



High Performance Computing & Informatie theorie met D-Hydro 1D2D in de Cloud

Demi de Rijke
Digishape 14-03-2023

Waarom HPC & informatie theorie in de cloud?

- Numerieke modellen tijdroverd!
- Modellen zullen benodigd blijven als training data voor data-gedreven oplossingen
- Informatie vs kosten

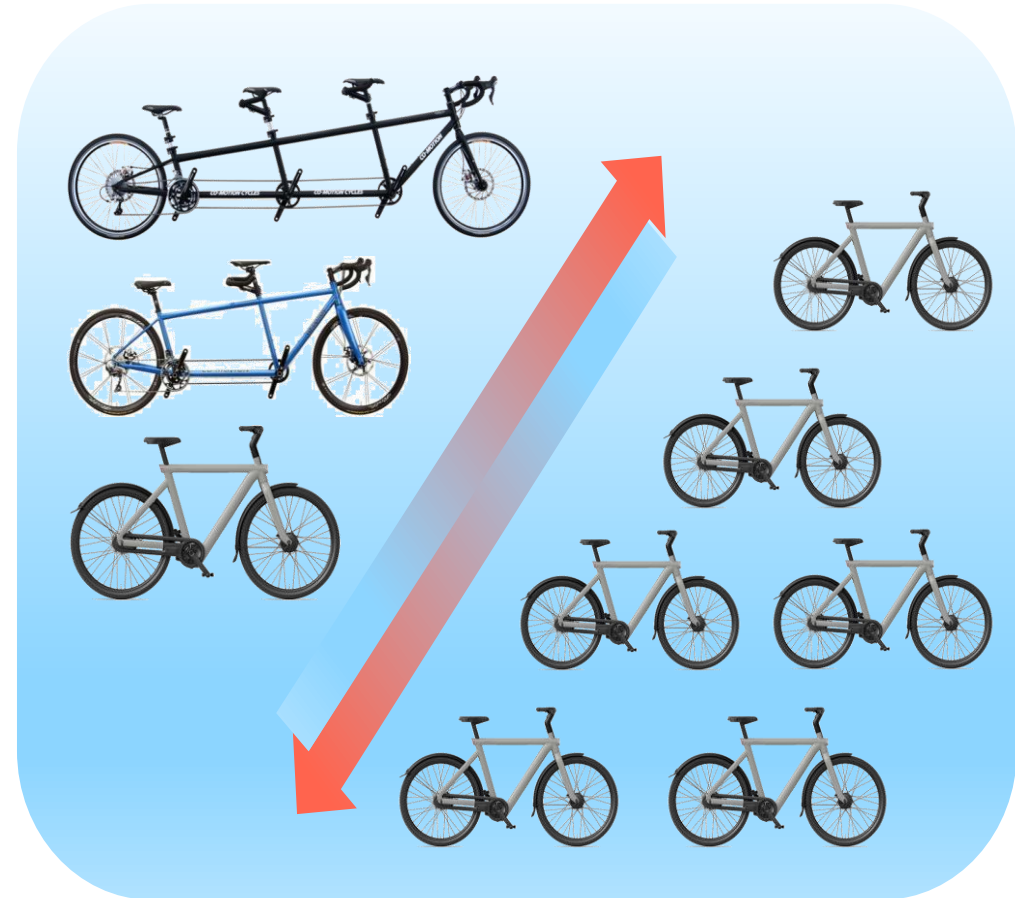
'Unlimited resources', maar je betaald voor alles wat je doet

