

# Adaptive Location of Replicas

## An algorithm based on the Ant Colony algorithm

Amadeo Ascó

SyncFree - **Work Package 1**

M12

21<sup>st</sup> October 2014



# Overview

## SyncFree

“Large-scale computation without synchronisation”



# Considerations

## Different considerations

Accessibility

Bandwidth consumption/availability

Access cost

Scalability

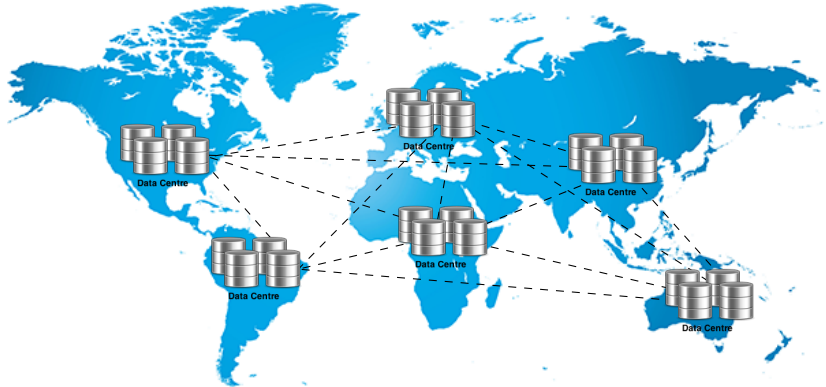
Execution time

Storage consumption

Makespan

# Adaptive Replication

## Adaptive Replication





# Current Considerations

Bandwidth consumption

Execution time

Accessibility



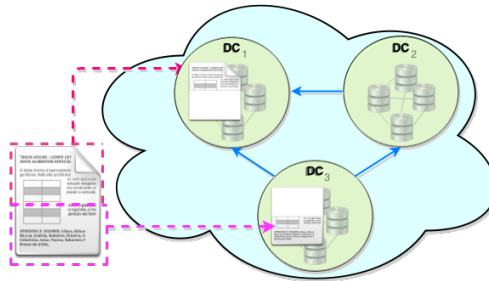
# Types of Replication

Partial Replication

Adaptive Location of Replicas

# Partial Replication

Avoid replicating large data structures

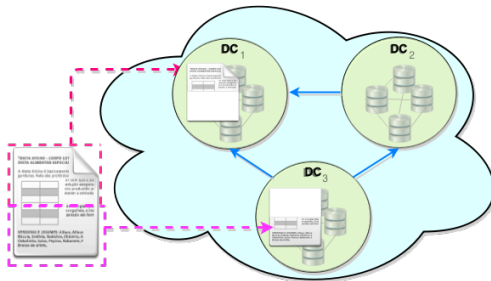


# Partial Replication

Avoid replicating large data structures

Not all the full data is required

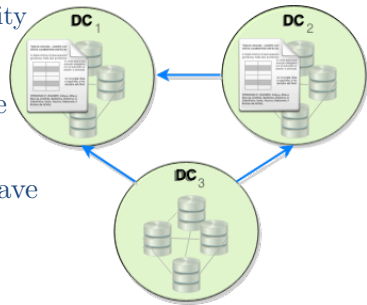
Data structures which allow breaking data into parts





# Adaptive Location of Replicas

- ◇ Location: On which DCs to place the replicas
  - Improve latency: reduce distance between user and replica
  - Improve data transmission quality
- ◇ Selection: Which data to replicate
- ◇ Number: How many replicas to have
  - Reduce unnecessary replicas
    - Reduce storage consumption
    - Reduce network bandwidth



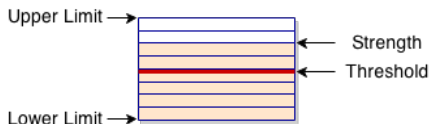


# Algorithm Overview

- ◇ Reads reinforce the strength of a replica
- ◇ Writes reinforce and weaken the strength of a replica
  - Reinforces replica in the DC the write was originally requested on
  - Weaken the strength of the data replicas in all other DCs
- ◇ The strength of the replica decay on time

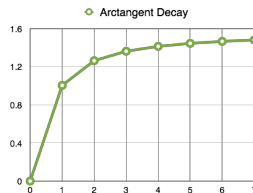
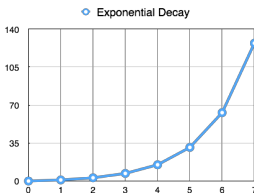
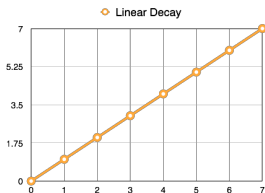
# Algorithm Overview

