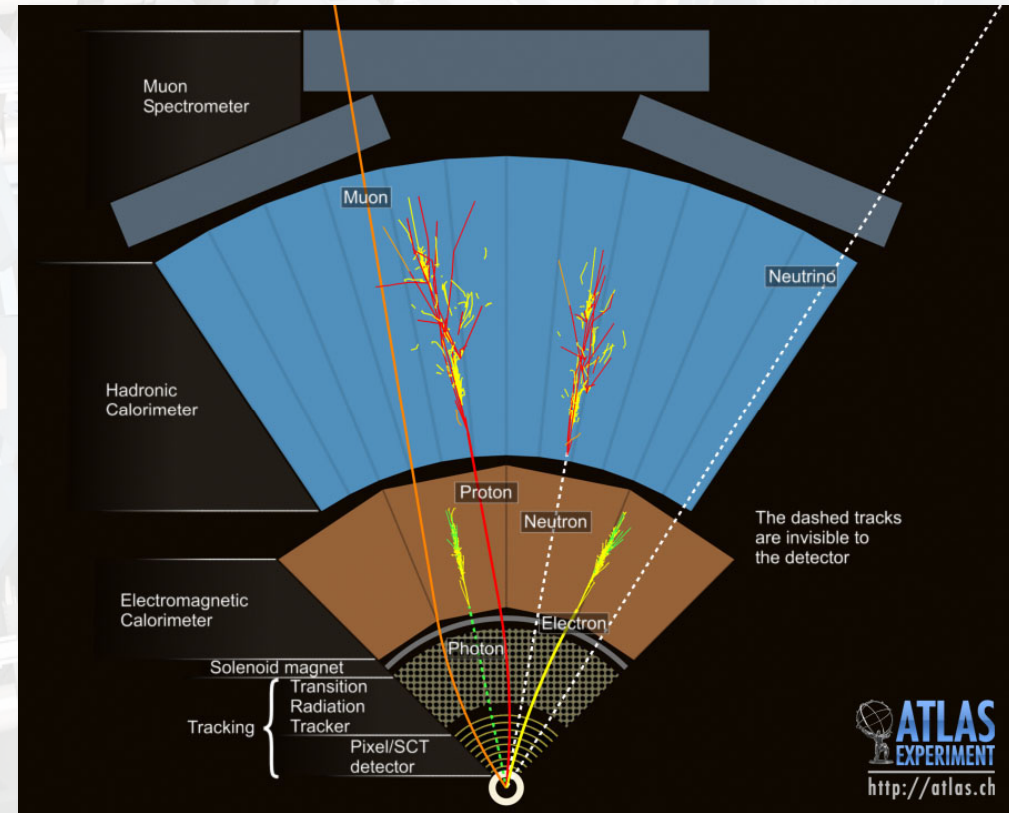
[Log Scale]

