

PRACE-GR - Ανάπτυξη Εθνικής Υπερυπολογιστικής Υποδομής και Παροχής Συναφών Υπηρεσιών στην Ελληνική Ερευνητική και Ακαδημαϊκή Κοινότητα

Ευάγγελος Φλώρος
Εθνικό Δίκτυο Έρευνας και Τεχνολογίας

Ανοιχτή ημερίδα ενημέρωσης – ΥΠΟΠΑΙΘ 14/9/2015

... μια μικρή ιστορική αναδρομή

HPC part of the
ESFRI Roadmap;
creation of a vision
involving 15
European countries



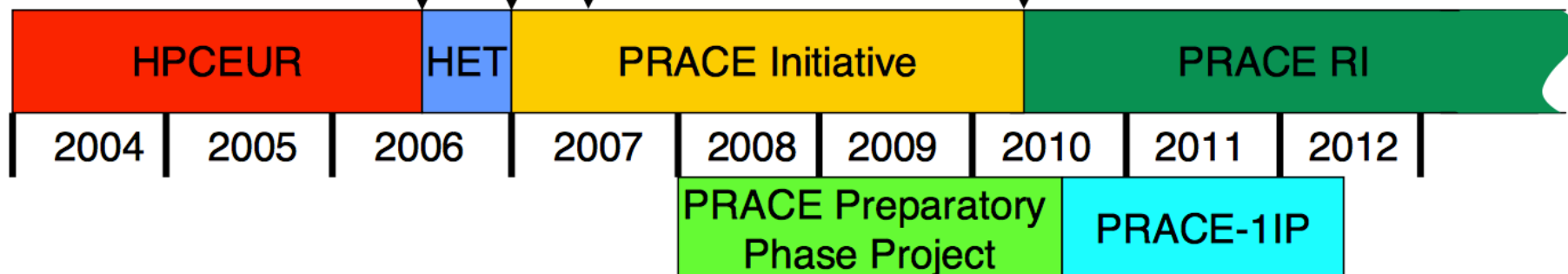
Creation of the
Scientific Case

Ελλάδα ιδρυτικό μέλος
ως non-hosting member

Signature of the MoU

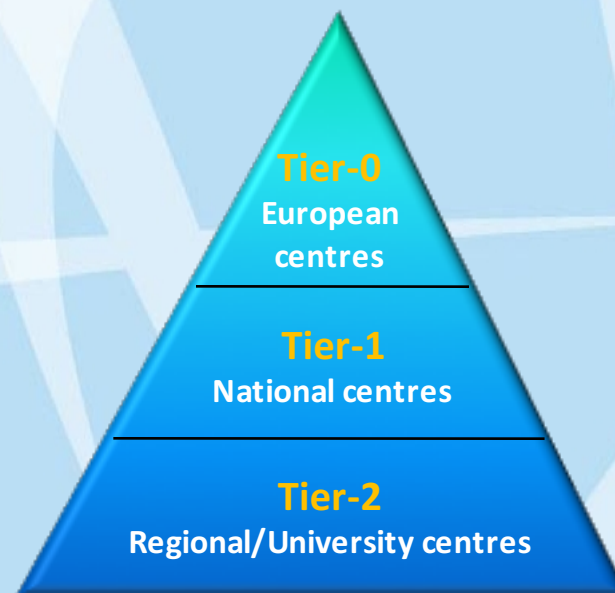
ΓΓΕΤ προκηρύσσει έργα για τη δημιουργία
ερευνητικών δικτύων για το ESFRI

Creation of the PRACE
Research Infrastructure



Partnership for Advanced Computing in Europe

- Διεθνής οργανισμός με έδρα τις Βρυξέλλες.
- Συντονίζει την ανάπτυξη των υποδομών HPC στην Ευρώπη
- Προσφέρει πρόσβαση σε υποδομές επιπέδου Petaflop (Tier-0)



35 εταιρείοι. Ακαδημαϊκά και Ερευνητικά Κέντρα από όλη την Ελλάδα
Διάρκεια 8 μήνες (6/2010-2/2011)

Αποτελέσματα:

- Ανάλυση απαιτήσεων, Αποτέλεσμα πανελλαδικής έρευνας σε όλη την Ελλάδα. 160+ Ερευνητικές Ομάδες, 200+ επιστημονικές εφαρμογές.
- State-of-the-art υπολογιστών υψηλών επιδόσεων
- Πρόταση Εθνικής Στρατηγικής για τις υποδομές HPC
- Μνημόνιο συνεργασίας των μελών του δικτύου

Κύρια συμπεράσματα

- Σημαντική υστέρηση στην Ελλάδα σε υποδομές υπολογιστών υψηλών επιδόσεων.
- Δυσκολία πρόσβασης σε αντίστοιχες υποδομές του εξωτερικού (PRACE Tier-0).
- Αυξημένες ανάγκες Ελλήνων ερευνητών για την πραγματοποίηση υψηλού επιπέδου έρευνας.

Γιατί HPC;

Υπολογιστική Επιστήμη: Ο 3^{ος} στυλοβάτης της επιστήμης (μαζί με Θεωρία και Πείραμα)

Χρήση μαθηματικών μοντέλων και υπολογισμών για την προσομοίωση περίπλοκων φυσικών φαινομένων:

Ασφάλεια, Οικονομία, Ευελιξία, Βελτίωση ακρίβειας, Ελάττωση χρόνου ανάπτυξης.

