Parallel and distributed calculations supported and managed by relational database.

Piotr Oramus

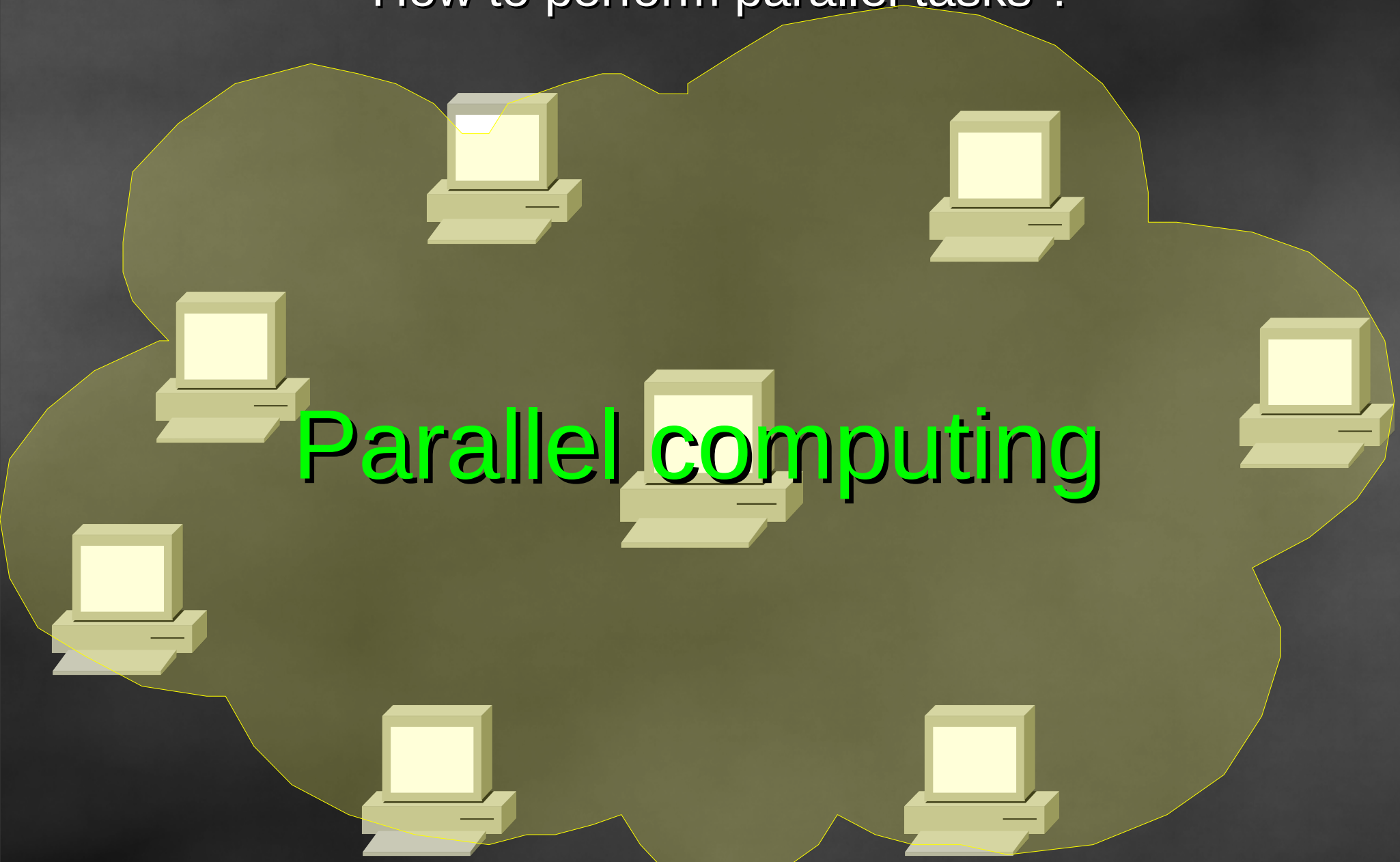
Department for Information Technology,
Faculty of Physics, Astronomy and Applied Computer Science