LHC Status

Rekonstrukció



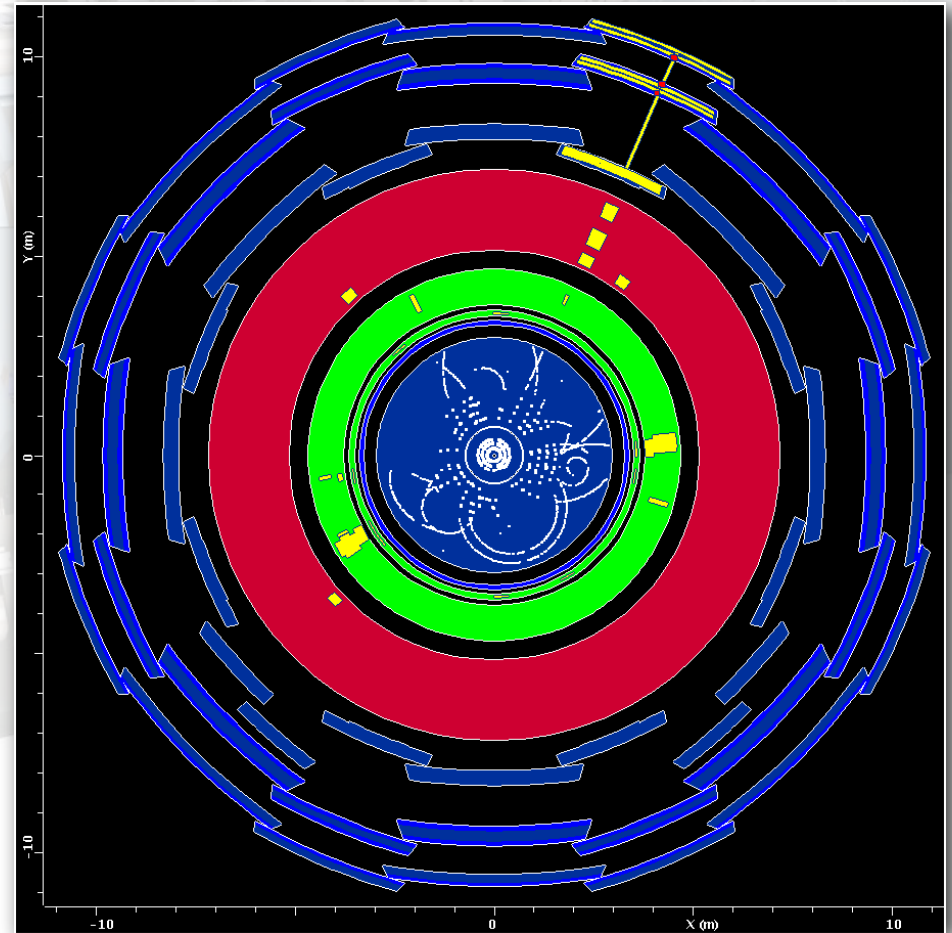
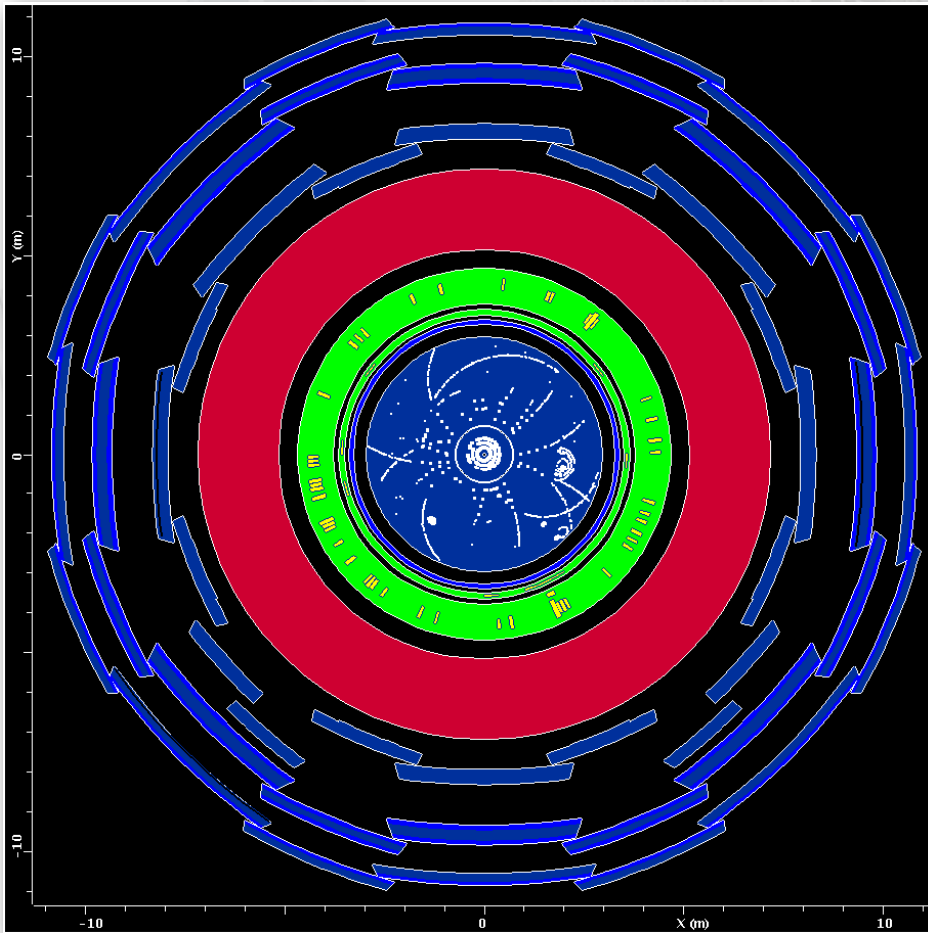
- A detektorokat úgy építjük meg, hogy a számunkra érdekes részecskék egymástól különböző jelet hagyjanak bennük
 - Így képesek vagyunk a részecskéket azonosítani és a tulajdonságaikat megállapítani
- Pl. egy elektron rekonstrukciójához:
 - Keresnünk kell egy töltött részecske nyomot a nyomkövető detektorban
 - Ehhez hozzá kell tudnunk rendelni egy energia-klasztert az elektromágneses kaloriméterben
 - A klaszter és töltött nyom tulajdonságaira sok technikai feltételt is szabunk
 - Nem szabad “aktivitást” találnunk a talált pályán a hadron kaloriméterben