“Today to out-compete is to out-compute”

ΕΠ «Αττική»

Πρόσκληση 88: Υποβολή Προτάσεων στο πλαίσιο της πράξης «CESSDA – EMSO – ARGO – INFRAFRONTIER – PRACE – Υποστήριξη και ανάπτυξη ελληνικών εταίρων για συμμετοχής στις Κοινοπραξίες των Ευρωπαϊκών Ερευνητικών Υποδομών στα πλαίσια του ESFRI/2006»

ΑΞΟΝΑΣ ΠΡΟΤΕΡΑΙΟΤΗΤΑΣ 3: «Ενίσχυση της ανταγωνιστικότητας της καινοτομίας και της ψηφιακής σύγκλισης»

Ημερομηνία Πρόσκλησης: **5/10/2011**

Υποβολή Τεχνικού Δελτίου Προτεινόμενης Πράξης από ΕΔΕΤ

- Τίτλος «**Ανάπτυξη Εθνικής Υπερυπολογιστικής Υποδομής και Παροχή Συναφών Υπηρεσιών στην Ελληνική Ερευνητική και Ακαδημαϊκή Κοινότητα**»

Ταυτότητα έργου

- Ημ. Ένταξης: **9/7/2012**
- Κωδικός Πράξης (ΟΠΣ): **379417**
- Προϋπολογισμός: **3.500.000,00** ευρώ
- Κωδικός ΣΑ: **E0138**
- Κωδικός ΣΑΕ: **2012ΣΕ01380017**

- Ημερομηνία Ολοκλήρωσης επιλεξιμότητας:
Σεπτέμβριος 2015

- Ανάπτυξη υπερυπολογιστικής υποδομής στην Ελλάδα για την πραγματοποίηση Έρευνας υψηλού επιπέδου.
- Εκμετάλλευση από μεγάλο εύρος επιστημονικών πεδίων.
- Ισχυροποίηση του ρόλου της Ελλάδας στον τομέα των Υπερυπολογιστών σε Πανευρωπαϊκό επίπεδο.

- Έμφαση στις επιδόσεις (TFlops) – Μεγάλος αριθμός κόμβων, Ελάχιστη επίδοση **160 Tflop/s**
- Ιδανικό για εφαρμογές υψηλής κλιμάκωσης (μεγάλη παραλληλία). MPI και OpenMP
- Δικτύωση υψηλών επιδόσεων/χαμηλής καθυστέρησης.
- Παράλληλο σύστημα αρχείων, υψηλών επιδόσεων
- Ικανοποιητικός αποθηκευτικός χώρος (Τουλάχιστον **500TB+**)
- Απαραίτητο λογισμικό ανάπτυξης – Επιστημονικές βιβλιοθήκες
- Επεκτασιμότητα – Σύστημα κλιματισμού και υποδομή H/M εγκαταστάσεων

Βήματα

1. **Σύνταξη RFI**
2. **Διαβούλευση**
Επικοινωνία με τις κορυφαίες εταιρείες του HPC
3. **Ολοκλήρωση Τεύχους Διαγωνισμού**
Έμφαση στα benchmarks και στην απόδοση των προτεινόμενων συστημάτων
4. **Διαγωνισμός - Παραλαβή / Αξιολόγηση Προσφορών**
2 προσφορές: COSMOS(IBM) & OTE/PERFORMANCE(HP)
5. **Κατακύρωση**
Ανάδοχος COSMOS Computers A.E.B.E.
Τελικός Προϋπολογισμός 3.197.522,77 Ευρώ (συμπ. ΦΠΑ)
6. **Υπογραφή σύμβασης**
7. **Υλοποίηση...**

1^η Φάση

- 2009 παραχώρηση χώρου στο κτήριο του Υπουργείου Παιδείας στο Μαρούσι
- Κατασκευή και λειτουργία από το 2010 (**DC1**)
- Φιλοξενία υποδομής cloud ~okeanos
- UPS, Μ/Σ μέση τάσης, Γεννήτρια

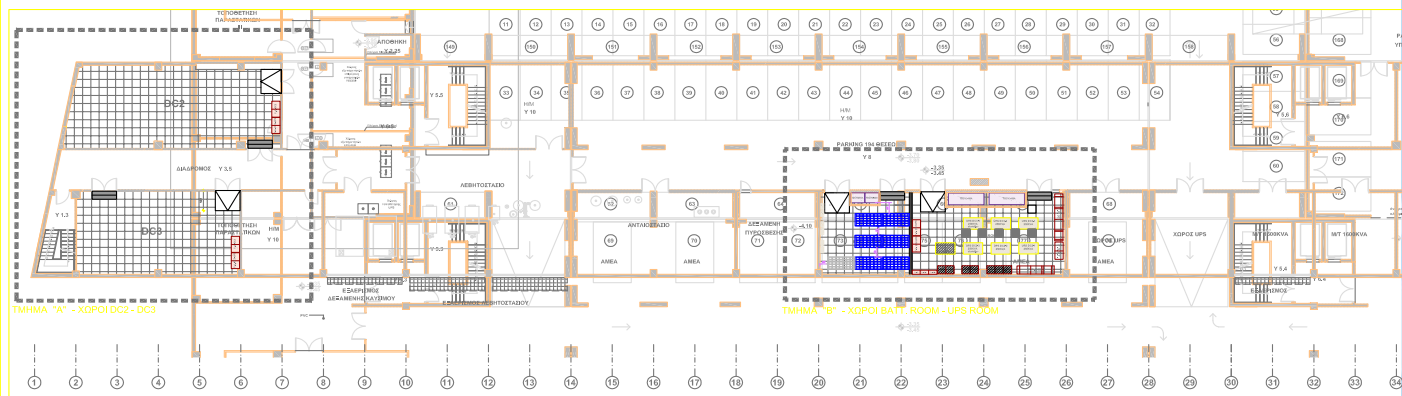
2^η Φάση

- Επέκταση υφιστάμενων υποδομών
- Υλοποίηση στο πλαίσιο του έργου «Παροχή ψηφιακών υπηρεσιών μέσω της δημιουργίας ενεργειακά αποδοτικού κέντρου δεδομένων» ΥΕ2, ΕΠ «Ψηφιακή Σύγκλιση» (Κωδ. ΟΠΣ 311568), Προϋπολογισμός 1 εκ Ευρώ.
- Ανάδοχος Lamda Hellix

