



UNIVERSITAS
INDONESIA

Veritas, Probitas, Justitia

FAKULTAS

ILMU
KOMPUTER



UAI

Universitas Al Azhar Indonesia

Penggunaan *Cloud Computing* untuk Riset Simulasi Dinamika Molekuler

Seminar Infrastruktur Cloud Ekonomi Digital Indonesia
UAI, 10 April 2018

Denny Hermawan
denny@uai.ac.id



Latar Belakang

- **Simulasi Dinamika Molekular pada Proses Perancangan Obat**
 - Struktur molekul
 - Protein folding



Latar Belakang (2)

$O(N^2)$ time complexity

Timesteps
(simulation time)

□ Tantangan Simulasi DM:

- **N (ukuran atom)**, semakin besar semakin lama proses yang dibutuhkan
- **Simulation Time** dalam Nano Second terlalu lama, perlu proses yang lebih cepat

Simulasi Dinamika Molekuler

Penggunaan lingkungan komputasi untuk menggambarkan posisi, kecepatan, dan orientasi molekul pada suatu waktu

Gromacs

Gromacs (*Groningen
Machine for Chemical
Simulation*)

Groningen University

Amber

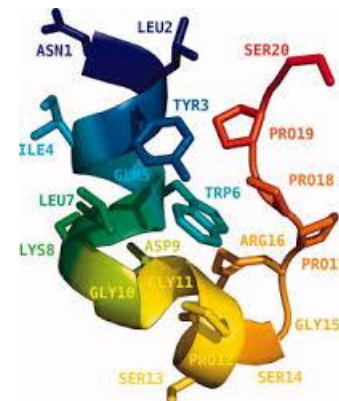
Assisted Model Building with
Energy Refinement

AmberTools

Amber

Protein Input

- **TRPCage** (tryptophan cage)
 - Jumlah atom 304
- **PfENR** ((Plasmodium falciparum Enoyl Acyl Carrier Protein Reductase)
 - Jumlah atom 37,873
- **Myoglobin**
 - Jumlah atom 2,942
- **DHFR / JAC** (Dihydrofolate Reductase)
 - Jumlah atom 23,558



Lingkungan Komputasi



□ Local Cluster

- 16 node 2,2 Ghz (dual core)
- 4 node (Quad Core 2.8 GHz) & GPU (GTX 465, GTX 470 dan Quadro 4000)

□ Barrine Cluster (UQ)

- University of Queensland, Australia
- 384 compute nodes with 3144 computing cores

□ Community Cloud (Future Grid)

- Dikelola oleh Digital Science Center, School of Informatics and Computing, Indiana University. <http://futuresystems.org>

- **XSEDE** (The Extreme Science and Engineering Discovery Environment (XSEDE)), didukung oleh National Science Foundation - <https://www.xsede.org>

□ Amazon EC2 Elastic GPU

Spesifikasi Future Grid



□ Delta Future Grid

16 compute nodes (192 cores):

Each node consist of 2 Intel X5560 6-core 2.8 GHz processor
192 GB DDR3 RAM

Graphic Processing Unit:

Each node equipped with 2 NVIDIA Tesla C2070 GPU (448 cores)

Operating environment:

Red Hat Linux;

Batch System TORQUE/PBS

Spesifikasi XSEDE



□ Cluster Stampede

Host	stampede.tacc.xsede.org
Site	tacc.xsede.org
CPU Type	Intel Xeon E5-2680
Operating System	Linux (CentOS)
Processor Cores	102400
Nodes	6400
Memory	200 TB
Peak Performance	9600 TFlops
Disk	14336 TB
Local Storage Per Node	50 GB
Primary Storage Shared	14 PB
Storage Network	FDR InfiniBand
Interconnect	FDR InfiniBand
Batch System	SLURM
Startup Allocation Limit	50000
Graphics Card	NVIDIA Kepler 2
Parallel File System	Lustre
Memory Per CPU	2 GB
CPU Speed	2.7 GHz
CPU Cores Per Node	16

