Introduction in High-Performance Computing (**HPC**)

Moldova

South Eastern Europe (SEE)

RENAM
IMI ASM

www.renam.md
www.math.md

Nicolai ILIUHA

Personal computer (PC) is any general-purpose computer which is intended to be operated directly by an end user.

Personal computer

Laptop
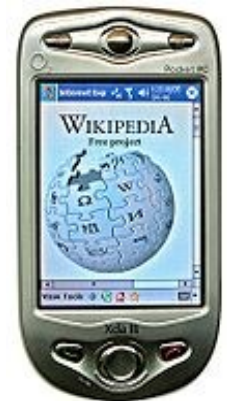
(Notebook)

Netbook

Pocket PC

Server (**hardware**) is a computer, dedicated to performing any service tasks without direct human intervention.

Servers are distinguished from a group of personal computers or from a group of specialized computer equipment.
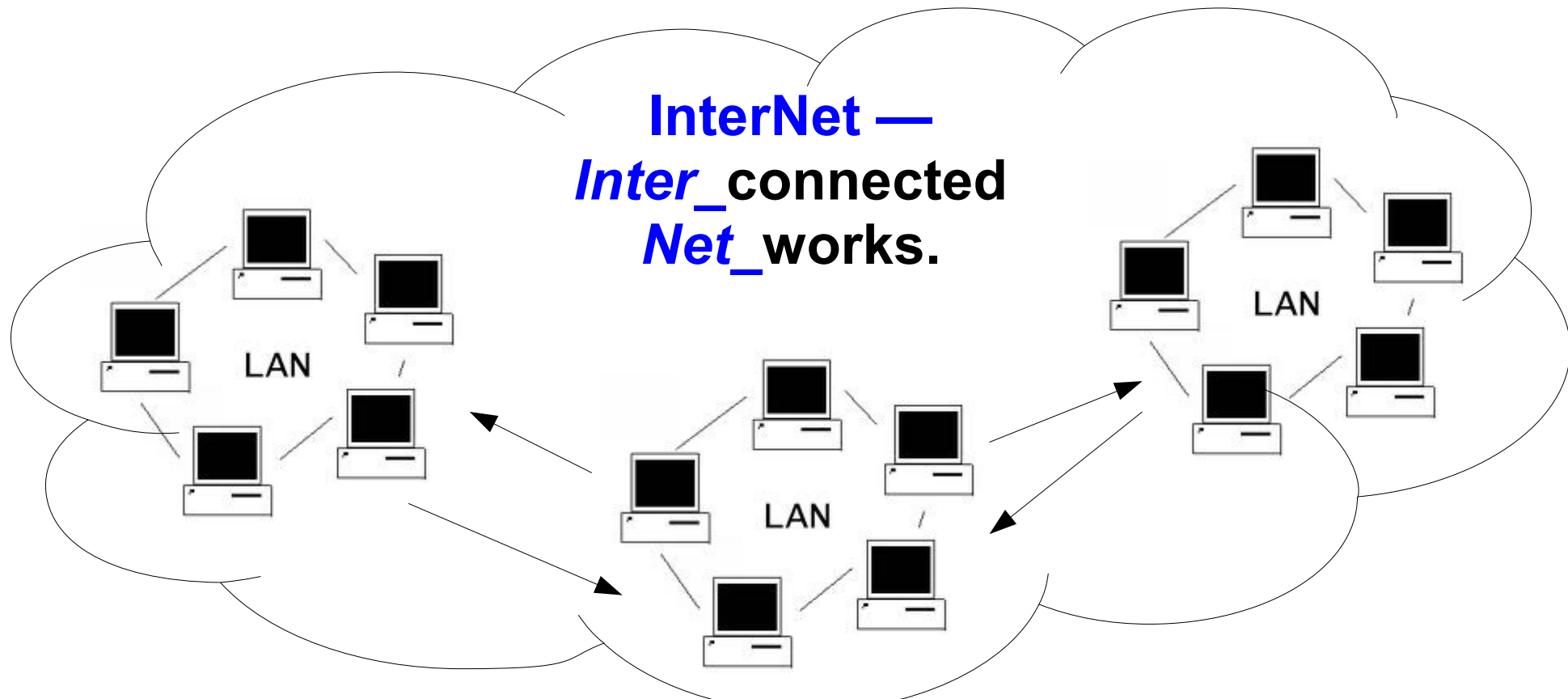
Personal computers and servers can be connected in **local area networks** (in the same room, one floor or one building).

Local networks are connected to global networks, creating the **Internet**.

InterNet —
***Inter*_connected**
***Net*_works.**

Using the included in **Internet** personal computer one can get access to the **Internet services.**

Internet Services - the software that processes requests from clients. Service is also called software server.

Client - a program on a PC, sending requests to the server.

Such interaction is called
client — server architecture.