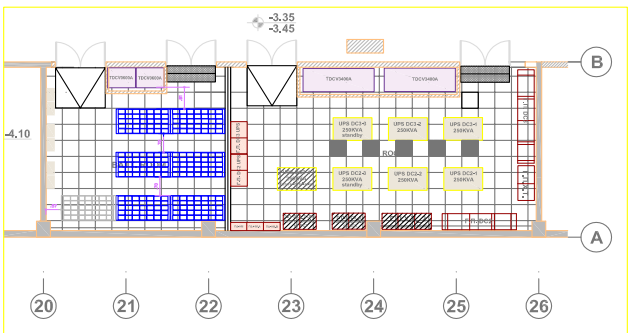
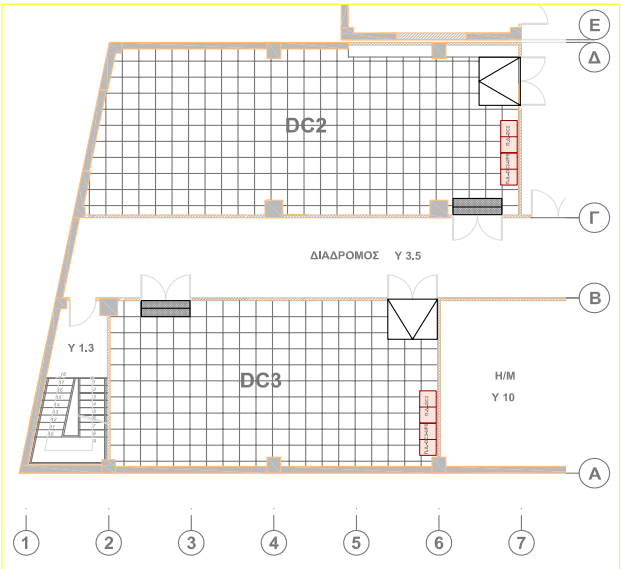
Αντικείμενο:

- Προσθήκη 2 επιπλέον Computer Rooms (**DC2, DC3**). Συνολικό εμβαδό για φιλοξενία IT ~180 m²
- 2 δωμάτια για UPS (6x250 KVA) + Συσσωρευτές
- Πεδία μέσης/χαμηλής τάσης
- Πυρασφάλεια / Έλεγχος πρόσβασης / Παρακολούθηση κλπ

Κάτοψη νέου Κέντρου Δεδομένων



ΚΑΤΟΨΗ ΥΠΟΓΕΙΟΥ
ΚΩΔΙΚΟΣ 1130



Κατασκευή Χώρου Φιλοξενίας

Πριν

Μετά: [//www.grnet.gr](http://www.grnet.gr)

DC2



DC3



UPS/Battery room



Υποδομή φιλοξενίας - UPS



- 6 x APC MGE Galaxy 7000
- 1,5 MW μέγιστη υποστηριζόμενη κατανάλωση
- 1500 KVA UPS