Verticaal schalen

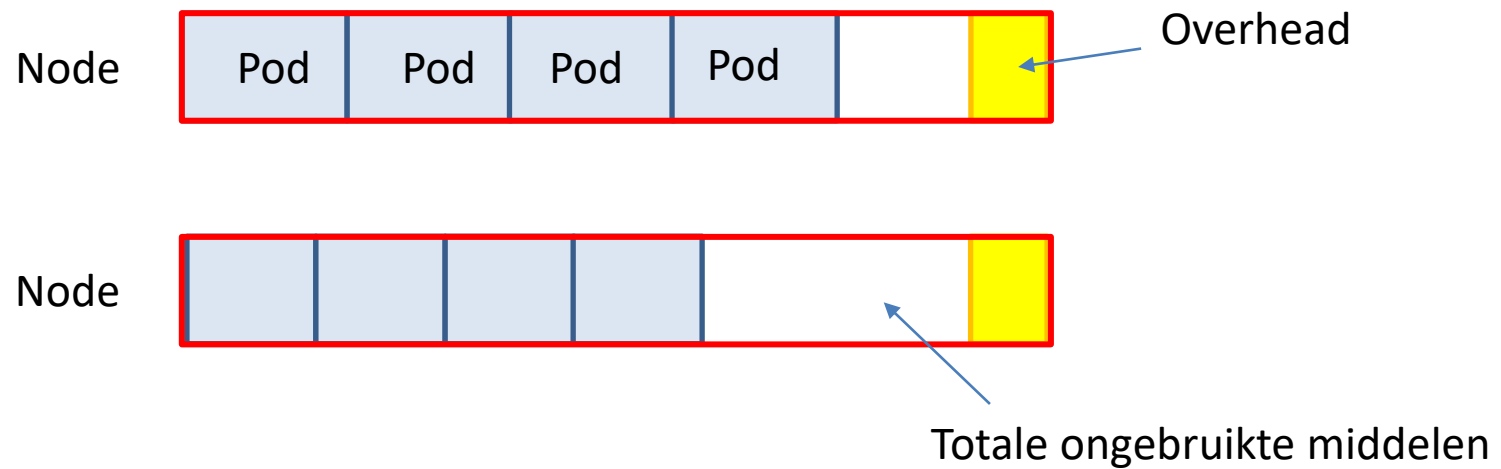
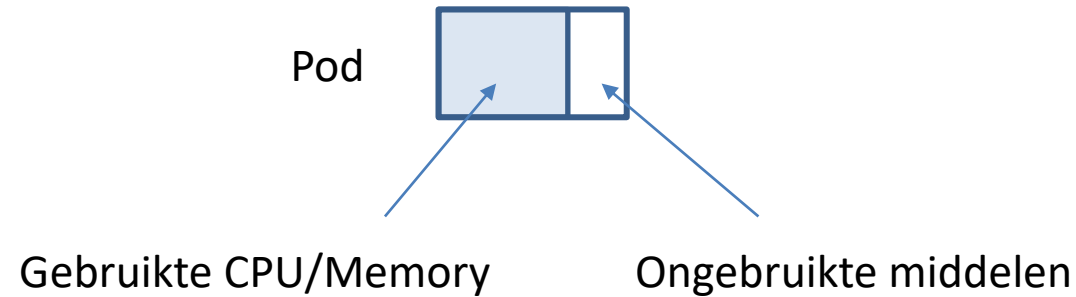
- De capaciteit van een individuele node aanpassen; vCPU, RAM, etc.

