



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# Scheduling of Genetic Analysis Workflows On GRID-like Environments (aihe-esittely)

*Arttu Voutilainen*

*3.3.2014*

*Ohjaaja: FT Lauri Eronen (Biocomputing Platforms Ltd.)*

*Valvoja: Prof. Harri Ehtamo*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

# Tausta

DNA:n kartoitus halpenee jatkuvasti

- Raakaa dataa saadaan todella paljon
- Datat analysointi on laskennallisesti raskasta

Analysointi on mahdollista jakaa osiin, joita suoritetaan rinnakkain ja peräkkäin

# Tausta

Jos laskentayksiköitä on vähemmän kuin rinnakkaisia tehtäviä, pitää suoritusjärjestys valita

NP-täydellinen ongelma!

Ratkaisuna erilaisia heuristiikkoja ja meta-heuristiikkoja (min-min, min-max, geneettinen algoritmi,..)

# Tavoitteet

Vertailla eri ajoitusmenetelmien (scheduler) sopivuutta tehtävään

- Ajoitukseen kuluva aika
- Yhden työn keskimääräinen valmistumisaika (turnaround time)

Valita ajoitusmenetelmä, jonka avulla työt voitaisiin suorittaa mahdollisimman tehokkaasti ja nopeasti

# Rajaukset

Ei oteta huomioon QoS-rajoituksia

Ei oteta kantaa työnkuvauksien suunnitteluun tai virheiden hallintaan

Tarkastellaan vain yhden työgraafin ajoittamista

Tiedon siirtämisessä ei huomioida yhtäaikaisten siirtojen vaikutusta kaistaan

# Tietolähteet

- Ian Foster and Carl Kesselman. ***The Grid 2: Blueprint for a new computing infrastructure.*** Morgan Kaufmann, 2003.
- H. Topcuoglu, S. Hariri, and M. Wu. **Performance-effective and low-complexity task scheduling for heterogeneous computing.** *Parallel and Distributed Systems, IEEE Transactions on*, 13(3): 260--274, 2002.
- J. D. Ullman. **Np-complete scheduling problems.** *Journal of Computer and System Sciences*, vol. 10, pages 384--393, 1975.
- M. Wiecek, R. Prodan, and T. Fahringer. **Scheduling of scientific workflows in the askalon grid environment.** *ACM SIGMOD Record*, 34(3): 56--62, 2005.
- J. Yu and R. Buyya. **A taxonomy of scientific workflow systems for grid computing.** *Sigmod Record*, 34(3): 44, 2005.
- J. Yu, R. Buyya, and K. Ramamohanarao. **Workflow scheduling algorithms for grid computing.** In *Metaheuristics for scheduling in distributed computing environments*, pages 173--214. Springer, 2008.
- H. Zhao and R. Sakellariou. **An experimental investigation into the rank function of the heterogeneous earliest finish time scheduling algorithm.** In *Euro-Par 2003 Parallel Processing*, pages 189--194. Springer, 2003.

# Menetelmät / työkalut

- Kirjallisuuskatsaus
- Algoritmien testaus
  - Java-kirjasto, joka implementoi yhden tai useita ajoitusmenetelmiä
  - Simulaattori, joka luo työt ja simuloi ajamisen (BC|Platforms vastaa)
  - Tarkastellaan sekä teoreettista aikataulua että simulaatiossa toteutunutta (simulaatio tuo suoritusaikoihin satunnaisuutta)

# Aikataulu

- Aiheen hyväksyntä toukokuussa 5/2013
- Kirjallisuuteen tutustumista ja koodausta 5/2013
- Aiheen esittely 3.3.2014
- Kandin kirjoitusta
- Valmiin työn esittely 16.6.2014