Trigger



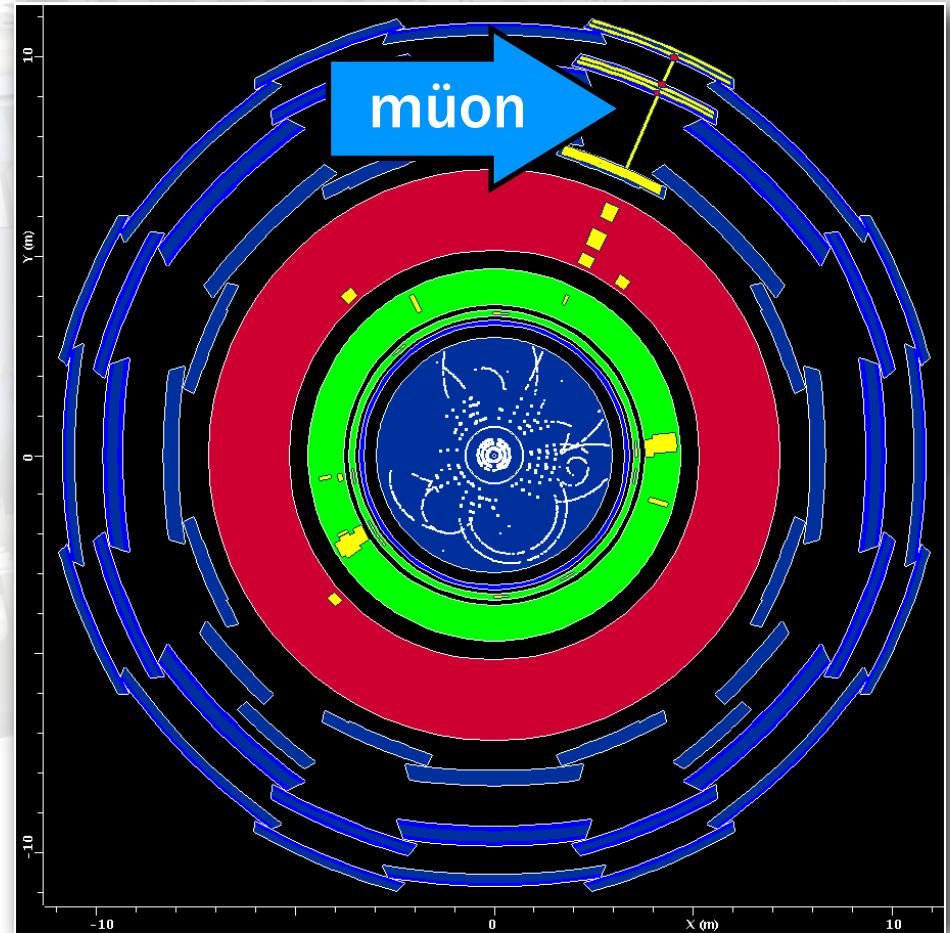
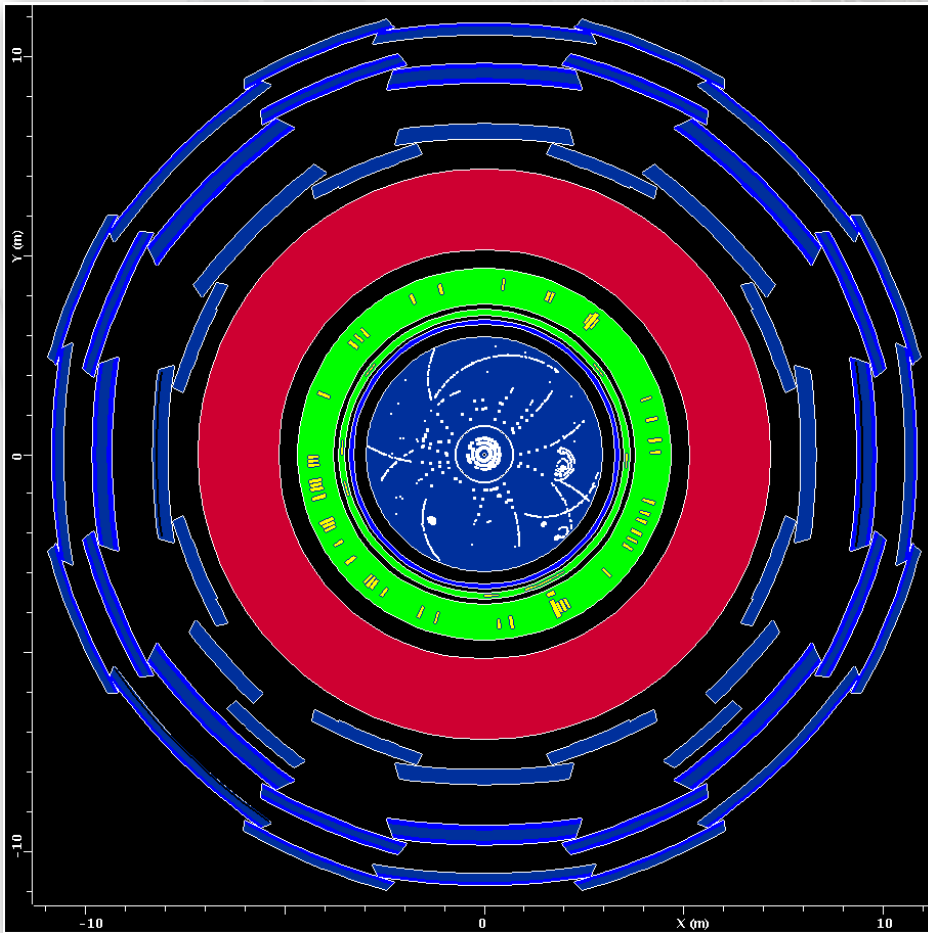
- Ez történik legelőször, de egyszerűbb a rekonstrukció után beszélni róla...
- Az érdekes eseményeket azok gyors rekonstrukciójával választjuk ki