* Οι φωτογραφίες είναι προσφορά του PCMagazine

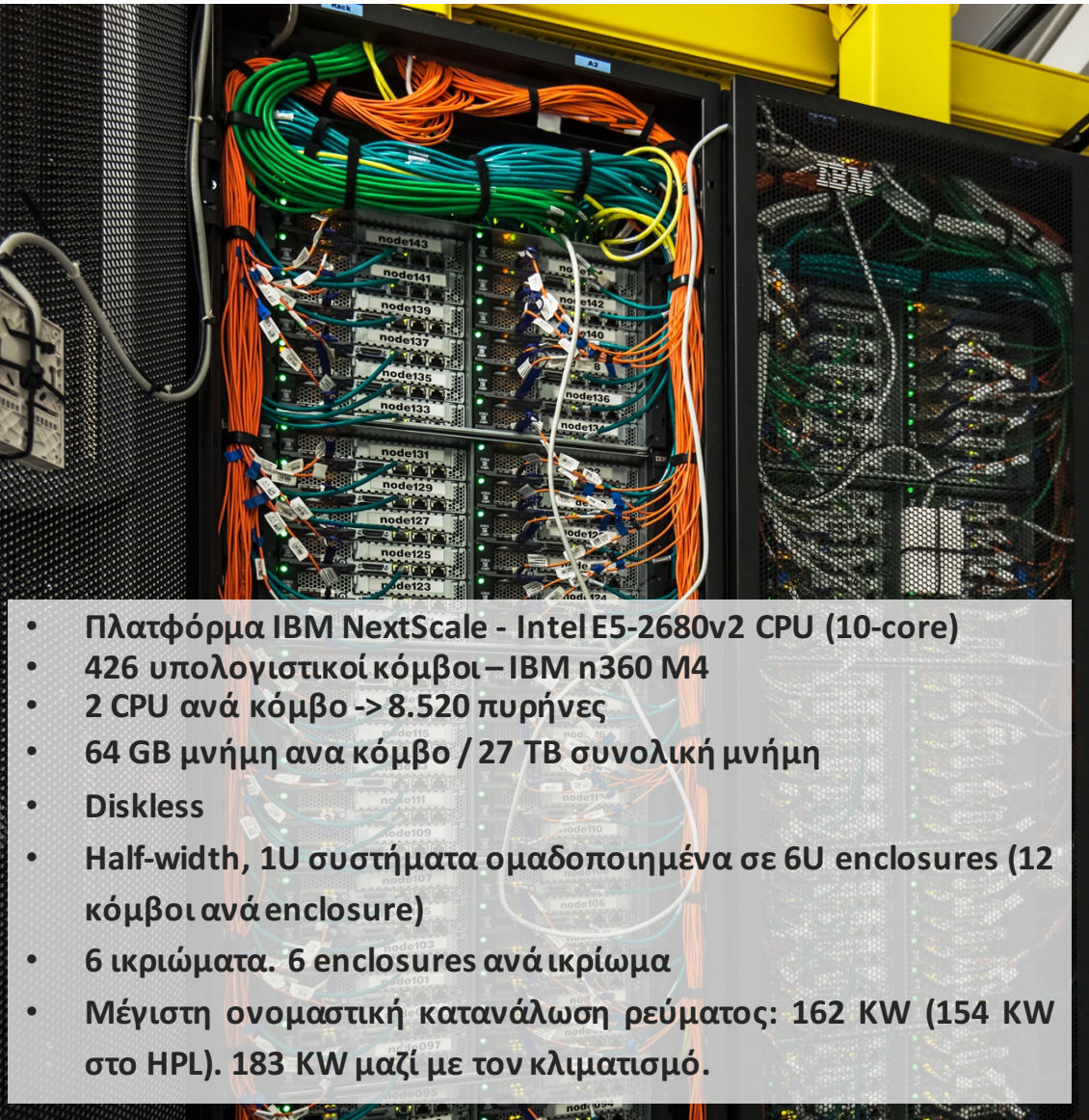


- 540 Συσσωρευτές 12V 69,5AH
- 15 τόνοι σε 6 ικρίωματα βαρέος τύπου

... και εγένετο ARIS



Σύστημα υπολογιστικών κόμβων



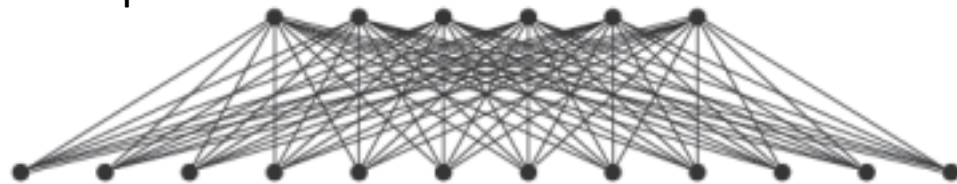
- Πλατφόρμα IBM NextScale - Intel E5-2680v2 CPU (10-core)
- 426 υπολογιστικοί κόμβοι – IBM n360 M4
- 2 CPU ανά κόμβο -> 8.520 πυρήνες
- 64 GB μνήμη ανα κόμβο / 27 TB συνολική μνήμη
- Diskless
- Half-width, 1U συστήματα ομαδοποιημένα σε 6U enclosures (12 κόμβοι ανά enclosure)
- 6 ικρίωματα. 6 enclosures ανά ικρίωμα
- Μέγιστη ονομαστική κατανάλωση ρεύματος: 162 KW (154 KW στο HPL). 183 KW μαζί με τον κλιματισμό.

Infiniband FDR-14 56 Gbps

- Mellanox Director Switch SX6536
- Υλοποίηση τοπολογία Fat Tree Full Non-blocking (1:1)
- ~ 450 καλώδια QSFP+ Optical
- ~ 5 Km οπτικής καλωδίωσης



24 x Spine switches



18 x Leaf switches



IBM GSS 26

- 2 x MD Servers, IBM System x3650 M4
- 6 JBODs x 58 HDDs x 3 TBs = 1.044 TB (~1PB) συνολικός αποθηκευτικός χώρος (raw)
- 2 x NFS Gateway Servers

IBM Elastic Storage (aka GPFS)

- Υψηλές επιδόσεις (> 10GB/s bandwidth)
- Μεγάλη αξιοπιστία. Υλοποίηση τεχνολογίας Declustered RAID. Χαμηλοί χρόνοι ανασύστασης δίσκων μετά από σφάλματα

Κόμβοι υποστήριξης



Κόμβοι υποστήριξης/υπηρεσιών

- IBM System x3650 M4, 2U, 2 x E5-2640v2 8C 2.0 GHz, 128 GB DDR3 Main Memory, 4 x 600 GB SAS HD.
- 2 x Κόμβοι διαχείρισης / Management Nodes
- 2 x Κόμβοι σύνδεσης / Login Nodes
- 10 x Κόμβοι υπηρεσιών - PRACE

Σύστημα κλιματισμού

- Χρήση τεχνολογίας κλειστού θερμού διαδρόμου (Hot Aisle Containment - HACS) APC by Schneider Electric
- 4 InRow μονάδες AC (πλεονασμός N+1) της APC. 2xACRP 502 + 2xACPC ARPP 502
- 4 Ψύκτες συνολικής ισχύος (πλεονασμός 2N). ERAF1222 της UNIFLAR. 4x117 KW ψυκτική ισχύς
- Υποστήριξη Intelligent Free Cooling

