

# Největší chrlič dat na světě

Vědci v ženevském výzkumném středisku CERN budou se svým novým kruhovým urychlovačem částic produkovat **MILIONY GIGABAJTŮ DAT DENNĚ**. Tím se tento vědecký experiment stane také prvním praktickým testem pro GRID – internet příští generace.

ANDREAS HENTSCHEL

**Z**ákladním kamenem celosvětové pavučiny WWW je černá kostka, na níž je nalepena malá cedulka. „Nevypínat! Toto je server!“ stojí na papírku – údajně jako varování pro příliš agilní a šetrné uklízečky. Tento první webový server uvedl do provozu fyzik Tim Berners-Lee roku 1990 v Evropském centru jaderného výzkumu CERN v Ženevě. Jeho pracovní stanici NeXT můžete v tamějším muzeu obdivovat ještě dnes. Vědec si tehdy potřeboval vyměňovat data s kolegy z jiných institutů a při řešení tohoto úkolu jaksi mimochodem vynalezl World Wide Web. Dnes už si tuto síť vůbec nedokážeme odmyslet ze svých všedních dnů.

Nyní vědci v CERN možná pokládají další základní kámen, a to pro novou generaci internetu. Opět si chtějí vyměňovat data s kolegy na celém světě. Jenomže tentokrát už ne dokumenty velké několik kilo- či megabajtů, nýbrž data z mimořádně rozsáhlých experimentů. Jejich objem se odhaduje na 90 milionů gigabajtů za rok, tedy 90 petabajtů, k nimž chtějí mít přístup

fyzikové celého světa. Pracovníci muzeí budoucnosti tedy asi budou mít své problémy, budou-li chtít novou generaci webu představit nějakým příkladným exponentem.

Tento neuvěřitelný datový objem totiž vyprodukuje největší stroj, jaký kdy lidská ruka postavila: je jím Large Hadron Collider (LHC). To je cca 26,7 kilometru dlouhý kruhový prstenec pro urychlování jaderných částic, který je uložen v systému tunelů vedených nedaleko Ženevy v hloubkách kolem sta metrů pod zemským povrchem. 10. září 2008 bylo zařízení uvedeno do zkušebního provozu a v tunelu do sebe začaly narážet první svazky protonů urychlených téměř na rychlost světla. Po zhruba týdenním provozu se bohužel objevila technická závada a systém byl dočasně odstaven – což však nic nemění na jeho obdivuhodných vlastnostech, s nimiž vás zde chceme seznámit.

Cíle zhruba 5 000 vědců, kteří v CERN pracují pro 300 institutů v 50 zemích, je na tomto místě možno načrtnout jen vel-



## Čidla obřího stroje

Čtyři detektory v urychlovači částic rozpoznají pomocí nejmodernější sensorové techniky všechny myslitelné elementární částice. Identifikují je podle jejich letových trajektorií, rychlostí a elektromagnetických vlastností. Náš schematický náčrt vpravo znázorňuje pohled do útrobu detektoru pro experiment nazvaný ATLAS.