Trigger



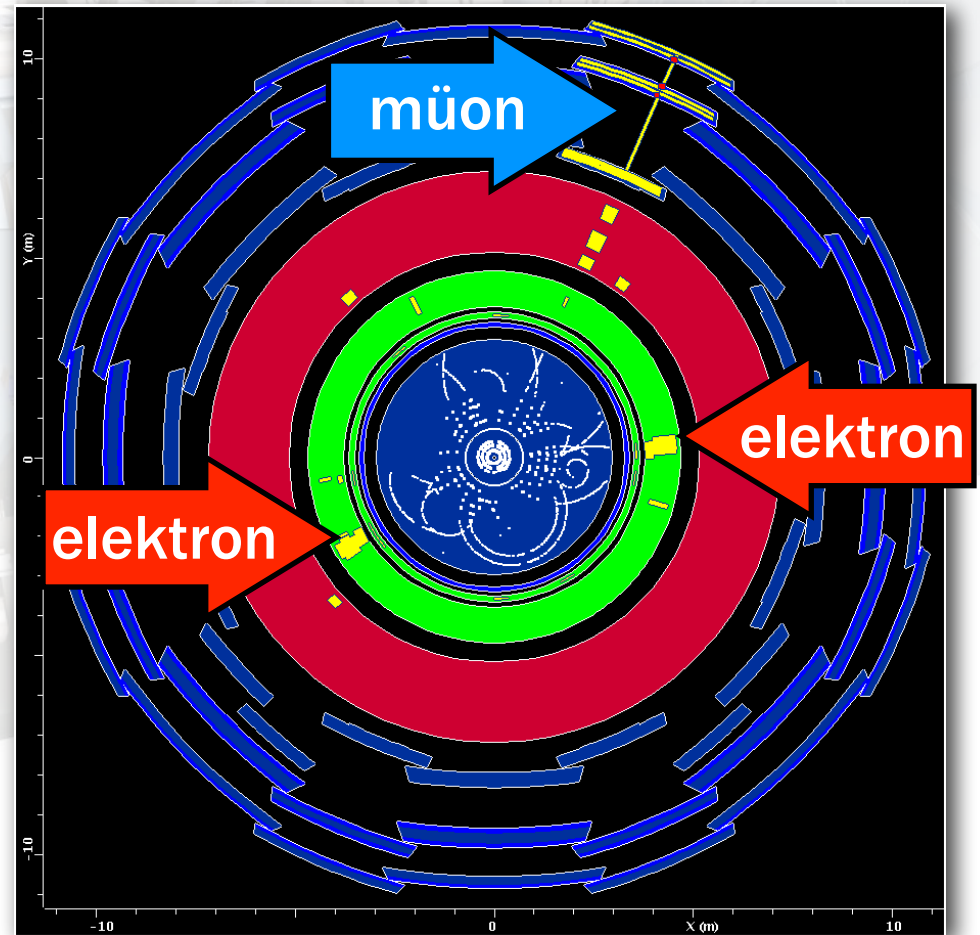
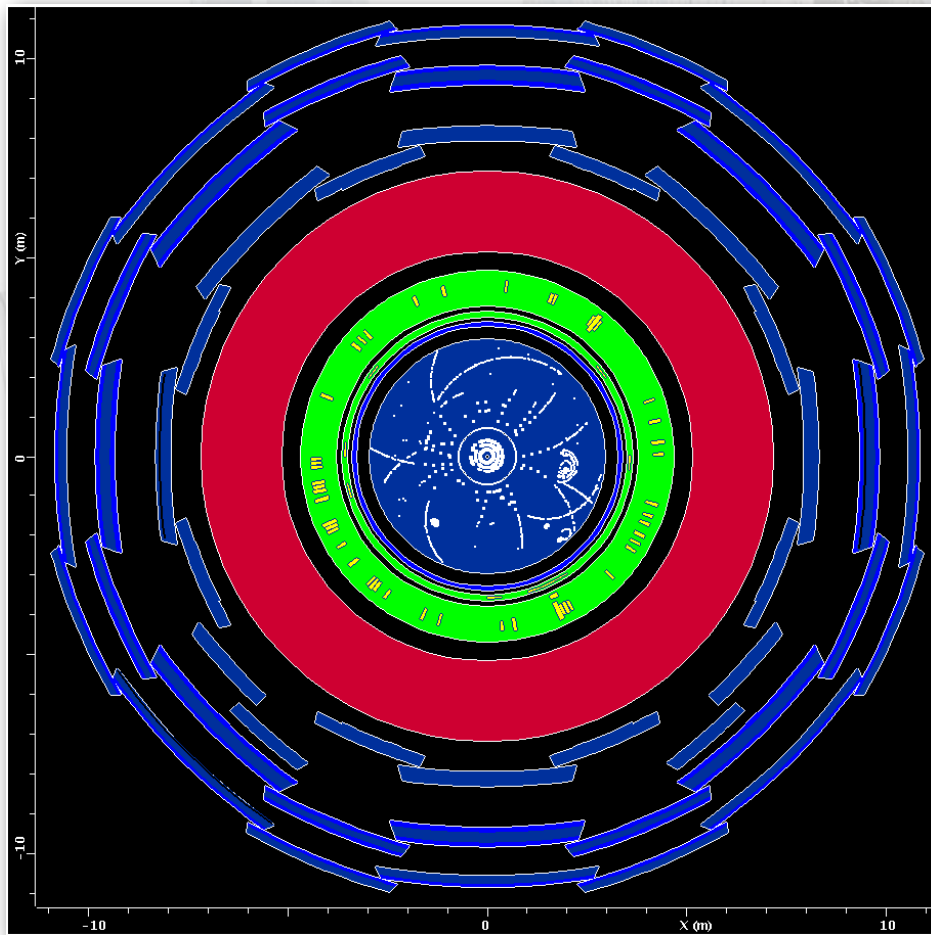
- Ez történik legelőször, de egyszerűbb a rekonstrukció után beszélni róla...
- Az érdekes eseményeket azok gyors rekonstrukciójával választjuk ki



