

# MSU HPC Center: From Universal Computing to Codesign Principles

*Dmitry Nikitenko*

*Head of the Scientific and Technical Department, Leading Researcher*

Research Computing Center  
Lomonosov Moscow State University

# Moscow State HPC Facilities

## MSU HPC Center

- “Lomonosov-2”
  - 4,9 PF peak
  - 1 696 nodes
  - 23 424 CPU cores
  - 40 960 GPU MPs
- Extensions
  - DGX-2 Extension
  - “ISTOK” SMP nodes

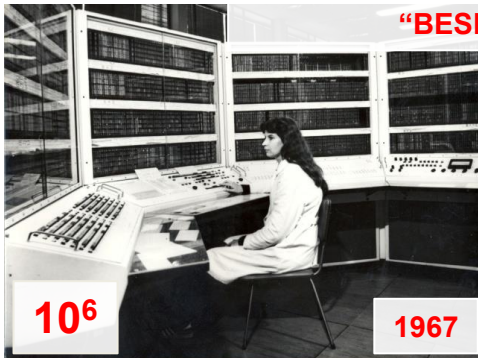
“Strela” – USSR 1st mass prod.



“Setun” – ternary logic



“BESM-6”



$10^{10}$



“Lomonosov-2”



$10^{15}$

“Lomonosov”



“Chebyshev”



$10^{13}$

# "Lomonosov-2": Top-Ranked HPC System (Top500 #12 in 2014)

## "Lomonosov-2" Supercomputer

Peak performance **4,95 Pflops**

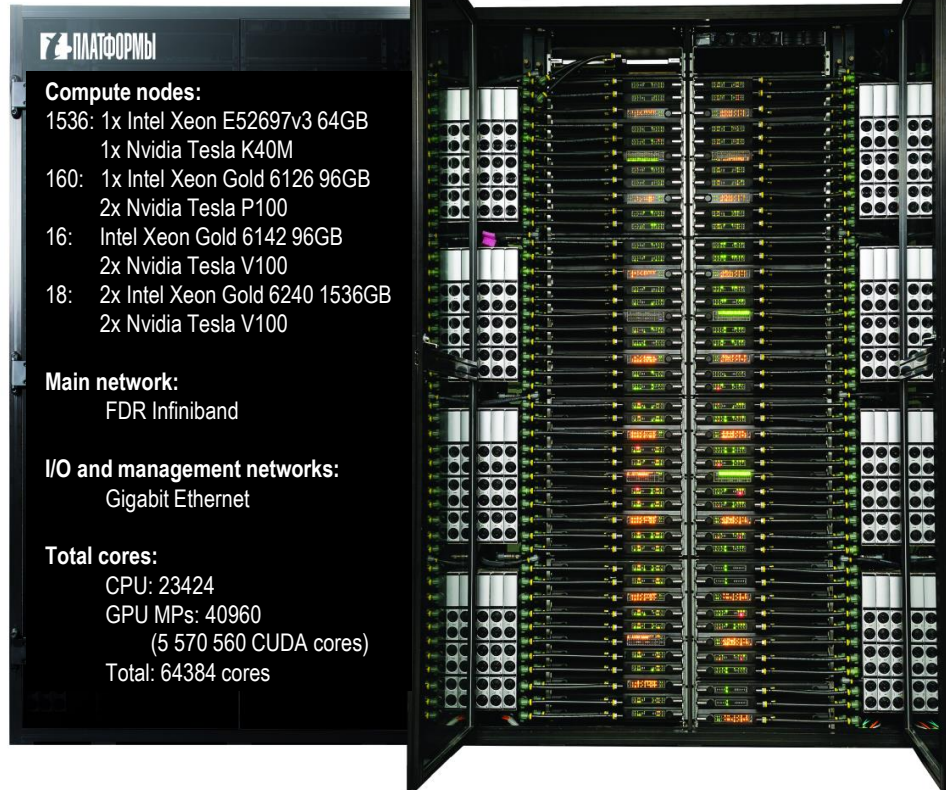
Linpack performance **2,48 Pflops**



Vendor: T-Platforms

#22 - 2014, November

#495 - 2025, June



ПАТФОРМЫ

### Compute nodes:

- 1536: 1x Intel Xeon E52697v3 64GB  
1x Nvidia Tesla K40M
- 160: 1x Intel Xeon Gold 6126 96GB  
2x Nvidia Tesla P100
- 16: Intel Xeon Gold 6142 96GB  
2x Nvidia Tesla V100
- 18: 2x Intel Xeon Gold 6240 1536GB  
2x Nvidia Tesla V100

### Main network:

FDR Infiniband

### I/O and management networks:

Gigabit Ethernet

### Total cores:

CPU: 23424  
GPU MPs: 40960  
(5 570 560 CUDA cores)  
Total: 64384 cores



# Flagship of Academia HPC for Many Years, Outstanding Long-liver

## "Lomonosov-2" Supercomputer

Peak performance **4,95 Pflops**

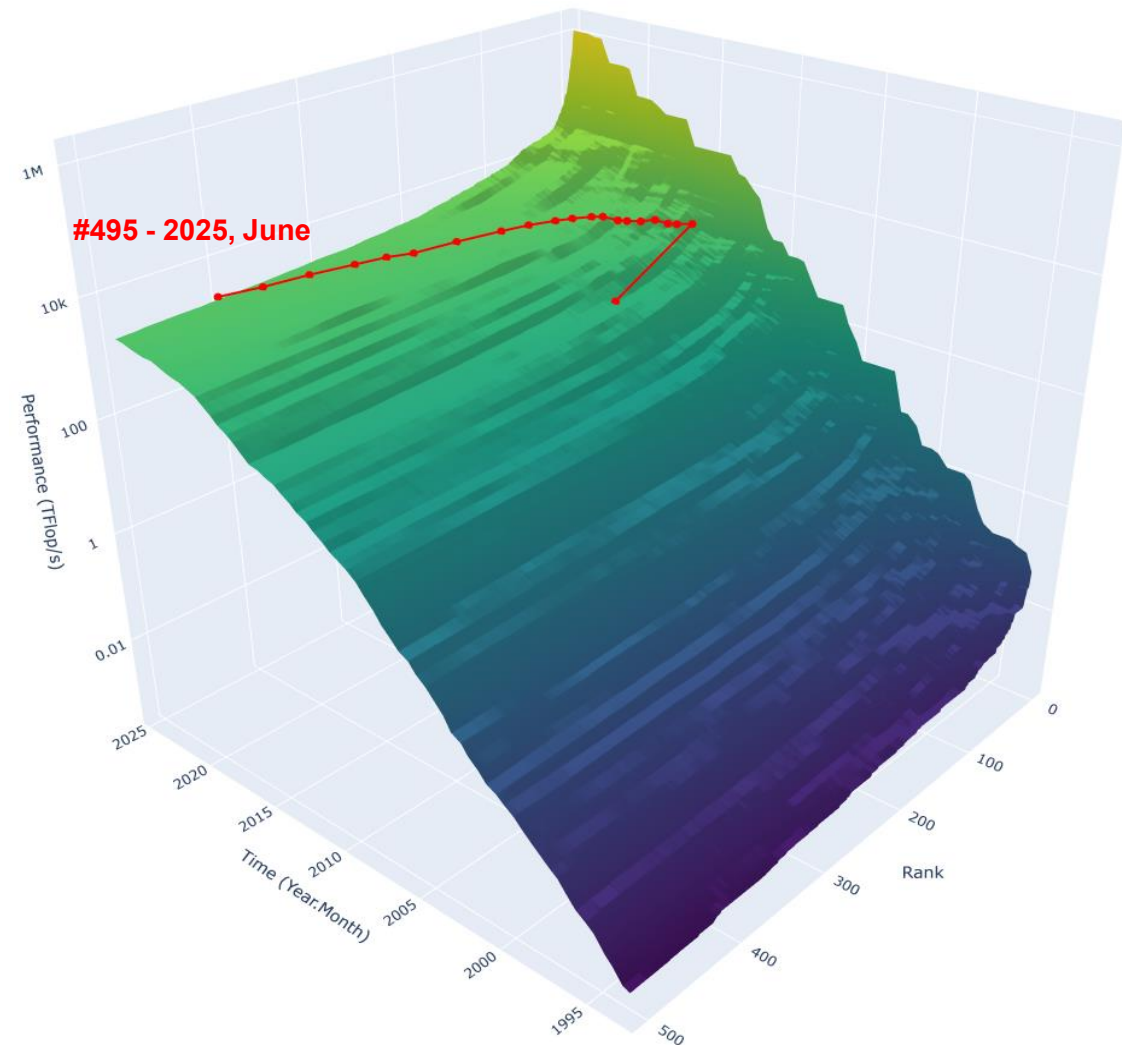
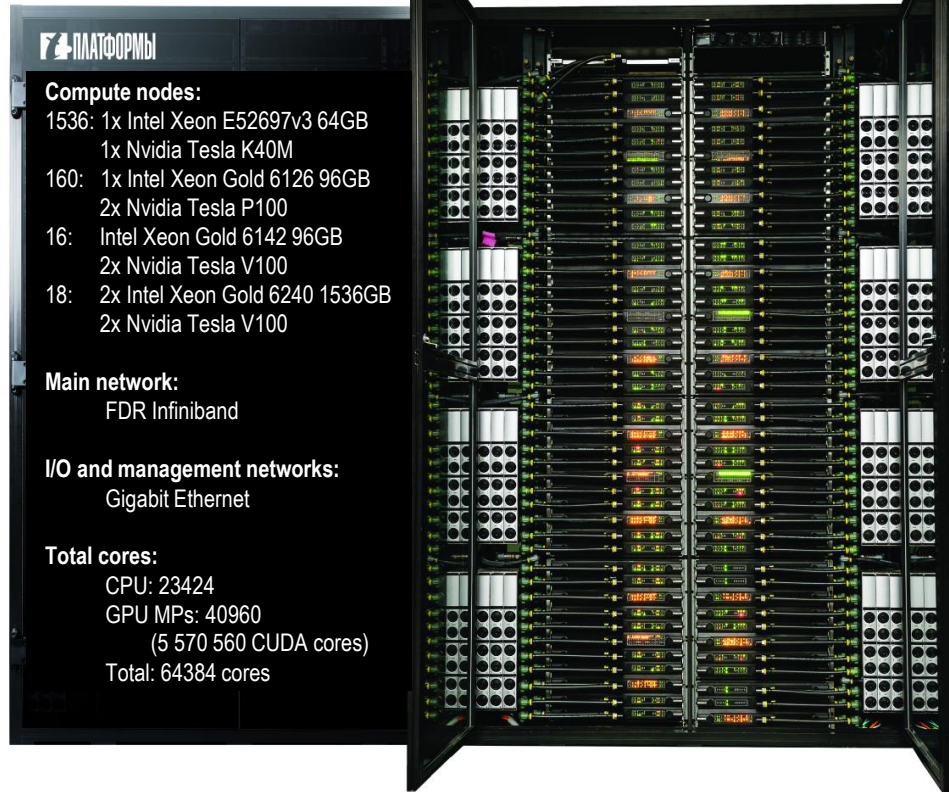
Linpack performance **2,48 Pflops**