Horizontaal schalen

- Het aantal nodes met een bepaalde configuratie vergroten/verkleinen

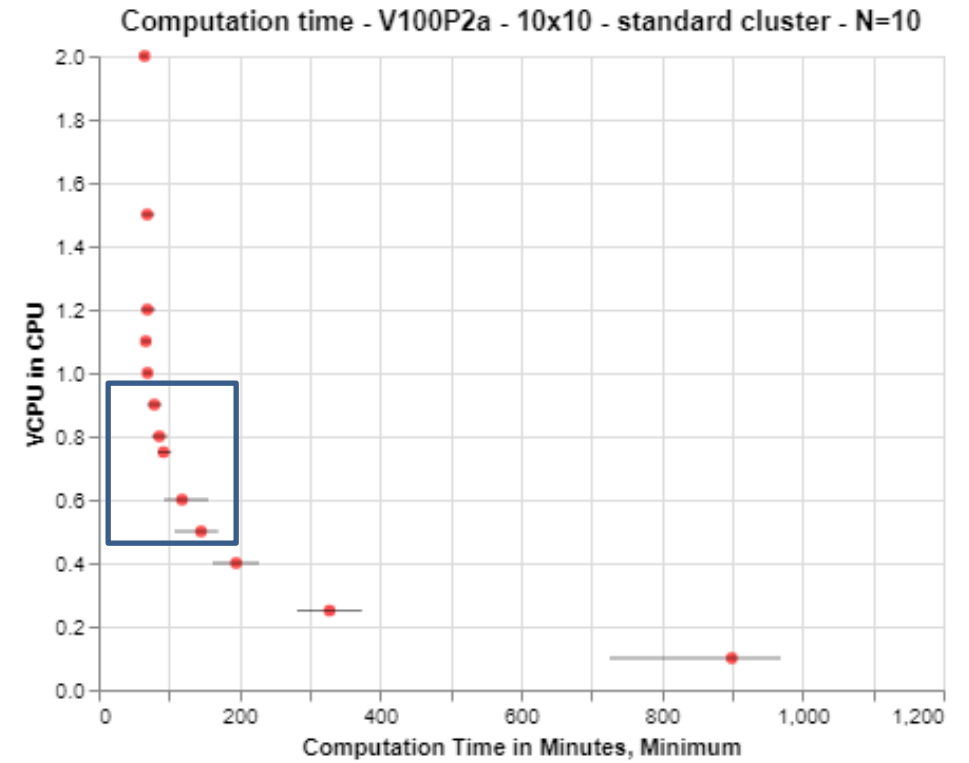


(In)efficiënt gebruik van middelen

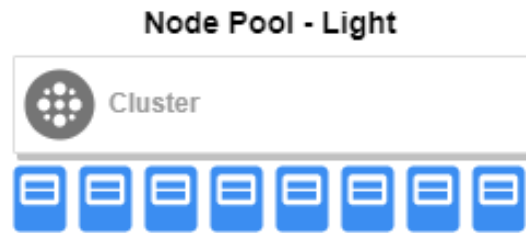


vCPU vs. Rekestijd

Water System Analyse – Delfland



Cluster configuratie



Cluster 16x 4vCPU + 16 Gb M

Total =

64vCPU

256 Gb Memory

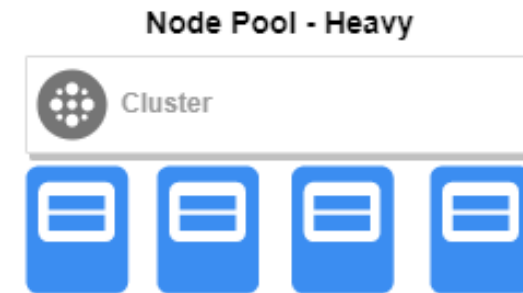


Cluster 8x 8vCPU + 32 Gb M

Total =

64vCPU

256 Gb Memory



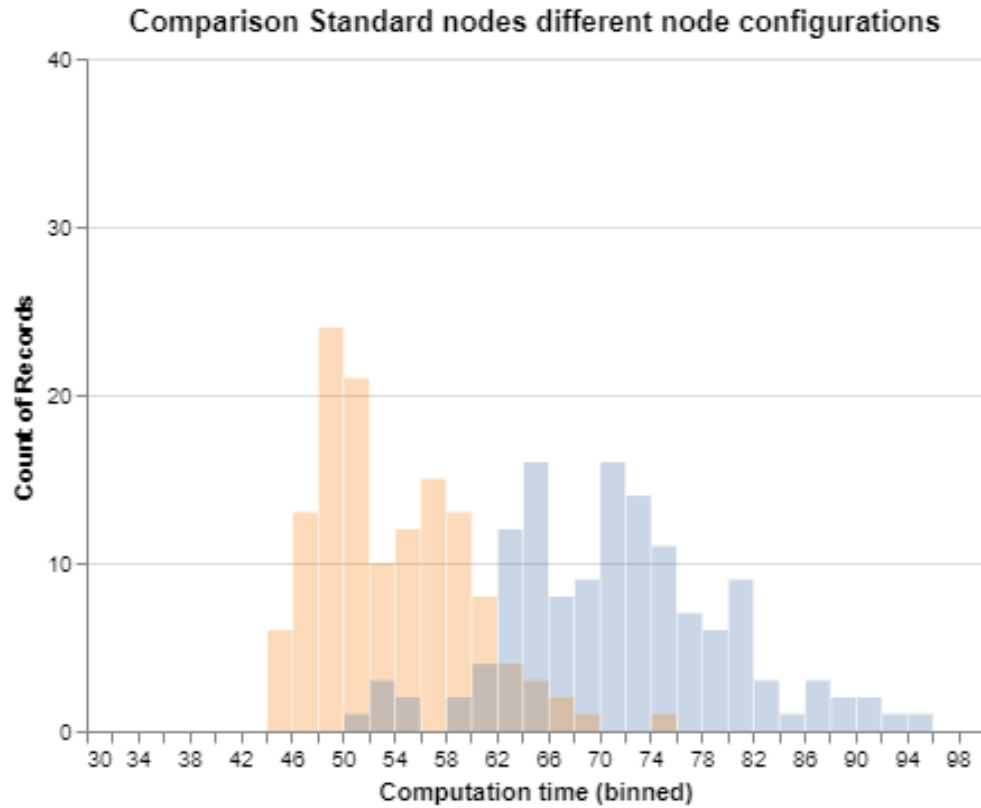