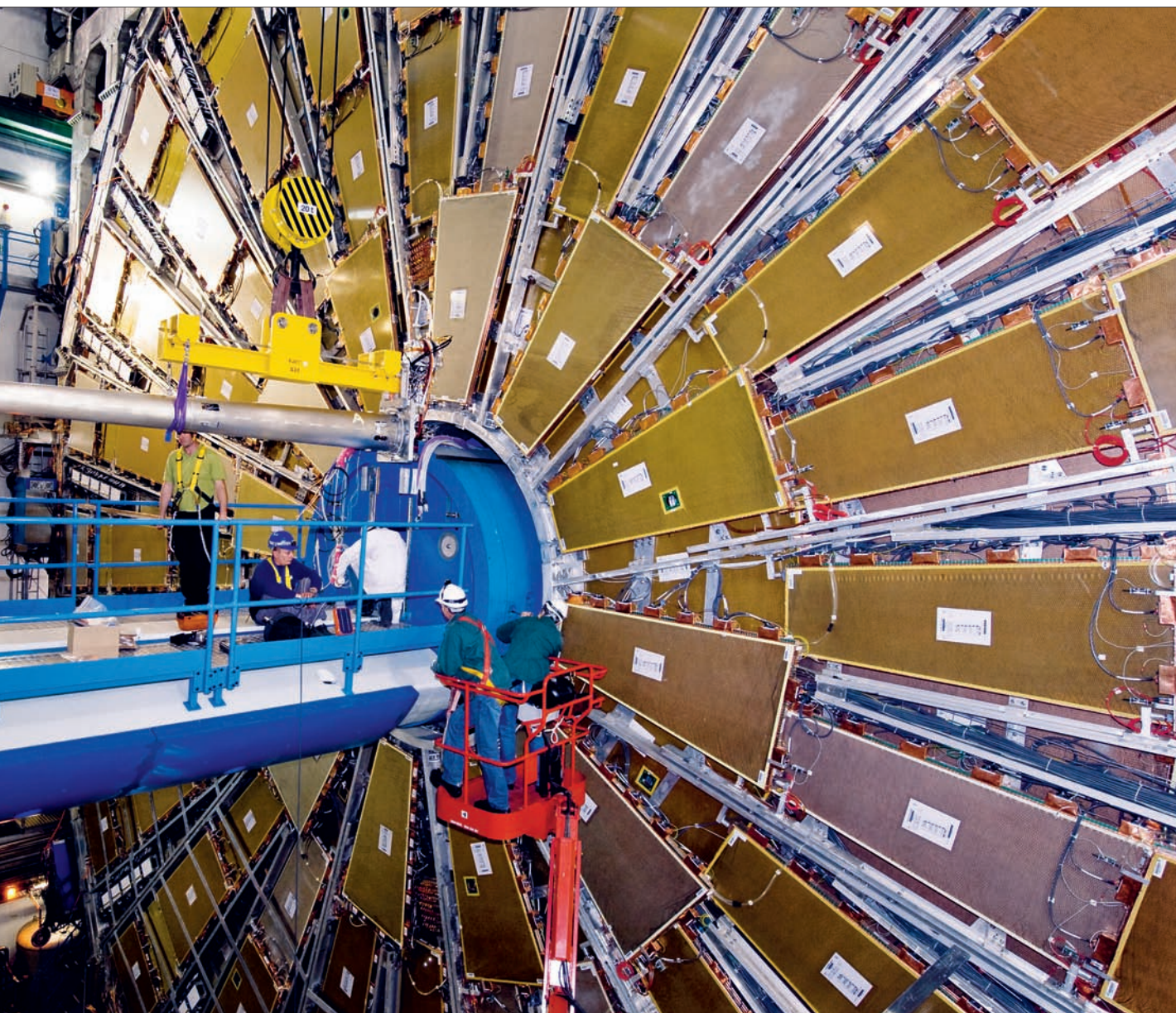


FOTO : CERN/MAXIMILIAN BRICE, CLAUDIA MARCELLONI

**Pixelový detektor**

Pixely umístěné na křemíkových čípcích proměřují křivky drah částic, které vznikají při kolizích, s přesností na 14 mikrometrů. Na vývoji tohoto senzoru se podíleli také čeští odborníci.

**Křemíkové proužkové čítače**

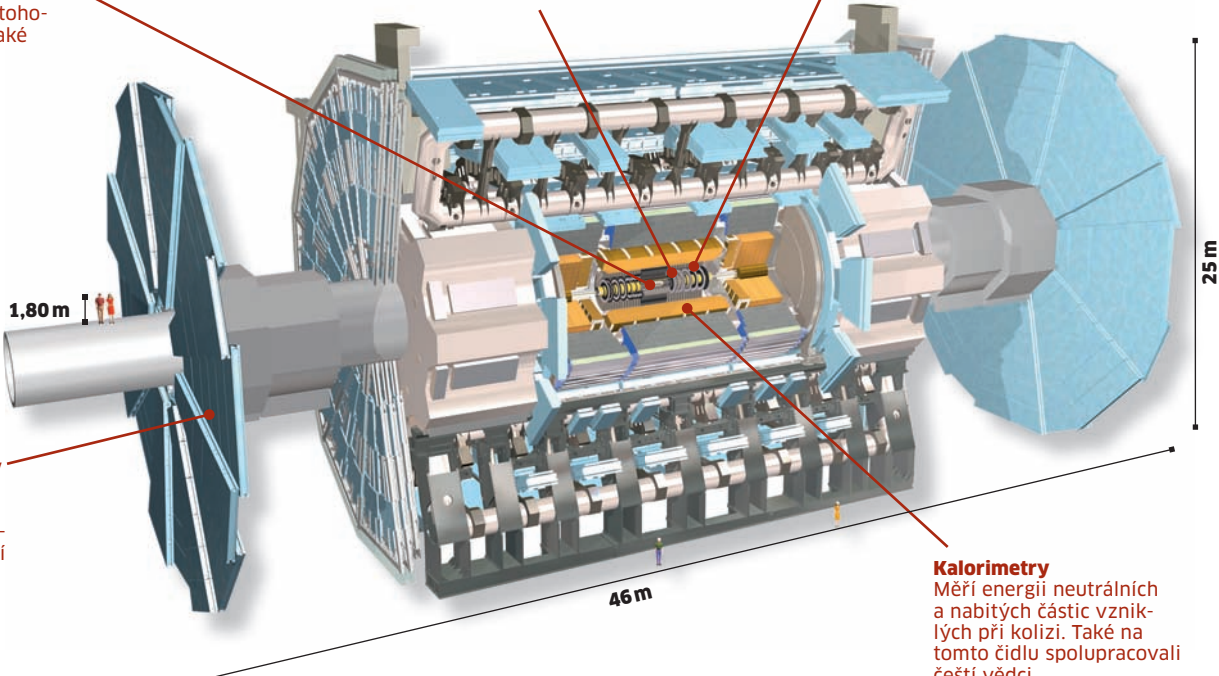
Jsou napojeny na pixelové detektory a měří další letovou dráhu částic - s ubývající přesností.