Jagiellonian University
Reymonta 4, 30-059 Krakow, Poland

e-mail: piotr.oramus@uj.edu.pl

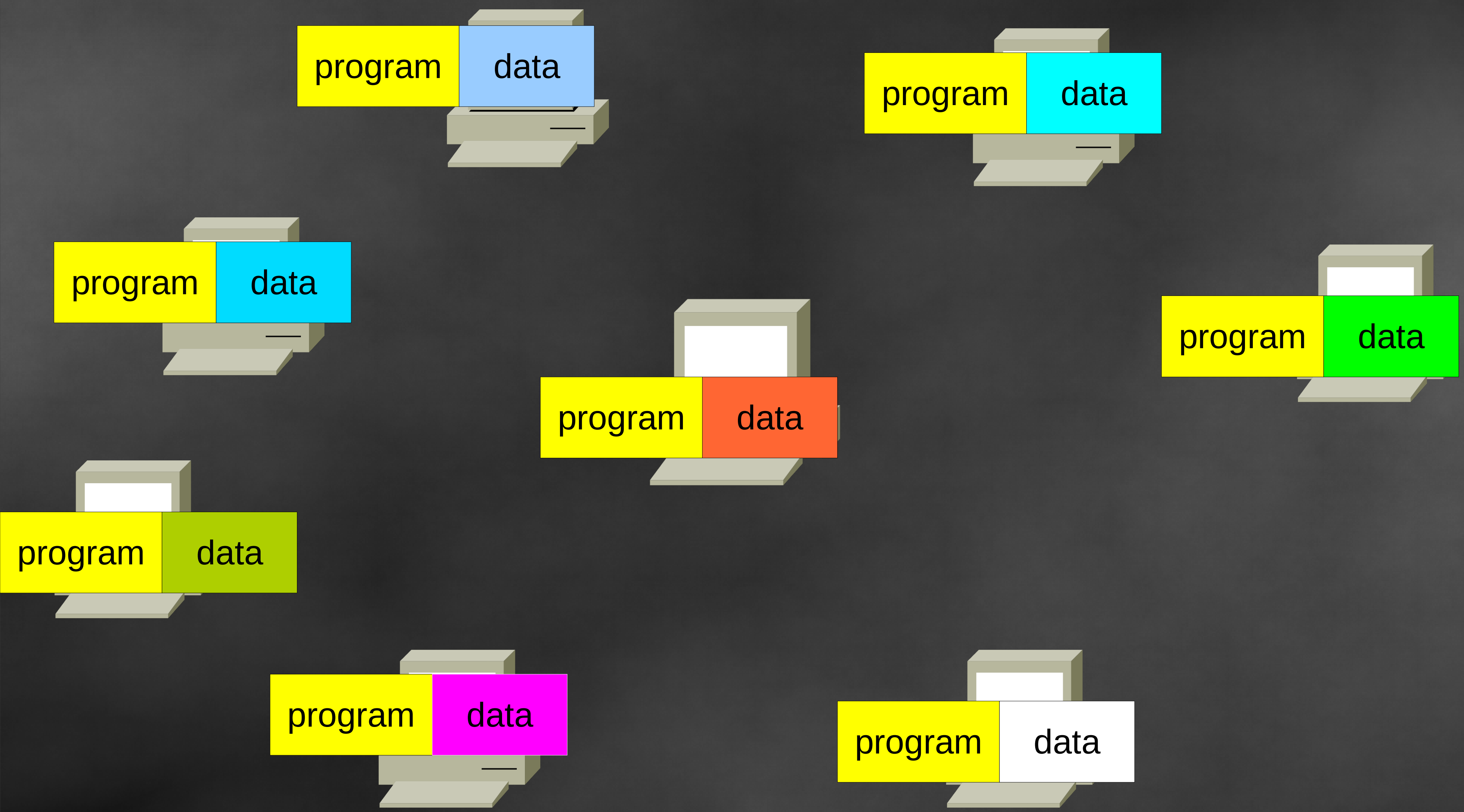
We can use many computers to perform our calculations

How to use them efficiently ?
How to perform parallel tasks ?



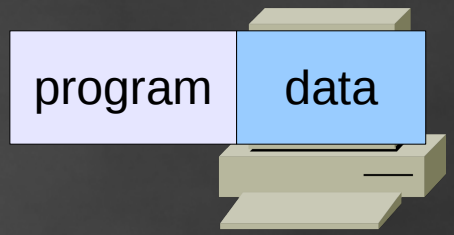
How do we use computers ?

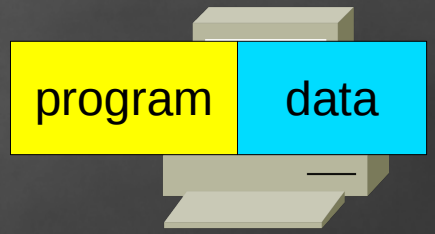
Single program + multiple data   Multiple program + multiple data

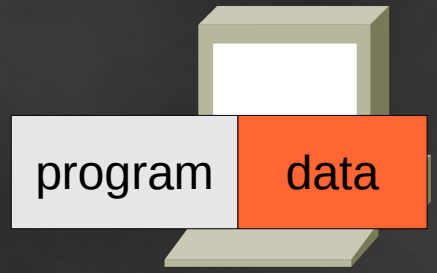


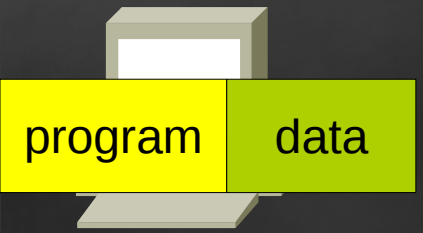
How do we use computers ?