Beberapa Hasil Simulasi

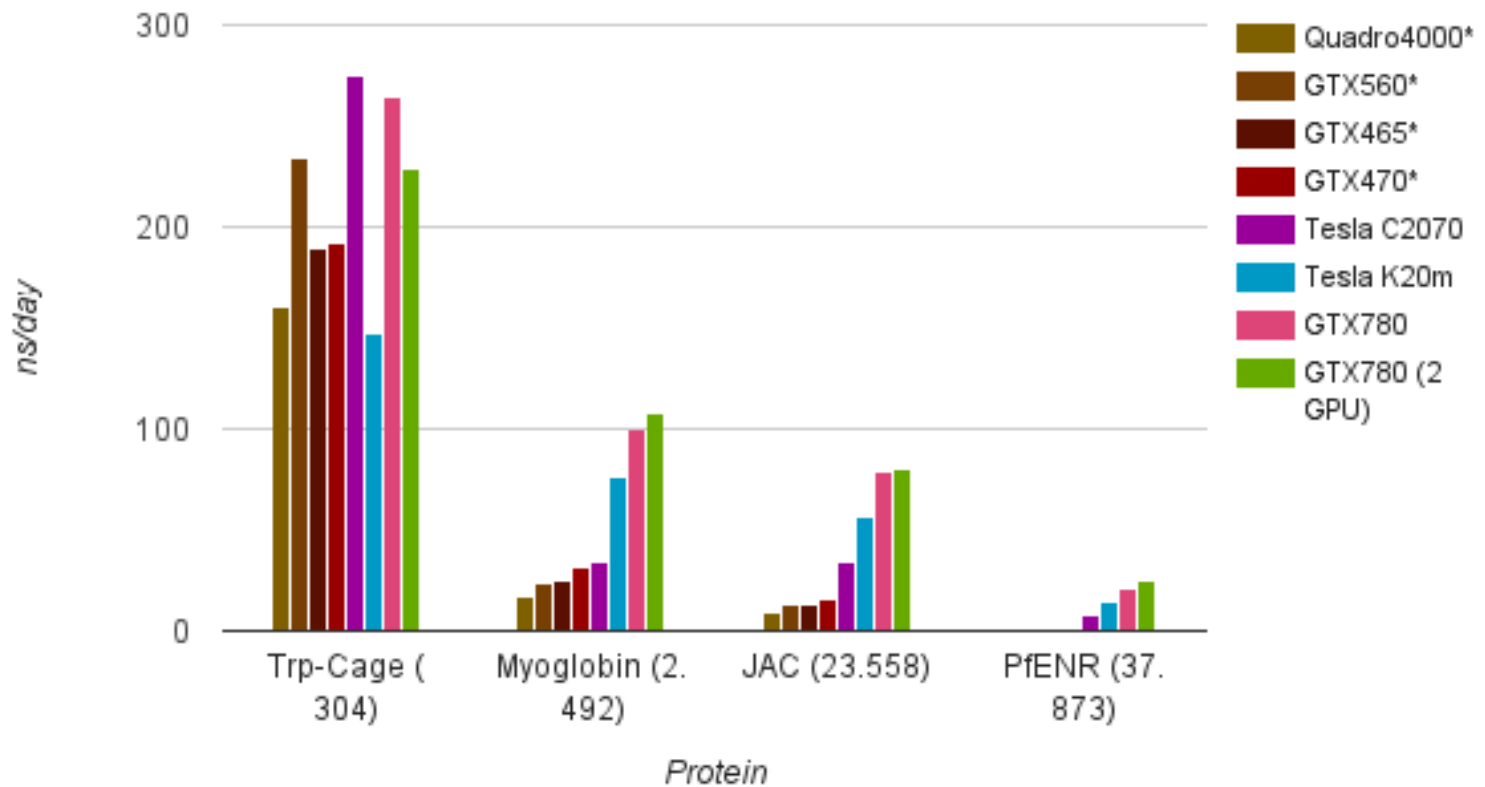


Time-steps	Barrine HPC Cluster										Delta Future Grid	
	4		8		16		32		64		GPU Tesla C2070	
	ns/d ay	time (hr)	ns/d ay	time (hr)	ns/d ay	time (hr)	ns/d ay	time (hr)	ns/d ay	time (hr)	ns/d ay	time (hr)
100ps	1.53	1.56	2.74	0.88	0.40	6.03	0.14	17.50	0.11	22.73	8.27	0.29
200ps	1.54	3.11	2.88	1.67	0.40	11.98	0.14	34.36	0.05	45.40	8.26	0.58
300ps	1.66	4.33	2.43	2.97	0.40	17.98	0.14	46.63	n/a	n/a	8.29	0.86
400ps	1.54	6.23	0.75	10.90	0.40	24.05	0.14	67.14	n/a	n/a	8.27	1.15