Trigger



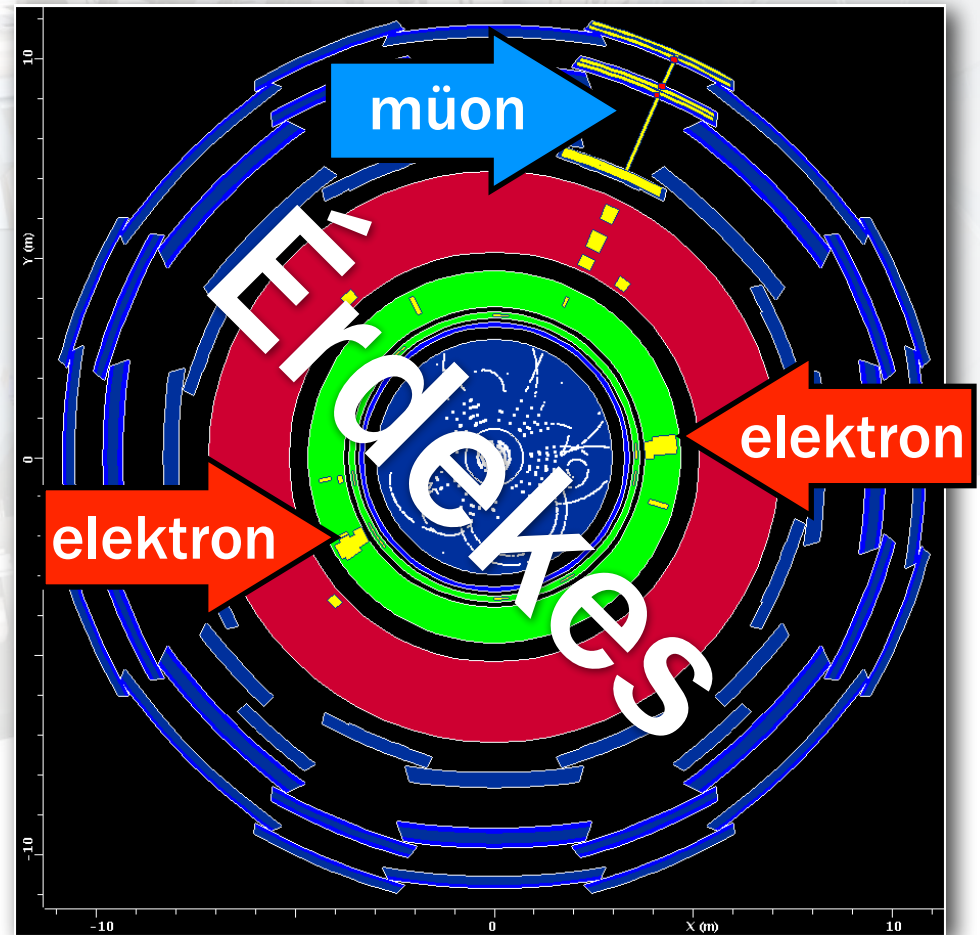
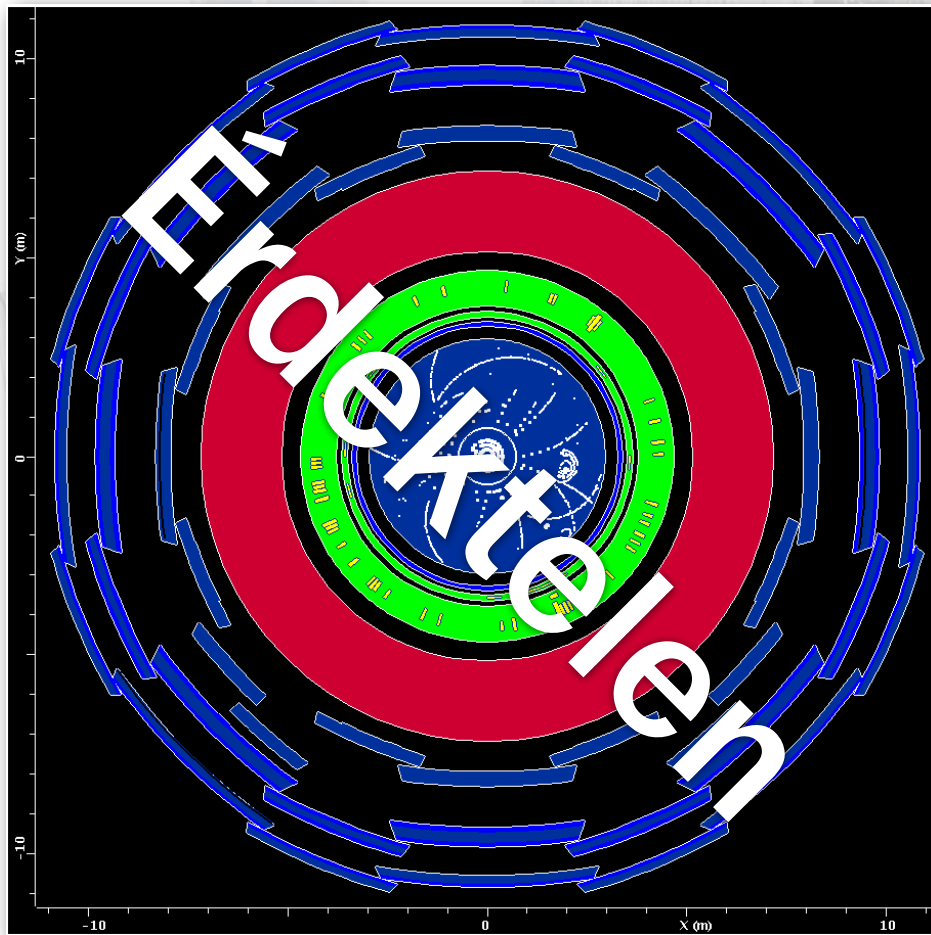
- Ez történik legelőször, de egyszerűbb a rekonstrukció után beszélni róla...
- Az érdekes eseményeket azok gyors rekonstrukciójával választjuk ki



Trigger



- Ez történik legelőször, de egyszerűbb a rekonstrukció után beszélni róla...
- Az érdekes eseményeket azok gyors rekonstrukciójával választjuk ki



