

Propositions

accompanying the dissertation

High-Performance Cluster-Scalable Computational Methods for Genomics Applications

by

Tanveer Ahmad

1. Apache Arrow is currently limited to data analytic workloads because of its immutable data structures. However, future transactional databases and other workloads will also benefit from this in-memory format. (This proposition pertains to this dissertation.)
2. Despite the high performance of distributed data processing frameworks like Apache Spark for genomics workflows (as demonstrated in this dissertation), parallel programming models like MPI/OpenMP are the ultimate solutions for bare-metal and performance scalability on computing clusters. (This proposition pertains to this dissertation.)
3. ASICs based computing solutions are vital for achieving the best performance in emerging data analytic applications including genomics. (This proposition pertains to this dissertation.)
4. Sequencing machines will generate exact and complete genome nucleotides of an organism in the foreseeable future, and all existing sequence aligners, assemblers and variant callers would by then be of no use.
5. In the coming decades, data centers and supercomputers will pose a momentous environmental impact due to significant energy consumption.
6. Natural to GMO and then back to Organic: We need to have a serious conversation because their impact is imminent.
7. PhD candidate should dedicate the first half of their tenure in achieving erudition and refining their knowledge. They can do better research afterwards.
8. Real leaders are those who leave behind trained next generation leaders and not a vacuum.
9. Create an enemy, frighten a nation, then provide security to secure power; a new norm of our times.

These propositions are regarded as opposable and defensible, and have been approved as such by the promotor prof. H.P. Hofstee and promotor dr. ir. Zaid Al-Ars.

Stellingen

behorende bij het proefschrift

High-Performance Cluster-Scalable Computational Methods for Genomics Applications

door

Tanveer Ahmad

1. Apache Arrow is op dit moment beperkt tot gegevensanalyse-workloads vanwege zijn onveranderlijke gegevensstructuren. Toekomstige transactiedatabases en andere workloads zullen echter ook profiteren van dit in-memory formaat.
2. Ondanks de hoge prestaties van gedistribueerde gegevensverwerkingsframeworks zoals Apache Spark voor genomics-workflows (zoals aangetoond in dit proefschrift), zijn parallele programmeermodellen zoals MPI/OpenMP de ultieme oplossingen voor bare-metal en prestatieschaalbaarheid op computerclusters.
3. Op ASIC's gebaseerde computeroplossingen zijn van vitaal belang voor het bereiken van de beste prestaties in opkomende data-analysetoepassingen, waaronder genomica.
4. Sequentiemachines zullen in de nabije toekomst exacte en volledige genoomnucleotiden van een organisme genereren, en daarmee zullen alle bestaande sequentie-aligners, assemblers en variant-callers tegen die tijd nutteloos worden.
5. Datacenters en supercomputers zullen de komende decennia een enorme impact op het milieu hebben vanwege het aanzienlijke energieverbruik.
6. Van natuurlijk eten naar GGO en dan terug naar biologisch: we moeten een serieus gesprek voeren hierover omdat de impact hiervan erg groot kan zijn.
7. Promovendi dienen de eerste helft van hun ambtstermijn te wijden aan het behalen en verfijnen van hun kennis. Ze kunnen daarna beter onderzoek doen.
8. Echte leiders zijn degenen die getrainde leiders van de volgende generatie achterlaten in plaats van een vacuüm.
9. Creëer een vijand, maak een natie bang en zorg dan voor veiligheid om de macht veilig te stellen; een nieuwe norm van onze tijd.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor prof. H.P. Hofstee and promotor dr. ir. Zaid Al-Ars.