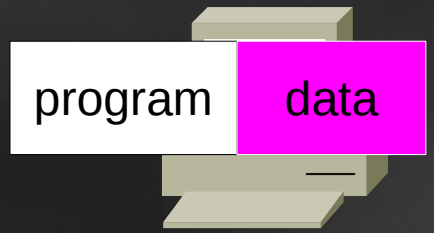
Single program + multiple data



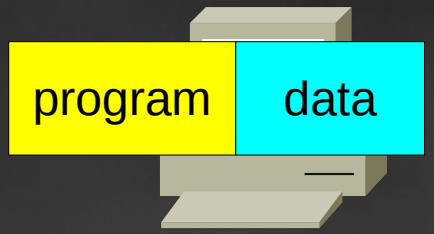


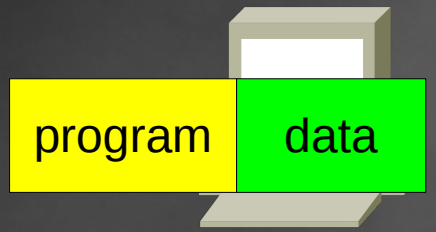


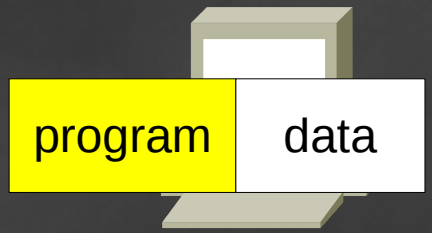




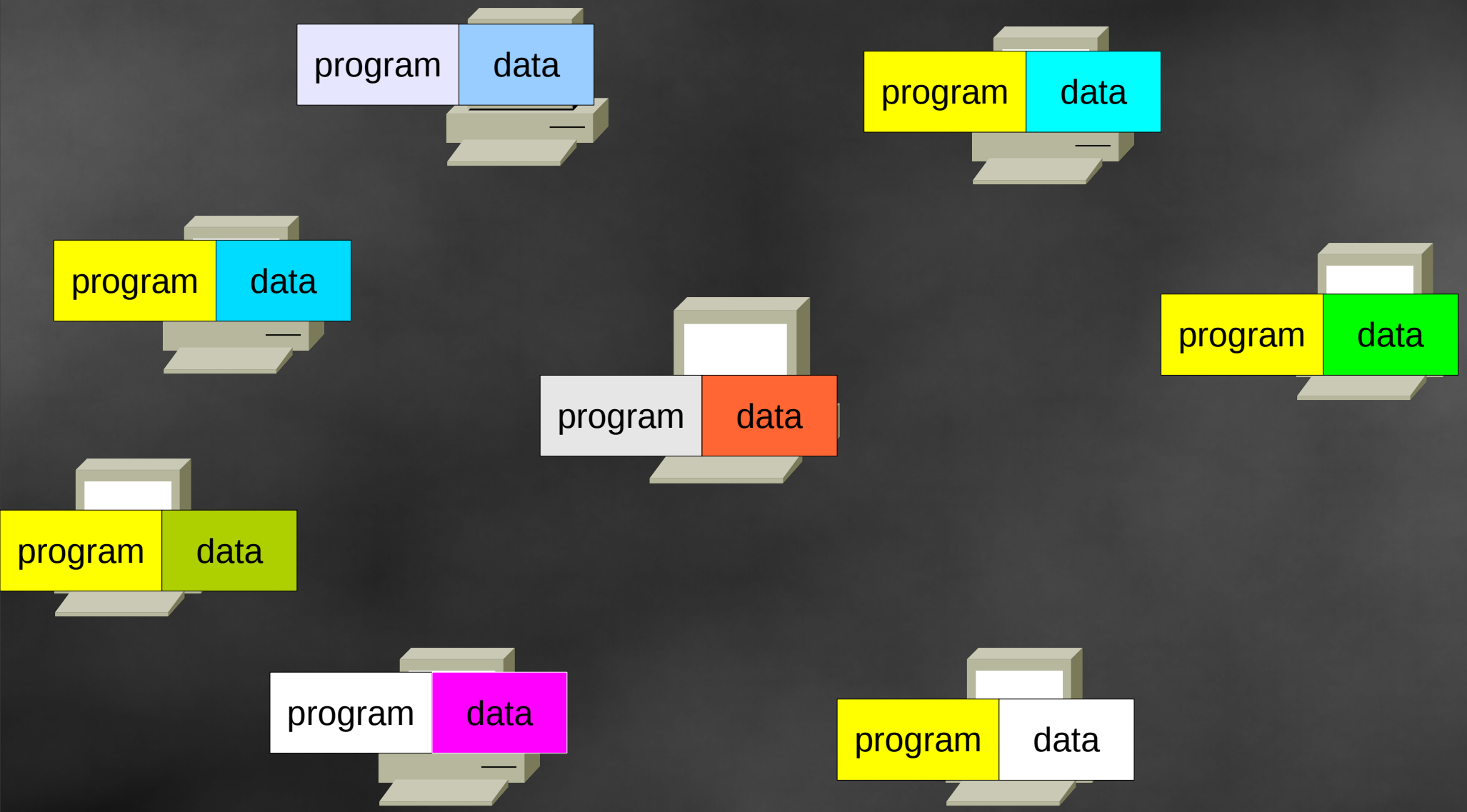
Multiple program + multiple data







What is missing ?
How to communicate with the program?



What is missing ?
How to communicate with the program?

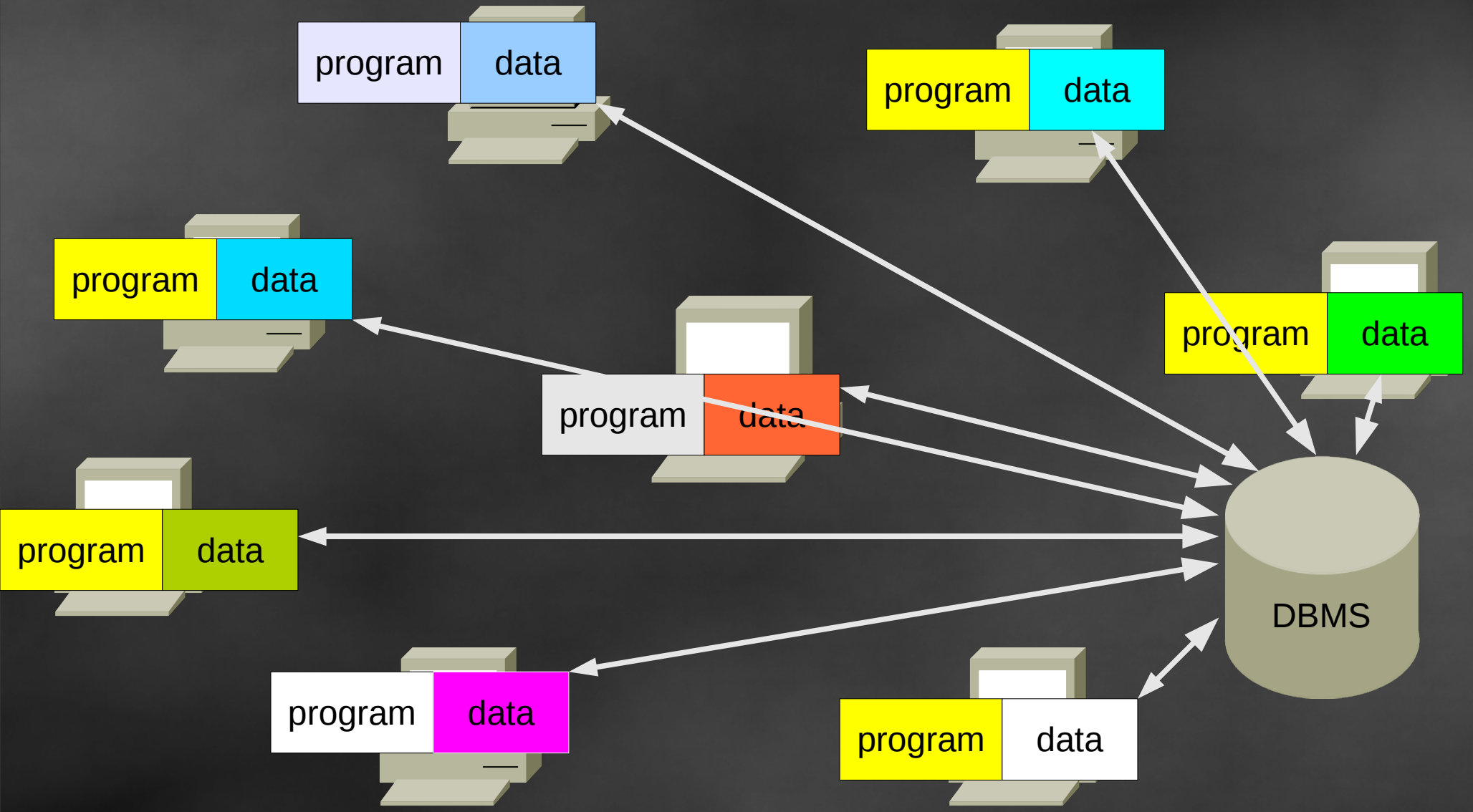
We do not communicate with our programs.
We do communicate with “file-system”..