Examples of Internet services

World Wide Web – service via the Internet provides access to information. The most widely used service today.

E-mail - E-mail.

HPC, **Grid** - access to computing and storage resources.

VoIP - (Voice over IP) — communication system that provides voice transmission over the Internet or any other IP-based networks.

Audio-, **Video-conferences** - the exchange of audio and video information in real-time between two or more remote users.

Telnet , FTP, Gopher , White pages,   Internet auction . . .

High Performance Computing (**HPC)**

High-performance computing is a branch of applied computer science that is dealing with the finding of solutions to problems that require a large amount of computing resources.

Computer users' applications can be divided into 3 categories (by the type of used computation):

1. **Sequential computation.**
2. **Parallel computing.**
3. **Distributed computing.**

Sequential computing

Application runs on one computer and its run time depends on the capacity of the computer. Applications in the scientific community generally fall into this category.

Application acceleration is achieved by:

- Using optimizing compilers – for getting a more efficient programm code.

- Using computers with huge processing power - for working with the applications that require the most intensive computing.

Parallel computing

Parallel computing is the simultaneous solution of various parts of one computational problem with the cores of one processor or by multiple processors (cores) of one or more computers.

This requires the solvable problem *to be segmented*, in other words divided into subtasks that can be evaluated in parallel.

Not every problem can be divided into subtasks that can be solved in parallel.

Not every problem *is worthwhile* to be separated into subtasks.

Supercomputers and **Clusters** are the best environment for the execution of parallel computations.

A **cluster** is a group of computers connected by high-speed communication channels and representing, from a user's point of view, a single hardware resource.

Client-server architecture.

56-core IMI-RENAM cluster

4096-core IBM iDataPlex cluster

A **supercomputer** is a multiprocessor complex with ultra-fast data transmission channels within the complex.

8192 cores

Distributed computing

Distributed or **Grid**-computing is a kind of parallel computing.

Computers of various capacities are united in a parallel computing system by the local and global networks.

Grid is a computational infrastructure that provides access through Internet to computing power and storage resources distributed across the globe.

- The increase of network bandwidth enhances the effectiveness of Grid.

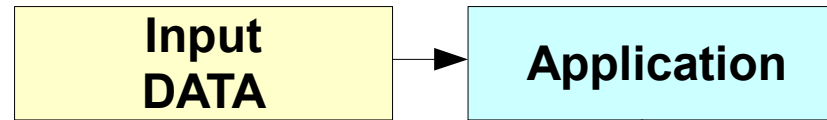- The duplication of network channels enhances the reliability of the infrastructure.

The logic of parallel computing

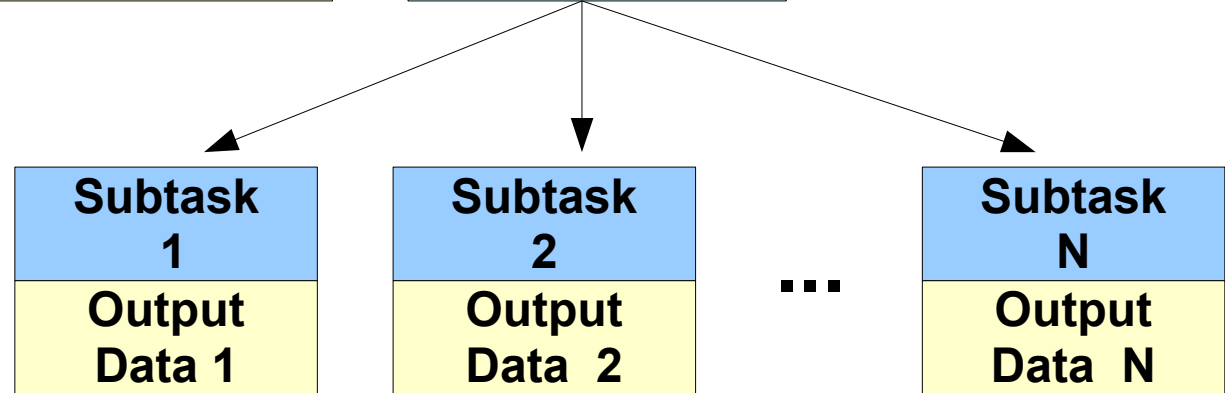Start of the application

| Input DATA | → | Application |

The application runs for the simultaneous execution of multiple subtasks

| Subtask 1 | Subtask 2 | ... | Subtask N |
| Output Data 1 | Output Data 2 | | Output Data N |