Vendor: T-Platforms

#12 - 2014, November

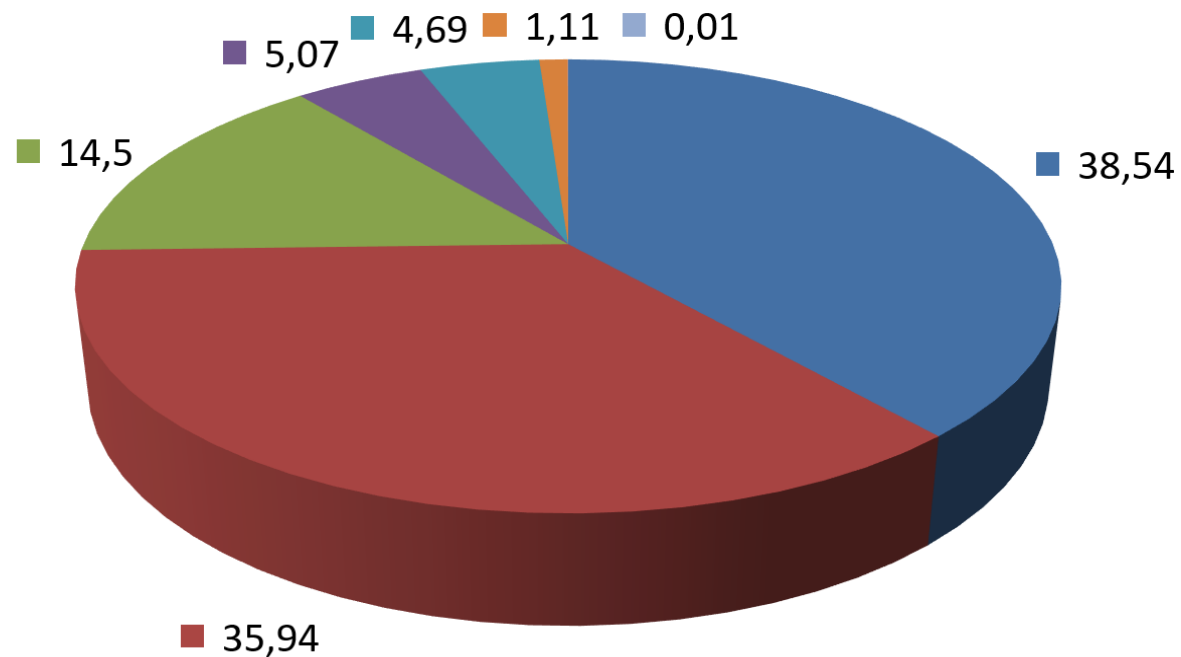
#495 - 2025, June



# Share of Consumed Nodehours by Field of Science

## TOTAL DIVERSITY:

- Programming Technologies
- Algorithms & Implementations
- Packages
- Hardware & Software



■ Chemistry

■ Physics and Astronomy

■ Fundamentals of Engineering

■ Biology and Medical Science

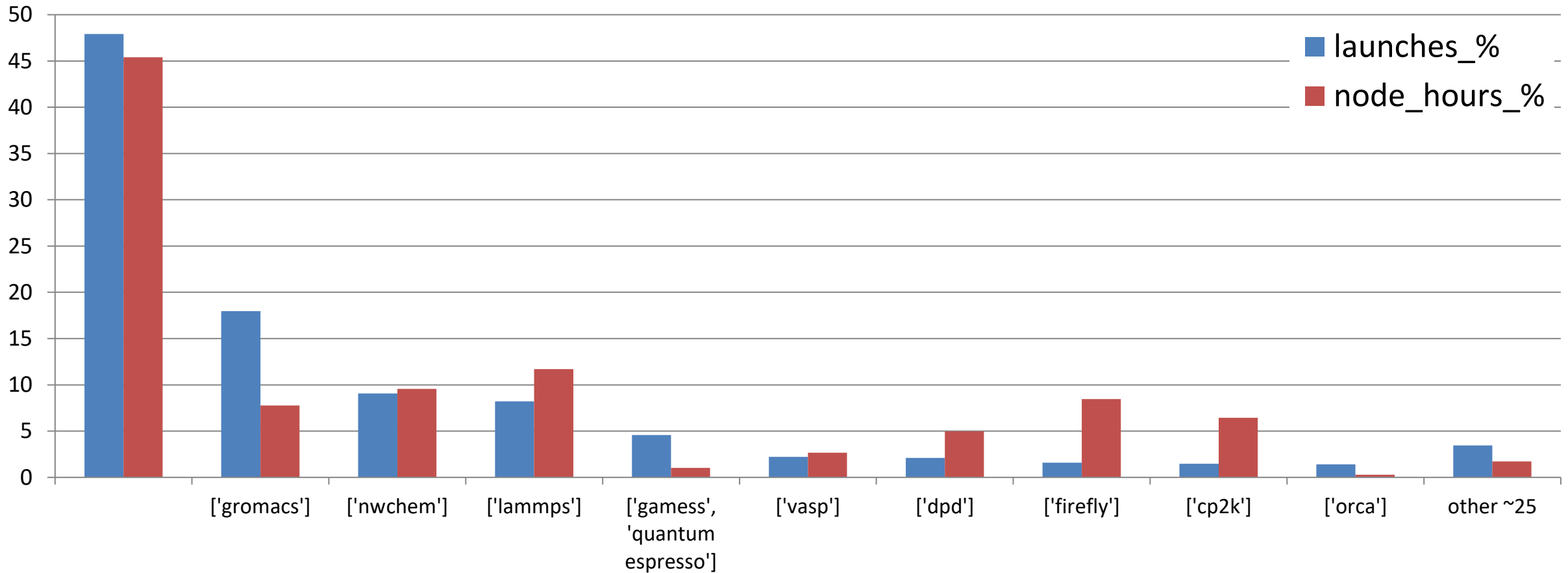
■ Mathematics, Computer Science, Mechanics