Files...

+ easy to use
+ highly efficient
- let's try to find something...
- is it a last copy ?

We can use **D**ata **B**ase **M**anagement **S**ystem !!!

We do not have to use files, at all...



Relational database
E.g. MySQL, SQLite, PostgreSQL, Oracle

Relational database keeps data in tables (relations)
data + structure = data base

MyExperiments(id, Temperature, Size, Steps)

Attribute

Row (tuple)

id	Temperature	Size	Steps
1	1000.0	16	10000
2	1100.0	16	10000
3	1200.0	16	10000

How to exchange data with the relational database?
MySQL + MySQL++

MySQL++ and Specialized SQL Structures

Relation:

```
MyExperiments( id, Temperature, Size, Steps )
```

Macro:

```
sql_create_4( MyExperiments, 1, 4,
    mysqlpp::sql_sql_int_unsigned, id,
    mysqlpp::sql_double, Temperature,
    mysqlpp::sql_sql_int_unsigned, Size,
    mysqlpp::sql_sql_int_unsigned, Steps );
```

C++ code:

```
mysqlpp::Query query = [.....];
query << "select * from MyExperiments where Size > 10";
vector< MyExperiments > res;
query.storein(res);
```

The system

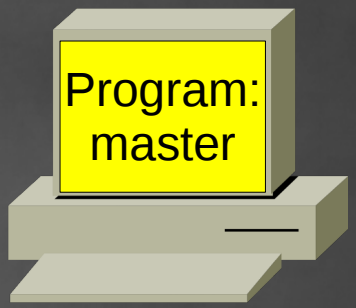
HOSTS			
host_id	hostname	role	timestamp

EXPERIMENTS		
experiment_id	host_id	parameters

JOBS				
job_id	experiment_id	host_id	status	parameters

RESULTS	
job_id	results

The system



You are responsible for the experiment 1

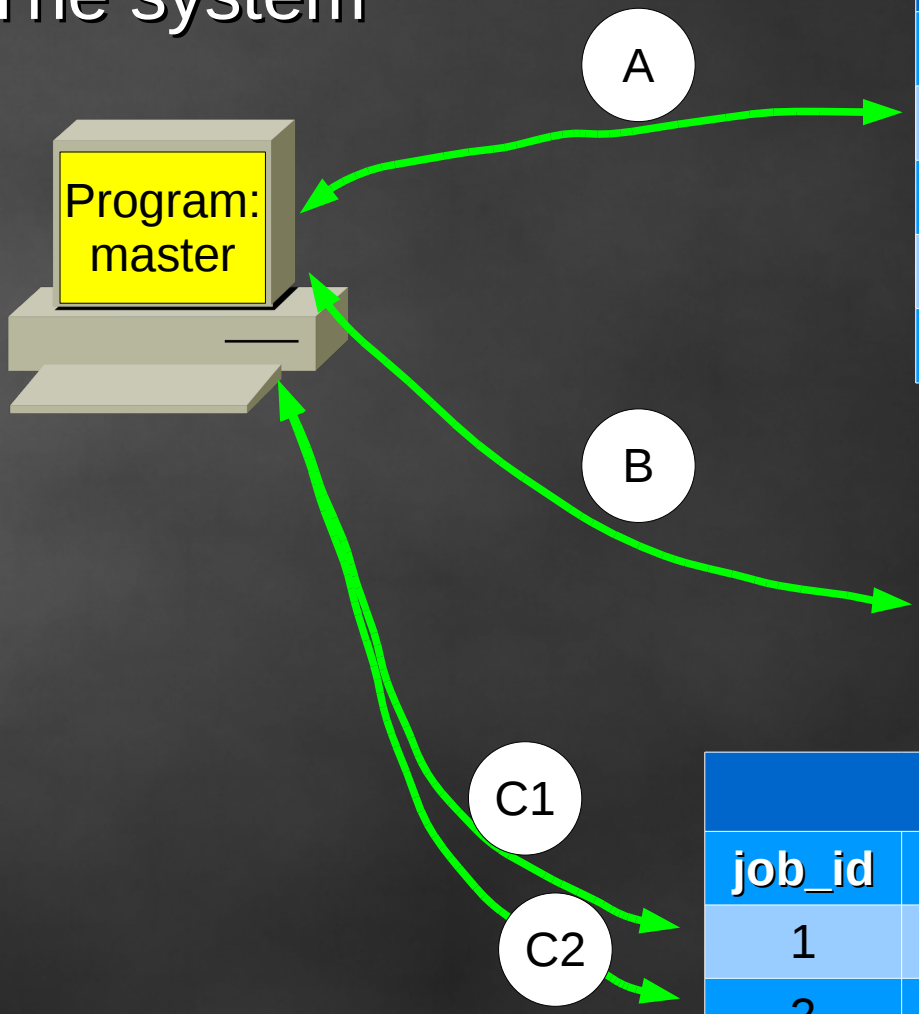
HOSTS			
host_id	hostname	role	timestamp