Εργαλεία ανάπτυξης – Compilers

- Intel Cluster Studio XE
- PGI Cluster Development Kit (CDK)
- GNU

Επιστημονικές εφαρμογές

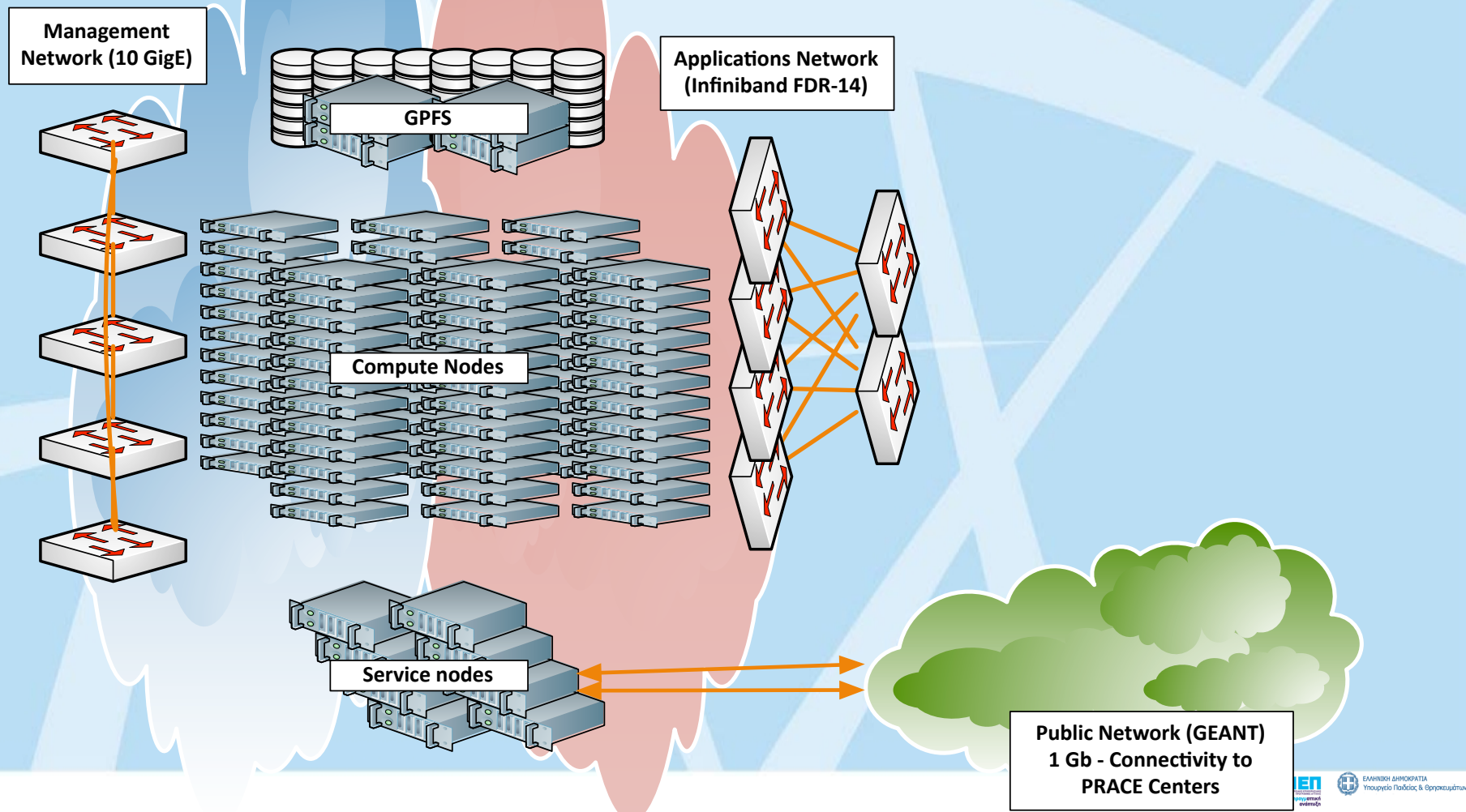
```
----- /apps/modulefiles/applications -----
abinit/7.10.4(default)      gopenmol/3.00(default)      molden/5.2(default)         nwchem/6.5(default)         towhee/7.1.0(default)
bigdft/1.7.6(default)      gromacs/4.5.7               molekel/5.4.0(default)     octave/4.0.0(default)      vmd/1.9.2(default)
cdo/1.7.0(default)         gromacs/4.6.7               mpqc/2.3.1(default)        octopus/4.1.2(default)     wrf/3.4.1/hybrid
code_saturne/4.0.1/intel   gromacs/5.0.5               namd/2.10/hybrid/memopt    openbabel/2.3.2(default)   wrf/3.4.1/purempi
cosmo-art/131108_5.00(default) gromacs/5.0.6(default)    namd/2.10/hybrid/normal   openmd/2.2(default)        wrf/3.7/hybrid
cp2k/2.6.1(default)       gromacs/5.1                 namd/2.10/purempi/memopt   paraview/4.3(default)      wrf/3.7/purempi
dlpoly/2.20                gromacs/5.1-rc1             namd/2.10/purempi/normal   qhull/2012.1               wrf-chem/3.7(default)
dlpoly/4.07(default)      lammps/15May15(default)     ncarg/6.3.0(default)      quantum-espresso/5.2.0(default) wrf-chem/3.7-hybrid
gamess-US/2014R1(default) mdynamix/5.2.7(default)    ncview/2.1.5(default)     R/3.2.1(default)
```

Επιστημονικές βιβλιοθήκες

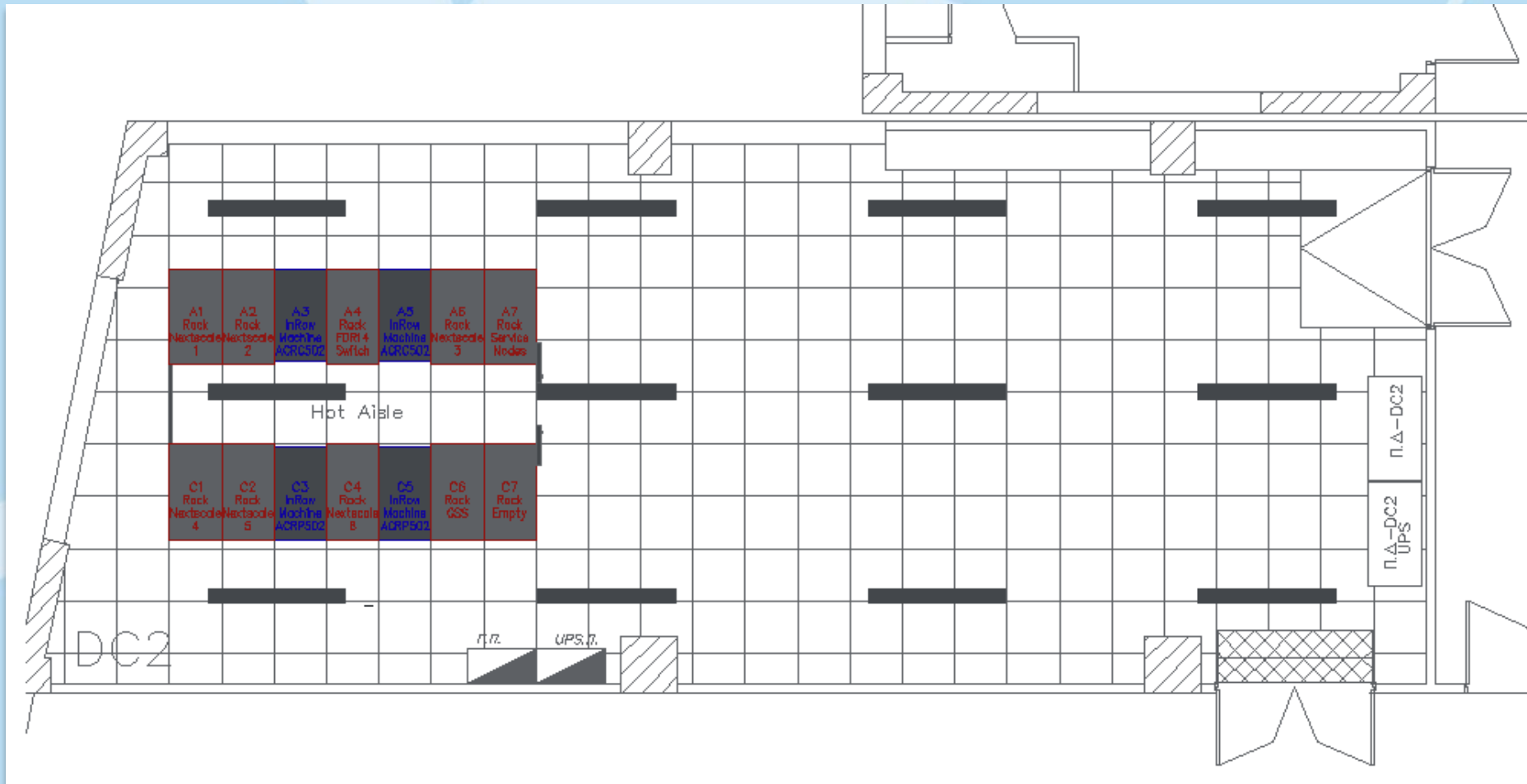
```
----- /apps/modulefiles/libraries -----
atlas/3.11.34(default)     glpk/4.55                   libsmm/gnu                  netcdf-c/4.3.3.1/intel      parmetis/4.0.3/intel
boost/1.58.0(default)     gsl/1.16/gnu(default)       libsmm/intel               netcdf-combined/4.3.3.1/intel pnetcdf/1.6.1/gnu
cgnslib/3.2.1/intel        hdf5/1.8.12/gnu             libxc/2.2.2                netcdf-fortran/4.4.2/gnu    pnetcdf/1.6.1/intel
elpa/2015.05.001/intel(default) hdf5/1.8.12/intel          med/3.0.8/intel            netcdf-fortran/4.4.2/intel  scalapack/2.0.2/gnu(default)
fftw/2.1.5                hdf5/1.8.15/gnu             metis/5.1.0                openblas/0.2.14/gnu/int4    scalapack/2.0.2/intel
fftw/3.3.4/avx(default)   hdf5/1.8.15/intel          netcdf/3.6.3/intel         openblas/0.2.14/gnu/int8    szip/2.1(default)
fftw/3.3.4/sse2           jasper/1.900.1(default)    netcdf/4.1.3/gnu           openblas/0.2.14/intel/int4  udunits2/2.2.19(default)
flame/5.0/gnu              libint/1.1.5               netcdf/4.1.3/intel        openblas/0.2.14/intel/int8  voro++/0.4.6(default)
flame/5.0/intel           libjpeg-turbo/1.4.1(default) netcdf-c/4.3.3.1/gnu      parmetis/4.0.3/gnu
```