**Stopový detektor**

Malé trubičky naplněné xenonovým plynem měří souřadnice nabitých částic s přesností na 120 mikrometrů.

**Rozměry**

Z porovnání s výškou člověka lze získat názornou představu o velikosti detektoru ATLAS.



**Myonové spektrometry**

Myony jsou elektricky nabitě částice, které se neregistrují uvnitř detektoru. Vnější senzory měří koordináty těchto myonů s přesností na 100 mikrometrů.

**Kalorimetry**

Měří energii neutrálních a nabitých částic vzniklých při kolizi. Také na tomto čidlu spolupracovali čeští vědci.



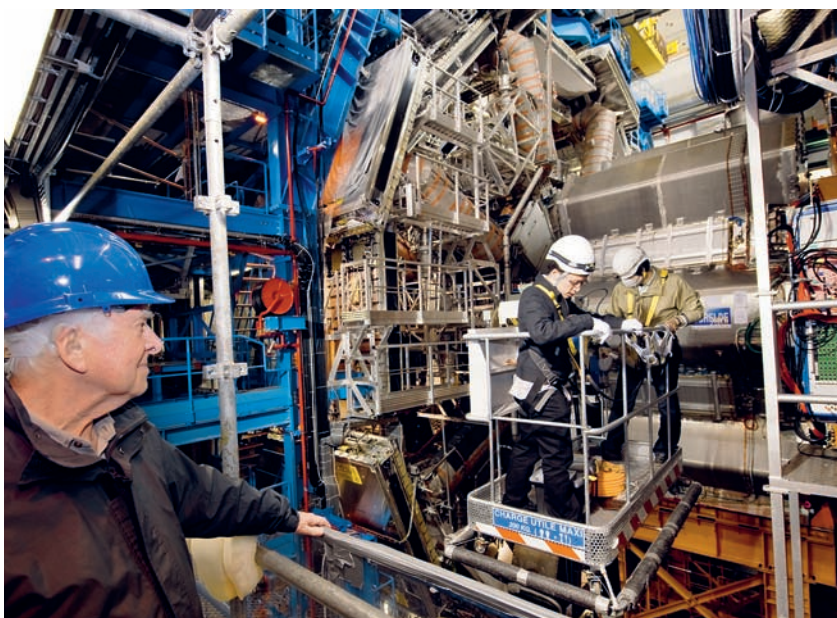
mi hrubě: ve čtyřech různých experimentech chtějí fyzikové prokázat existenci částic, které dosud znali jenom na papíře – například Higgsovy částice. Kromě toho mají být v dalším pokusu nastoleny stavy, jaké panovaly bezprostředně po velkém třesku; pak by bylo možno studovat tzv. kvark-gluonovou plazmu. Od dalších experimentů si výzkumníci dokonce slibují vznik malých černých děr (doufejme, že opravdu jen malých...) nebo na Zemi dosud neprokázané temné hmoty.

### Datové tsunami: Jak zpracovat 600 TB za sekundu?

Černé díry a temná hmota – takové experimenty snad musí nadchnout i toho, kdo z fyziky ve škole propadal. Fascinace je zde zcela namístě. Dlouhá léta budují vědci pokusná zařízení, aby na nich prováděli experimenty, jejichž výsledky přetrvávají jen pikosekundy. Díky výkonné počítačové technice to však pro fyziky není žádný problém: částice vznikající při kolizích protonů jsou registrovány – prakticky vzato nafilmovány – obrovskými detektory. Stopy částic se ukládají v podobě surových digitálních dat a později používají jako základ pro rekonstrukci experimentů, které se pak simulují v počítači.