EXPERIMENTS		
experiment_id	host_id	parameters
1		XXXXXXXXXX

JOBS				
job_id	experiment_id	host_id	status	parameters

RESULTS	
job_id	results

The system



HOSTS

host_id	hostname	role	timestamp
1	server1	master	12:00

EXPERIMENTS

experiment_id	host_id	parameters
1	1	XXXXXXXXXX

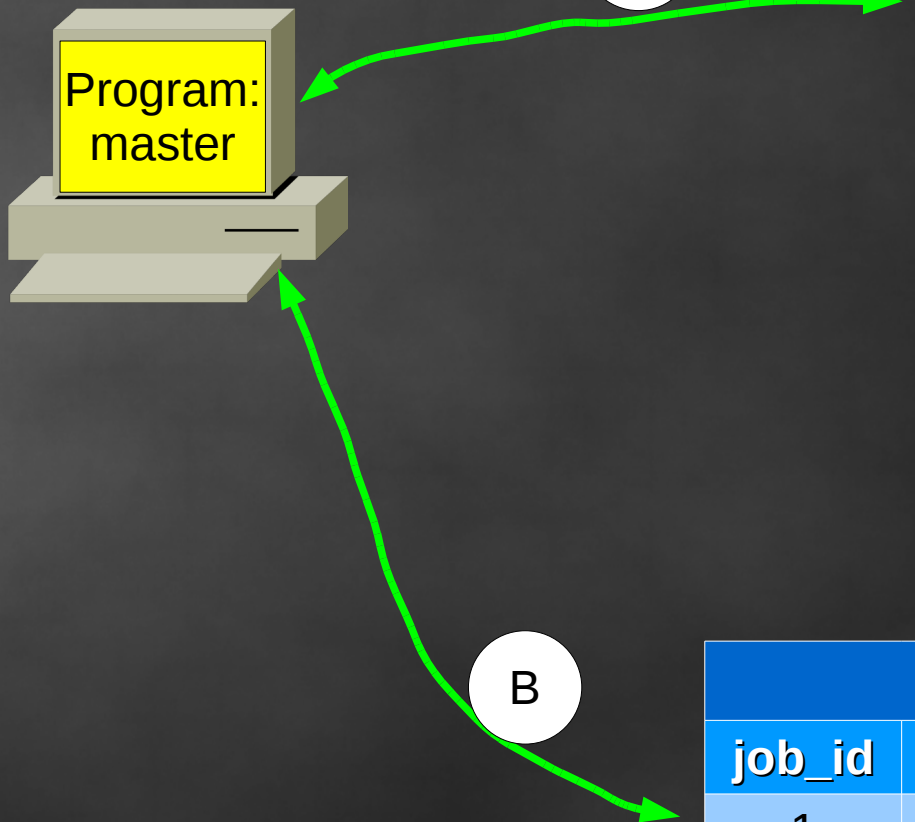
JOBS

job_id	experiment_id	host_id	status	parameters
1	1		waiting	xxxxx1111
2	1		waiting	xxxxx2222

RESULTS

job_id	results

The system



HOSTS

host_id	hostname	role	timestamp
1	server1	master	12:15

EXPERIMENTS

experiment_id	host_id	parameters
1	1	XXXXXXXXXX

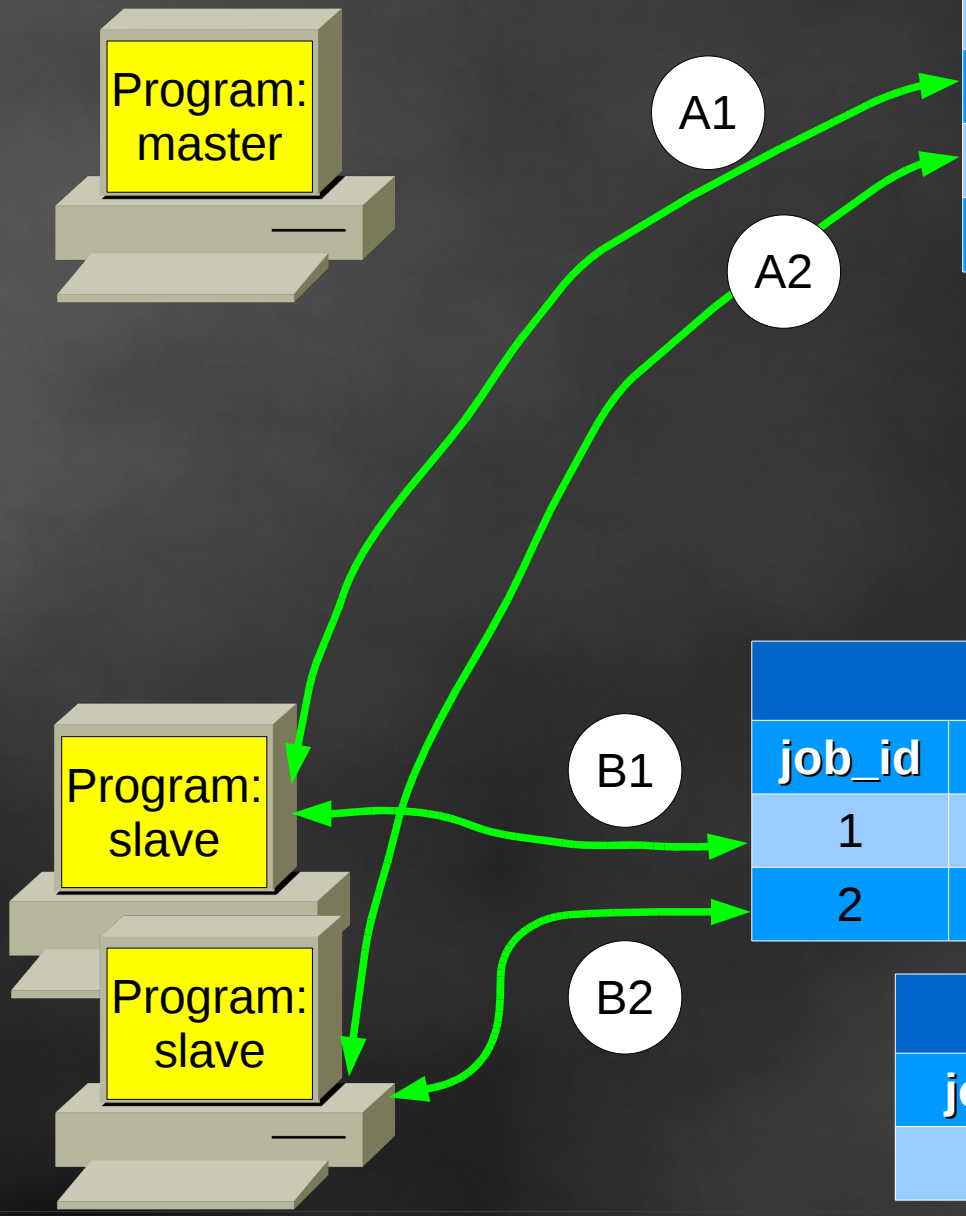
JOBS

job_id	experiment_id	host_id	status	parameters
1	1		waiting	xxxxx1111
2	1		waiting	xxxxx2222

