

# Mathematical Modelling of Distributed Heterogenous Data Storages

Olga M. Gerget  
Valeriy V. Parubets

National Research Tomsk Polytechnic University

# Distributed heterogeneous storages

- Unstructured data
- Data is collected from different sources
- Data is located on different physical units

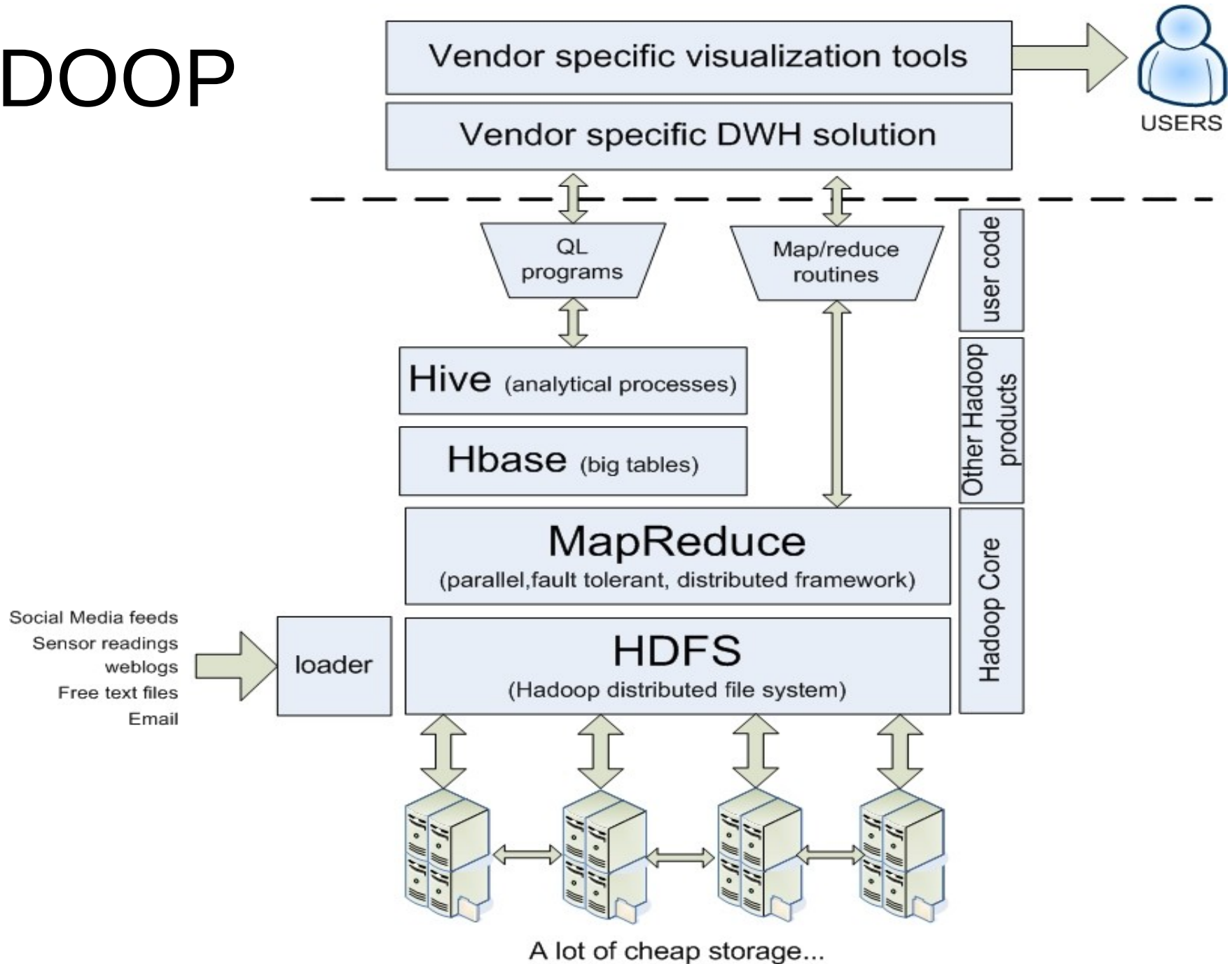
# Purposes

- Rough prediction of data traffic flow
- Defining of «bottle-necks» in cluster
- Probable way to build custom balancer
- Prediction of problems that may be caused by node failure