- In-band + Out-of-band Management Δίκτυο (10G Ethernet)
- Public δίκτυο
- Λογισμικό Διαχείρισης / Παρακολούθησης
 - xCat
 - Nagios
 - Ganglia
- Λογισμικό διαχείρισης εργασιών – SLURM
- BMS (Business Management System)
- OS Red Hat Linux 6.6

Αρχιτεκτονική συστήματος



Κάτοψη χώρου



Peak Performance (R_{Peak}) vs Sustained Performance (R_{Max})

Υπολογισμός μέγιστης θεωρητικής απόδοσης:

$$\text{\#Nodes} \times \text{\#CPUperNode} \times \text{\#CoresPerCPU} \times \text{CPUFreq} \times \text{InstrPerCycle} = \text{Gflop/s}$$

Για το ARIS:

$$426 \times 2 \times 10 \times 2,8 \text{ GHz} \times 8 = 190.848 \text{ Gflops (190,85 Tflops)}$$

High-Performance Linpack (HPL)

- Επίλυση πυκνού συστήματος γραμμικών εξισώσεων. Καταγράφει την σταθερή απόδοση (sustained performance) που επιτυγχάνει ένα υπολογιστικό σύστημα κατά την επίλυσή του.
- Χρήση MPI/OpenMP για την παράλληλη εκτέλεση σε μεγάλης κλίμακας συστήματα.
- Το πιο δημοφιλές benchmark για την αξιολόγηση υπερυπολογιστών παγκοσμίως.

Top500.org

- Λίστα με τα 500 ισχυρότερα συστήματα του κόσμου
- Διατηρείται από το 1993. Ανακοινώνεται 2 φορές το χρόνο στα συνέδρια SuperComputing (Νοέμβριος) και International SuperComputing (Ιούνιος)

Επιδόσεις ARIS

- Τιμή που κατατέθηκε: 169,73 Tflops
- Τελική τιμή που έχει επιτευχθεί: 179,83 Tflops