RESULTS

job_id	results

The system



HOSTS

host_id	hostname	role	timestamp
1	server1	master	12:15
2	node01	slave	12:16
3	node10	slave	12:14	

EXPERIMENTS

experiment_id	host_id	parameters
1	1	XXXXXXXXXX

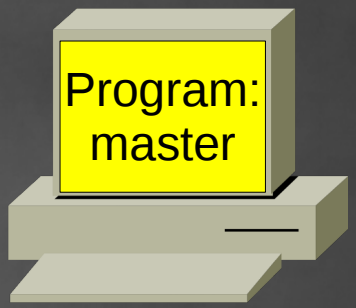
JOBS

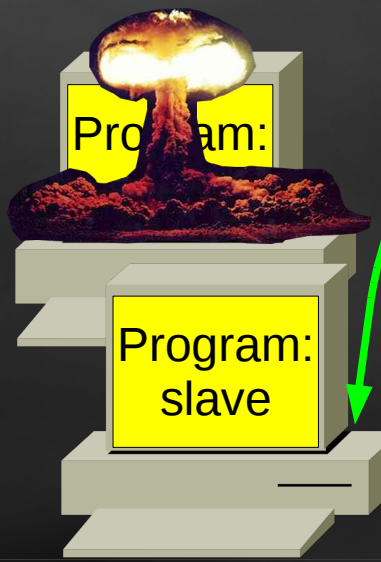
job_id	experiment_id	host_id	status	parameters
1	1	2	working	xxxxx1111
2	1	3	working	xxxxx2222

RESULTS

job_id	results

The system





A

B

C

HOSTS

host_id	hostname	role	timestamp
1	server1	master	12:45
2	node01	slave	12:20
3	node10	slave	12:46

EXPERIMENTS

experiment_id	host_id	parameters
1	1	XXXXXXXXXX

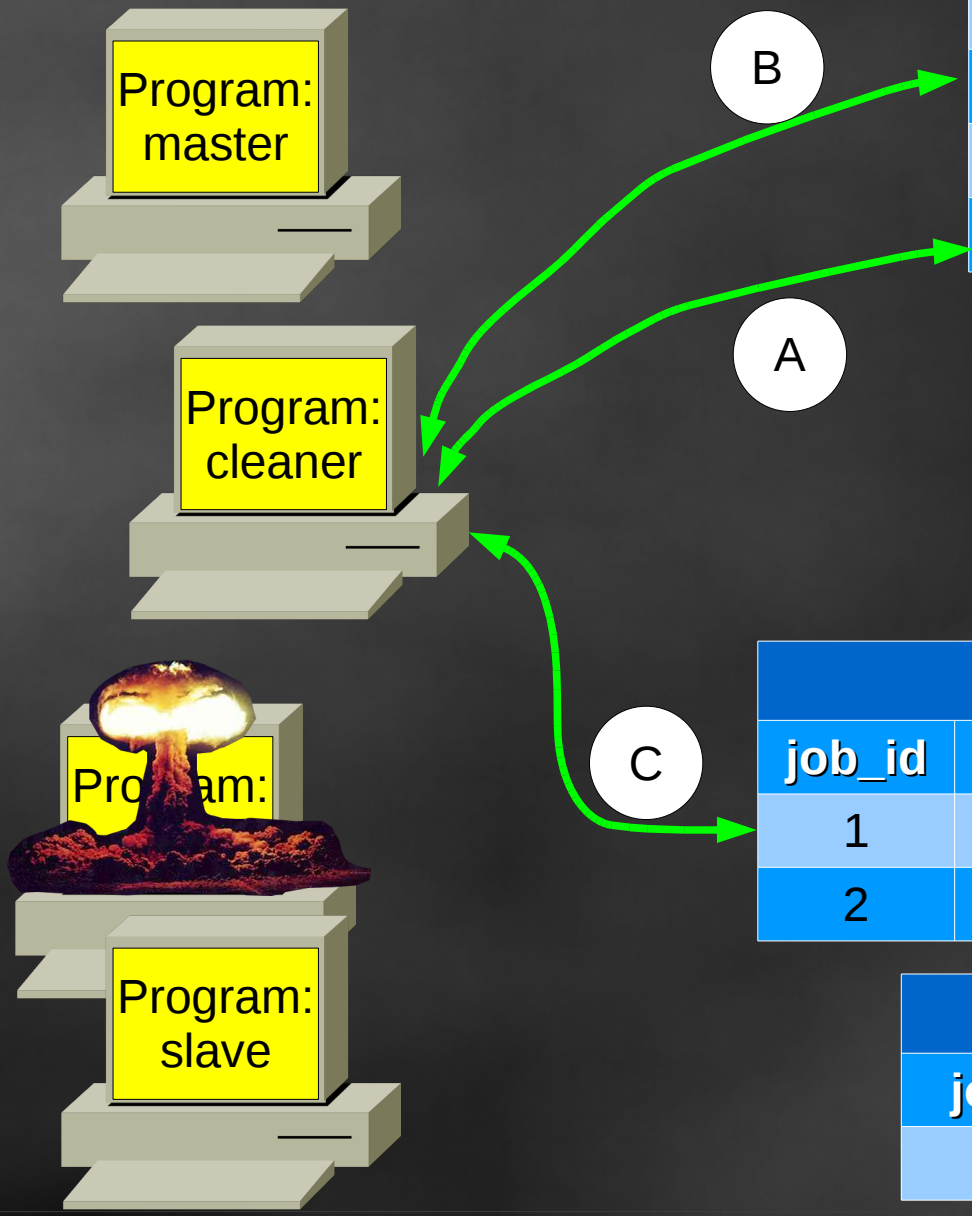
JOBS

job_id	experiment_id	host_id	status	parameters
1	1	2	working	xxxxx1111
2	1	3	done	xxxxx2222