■ Earth Sciences



# The Share of Self-Written User SW and Packages

The share of "standard" application packages is >50%



# Operability and Productivity

## Monitoring and automated reactions needed

Idle system: **expensive**  
Inefficient load: **expensive**  
Late decisions: **expensive**  
**dangerous**

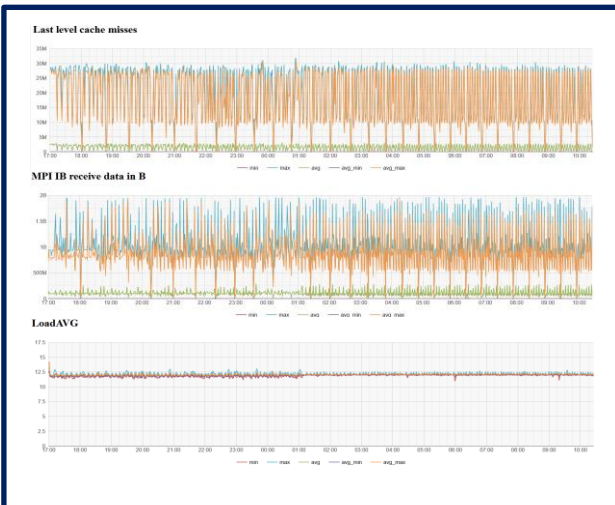


**It is imperative to permanently keep track on  
all components that influence efficiency of  
large-scale system output**

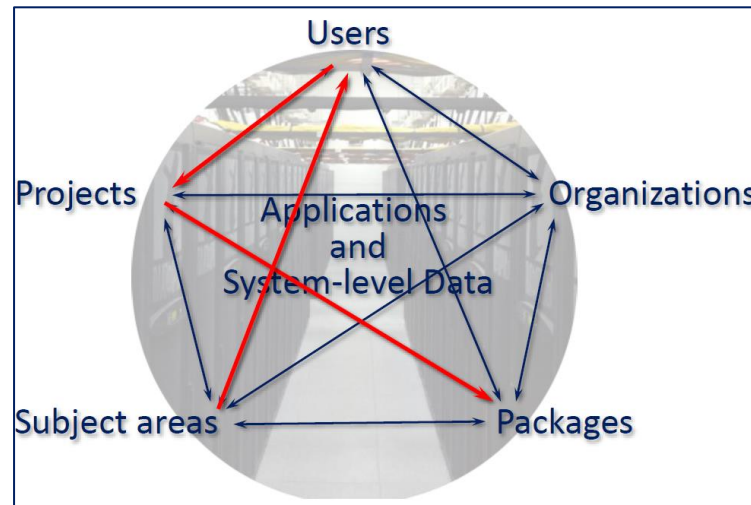
## Every component matters - total monitoring

- **Computing HW:** nodes, CPUs, memory stack, disks and storage hierarchy, networks, etc.
- **Infrastructure HW:** much more fault-tolerance critical. It includes cooling system: chillers, heat exchangers, air conditioners; piping, pumps; a set of components of the power system in conjunction with an uninterruptible power supply; fire safety systems and smoke removal; access control; etc.
- **Whole SW stack:** OS, package and license usage rates and peculiarities, etc.
- **Job scheduling** and queuing details from various points of view.
- **Job execution details:** dynamics and peculiarities of resource utilization by any and every executed user job.
- **Users.** Peculiarities of resource utilization, using available SW, cheating, etc.

# Operational Data Analytics (ODA) of Supercomputer Functioning



*Job dynamics during execution  
(**JobDigest**)*



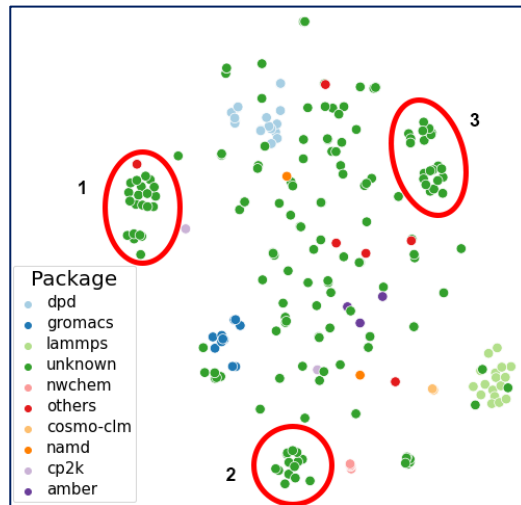
*“Show statistics for all packages used in materials science projects, then for each package breaks it down by individual users” (**TASC** endless analysis)*

| Description                                                                                | Supposition                                                                   | Recommendation                                                                                                              |
|--------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| The share of useful CPU computations is small.                                             | Not enough information to determine the reasons for this issue.               | Try performing the following type of more detailed analysis:<br>General program analysis<br>VTune Amplifier                 |
| The usage of I/O network presumably interferes with useful computations.                   | It is likely that the file system usage is not organized efficiently.         | Try performing the following type of more detailed analysis:<br>I/O analysis<br>VTune Amplifier                             |
| Processor are idle for a notable time while waiting for memory operations to be performed. | Memory usage is presumably organized inefficiently and requires optimization. | Try performing the following type of more detailed analysis:<br>Analysis of memory usage efficiency<br>Valgrind (Callgrind) |

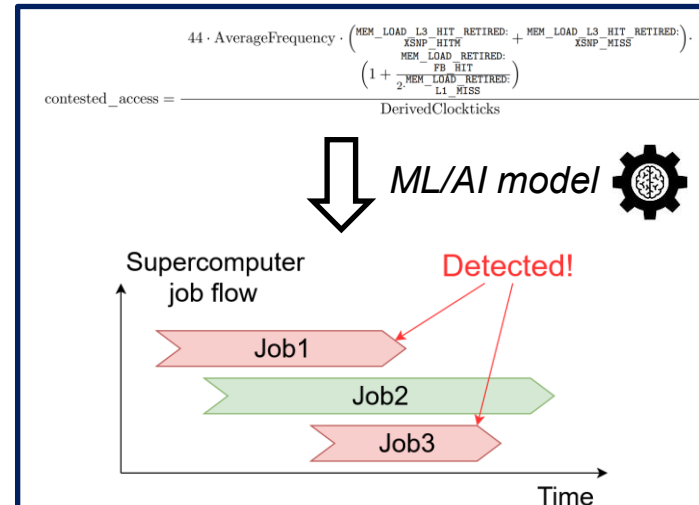
*Job issues and recommendations  
(**TASC** assessment system)*

# Operational Data Analytics (ODA) of Supercomputer Functioning

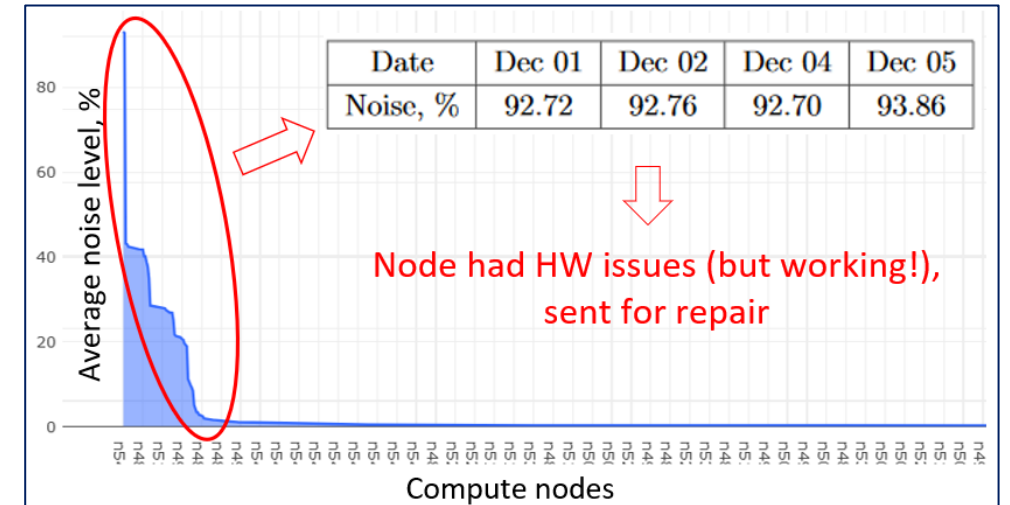
- A rich toolset for describing and analyzing supercomputer behavior
  - JobDigest, TASC (“endless” analysis system, assessment system, primary rules), noise evaluation, ML/AI-based similar job detection & detection of performance issues