Szimuláció

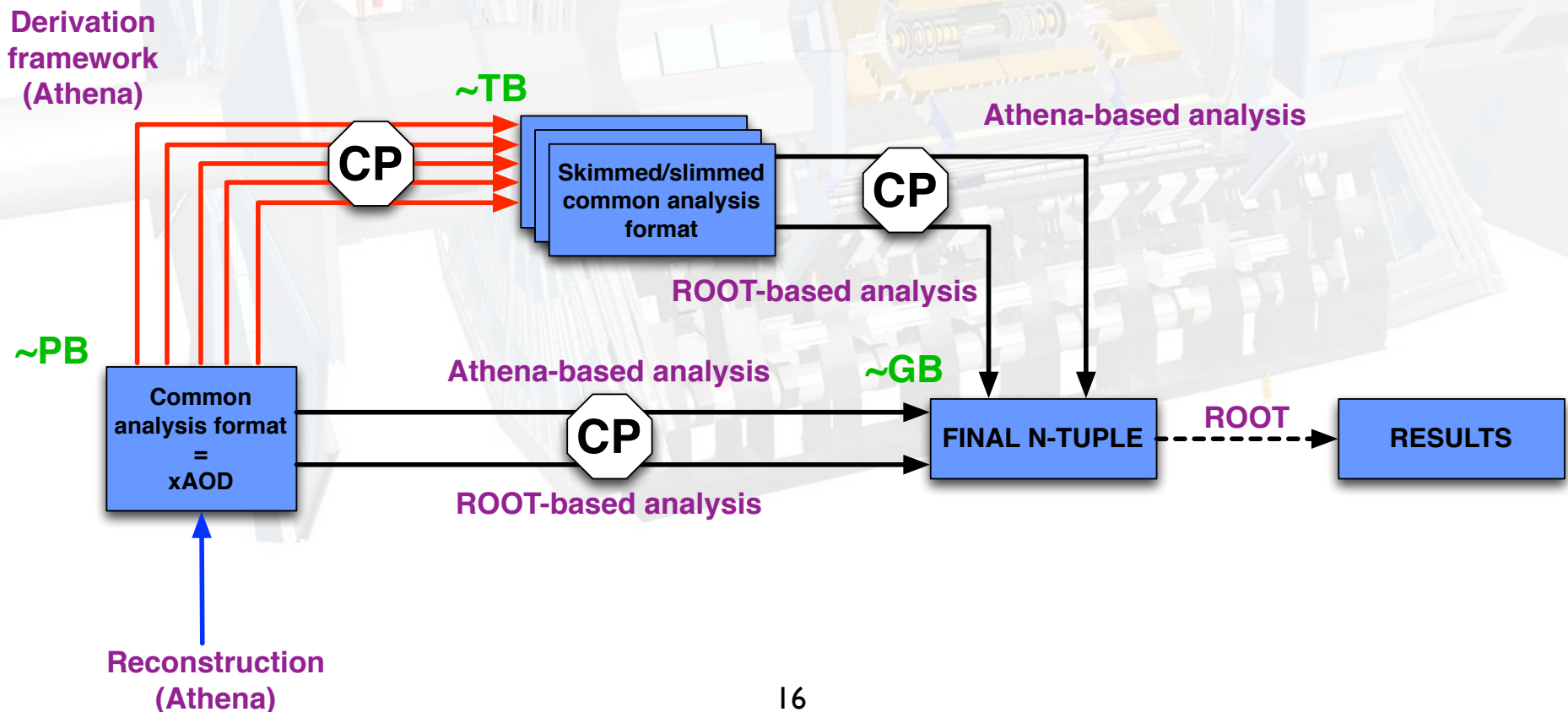


- A mért adatokat az esetek döntő részében csak szimulációk segítségével tudjuk értelmezni
 - Még a legegyszerűbb fizikai folyamatok is annyira bonyolult módon hatnak kölcsön a detektorokkal, hogy szimulációkkal kell meggyőzzük magunkat, értjük mi történik a kísérletben
- Sok lépésből áll:
 - **Esemény generálás:** Az elméleti fizikus kollégák írnak rá szoftvert, hogy véletlenszerű statisztikával valamilyen fizikai folyamathoz eseményeket gyártsunk.
 - **Detektor szimuláció:** A generált események stabil részecskéinek a kölcsönhatását modellezzük a teljes detektorral. Megállapítva, hogy melyik részecske mekkora energiát adott le a detektor különböző pontjain.
 - **“Digitalizáció”:** Szimuláljuk, hogy ekkora energialeadásokra a detektor kiolvasó elektronikája hogyan reagálna a valóságban. Kimenatként ugyanolyan formátumú adatot generálva mint amit a detektorból kiolvasunk.
 - Itt figyelembe vesszük azt is, hogy a “jel” eseményünkön kívül milyen más proton-proton ütközések történtek ugyanabban a proton-csomag ütközésben, és valamivel korábban.
 - **Trigger és rekonstrukció:** Ugyanazt a trigger és rekonstrukciós szoftvert futtatjuk a digitalizáció kimenetén amit a valódi adatokra is alkalmaztunk.

Analízis modell



- Nem engedhetjük meg, hogy minden fizikus szervezetlenül dolgozza fel a több petabájtnyi valódi és szimulált adatot az analízisében
 - Ehhez nem elegendő a grid kapacitása
- Az analízis eleje szervezett formában zajlik a legtöbb kísérletben
 - “Közös” analízis lépéseket egységesen végzünk, hogy ne kelljen a grid infrastruktúrának ugyanazt csinálnia többször is



Mit csinál az anlizátor?



- Kiválasztja, hogy melyik adatszettet fogja feldolgozni
 - Egyúttal azt is, hogy milyen triggeret fog használni az adatok kiválogatásához, ami hatékonyan őrizte meg a számára érdekes fizikai folyamat eseményeit
- Kiválasztja, hogy mely Monte Carlo szimulációkat kell figyelembe vennie az analízisében
 - Ez lefedi az analízis által keresett folyamat szimulációját, és minden más “háttér” folyamat szimulációját amire érzékeny lehet az analízis
- Lefuttatja az analízis szoftverét ugyanolyan beállításokkal az adat és szimulációs adatszettekre is, nagyon kis adatfájlokat létrehozva
 - Sokszor ez több lépésben történik, de végül különböző mennyiségek eloszlásait határozzák meg az analízisek a legtöbb esetben
- “Statisztikai analízist” végez a kapott eloszlásokon
 - Megméréendő valamilyen paramétert (részecske tömeget vagy egy folyamat hatáskeresztmetszetét), vagy felfedezési/kizárási valószínűséget meghatározva
- Elvégzi az egészet N alkalommal, a detektor és adatfeldolgozás minden szisztematikus bizonytalanságának hatását megállapítva az analízis végső eredményére.

Az ATLAS offline szoftvere



- Kb. 4 millió sor C++ és kb. 1 millió sor Python
 - Publikusan sajnos csak részei elérhetőek, ezen még dolgozunk...
 - Óvatos becslések is > 100 millió dollárra teszik a szoftver "értékét" (több száz ember több évtizedes munkája)
- Folyamatosan fejlesztjük
 - Jelenleg a szoftver 21. kiadásánát használjuk produkcióban
- A teljes szimulációs + rekonstrukciós szoftvert kb. 5 óra megépíteni egy 16 processzoros gépen
 - Az analízis szoftverkiadás csak kb. 20 percbe kerül
 - Így a szoftverfejlesztést is egyedi módon kell végeznünk, ahol az egyes fejlesztők mindig csak a szoftver egy kis részét építik meg magukban
- Egy világméretű elosztott fájlrendszeren (CVMFS) biztosítjuk a felhasználóknak
 - És a grid site-oknak



Az ATLAS offline szoftvere




atlas / athena

This project Search

Project Repository Registry Issues Merge Requests 100 Pipelines Settings

Home Activity Cycle Analytics



athena


ATLAS Experiment main repository for Athena code

★ Unstar 73 Fork 556 KRB5 https://:@gitlab.cern.ch:844: [Download] [Add] [Global]