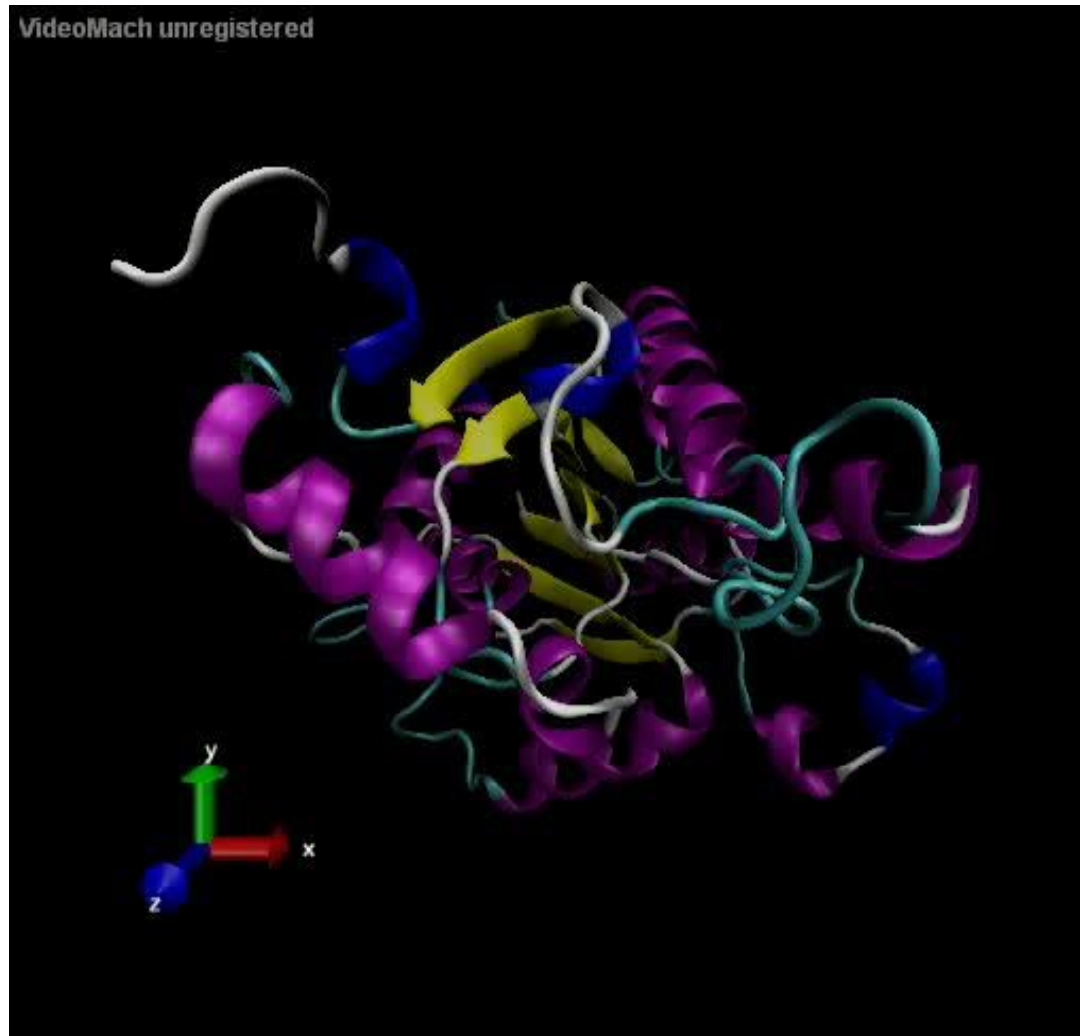
(ps = picoseconds, 1ps = 10^{-12} seconds)