*ML-based similar job detection*



*ML/AI-based automatic detection of performance issues*



*Noise evaluation*

- It helps to study major aspects of HPC system behavior

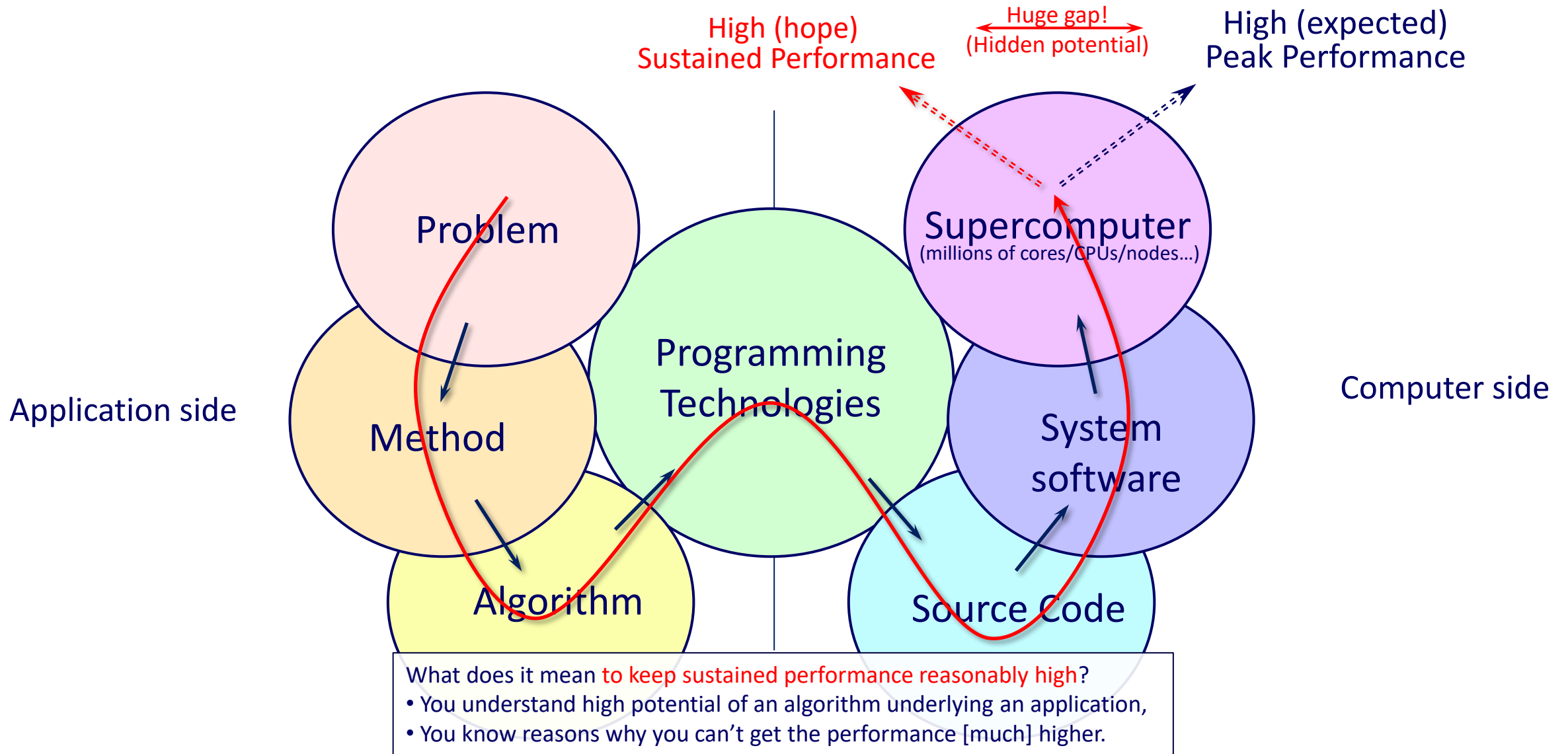
# Operational Data Analytics (ODA) of Supercomputer Functioning

Goal: Continuously collect, store and analyze data on performance and properties of supercomputer components in order to grant efficient HPC system productivity

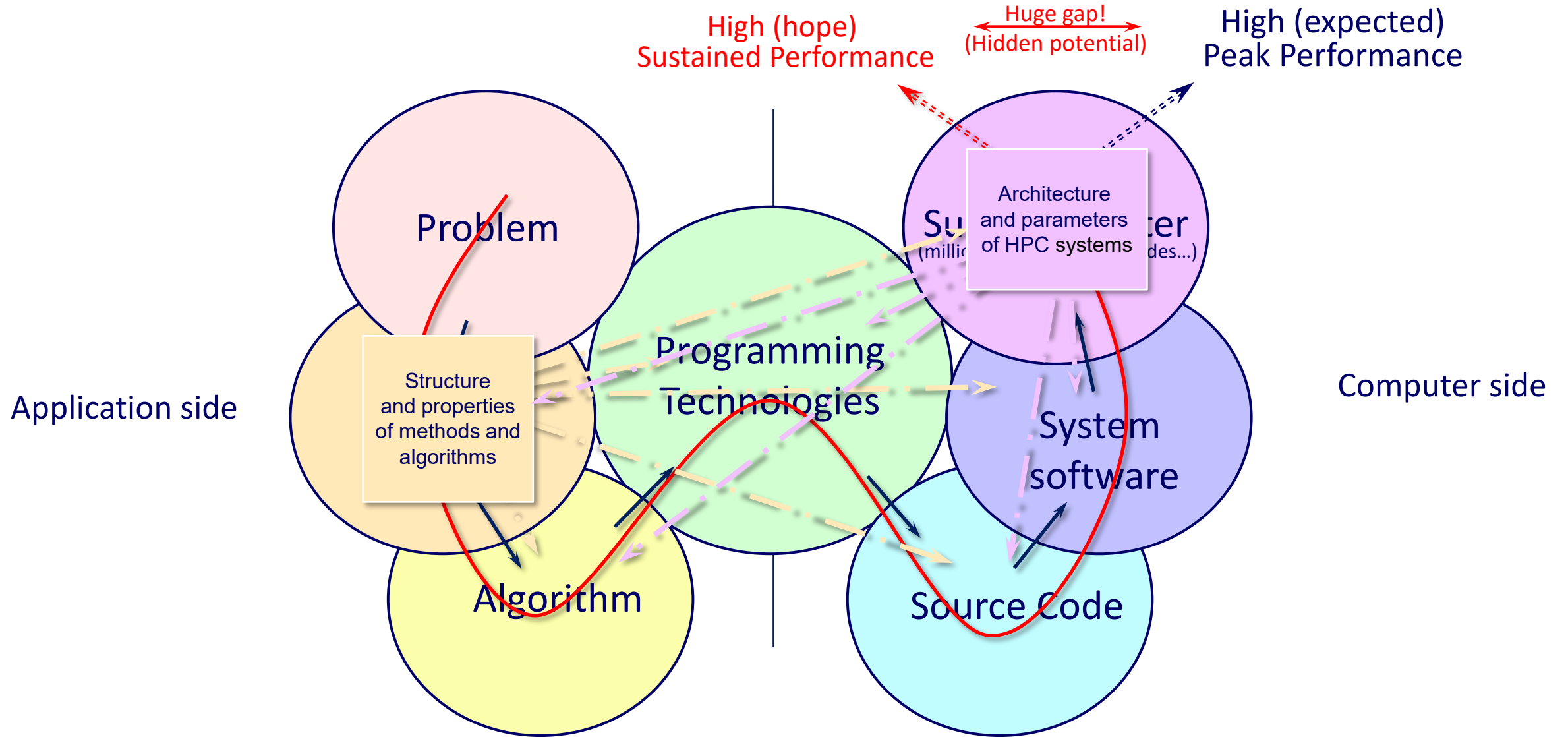
- Understand, what is going on and has been going on up to NOW
- Helps keep utilization of existing resources effective
- Highlight bottlenecks
- Sets motivation for development
  - Most valuable resource types
  - Rooflines of all kinds