Cluster 2x 32vCPU + 128 Gb M

Total =

64vCPU

256 Gb Memory

Rekentijd individuele pods

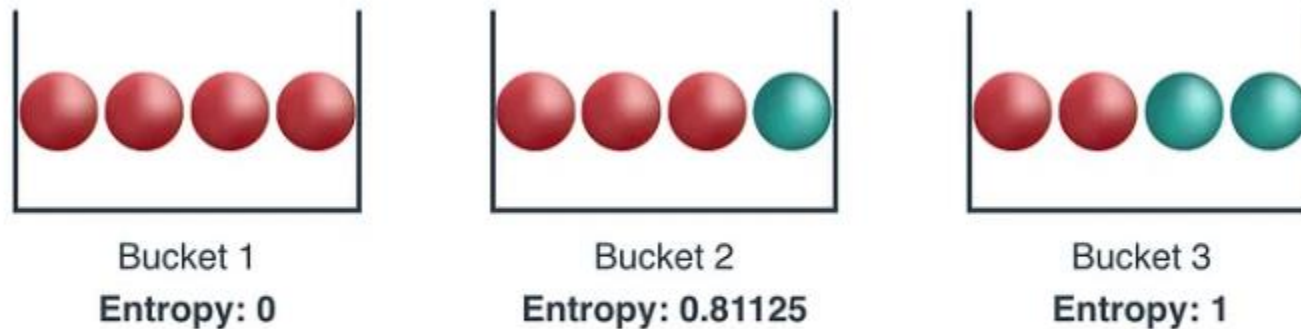


Cluster type	No. of nodes	vCPU	Memory
Standard light	16	4	16
Standard heavy	2	32	128

Shannon's Entropy

$$H(x) = - \sum_{i=1}^n p(x) \cdot \log(p(x))$$

- Quantificeer 'informatie', 'verassing' of 'onzekerheid' met de eenheid van bits.

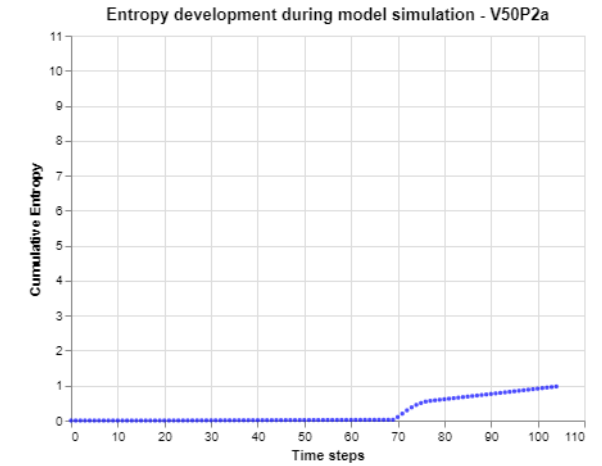
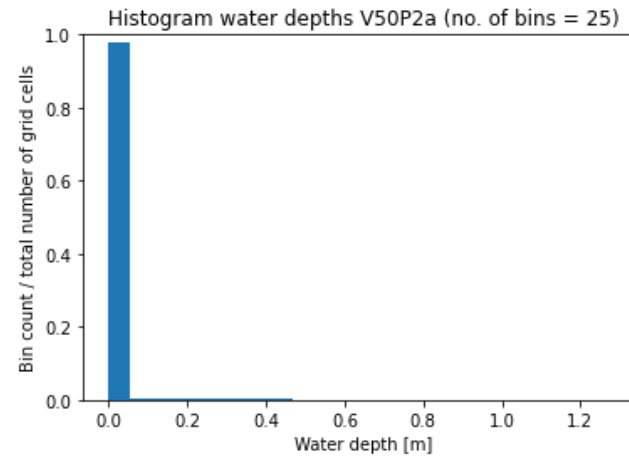


Entropies for the different buckets.