Θέση στη λίστα June 2015: 468

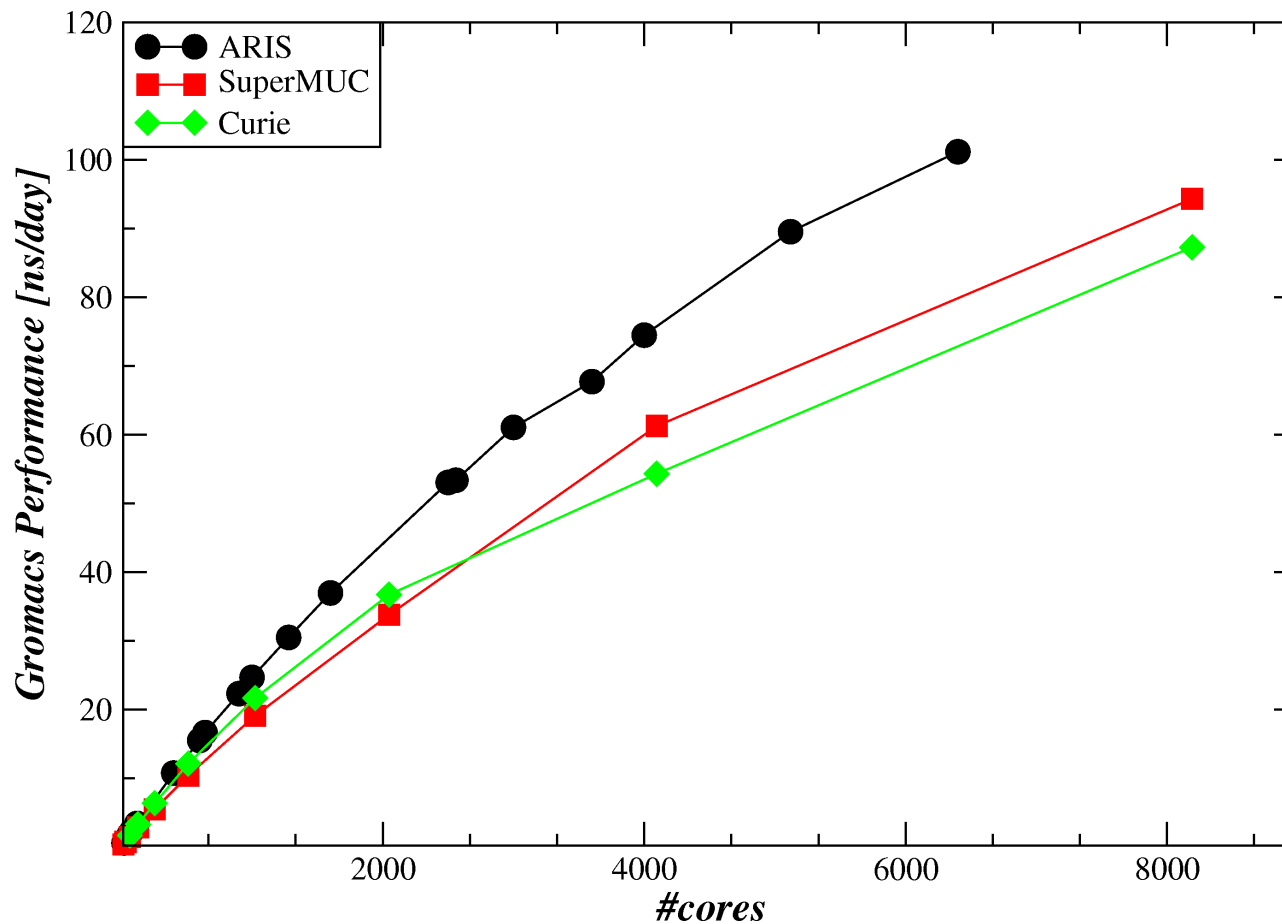
- Πολύ καλή αποδοτικότητα(efficiency).



$$R_{\text{Max}}/R_{\text{Peak}} = 179,83 / 190,85 = 94\% \text{ Αποδοτικότητα}$$

- Εξαιρετικό για την εκτέλεση εφαρμογών με μαθηματικά παρόμοια του HPL
- Αλλά... Η αλήθεια είναι ότι το HPL λέει λίγα πράγματα για την πραγματική απόδοση ενός συστήματος (και θεωρείτε και μάλλον ξεπερασμένο ως benchmark)

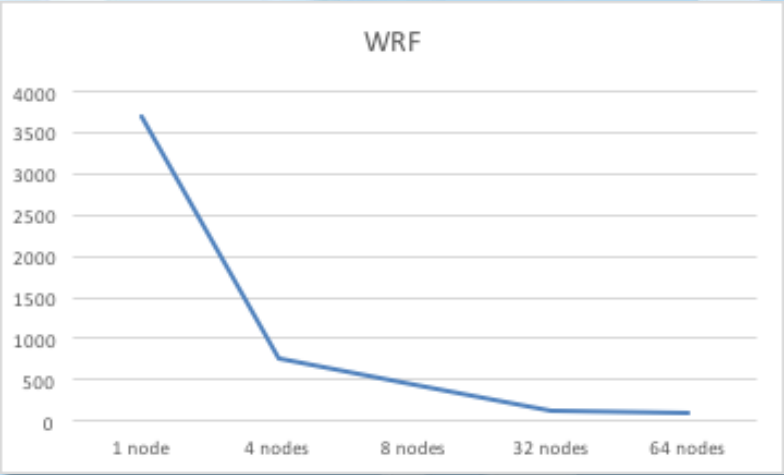
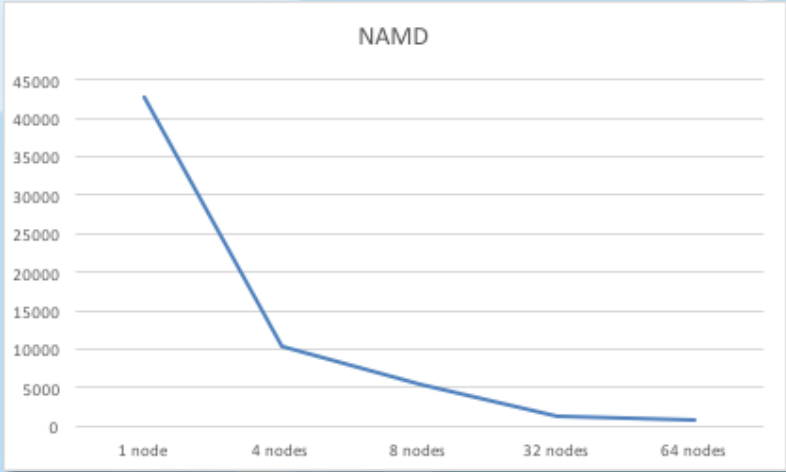
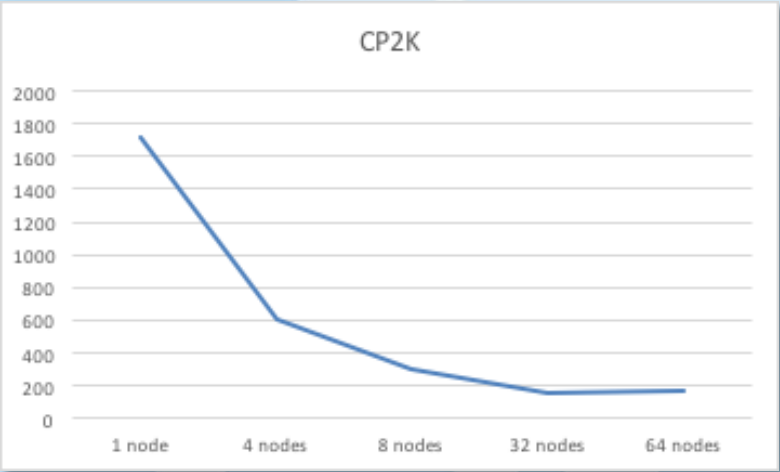
Πόσο γρήγορο είναι το σύστημα πραγματικά;