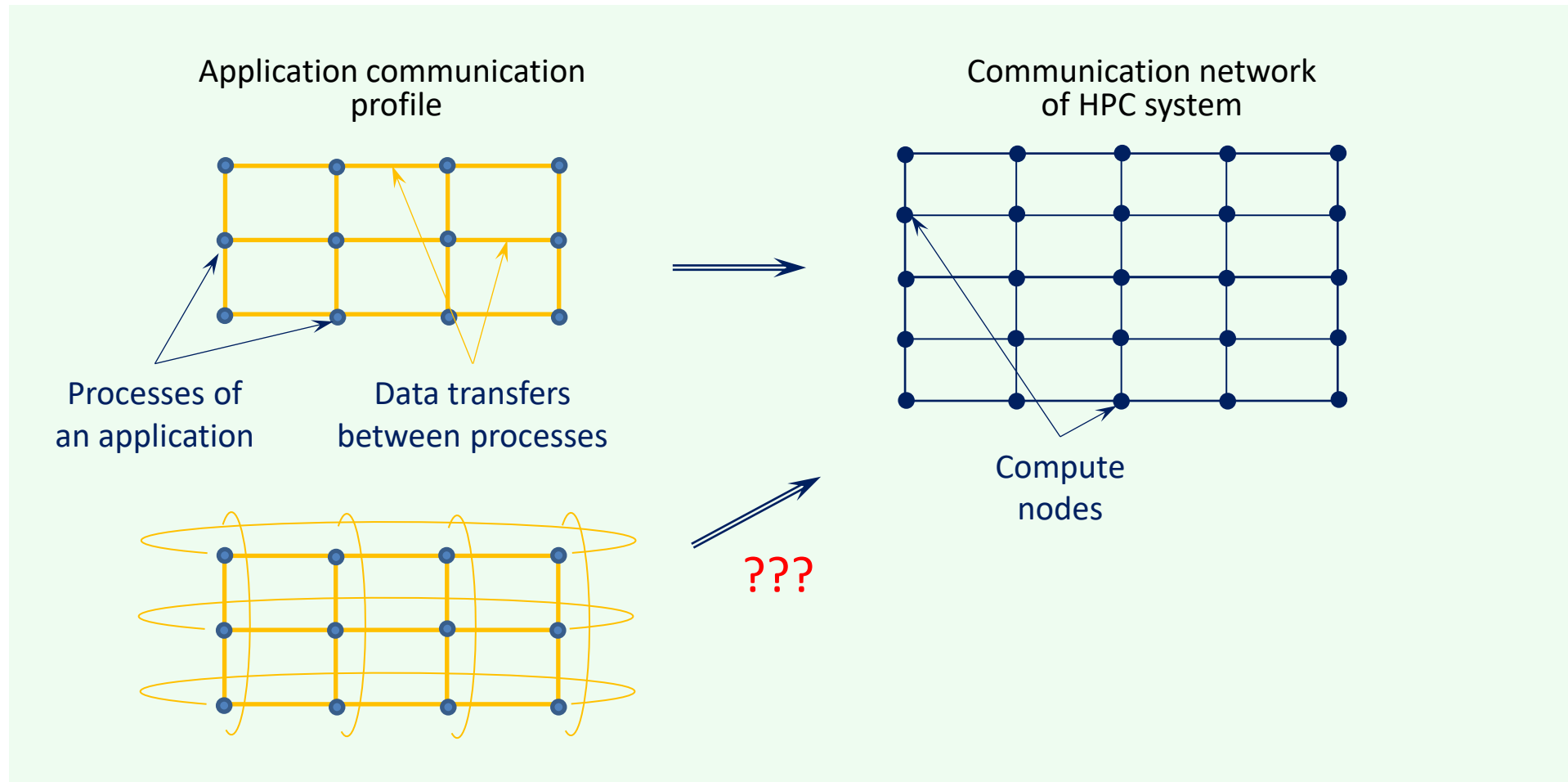
# Solving Problems using HPC Systems (Main Stages)