# Assigned projects

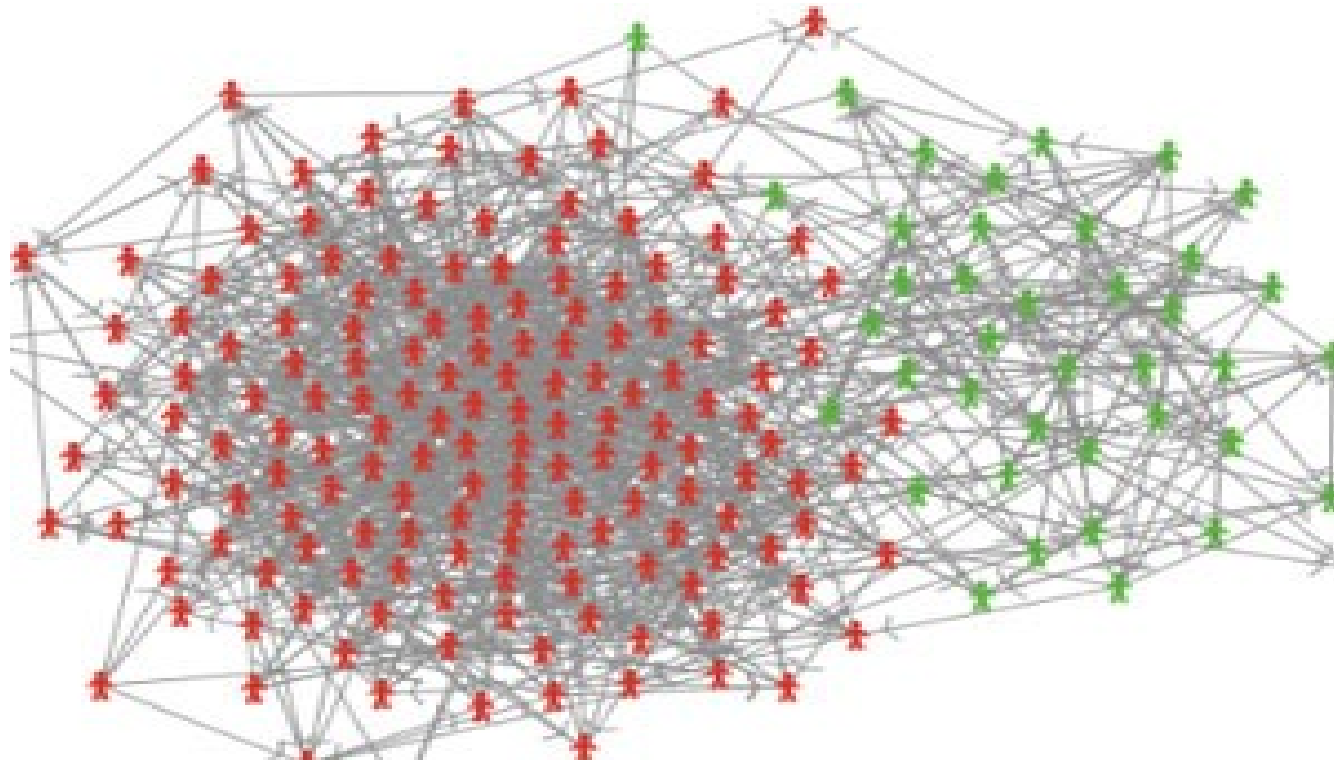
- Geological seismic exploration
- Medical data
- CERN-TPU projects (BigData Lab)

# HADOOP

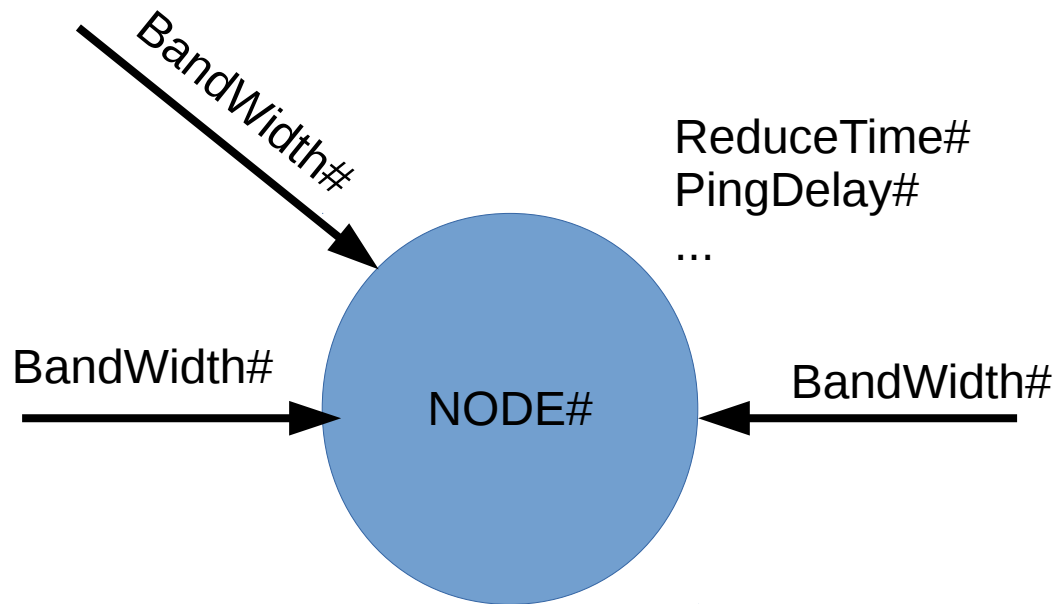


# Agent-based models

- Similar to SOM
- Easy to use, describe and develop



# Node



- Bandwidths
- Map time (as a function of type of traffic)
- Reduce time (as a function of type of traffic)

# ABM Optimization

- General genetic algorithm
- Multi-parameter optimization
- Easy to develop and use



# Results

- It works
- Method can be used for rough estimation of complex system
- Method can be easily implemented on HADOOP cluster for using as a balancer
- Method can be improved with forecasting system for traffic flow

Thank you for attention