The application processes the results of completed subtasks
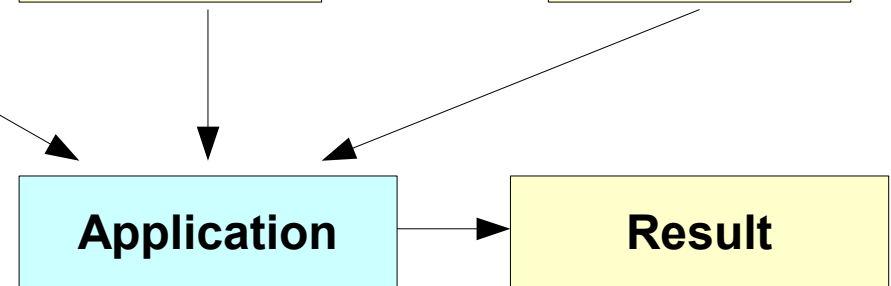
| Application | → | Result |

The logic of distributed computing
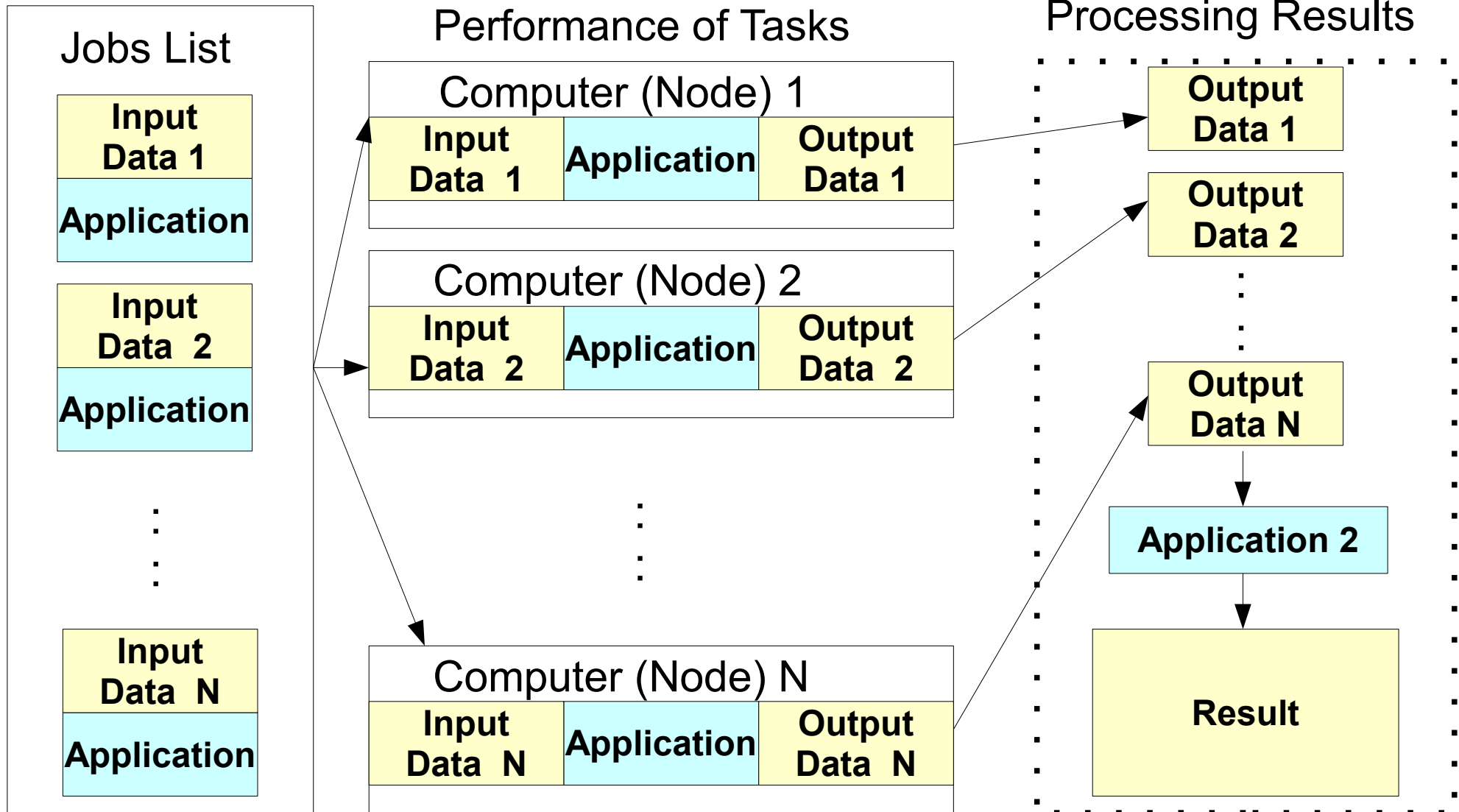
Conclusion

To reduce the physical time of the program work (walltime) the possibility of HPC will be increasingly used.

The following reasons influence the growth of the number of applications using parallel, distributed and combined calculation:

- The increase of the number of cores in a single processor.
- Constant growth of the rate of data transmission in local and global networks.
- Using the GPU video card for general computing.
- Improving programmers tools.