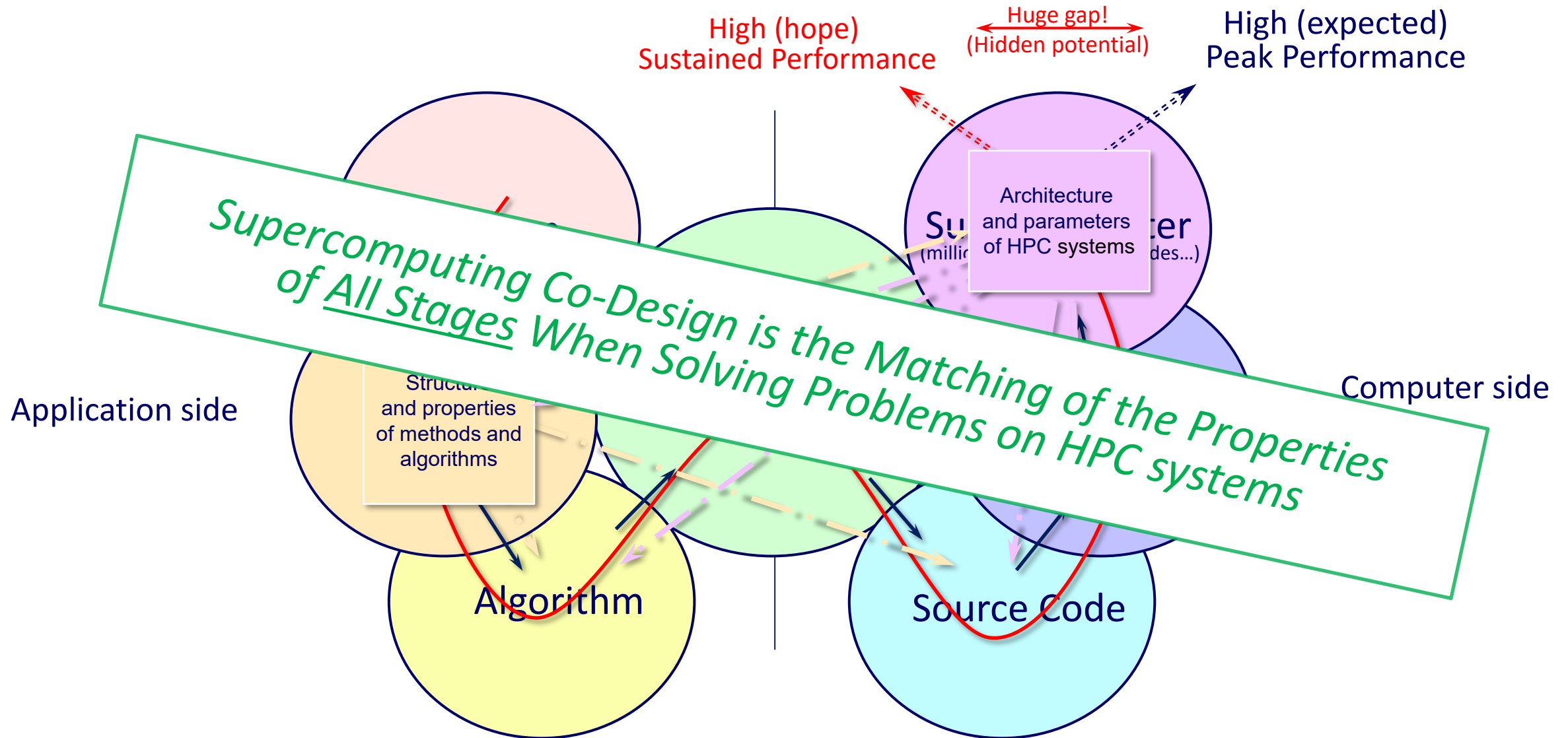
# Supercomputing Co-Design



# Supercomputing Co-Design – Matching the Properties



# Supercomputing Co-Design

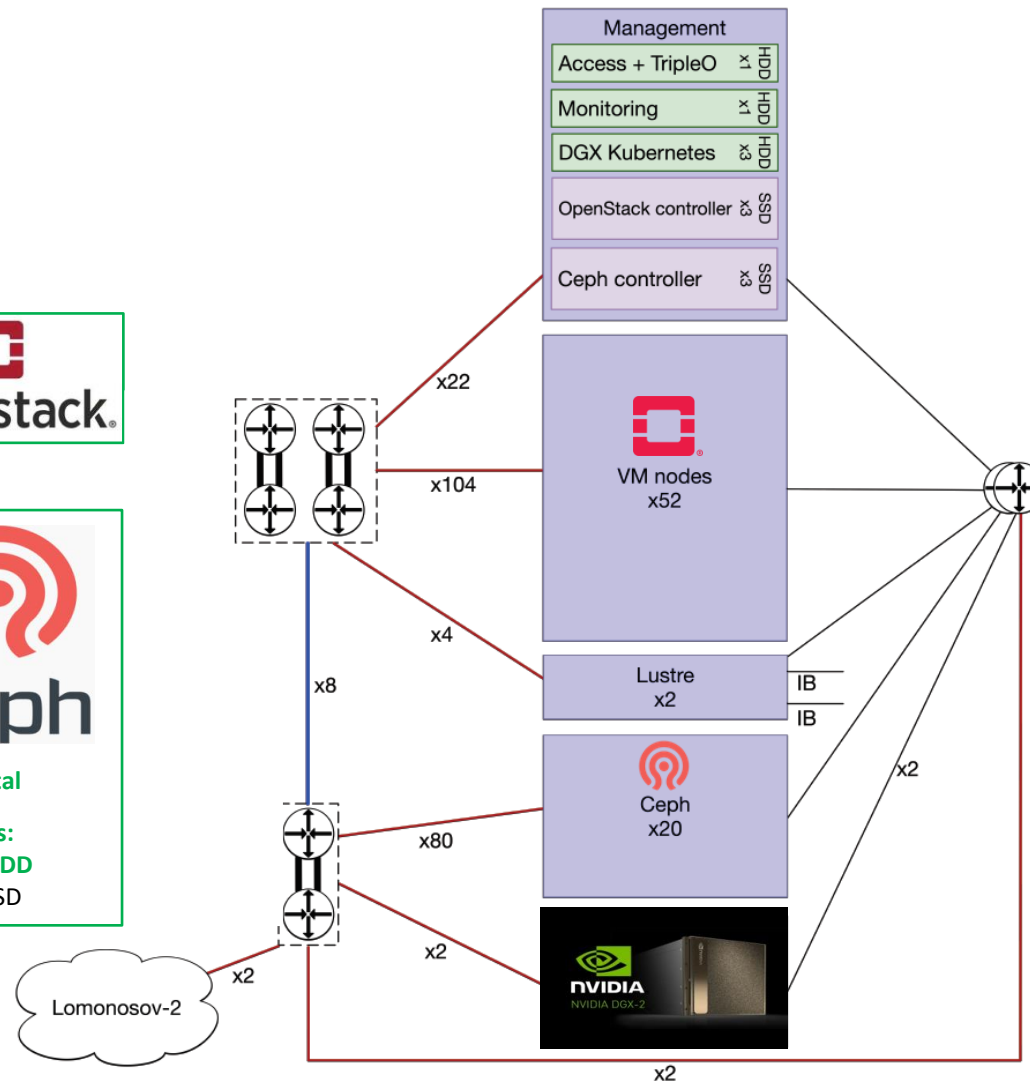
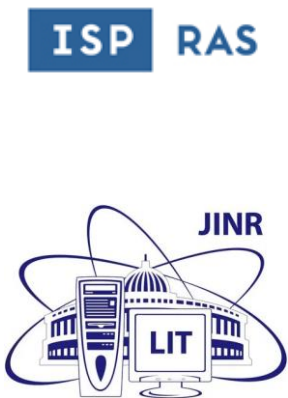


# Bringing Co-Design Principles to Life of HPC Center

- Basics
  - Single-node and HTC application should not compete for resources with HPC applications
  - CPU-only apps should not be run in GPU partitions in most cases
  - May be we could advise which partition of heterogeneous system should be used by user, taking into the consideration history of his runs (same project, etc.)?
- Large supercomputers (originally homogeneous by node types) can not be efficient for every application – there's always a trade off
- What are the possible ways of development of HPC facilities?



# Big Data Storage and Analysis Center – “ЦХАБД” (Lomonosov-2 extension, 2019-2021)

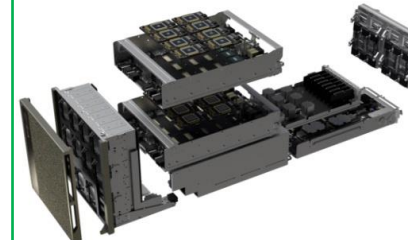


## VM Applications:

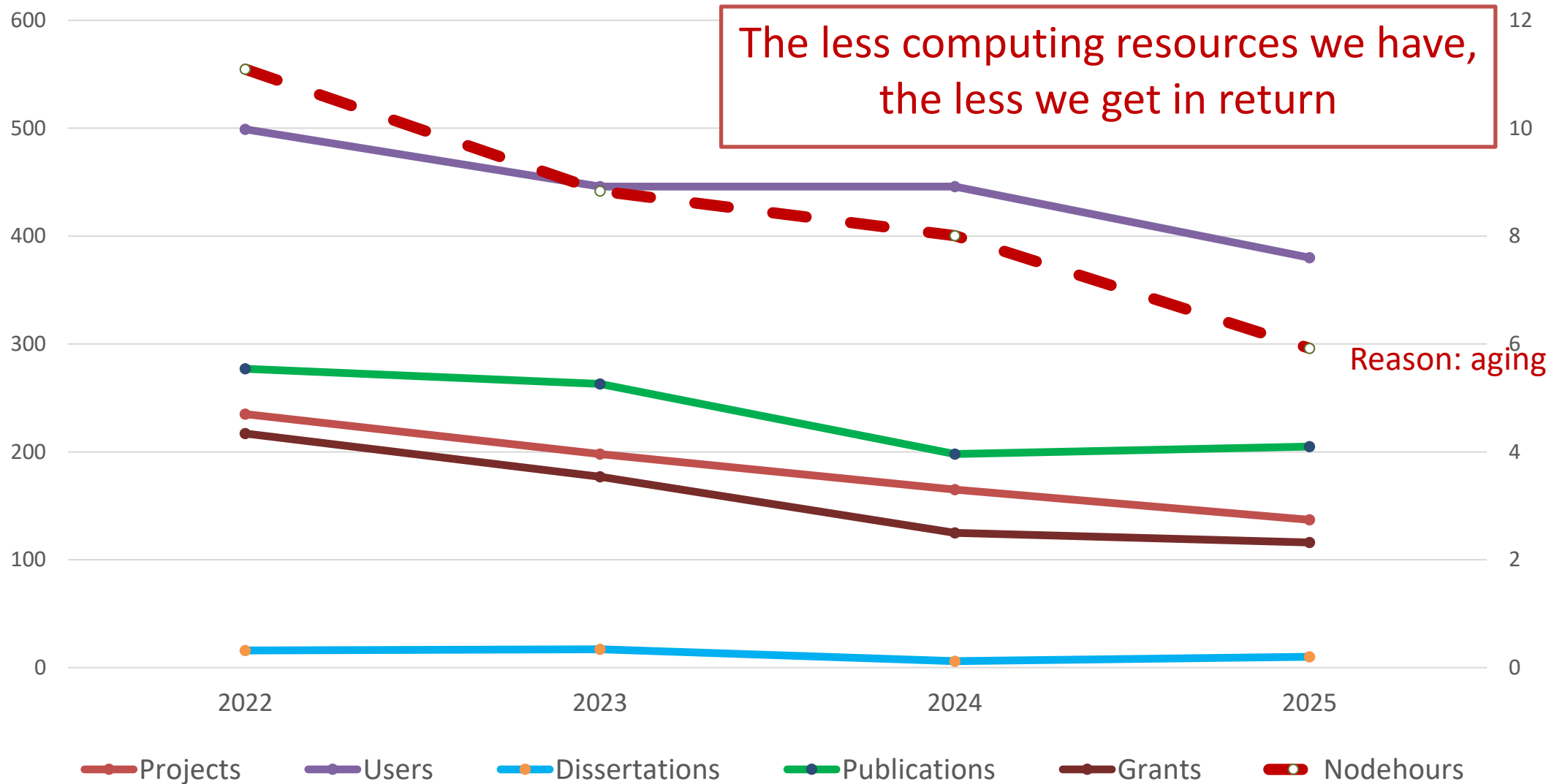
- World Ocean Database
- Soil and Geographical Database
- etc.