Files (1.4 GB) Commits (17,478) Branches (54) Tags (2,044) Readme LICENSE

master athena / [Add]

Find file History [Download]

 Merge branch 'cherry-pick-3bb68a8b-master' into 'master' ... c8d74b41 [Download]
Tim Martin committed about an hour ago

Name	Last commit	Last Update
AsgExternal/Asg_Test	Merge branch 'svncopy-Asg_Test-02-05-03' into '21.2'	a month ago
AtlasGeometryCommon	Merge branch 'remove-TruthSvc-warnings-in-forward-dete...	2 months ago
AtlasTest	AthenaPoolMultiTest: Fixing ATN test failures.	a week ago

Analízis program-kód




Attila Krasznahorkay / HSSIPAnalysis

This project Search

Project Repository Registry Issues 0 Merge Requests 0 Pipelines Settings

Home Activity Cycle Analytics




HSSIPAnalysis

Analysis code prepared for the HSSIP 2017 students.

Star 0 Fork 0 KRB5 https://:@gitlab.cern.ch:844: Global

Files (952 KB) Commits (21) Branch (1) Tags (0) Readme Add Changelog Add License Add Contribution guide Set up CI

master HSSIPAnalysis / Find file History

 Removed the TDirChanger class from the repository, as by now it ... 7806ba6f
Attila Krasznahorkay committed 2 months ago

Name	Last commit	Last Update
CPPAnalysis	Removed the TDirChanger class from the repository, as by now it	2 months ago
PythonAnalysis	Started writing some Python example code as well. Demonstrating the	2 months ago
ROOTScripts	Added a small ROOT macro for making a nice looking plot out of the ...	2 months ago

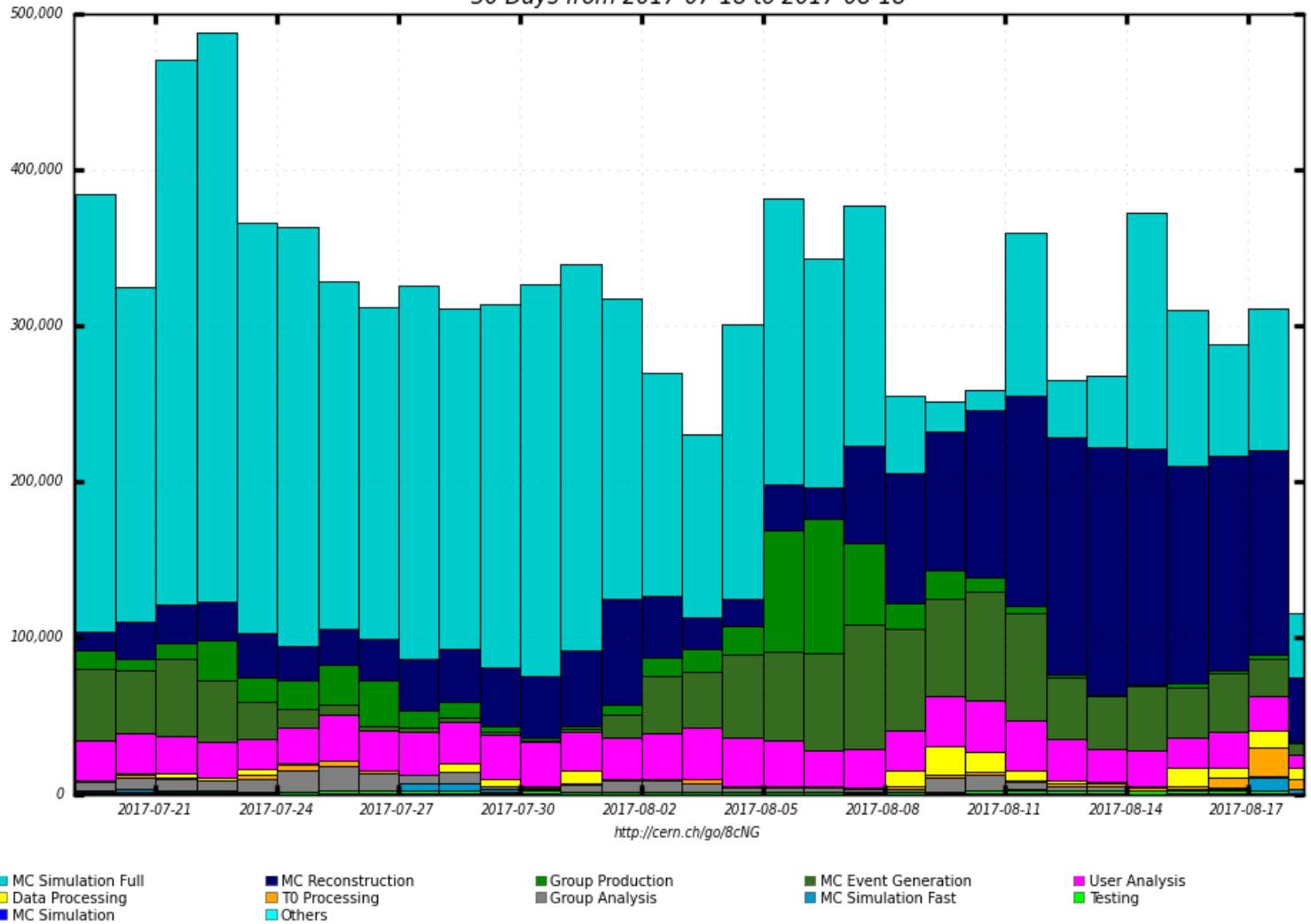
Ez a High-School Student Internship Program-ra készített nagyon egyszerű analízis kódja.