Beberapa Hasil Simulasi

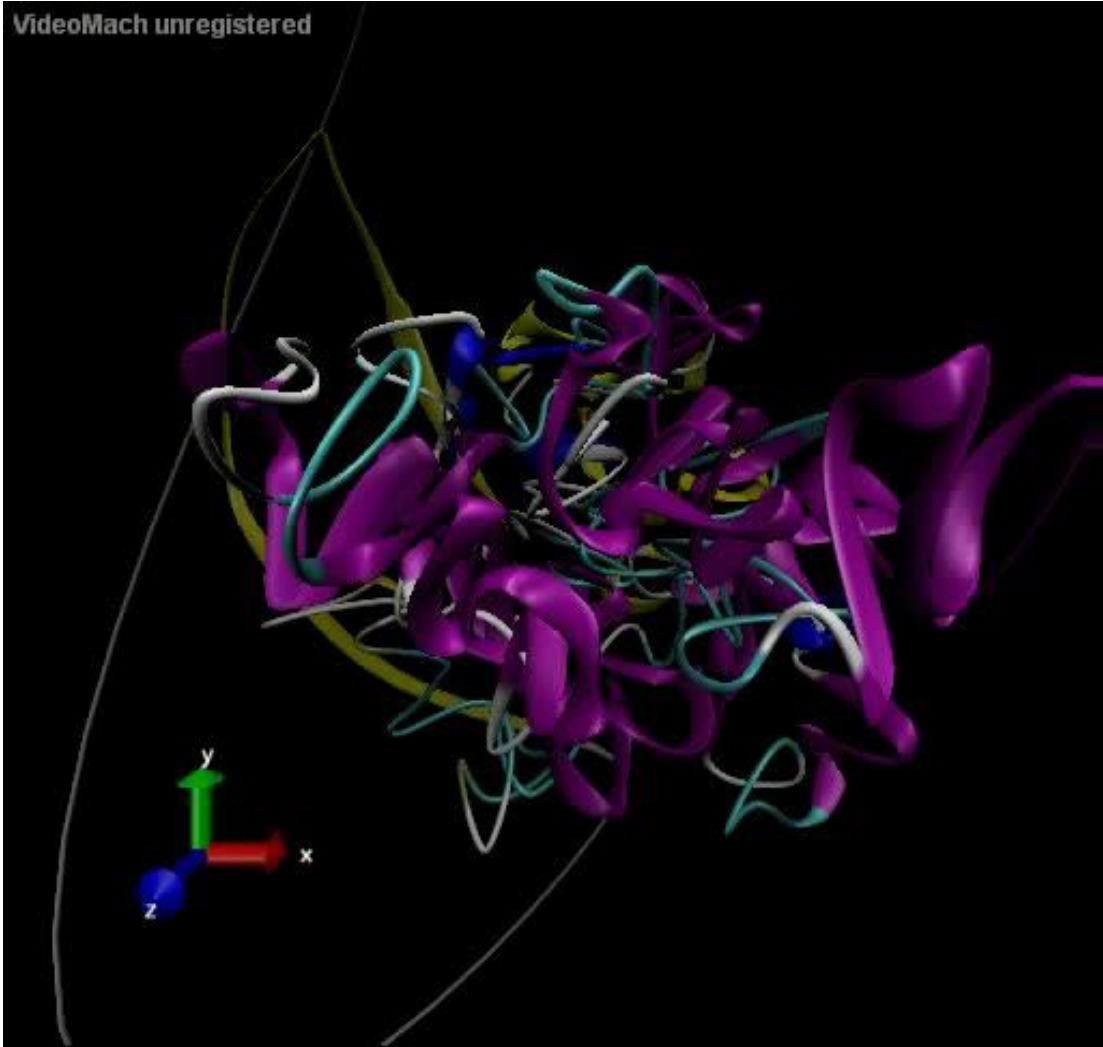
Perbandingan Kinerja ns/day GPU



Trajectory



Rock and roll



Kesimpulan

- Simulasi pada cluster cloud mendapatkan percepatan 11 hingga 12 kali
- Simulasi pada cluster GPU mendapatkan percepatan 14 hingga 39 kali
- Biaya yang diperlukan (benchmark dengan AWS EC2) sekitar USD 0.02 atau Rp260 per simulasi (benchmark dengan AWS EC2)

Parameter	Stampede	Delta	Riset 970	Riset 970 (2 GPU)
Total exectime (h)	1.140	1.137	0.60	0.648
Avg. Time (h)	0.285	0.284	0.151	0.162
Cost (USD)	0.082	0.082	-	-
Avg. Cost (USD)	0.020	0.020	-	-

Kesimpulan (2)



- Menggunakan infrastruktur cloud untuk riset DM mendapatkan keuntungan sbb:
 - *Cost effective*, hanya membayar yang benar-benar digunakan
 - Data aman
 - Menyediakan *unlimited resources* computation, storage, networking dan data management (secara virtual)
 - Mengurangi kebutuhan pembelian, pengelolaan dan perawatan peralatan IT
 - Fleksibel, kita dapat menambah resources ketika beban komputasi meningkat, dan sebaliknya.

Kesimpulan (3)

- Terdapat kebutuhan infrastruktur cloud khususnya SAAS untuk riset:
 - AMBER, GROMACS (Molecular Dynamic)
 - Python/Anaconda (AI, Text Mining, dll)
 - Rstudio (Statistic)
 - OpenCV (Computer Vision)
 - Hadoop (Big Data)
 - dll



Terima kasih