

Brugerkrav til DeIC

Jan Gorodkin

Center for non-coding RNA in Technology and Health (<http://rth.dk>)

University of Copenhagen

HPC for komplekse problemer i RNA bioinformatik

- Algoritmer $O(n^3)$ – $O(n^6)$
- Store mængder input data
- Store mængder output data
- Parallelisering ved data split up

CAGUCUCAGGUGGUUGGGCU
.((((.((((.....)))))))

UAGCUGAGGUGUCGUGCUA
(((((((.....)))..)))

Sequence alignment

CAGUCUCAGGUGGUUGGGCU-
UAG-CUCAGGUG-UCGUCUA
.((((.((((.....)))))))
(((((.....))))..)))

Structure alignment

CAGUCUCAGGUGGUUGGGCU-
-UAGC-UGAGGUGUCGUCUA
.((((.((((.....)))))))
-(((.....))))..)))

Fagområde i E-science kontekst

- Data i biologi bliver meget billigere end prisen for harddiske.
- Omkostningen er flyttet fra at generere data til at analysere data.
- Ikke-standard værktøjer kræver *on demand* scripts (data parsing) og programmer / algoritmer (bevæge sig hurtigere / mere effektivt igennem data).

E-science er helt centralt for forståelse af data. Ofte bliver kun en lille del af data fra de store internationale konsortie projekter analyseret.



Hvad DeiC også kunne

- Blive større.

Hvad DeiC også kunne

- Blive større.
- *Selvstændig* organisation, der *uafhængigt* kan dele HPC tid ud (også selv om det endnu ikke er gået op for den mere traditionelle del af funding systemet at det er en god ide).
- Et success kriterie: Katalysere at nye forskningsprojekter via E-science støtte kan startes op (hvor det ellers aldrig var sket).
- DeiC skal kunne se behovet for E-Science før alle andre ...
- Sikre at behovet for E-Science er synligt.

Kø-system

De er 3 vigtige ting der **skal** fungere:

Kø-system

De er 3 vigtige ting der **skal** fungere:

Køsystem, Køsystem og Køsystem



Kø-system

De er 3 vigtige ting der **skal** fungere:

Køsystem, Køsystem og Køsystem

Nogle ønsker:

- Hver enkelt bruger skal kunne submitte jobs med forskellig prioritet.
- En god baggrundskø, vil kunne sikre 100% udnyttelse af resourcen.
- Godt system for job-prioritering mellem brugere indenfor samme gruppe og på tværs af grupper.
- Håndtering af large scale submission (så brugere ikke skal skrive ad hoc scripts for at dumpe den næste batch til køen...)



Mere teknisk

- Hvem skal installere hvilket software hvor?
- Hvor hurtigt kan man få installeret software der kræver installation af root?
- Data. Store mængder data (databaser): Kopi på hver enkelt node.
- Storage og backup. Midlertidig plads på den enkelte node er hensigtsmæssigt.
- Solide kabler fra HPC faciliteten til egen server. (I praksis: nogen på universiteterne skal betale for at udskifte switche i bygningen. Uddelivering på universitet til den enkelte forsker ...)

Afsluttende bemærkninger

DeiC/DCSC støtte i vores område har sikret og katalyseret

- Internationale forskningsresultater.
- *Efterfølgende* funding til forskning der tager direkte udgangspunkt i resultater fra HPC.
- Hvem er bedst til at få fat i de skæve ideer med E-science?