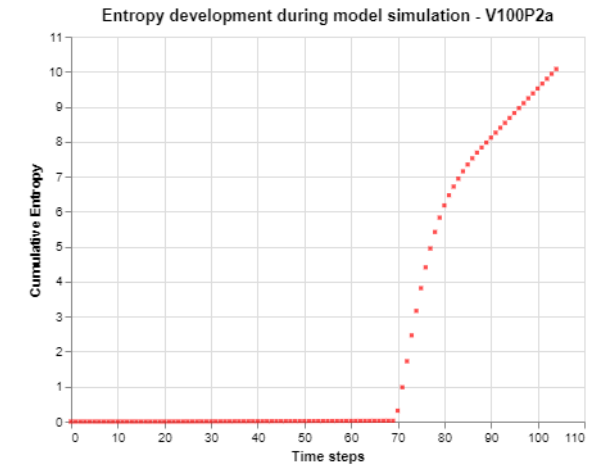
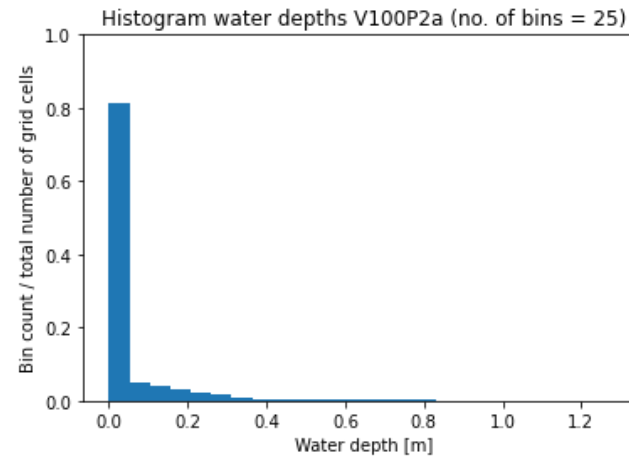
<https://medium.com/udacity/shannon-entropy-information-gain-and-picking-balls-from-buckets-5810d35d54b4>

Vergelijking twee neerslag combinaties

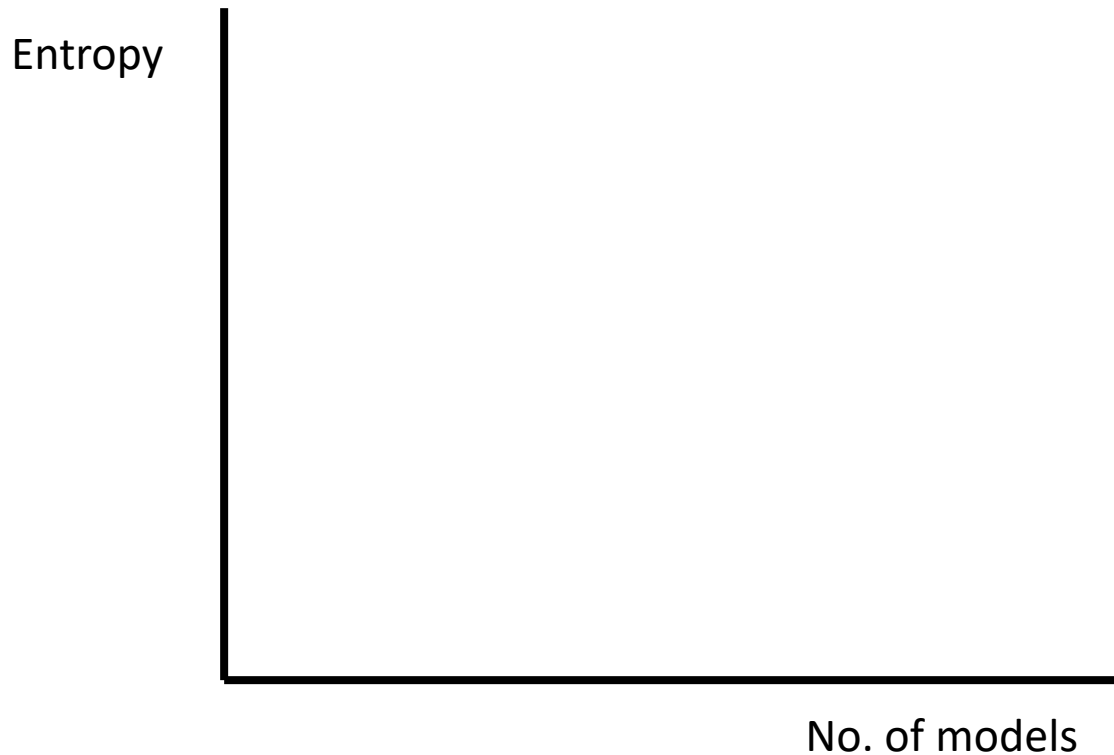
V50P2a



V100P2a



Cost Benefit Analysis



- Bij elk model dat wordt berekend wordt informatie gegenereerd.
- Aantal model runs gelinkt aan kosten van cloud computing
- Kan een som zo worden benaderd?

Bedankt voor de sessie 😊