



MapReduce

Jan Cychnerski
2017



POLITECHNIKA
GDAŃSKA

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI

Parallel algorithms

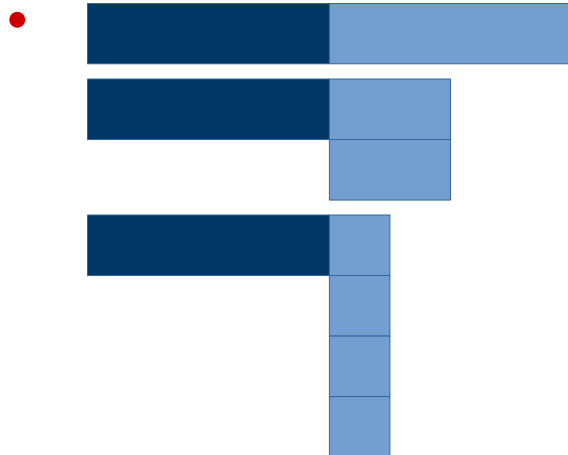
- Algorithm optimization for parallel execution
- The same problem, the same data



Amdahl's law

- More resources →
 - faster parallel part
 - the same sequential part
- Optimization of sequential part

- $$S_n = \frac{n}{1 + (n-1)\alpha}$$



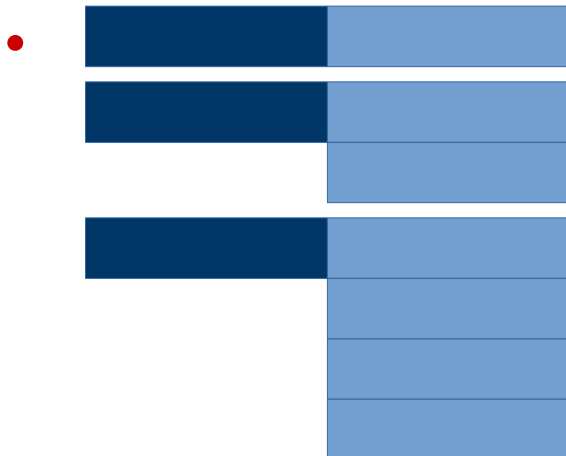


POLITECHNIKA
GDAŃSKA

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI

Gustafson's law

- More resources →
 - more data →
 - bigger parallel part
 - smaller sequential part
- Optimization of resource utilization
- $S'_n = \alpha + (1 - \alpha)n$





POLITECHNIKA
GDAŃSKA

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI

Algorithm parallelization

- Shorter end-to-end time
- Longer total computation time
- Scalability
- Implementation cost & time
- Execution cost



POLITECHNIKA
GDAŃSKA

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI

Today problems

- Big Data
- Energetic efficiency
- Cost (in money)



POLITECHNIKA
GDAŃSKA

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI

Big Data

- Buzzword
- 3V
 - Volume – huge amount of data
 - Velocity – changes quickly
 - Variety – different types



POLITECHNIKA
GDAŃSKA

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI

Big data processing

- Clusters, grids, clouds
- Failure recovery
- Efficiency
- Simplicity



HISTORIA MĄDROŚCIĄ
PRZYSZŁOŚĆ WYZWANIEM