## NVIDIA DGX-2:

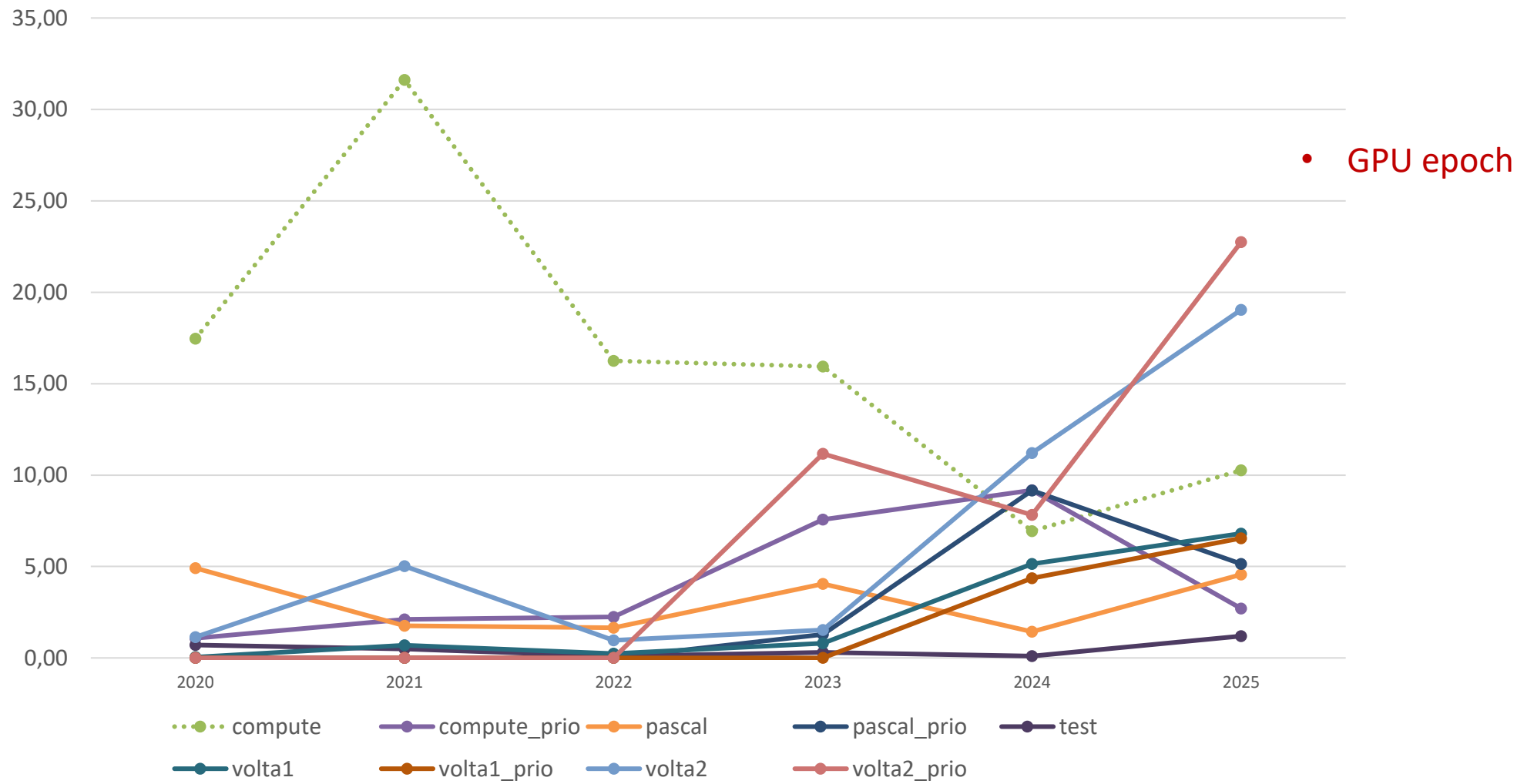
GPUs: 16x NVIDIA Tesla V100 32Gb  
Сеть: NWSwitch/NVLink  
Perf: 2 PFlops (tensor ops)



# Computing Resource-Driven KPIs



# Queue wait times



# What About Something New and Big?

- ~ 1 Eflop/s system:
  - ~ 20-50 MW
  - 1+ Trillion rubles (USD 10+ billion)
- x2, x3, x5 pricing for chips during 2025, and still growing
- Top GPUs are AI-oriented, not HPC

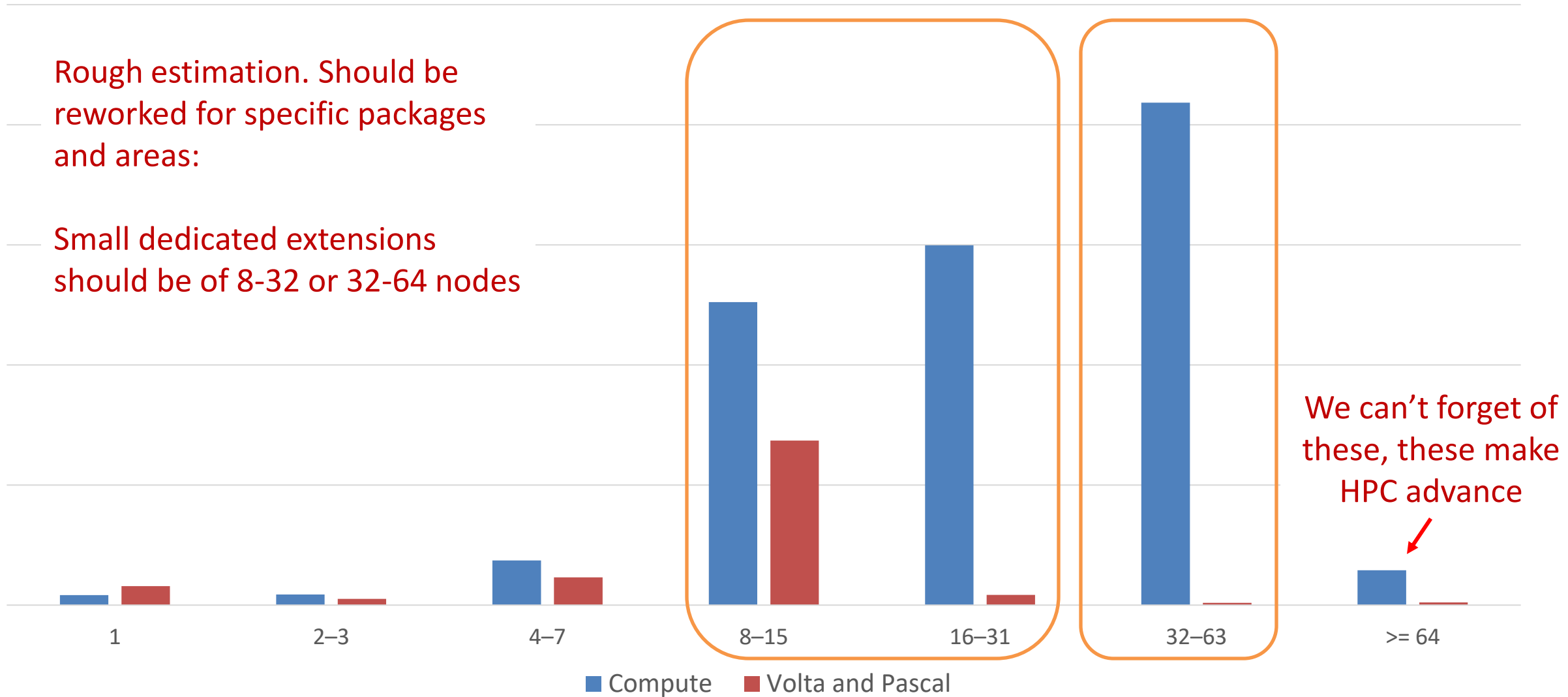
|             | H100 PCIe<br>vs A100 | H100 (H200) SXM<br>vs A100 | B200<br>vs H100 | B300<br>vs H100 |
|-------------|----------------------|----------------------------|-----------------|-----------------|
| FP64        | x2,7                 | x3,5                       | -40%            | -98%            |
| FP64 Tensor | x2,6                 | x3,4                       | -70%            | -99%            |
| FP32        | x2,6                 | x3,4                       | -40%            | -40%            |

- We are forced to think of diversification of computing resources, but it's not that bad really

# Common Application Scales

Rough estimation. Should be reworked for specific packages and areas:

Small dedicated extensions should be of 8-32 or 32-64 nodes



# 2025-2026 “Istok” system

“ISTOK” (rus) is the place where a water flow begins, and also the figurative beginning or source of something

2025 - Call for extending shared use centers by Ministry of Education and Science – WINNER!