Mekkora az ATLAS grid?



dashboard

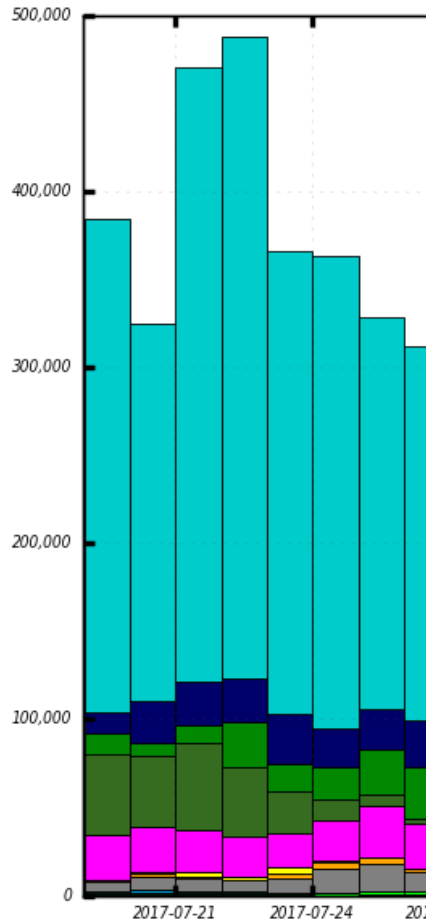
Slots of Running Jobs
30 Days from 2017-07-18 to 2017-08-18



Mekkora az ATLAS grid?



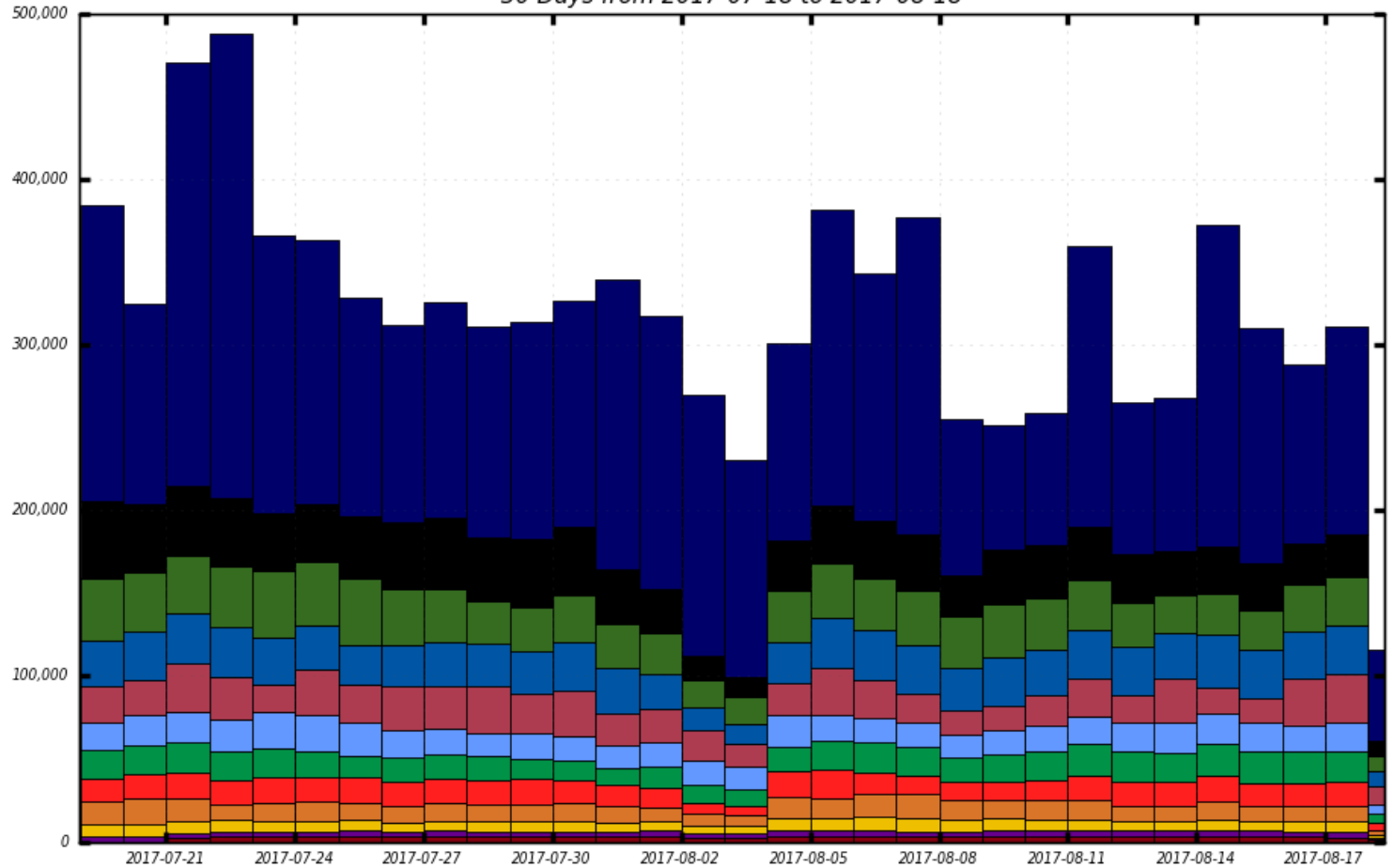
dashboard



■ MC Simulation Full
■ Data Processing
■ MC Simulation
■ MC Reconstru
■ TO Processing
■ Others

Slots of Running Jobs
30 Days from 2017-07-18 to 2017-08-18

dashboard



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

Maximum: 488,290 , Minimum: 0.00 , Average: 310,521 , Current: 115,626

Összefoglalás



- Mostanra a szoftverfejlesztés egy nagyon lényeges elemévé vált a fizikának
- Gyakorlatilag semmilyen kísérlet nem létezhez saját adatgyűjtő és feldolgozó szoftver nélkül manapság
- Az adatok fizikai feldolgozását programok írásával végezzük
 - Minden LHC fizikusnak muszály legalább “elfogadhatóan” programozni tudni
- Az LHC több száz PB adatát a világméretű grid sok-százezer processzorán dolgozzuk fel