Každý ze čtyř experimentů v CERN má svůj vlastní detektor, ten největší z nich je 25 metrů vysoký a 46 m dlouhý, což představuje zhruba poloviční velikost pařížské katedrály Notre Dame. Sensory v různých obvodových vrstvách těchto gigantických měřících přístrojů zaznamenávají chaos částic po srážce protonů. Mají při tom skutečně napilno: podle intenzity svazku protonů či jiných zkoumaných částic dochází za sekundu až k miliardě takových kolizí. V částicovém detektoru CMS (Compact Muon Solenoid), v němž má být mimo jiné prokázána Higgsova částice, zaznamenává informace 15 milionů senzorů s frekvencí 40 MHz – výstup představuje 600 TB dat za sekundu.

„Je to opravdové datové tsunami,“ přiznal vědecký pracovník CERN Pere Mató před několika měsíci na sympoziu Euroscience Open Forum v Barceloně. Žádné výpočetní středisko by se s ním nedokázalo vypořádat. To však ani není nutné, neboť největší podíl tohoto datového objemu představuje v pravém slova smyslu datový balast. Ten popisuje chybně směrované nebo již známé částice. Z deseti milionů srážek, odhadují badatelé, jsou zajímavé dvě až tři, při nichž pokaždé vzniká kolem stovky nových částic. Tajemnou Higgsovu částici očekávají fyzikové dokonce jen jednou za bilion kolizí.



**Detektory:** Jsou plné senzorů a procesorů, které mají pomoci při pátrání po dosud neznámých částicích.

## „Je to skutečné datové tsunami!“

Pere Mató, vědecký pracovník CERN

### Digitální síť: Po částicích pátrá 60 000 procesorů

Hodně přes 99 % zaznamenaných dat tedy může putovat rovnou do koše. O to se v prvním kroku starají speciální, pro tuto aplikaci vyvinuté procesory a logické členy, které pracují přímo v detektorech – Pere Mató jim říká Level-1-Trigger. Tyto součásti analyzují jednoduchá data, jako rychlost částice, její elektromagnetické vlastnosti nebo energetickou úroveň, a na základě těchto informací ihned vyřadí mnoho nezajímavých údajů. Během tří až čtyř mikrosekund tyto čipy rovnou zavrhnou 9 999 z 10 000 událostí.

Druhý krok, Level-2-Trigger, ještě jednou proseje zbylé události. Zde opět speciální čipy navzájem porovnávají výsledky měření různých senzorů a dále vyloučí devět z deseti událostí – během jediné milisekundy. Podobně důkladně prověří zbylá data High-Level-Trigger, poslední kontrolní stupeň. Zde musí procesory do jedné sekundy pomocí komplexních algoritmů



**100 000 procesorů:** Vyhodnocují extrémně velké datové objemy z různých detektorů a experimenty znovu simulují.

údaje zrevidovat a jednotlivá data zařadit do celkové události. Pro každý experiment jsou tu k dispozici farmy 2 500 PC – aby na konci ještě jednou vyřadily 99 % už tak silně prořídlých dat. Celkem 60 000 procesorů tak nakonec mezi deseti miliony událostí najde jednu jedinou, která se uloží. I tak jde o slušné množství: jenom z experimentů v urychlovači částic se na pevné disky v CERN bude ročně ukládat 15 petabajtů dat.

Fascinující je však nejen množství získaných informací. Celé zařízení uvádí každého byt jen zpola technicky nadaného člověka v bezmezný úžas, jaký je snad vyhrazen jen malým dětem. Od prvních projektů tohoto zařízení uplynulo 24 let, konečné náklady se pravděpodobně vyšplhají na tři miliardy eur, bylo spotřebováno 80 000 tun materiálu, jenom 34 000 tun spolykaly čtyři zmíněné detektory. V trubici stočené do kružnice o délce 26 659 metrů panují teploty těsně nad absolutní nulou. Částice tak nepatrné, že nejsou vidět ani v elektronovém mikroskopu, pro-





**27 kilometrů:** Tolik měří tunelový systém urychlovače Large Hadron Collider výzkumného centra CERN nedaleko Ženevy.



**LHC GRID:** Vědcům v CERN umožní využívat volné výpočetní kapacity po celém světě.

létávají trubici téměř rychlostí světla – dokážete si něco takového vůbec představit?