Valuable support by MSU rector's office and key scientific groups of target subject areas:

**1) Adaptation to Climate Change, Conservation and Rational Use of Natural Resources:**

- Lomonosov Moscow State University (Research Computing Center, Faculty of Geography, Faculty of Physics);
- Marchuk Institute of Computational Mathematics, Russian Academy of Sciences;
- Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences;
- Water Problems Institute, Russian Academy of Sciences;
- Hydrometeorological Center of Russia.

**2) Preventive and Personalized Medicine, Ensuring Healthy Longevity:**

- Lomonosov Moscow State University (Faculty of Chemistry, A.N. Belozersky Research Institute of Physical Chemistry and Biology, Faculty of Bioinformatics, Faculty of Physics);
- V.N. Orekhovich Institute of Biomedical Chemistry, Russian Academy of Sciences;
- Federal Research Center of Biotechnology, Russian Academy of Sciences;
- V.I. Kulakov National Medical Research Center of Obstetrics, Gynecology, and Perinatology;
- Institute of Bioorganic Chemistry, Russian Academy of Sciences;
- Institute of Biochemical Physics, Russian Academy of Sciences.

**3) Data Analysis Technologies for Large-Scale Scientific Facilities:**

- Moscow State University Research Computing Center;
- Nuclear Physics Research Institute, Moscow State University;
- State Astronomical Institute;
- Faculty of Physics, Moscow State University;
- Faculty of Bioengineering and Bioinformatics, Moscow State University.



# 2025-2026 “Istok” system



## Compute: 22 nodes

2 x AMD EPYC 9534 64c  
768GB DDR5, SSD 240GB

## Access and management: 3 nodes

## Interconnect: Infiniband HDR 200 Gb/s, 25 Gb/s ETH

## Storage:

10 nodes: Baikal-S BE-S1000 ARM Cortex-A75, 48c  
384GB DDR4

12 x 15.36TB SSD, 1 x SSD960 GB  
8 nodes: Baikal-S BE-S1000 ARM Cortex-A75 , 48c  
384GB DDR4

24 x 8TB HDD, 1 x SSD 960GB  
3 nodes: Baikal-S BE-S1000 ARM Cortex-A75 , 48c  
384GB DDR4  
12 x 3.8TB HDD, 1 x SSD 960GB

## Storage:

- Ceph ~1,35PB
- TROK
- [ BeeGFS / Lustre ]



Initially dedicated system for the problems of:

- Adaptation to climate change, conservation and rational use of natural resources
- Preventive and personalized medicine, ensuring healthy longevity

# Challenges of Further Development

- Power, power, and power.
  - Stable, sufficient, quality power supply is vital. Not only for uptime.
- Hierarchy of data centers and HW location - TBD
  - Laboratory-level IaaS with dedicated HW to the project for its lifetime
  - Subject-area PaaS and/or SaaS
  - Unified computing facilities
- Theoretical research
  - New algorithms, architecture, matching them with real-life tasks
  - We need to support research in these areas
- Users and Staff
  - All of the above is useless until users and staff are professional enough – need education activities at all levels. Administrative/governmental support is vital.



Take care of your users!  
Take care of your staff!

Without them, all the supercomputers would be useless,  
expensive piles of hardware.



# JSFI: Annual Special Issue Series in Memory of Vladimir Voevodin



## SUPERCOMPUTING FRONTIERS AND INNOVATIONS

AN INTERNATIONAL OPEN ACCESS JOURNAL

**Founding Editor-in-Chief:** Vladimir Voevodin (1958–2026), Moscow State University, Russia

### Editors-in-Chief:

Jack Dongarra, University of Tennessee, Knoxville, USA

Dmitry Nikitenko, Moscow State University, Russia

Leonid Sokolinsky, South Ural State University, Chelyabinsk, Russia

### Editors-in-Chief Foreword

**Editorial Director:** Mikhail Zymbler, South Ural State University, Russia

### Editorial Board

**Production:** South Ural State University (Chelyabinsk, Russia)

**ISSN:** 2313-8734 (online), 2409-6008 (print)

**Publication Frequency:** 4 issues (print and online) per year

**Current Issue:** Volume 13, Number 2 (2026)

**Abstracting and Indexing:** Scopus, SJR, OJS, etc.

**Annual Special Issue Series in Memory of Vladimir Voevodin**

**Most Viewed** **Most Downloaded** **Most Cited**

| # | Article title                                                                                                                              |
|---|--------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | <a href="#">Toward Exascale Resilience: 2014 update</a><br>Franck Cappello, Al Geist, William Gropp                                        |
| 2 | <a href="#">Predicting the Energy and Power Consumption of Data Centers</a><br>Hayk Shoukourian, Torsten Wilde, Axel                       |
| 3 | <a href="#">Extreme Big Data (EBD): Next Generation Data Analytics</a><br>Hitoshi Sato, Michihiro Koibuchi, Satoshi Masanori Kakuta et al. |
| 4 | <a href="#">Towards a performance portable architecture for climate models</a><br>Thomas Christoph Schulthess, Oliver P. Cummings et al.   |
| 5 | <a href="#">Exascale Storage Systems – An Analytical Study</a><br>Julian Martin Kunkel, Michael Kuhn, Th                                   |



### Editors-in-Chief Foreword

Parallel scientific computing has entered a new era. Multicore processors on desktop computers make parallel computing a fundamental skill required by all computer scientists. High-end systems have surpassed the Petaflop barrier, and significant efforts are devoted to the development of the next generation of hardware and software technologies towards Exascale systems. This is an exciting time for computing as we begin the journey on the road to exascale computing. 'Going to the exascale' will mean radical changes in computing architecture, software, and algorithms – basically, vastly increasing the levels of parallelism to the point of billions of threads working in tandem – which will force radical changes in how hardware is designed and how we go about solving problems. There are many computational and technical challenges ahead that must be overcome. The challenges are great, different than the current set of challenges, and exciting research problems await us.

This journal, Supercomputing Frontiers and Innovations, gives an introduction to the area of innovative supercomputing technologies, prospective architectures, scalable and highly parallel algorithms, languages, data analytics, issues related to computational co-design, and cross-cutting HPC issues as well as papers on supercomputing education and massively parallel computing applications in science and industry.

This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. We hope you find this journal timely, interesting, and informative. We welcome your contributions, suggestions, and improvements to this new journal. Please join us in making this exciting new venture a success. We hope you will find Supercomputing Frontiers and Innovations an ideal venue for the publication of your team's next exciting results.



Jack Dongarra  
University of Tennessee



Vladimir Voevodin  
(1958–2026)  
Moscow State University

Dear colleagues, dear authors and readers,

The launch of this journal, dedicated to the world of high-performance computing, is thanks to Vladimir Voevodin. Much has already been said, and much more will be said, about this outstanding individual, and many of those who were fortunate enough to work with him continue his legacy with pride, guided both by the direction set by Vladimir Voevodin and by the highest standards for the quality of results.

For its part, the editorial board is convinced that the best way to honor his memory within the framework of this journal is to organize a special series of annual issues featuring the best papers from various scientific conferences in Russia and the BRICS countries in the fields of HPC, parallel programming, and data analysis, to which the final issue of the journal each year will be dedicated.

We wholeheartedly invite organizers of relevant conferences to recommend our journal for their best papers and encourage authors to participate in major conferences, receive recommendations, and publish in our journal.

It is no secret that Vladimir excelled at bringing professionals together even in interdisciplinary projects, organizing applied and fundamental research, and facilitating the exchange of expertise through conferences, seminars, and other educational activities. That is precisely why we believe that by bringing together the best papers from relevant conferences, we are, in a sense, continuing the work of our dear teacher and colleague.

*The Journal of Supercomputing Frontiers and Innovations*  
*Editorial Board*

<https://superfri.org>

Russian Supercomputing Days, Moscow State University  
September 28, 2026