- ◇ DC without replica
  - Strength must exceed a threshold to generate a new replica
- ◇ DC with replicate
  - The strength has some limits
    - Upper
    - Lower, i.e. zero
  - Remove replica if strength become zero\*



# Algorithm Overview

- ◇ The strength of a replica decay in time
  - Multiple possible approaches



- ◇ There must be at least a minimum number of replicas



# Algorithm Overview

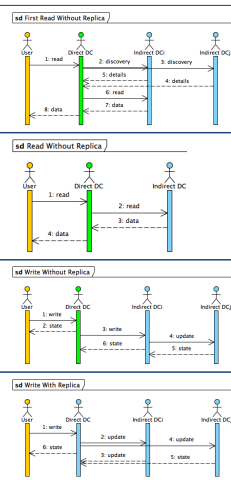
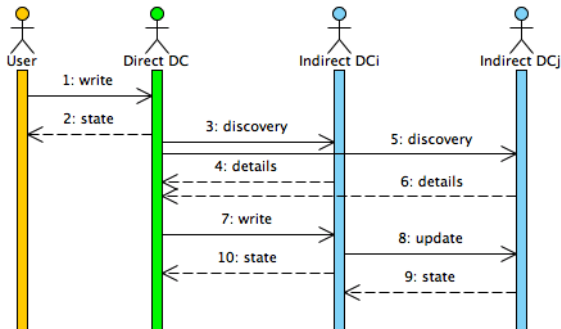
Other possible constraints:

No replicas of personal data outside of Europe

At least one replica of the data at a pre-set DC

# Sequence Diagrams

**sd First Write Without Replica**





# Characteristics

- ◇ Not really an ant colony algorithm
- ◇ Reads only need to be known on the DC the read is requested
- ◇ Writes need to be known by all DCs with replicas
  - Use already available data, e.g. DCs with replicas
  - Use operations that would be already sent (updates)



## Any Questions?

Thank You!

- Ouri Wolfson. A distributed algorithm for adaptive replication of data. Technical report, Department of Computer Science, Columbia University, 1990
- Iwan Briquemont. Optimising client-side geo-replication with partially replicated data structures. Masters thesis, Louvain-la-Neuve, September 2014
- Aimee Chanthadavong. Internet of things to drive explosion of useful data: Emc. Technical report, ZDNet, April 2014
- R. Kingsy Grace and R. Manimegalai. Dynamic replica placement and selection strategies in data grids a comprehensive survey. J. Parallel Dis- trib. Comput., 74(2):20992108, February 2014. ISSN 0743-7315. doi: 10.1016/j.jpdc.2013.10.009
- Cisco. The zettabyte era-trends and analysis. Technical report, Cisco, June 2014
- Noriyani Mohd. Zin, A. Noraziah, AinulAzila Che Fazi, and Tutut Herawan. Replication techniques in data grid environments. In Jeng-Shyang Pan, Shyi-Ming Chen, and NgocThanh Nguyen, editors, Intelligent Information and Database Systems, volume 7197 of Lecture Notes in Computer Science, pages 549-559. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-28489-2. doi: 10.1007/978-3-642-28490-8\_57
- T. Hey, Tansley S, and K. Tolle. The fourth paradigm: Data-intensive scientific discovery. Technical report, Microsoft Research, 2009
- Shaik Naseera and K.V. Madhu Murthy. Agent based replica placement in a data grid environment. Computational Intelligence, Communication Systems and Networks, International Conference on, 0:426-430, 2009. doi: <http://doi.ieeecomputersociety.org/10.1109/CICSYN.2009.77>
- Xiaohua Dong, Ji Li, Zhongfu Wu, Dacheng Zhang, and Jie Xu. On dynamic replication strategies in data service grids. In Object Oriented Real- Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on, pages 165161, May 2008. doi: 10.1109/ISORC.2008.66
- Sushant Goel and Rajkumar Buyya. Data replication strategies in wide area distributed systems. In Robin G. Qiu, editor, Enter- prise Service Computing: From Concept to Deployment, pages 211241. Idea Group Inc, 2006
- M. Dorigo. Optimization, Learning and Natural Algorithms. PhD thesis, Politecnico di Milano, Italy, 1992