

Όγκος δεδομένων LHC & το GRID



Ο όγκος των δεδομένων

Τα δεδομένα που θα προκύπτουν από όλα τα πειράματα του LHC μέσα σε ένα χρόνο, θα έχουν όγκο 15 petabytes δηλαδή 15 εκατομμύρια GB¹. Ένας τέτοιος όγκος δεδομένων ισοδυναμεί με 1.700.000 DVD διπλής επίστρωσης (dual layer) ή 20.000.000 CD. Αν κανείς μπορούσε να τοποθετήσει όλα αυτά τα CD το ένα πάνω από το άλλο, η στοίβα που θα σχηματιζόταν θα είχε ύψος 20km! Κι όλα αυτά μετά το «φιλτράρισμα» που υφίστανται τα δεδομένα ώστε να κρατηθεί το χρήσιμο κομμάτι της πληροφορίας· αλλιώς αν η πληροφορία από ένα μόνο ανιχνευτή γραφόταν σε δίσκους DVD και έμπαινε το ένα πάνω στο άλλο θα έφθαναν σε έξι μόνο μήνες λειτουργίας μέχρι το φεγγάρι.

Χιλιάδες επιστήμονες χρειάζονται πρόσβαση σε αυτά τα δεδομένα για να τα αναλύσουν. Η επεξεργασία, η μεταφορά και η αποθήκευση των δεδομένων αυτών είναι αδύνατο να γίνει σε τοπικό επίπεδο στο CERN όσο εξοπλισμένο και να είναι. Το CERN λοιπόν σε συνεργασία με 33 χώρες, δημιούργησε μία ειδική δικτυακή δομή για διαμοιρασμένη υπολογιστική ισχύ και αποθήκευση των δεδομένων. Το δίκτυο αυτό είναι το λεγόμενο LHC Computing GRID ή απλώς GRID και είναι ουσιαστικά μια τεράστια συστοιχία υπολογιστών. Το δίκτυο αυτό εκτείνεται σε όλο σχεδόν τον πλανήτη. Από την Ιαπωνία μέχρι τον Καναδά. Χιλιάδες υπολογιστές σε ερευνητικά κέντρα και Πανεπιστήμια είναι συνδεδεμένοι σε αυτό.

Αν και τα δεδομένα θα διαμοιράζονται σε πολλές χώρες, θα γίνεται και μια εφεδρική απόθηκευση όλων των δεδομένων (backup) στο CERN όχι όμως σε σκληρούς δίσκους αλλά σε ειδικές ταινίες. Μετά από μια προεπεξεργασία, τα δεδομένα θα διαμοιράζονται σε 11² μεγάλα υπολογιστικά κέντρα. Αυτά τα 11 κέντρα θα αποθηκεύουν τα δεδομένα και θα παρέχουν πρόσβαση σε αυτά από 120 επιμέρους κέντρα

¹ Αντιστοιχεί σε 10.000 φορές τον όγκο των πληροφοριών που περιέχονται σε όλα τα βιβλία που τυπώνονται σε όλο τον κόσμο μέσα σε ένα χρόνο.

² Σε Καναδά, Γαλλία, Γερμανία, Ιταλία, Ολλανδία, Δανία & Σκανδιναβικά κράτη, Ισπανία, Ταϊπέι, Ηνωμένο Βασίλειο και σε 2 κέντρα στις ΗΠΑ.

που με τη σειρά τους δίνουν πρόσβαση σε τοπικά υποδίκτυα και απλούς υπολογιστές. Συνολικά υπάρχουν πάνω 200 τέτοια κέντρα σε 30 χώρες.

Να επισημάνουμε ότι εκτός από τη συλλογή και αποθήκευση των δεδομένων, θα γίνεται και **επεξεργασία** τους. Η επεξεργασία απαιτεί πολλές εκατοντάδες χιλιάδες επεξεργαστές. Αυτήν την στιγμή υπάρχουν 100.000 υπολογιστές συνδεδεμένοι στο GRID και ο αριθμός αυτός αναμένεται να διπλασιαστεί μέσα στα επόμενα χρόνια.



Ο τεράστιος ρυθμός παραγωγής δεδομένων όμως δεν μπορεί να απορροφηθεί μόνο με την χρήση πολλών υπολογιστών αλλά απαιτεί και τεράστια ταχύτητα μετάδοσης δεδομένων. Για τον λόγο αυτό όλη η τεχνολογία που χρησιμοποιούμε σήμερα και είναι βασισμένη σε συστήματα τηλεφωνίας απορρίφθηκε. Το GRID είναι βασισμένο εξολοκλήρου σε γραμμές οπτικών ινών και επιτυγχάνει μέχρι και 10.000 φορές μεγαλύτερη ταχύτητα από τις συνηθισμένες ευρυζωνικές συνδέσεις (ADSL).

Πηγές: <http://public.web.cern.ch/public/en/LHC/Computing-en.html>

<http://gridcafe.web.cern.ch>

<http://spectrum.ieee.org/print/4117>