RESULTS

job_id	results
2	rrrrrr-xxxx-2222

The system



HOSTS

host_id	hostname	role	timestamp
1	server1	master	12:59
~~2~~	~~node01~~	~~slave~~	~~12:20~~
3	node10	slave	12:58
4	PCxxx	cleaner	13:00

EXPERIMENTS

experiment_id	host_id	parameters
1	1	XXXXXXXXXX

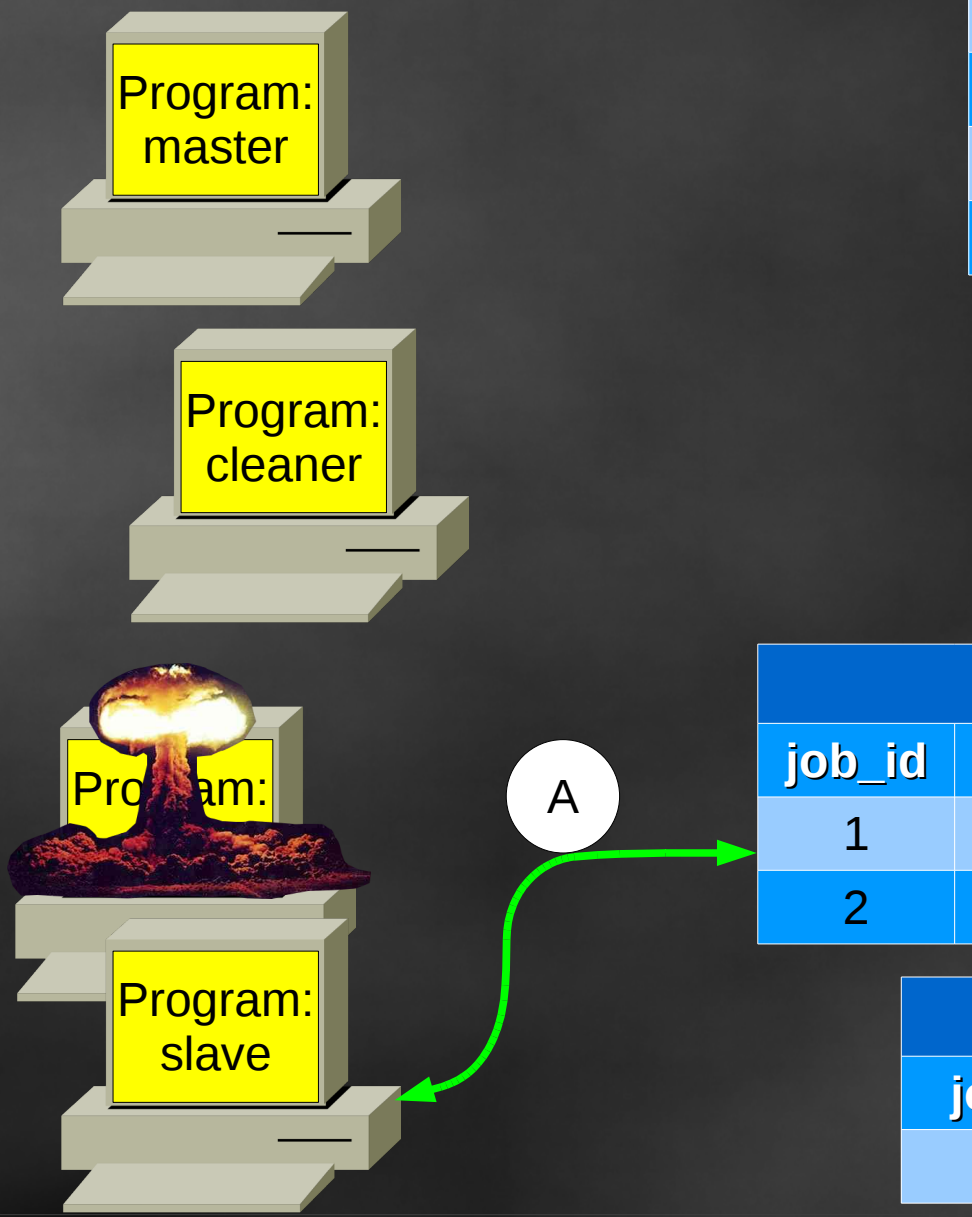
JOBS

job_id	experiment_id	host_id	status	parameters
1	1		waiting	xxxxx1111
2	1	3	done	xxxxx2222

RESULTS

job_id	results
2	rrrrrr-xxxx-2222

The system



HOSTS

host_id	hostname	role	timestamp
1	server1	master	13:10
3	node10	slave	13:08
4	PCxxx	cleaner	13:11