Lignocellulose, 3.3 M atoms, Reaction Field

Πόσο γρήγορο είναι το σύστημα πραγματικά;

Κλιμάκωση WRF, NAMD και CP2K



Το υπολογιστικό σύστημα αναμένεται παίξει καταλυτικό ρόλο στο ερευνητικό και επιστημονικό γίνεσθαι της Ελλάδας, με σημαντικό κοινωνικό-οικονομικό αντίκτυπο

Έρευνα. Προαγωγή της επιστημονικής αριστείας. Ένα ισχυρό εργαλείο στα χέρια του Έλληνα επιστήμονα – ακαδημαϊκού – ερευνητή.

Διεθνείς συνεργασίες. Συνεισφορά πόρων στην υποδομή του PRACE → Tier-1 σύστημα

Στρατηγική. Κέντρο βαρύτητας HPC στη Νοτιο-Ανατολική Ευρώπη

Κοινωνικό-οικονομικό αντίκτυπο. Μεσοπρόθεσμη εκμετάλλευση των αποτελεσμάτων από την ελληνική οικονομία. Βελτίωση ποιότητας αποτελεσμάτων σε κρίσιμα πεδία εφαρμογών: **Μετεωρολογία, Σεισμολογία, Κλιματολογία, Βιοϊατρική, βιομηχανία υλικών.**

- Σημαντικός παράγοντας εξασφάλισης **βέλτιστης χρήσης** του συστήματος
- **Δίκαιη και ισότιμη πρόσβαση** στο σύνολο της ελληνικής ερευνητική κοινότητας
- Βασισμένη σε **διεθνείς βέλτιστες πρακτικές** και συστάσεις

Βασικοί άξονες:

- Περιοδικές προσκλήσεις για εφαρμογές παραγωγής
- Αξιολόγηση σε τεχνικό και επιστημονικό επίπεδο με χρήση peer-review
- Συνεχείς προσκλήσεις για εφαρμογές προετοιμασίας
- Διαθεσιμότητα πόρων για δράσεις εκπαίδευσης
- Διάθεση πόρων στα πλαίσια διεθνών συνεργασιών

- Ομάδα Διαχείρισης και υποστήριξης HPC
 - Διαχείριση λειτουργία υποδομής
 - Υποστήριξη χρηστών σε λειτουργικά προβλήματα
 - Υποστήριξη εφαρμογών - Μεταφορά και βελτιστοποίηση εκτέλεσης εφαρμογών στο σύστημα.
 - Ανάπτυξη τεκμηρίωσης
 - Εκπαίδευση
 - Συμβουλευτικές υπηρεσίες

Χρήσιμοι Σύνδεσμοι

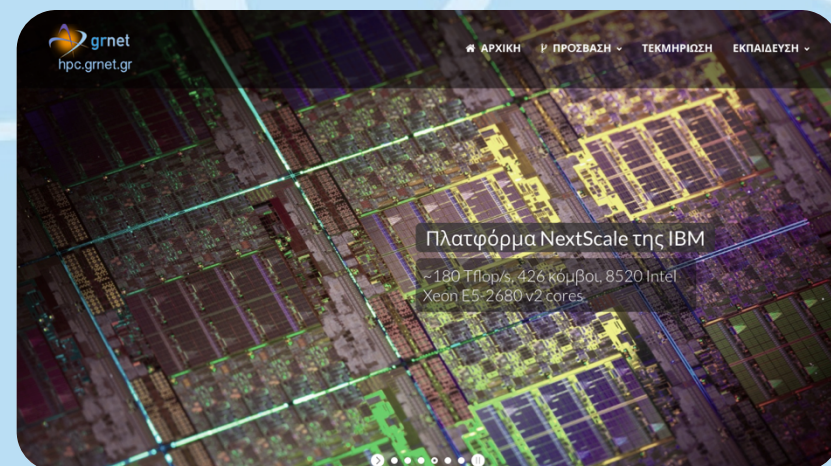
Κεντρική σελίδα υπηρεσίας: <http://hpc.grnet.gr>

Τεκμηρίωση συστήματος: <http://doc.aris.grnet.gr>

Γενικές πληροφορίες: hpc-info@lists.grnet.gr

Πληροφορίες για την πρόσβαση: hpc-access@lists.grnet.gr

Τεχνική υποστήριξη: support@hpc.grnet.gr



- **PRACE-GR: Πρώτη φάση ανάπτυξης HPC υποδομών**
- **Δεύτερη φάση – Νέες δυνατότητες**
 - Ικανοποίηση εφαρμογών διαφορετικού υπολογιστικού προφίλ και **Big Data**
 - Μεγάλη κεντρική μνήμη (Fat Nodes)
 - Νέες αρχιτεκτονικές: GPUs, Accelerators
 - Επέκταση αποθηκευτικών δυνατοτήτων
 - Μακρόχρονη αρχειοθέτηση δεδομένων (Archiving)
- **Βιωσιμότητα υποδομής και παρεχόμενων υπηρεσιών για τα επόμενα χρόνια**

- **Υπουργείο Πολιτισμού, Παιδείας και Θρησκευμάτων**
 - Αρωγός σε κάθε μας βήμα
- **COSMOS/IBM & Lamda Hellix**
 - Εξαιρετική συνεργασία
 - Άμεση απόκριση στις ανάγκες μας
 - Επιμονή και υπομονή σε όλα τα προβλήματα που προέκυψαν
- **Τα μέλη του δικτύου HellasHPC (όλοι εσείς!)**

ΕΡΩΤΗΣΕΙΣ