

Data Planning/Management

Δρ. Δημήτρης Ντελλής
ntell [at] grnet.gr

GRNET

- ▶ Γενικά το κόστος του αποθηκευτικού χώρου πέραν των πολύ μικρών απαιτήσεων, προσθέτει επιπλέον 50% περίπου στο κόστος.
- ▶ Αυτό το κόστος αφορά τον χρόνο κατά τον οποίο γίνονται οι υπολογισμοί.
- ▶ Τυπική περίπτωση
 - ▶ Σχεδιάζουμε πολύ καλά τι θέλουμε να υπολογίσουμε, τι χρόνους κλπ., γίνονται οι υπολογισμοί, όλα καλά, τελειώνει το project.
 - ▶ Τι θα γίνει με τα data που παράχθηκαν ?
 - ▶ Σε αρκετά μεγάλο ποσοστό των περιπτώσεων, δεν έχει γίνει planning ώστε
 - ▶ Κατά τη διάρκεια των υπολογισμών
 - ▶ Χρειάζονται όλα αυτά τα data που παράγονται ?

- ▶ Θα μπορούσε το software να μην αποθηκεύει τα πάντα αλλά όσα σχεδιάζουμε ή και κάποια που δεν είμαστε σίγουροι ?
- ▶ Πραγματική περίπτωση : Πραγματικά usable data 40 GB, μέγεθος μαζί με τα εντελώς άχρηστα 2 TB.
- ▶ Αν αγοράζαμε το storage για να γίνουν οι υπολογισμοί, θα μπορούσαμε να αγοράσουμε αναλογικά το 1/50 χωρίς έκπτωση στην ποιότητα των υπολογισμών.
- ▶ Μετά το τέλος της χρήσης της υποδομής
 - ▶ Ασχέτως μεγεθών, έχει προβλεφθεί χώρος για την μακροχρόνια αποθήκευση ?
 - ▶ Ειδικά για μεγάλες ποσότητες δεδομένων, συνήθως όχι.
 - ▶ Στον χώρο του industry αλλά πλέον και στον Ακαδημαϊκό/Ερευνητικό χώρο, είναι απαραίτητο να διατηρούνται τα δεδομένα με τα οποία παράχθηκαν είτε προϊόν είτε δημοσιεύσεις για αρκετά μεγάλο χρονικό διάστημα.

Σχεδιασμός/Διαχείριση δεδομένων III

- ▶ Έχουν ληφθεί μέτρα σχετικά με disaster recovery (disk failures, φωτιά στο χώρο, πλημμύρα κλπ.) ?
- ▶ Σπάνια.

- ▶ Καλή σχεδίαση του αποθηκευτικού χώρου που θα χρειαστεί.
- ▶ Αποφυγή καταγραφής δεδομένων που συνήθως καταγράφονται για λόγους debug. Όταν κάποιος χρησιμοποιεί τέτοια συστήματα, λογικά δεν χρειάζεται να κάνει debug αλλά παραγωγή.
- ▶ Χρήση εφόσον γίνεται συμπίεσης δεδομένων
- ▶ Σχεδιασμός αποθηκευτικών χώρων για διατήρηση δεδομένων μετά το τέλος του έργου.
- ▶ Για αποφυγή απώλειας δεδομένων λόγω αποτυχίας υλικού, σε τουλάχιστον 2 διαφορετικούς χώρους.

- ▶ Για αποφυγή απώλειας δεδομένων λόγω καταστροφών, π.χ. φωτιά στο κτήριο, διατήρηση αντιγράφων σε άλλο κτήριο.
- ▶ Για πολύ σημαντικά δεδομένα, για αποφυγή απώλειας δεδομένων λόγω μεγάλων φυσικών καταστροφών π.χ. πλημύρα, τσουνάμι κλπ. διατήρηση των δεδομένων σε άλλη περιοχή/πόλη/χώρα. Αφορά κυρίως αρκετά μεγάλες επιχειρήσεις.
- ▶ Επειδή η διατήρηση αποθηκευτικών χώρων για π.χ. 20 χρόνια στο ίδιο υλικό συνήθως δεν είναι εφικτή, ανα περιόδους αρκετά πριν τον MTBF του υλικού ή/και τον χρόνο που το υλικό παύει να είναι διαθέσιμο στην αγορά, ανανέωση του υλικού με νεώτερο.

- ▶ Αναπαραγωγή των δεδομένων. Εφόσον το κόστος συντήρησης είναι αρκετά μεγάλο ως προς την "αξία" κάποιων δεδομένων, κάποιος μπορεί να κρατήσει τα συνήθως λίγα "συμπερασματικά" δεδομένα για μεγάλο χρόνο, σε πολλά αντίγραφα κλπ. αλλά θα πρέπει να φροντίσει και την πιθανή μελλοντική αναπαραγωγή των πολλών αλλά λιγότερο σημαντικών δεδομένων με το ίδιο environment.
- ▶ Μπορεί κάποιος να κρατήσει τα αποτελέσματα του post-processing και τα πολλά με τα οποία έγινε το post-processing να μπορούν να αναπαραχθούν εφόσον χρειαστεί στο μέλλον.

- ▶ Γενικά, με χρήση containers τύπου current snapshot/image και όχι rolling updates (singularity vs docker κλπ).

Ερωτήσεις ?