EXPERIMENTS

experiment_id	host_id	parameters
1	1	XXXXXXXXXX

JOBS

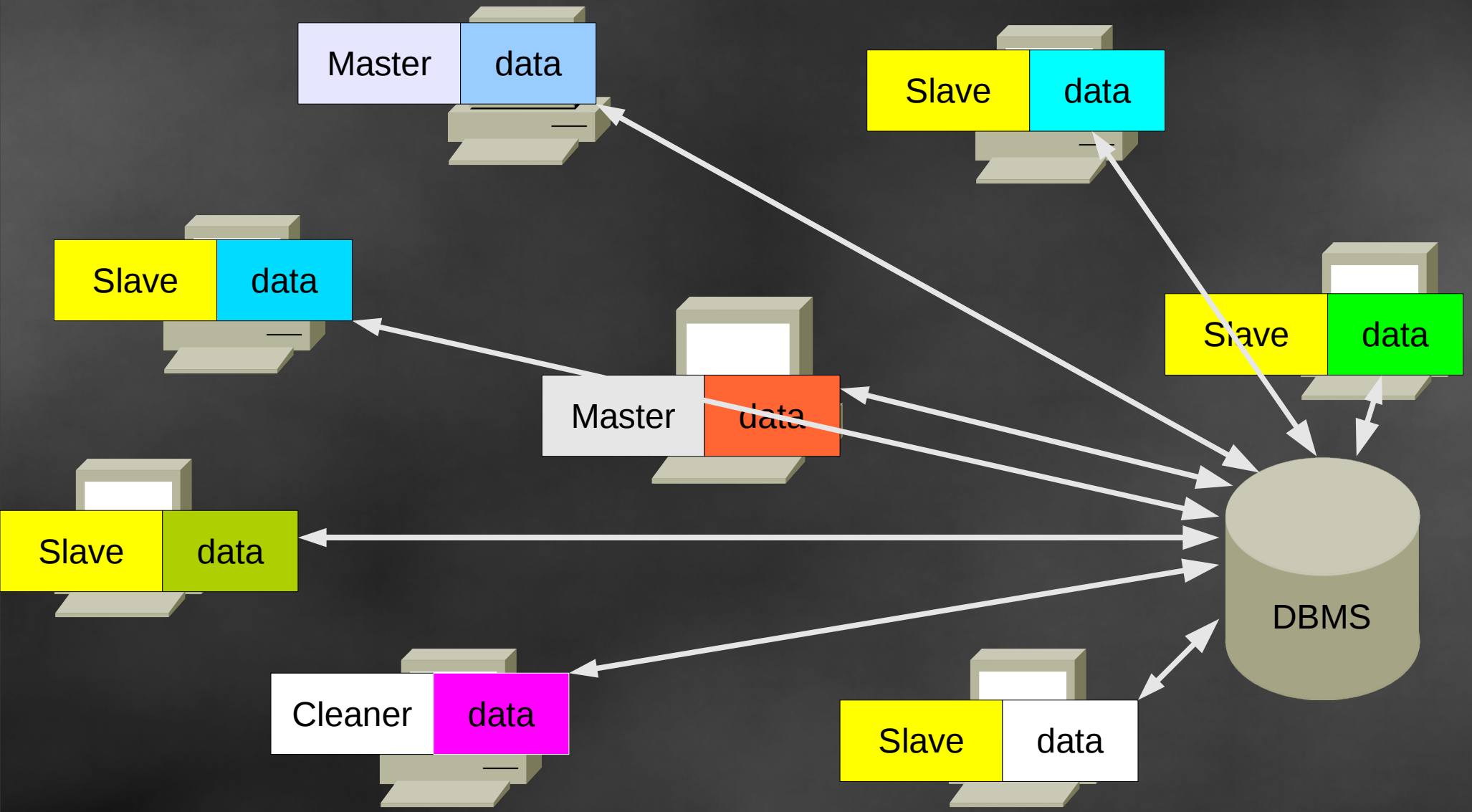
job_id	experiment_id	host_id	status	parameters
1	1	3	working	xxxxx1111
2	1	3	done	xxxxx2222

RESULTS

job_id	results
2	rrrrrr-xxxx-2222

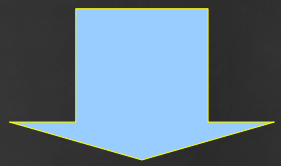
System can perform simultaneously several different tasks

We can start several master programs...

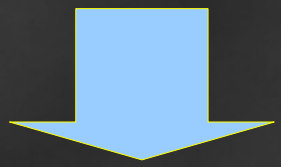


Application: Optimization by means of Parallel Genetic Algorithm

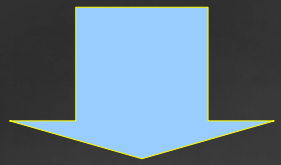
Master generates initial chromosome population



Slaves calculate chromosome evaluation



Master generates next generation
(selection, crossover and mutation)



Slaves calculate chromosome evaluation

HOSTS relation:

host_id	hostname	role	status	up_time	timestamp
222	ztis5.if.uj.edu.pl	master	waiting	554055.588951	2009-03-05 12:03:32
224	pc1	slave	working	554002.773231	2009-03-05 12:03:10
231	mars47	slave	working	552981.694843	2009-03-05 12:03:28
233	mars41	slave	working	552973.877192	2009-03-05 12:03:20
242	mars37	slave	working	544630.061523	2009-03-05 12:03:28
243	mars36	slave	working	544628.034187	2009-03-05 12:03:28
244	mars37	slave	working	544631.523268	2009-03-05 12:03:32
245	mars36	slave	working	544625.306000	2009-03-05 12:03:26
246	mars35	slave	working	544610.998289	2009-03-05 12:03:27
248	mars43	slave	working	543725.967278	2009-03-05 12:03:18
249	mars43	slave	working	543709.426398	2009-03-05 12:03:18
250	mars43	slave	working	543672.438118	2009-03-05 12:03:16
252	mars50	slave	working	543663.246544	2009-03-05 12:03:14
253	mars49	slave	working	543642.270812	2009-03-05 12:03:32
254	mars39	slave	working	543607.092386	2009-03-05 12:03:15
255	mars41	slave	working	543596.116262	2009-03-05 12:03:20
262	pc4	slave	working	245031.627156	2009-03-05 12:03:08
263	ztis5.if.uj.edu.pl	master	waiting	219923.809201	2009-03-05 12:03:26
264	ciserv	slave	working	219829.137327	2009-03-05 12:03:20
265	pc6	slave	working	219756.796865	2009-03-05 12:03:10
266	ztis5.if.uj.edu.pl	cleaner	working	136026.507011	2009-03-05 12:02:31
267	pc3	slave	working	37724.719792	2009-03-05 12:03:07
268	pc5	slave	working	37700.237025	2009-03-05 12:02:56

Neural network ensemble

The same test patterns
The best results:

Single neural networks : **2537** wrongly classified patterns
Ensemble composed of 15 NN : **355** wrongly classified patterns