A přesto jsou experimenty zcela reálné. V LHC se nejprve protonům v předřazeném lineárním urychlovači pomocí elektrických polí udělí vysoká rychlost. Pak se svazky protonů zavedou do dvou tzv. radiačních trubec, v nichž protiběžně krouží kolem dokola, přičemž jejich rychlost se zvýší až téměř na hodnotu rychlosti světla. Tyto trubice jsou přibližně 5,5 cm silné ocelové roury, v nichž panuje téměř dokonalé vakuum. Protony jsou v nich ve správné dráze udržovány magnetickým polem vytvářeným supravodivými magnety, které je asi 180 000krát silnější než magnetické pole Země. Pro udržení jejich supravodivosti musí být magnety intenzivně chlazeny – na 1,9 kelvina, tedy na  $-271,25^\circ\text{C}$ . To je jen těsně nad absolutní nulou, a dokonce ještě o něco chladněji než ve vesmíru. Jenom ochladit všech 8 000 magnetů rozmístěných podél prstence na jejich provozní teplotu trvalo 16 měsíců – a kilometry

dlouhý prstenec se přitom smrští o deset metrů.

### **Tuny bitů: CERN vyprodukuje 90 petabajtů za jediný rok**

Enormní náklady jsou ovšem nezbytné. Ačkoliv protony v urychlovači mají nepředstavitelně malou hmotnost, díky své rychlosti disponují stejnou kinetickou energií, jakou má 40tunový kamion při rychlosti 150 km/h. Co se stane, když se protony vymknou kontrole, poznali výzkumníci při jedné testovací nehodě: protony jeden dílec ocelové roury doslova rozpáraly. Za normálních okolností se však jejich svazky zkříží uvnitř detektorů – a protony narazí do sebe. Pokud se pokus zdaří, kolidující protony se rozpadnou v záblesku záření a vzniknou nové částice.


Ovšem zda se experiment skutečně podaří, to se vědci dozvědí až po měsících vyhodnocování. Pak z vyfiltrovaných dat konečně počítače zrekonstruují průběh pokusu, a to hned několikrát: během času jsou detektory justovány a kalibrovány stále

jemněji – a odpovídající změny zohledňovány při rekonstrukcích v počítačích. Pokusy se virtuálně opakují až třikrát za rok – pro tento úkol je vyčleněno 20 000 procesorů. CERN se tak stává největším producentem dat na světě. Surová data, záložní kopie, náročné simulace – to všechno představuje asi 90 petabajtů dat, která ženevský institut vygeneruje za jediný rok.

„Jsou to tuny bitů,“ říká Gonzalo Merino ze španělského inženýrského střediska Pic a hned také formuluje problém: „Je toho mnohem více, než kolik můžeme výpočetními kapacitami v CERN zvládnout.“ To pochopily instituce podílející se na LHC už v 90. letech a investovaly do nové, decentralizované počítačové struktury – GRIDu. „World Wide Web umožňuje přístup k informacím, které leží roztroušeny po celém světě,“ vysvětluje Merino, „GRID nyní přichází s novou infrastrukturou, která dovozuje také přístup k výpočetním kapacitám roztroušeným po zeměkouli.“

### **Tajemný GRID: Hračka pro badatele, nebo nový internet?**

LHC GRID představuje hierarchický systém s různými úrovněmi – tzv. „Tiers“. První úroveň je samotný CERN, kde se především ukládají surová data z experimentů a vypočítávají první rekonstrukce. Jedenáct výpočetních středisek úrovně Tier 1 ukládá mj. zálohy experimentů a vypočítává zjevně rekonstrukce. Každé z těchto středisek je spojeno s Ženevou optickou sítí ze světlovodných vláken – o kapacitě 10 Gb/s. Úroveň Tier 2 pak zahrnuje asi 120 výpočetních středisek ve 35 zemích – ta zajišťují distribuovaný výpočetní výkon pro výpočet simulací. Pro vědce je přitom velmi důležitá stavba této sítě: LHC GRID je koncipován tak, aby každý na projektu zúčastněný výzkumník měl ze svého počítače přístup k datům z experimentů – bez ohledu na to, kde jsou uložena nebo vypočítávána.

Zda bude tato infrastruktura někdy také hrát důležitou roli mimo okruh vědeckých institucí, to se Gonzalo Merino neodvažuje předpovídat. Momentálně ho napadají jen příklady z výzkumu, například výpočty obrazů z astronomických pozorování nebo lékařské obrazové databanky. Jedním si je však jist: „GRID bude do tří let vypadat úplně jinak než dnes.“ Koneckonců, Tim Berners-Lee si začátkem 90. let také nepředstavoval World Wide Web tak jako to, co se z jeho vynálezu během doby vyvinulo. Na otázku, co ho dnes na WWW nejvíce překvapuje, praotec webu odpovídá: „Google!“ 

AUTOR@CHIP.CZ