EXPLANATION OF PROCUREMENT DOCUMENTS No. 2

IDENTIFICATION INFORMATION OF THE CONTRACTING AUTHORITY

I.

Name of the Contracting Authority	VSB – Technical University Ostrava
Registered office	17. listopadu 2172/15, 708 00 Ostrava-Poruba, CZ
Corporate ID	61989100
Person authorised to act on behalf	prof. RNDr. Václav Snášel, CSc. – Rector
Contact person	Ing. Jan Juřena, e-mail jan.jurena@vsb.cz
Profile of the contracting authority	https://zakazky.vsb.cz/

(hereinafter "VSB – TUO")

II.

Name of the Contracting Authority	The European High-Performance Computing Joint Undertaking
Registered office	12, Rue Guillaume J. Croll, L-1882 Luxembourg, LUX

(hereinafter "EUROHPC JU")

(VSB – TUO JU hereinafter jointly as the "Contracting Authority")

VSB – TUO is "the lead contracting authority" and the only contact point between the Contracting Authorities and economic operators for the purposes of the procurement.

IDENTIFICATION INFORMATION OF THE PROCUREMENT PROCEDURE

Public contract	**EURO_IT4I Supercomputer**
TED reference number of the procurement	Z2019-021999
File number	9600/2019/01
Type of public contract	supplies

On 15 and 18 May 2020 the Contracting Authority received questions regarding the procurement documents delivered by an economic operator. Therefore, according to Section 98 (3) of Act No. 134/2016 Coll., on Public Procurement, as amended (hereinafter the "Act") the Contracting Authority provides the explanation of procurement documents below.

The acquisition and operation of the EuroHPC supercomputer is funded jointly by the EuroHPC Joint Undertaking, through the European Union's Connecting Europe Facility and the Horizon 2020 research and innovation programme, as well as Czech Republic.
The supercomputer is also supported by the project „IT4Innovations national supercomputing center - path to exascale" (CZ.02.1.01/0.0/0.0/16_013/0001791), funded by European Structural and Investment Funds as well as Czech Republic.

Introductory information

First of all, the Contracting Authority would like to inform the economic operators that there are certain objective restrictions that currently prevent alteration or supplementation of the procurement documents. However, this fact does not mean that the Contracting Authority would like to resign upon the proper answering of questions regarding the procurement documents raised by economic operators.

Even if these iterations do not lead to the successful performance of this public contract, the contracting authority may use them in the event of the need for a repeated procurement procedure.

Question No. 1

There were some changes in the market from the previous round of PMC of EURO_IT4I procurement.

Primarily Intel updated CPU roadmap for dual-socket servers. The current Intel CPU options related to EURO_IT4I procurement are:

- current Intel Xeon SP refresh (Cascade Lake SP), dual-socket – does not fulfill SPEC_48 (memory speed), limited number of cores for optimal Linpack performance,
- current Intel Xeon AP (Cascade Lake AP), dual-socket – does not fulfill SPEC_48 (memory speed) and SPEC_245 (replacement of CPUs),
- coming Intel Xeon (Cooper Lake), 4 -socket – does not fulfill SPEC_48 (two CPU sockets per node) and SPEC_340 (limited support of DLC systems),
- coming Intel Xeon (Ice Lake), dual-socket – fulfill all requirements, estimated product availability is outside of the current expected timeline.

In summary, there is no reasonable Intel processor option for EURO_IT4I procurement based on current RFP requirements and timeline. The only option might be Ice Lake. The most critical for Intel Ice Lake processors proposal seems to be time schedule. Based on provided procurement schedule the contract award is planned to 07/2020. As described in Annex 2 of binding draft of contract document the milestone for "Successful performing of Acceptance Tests of the part of the System" (D+5 months) is therefore not realistic for Intel Ice Lake processors while D+8 months (the whole cluster delivery) would make big change towards to Intel Ice Lake proposal.

Question: Would it be possible to reflect the recent market changes in the RFP documentation? e.g. postpone contract award date, adjust list of Compute partitions for initial hardware delivery milestones, adjust timeline for partial milestones related to hardware delivery or remove penalties for partial milestones to allow Intel Ice Lake processors in the proposal?

Contracting's Authority response to Question. No. 1

The Contracting Authority has repeatedly dealt with this issue during the preparation of the award criteria. Based on the conclusions of the preliminary market consultation as well as taking into account the Contracting Authority's preferences regarding the preferred technologies, the solution set out in the procurement documents was determined. The contracting authority intends to use the most appropriate and best possible technology available on the market for its computing system, although it may not be available from the manufacturer, which is mentioned by the inquirer.

At the moment, the Contracting Authority will not change the related technical specifications. Given the specific subject matter of this public contract, it is possible that direct competition will not be maintained across all parts of the solution, but the contracting authority has made every effort to maintain, in particular, competition between suppliers of accelerated and universal partitions. Likewise, the contracting authority does not want or cannot, with regard to the rules of the subsidy provider, to extend the delivery date of the subject of performance at the moment, even with regard to the fact that the launch date of the mentioned Ice Lake CPUs has been repeatedly postponed by the manufacturer. In addition to the technical specifications set in the procurement documents, the

contracting authority is also bound by the fulfilment of the schedule for putting the new computing system into operation.

Question No. 2

There is SPEC_105 in the Technical requirements specification for EURO_IT4I system in SCRATCH Storage section:

SPEC_105 SCRATCH storage must be flash based. All data (including metadata) must be stored on SSD or NVMe disks. The disks must be suitable for their designation and expected load.

There are several flash (NAND) technologies available in the market. We would like to understand projected five years write / turnover rate of the SCRATCH storage. It may hugely influence selection of suitable flash technology.

Question: Would it be possible to explain in more details what does "expected load" mean? What is projected average write / turnover rate of SCRATCH storage per day?

Contracting's Authority response to Question. No. 2

As the Contracting Authority does not run similar computing system, it does not possess the information about the "expected load" of SCRATCH storage. The contracting authority is also of the opinion that this value cannot be objectively determined.

In fact, the SPEC_105 means that the economic operator itself should choose the proper SSD or NVMe disks according to its own technical solution and regarding the technical requirements and warranty conditions set in the procurement documents, especially in the Annex 3: Business Terms and Conditions – Binding Draft Contract. That is why the Contracting Authority did not set specific requirements on the SSD/NVMe disks mentioned in SPEC_105.

Question No. 3

In the RFP documentation, the following SPEC is given:

SPEC_61 The Data analytics node must provide fast memory access; the latency of remote NUMA node memory access must not be greater than six times the latency of local NUMA node memory access, idle latencies are considered. The memory latency will be evaluated using Intel Memory Latency Checker or similar tool.

In Tenderer's opinion, the required memory latency parameter given in this SPEC is a remnant from the previously considered 8-socket systems requirements and was not updated accordingly when the requirement was changed to a single 32-socket server.

In Tenderer's opinion, this required value is not possible to meet with 32 socket system and we kindly ask Contracting Authority to change latency to higher, realistic number or remove this request or accept solution based on two 16 socket servers? Could the Contracting Authority kindly clarify this requirement?

Contracting's Authority response to Question. No. 3

The ratio between the latency of remote NUMA node memory access and the latency of local NUMA node memory access, as it is set in SPEC_61 is based on the information acquired during the preliminary market consultation. In this case, the contracting authority raised a question for the participants in the market consultation, which was subsequently answered and this answer became an objective source of information stated in the SPEC_61 finally.

Currently the Contracting Authority does not hold any other objective data that the requirement cannot be met. The proposed amendment would constitute a fundamental revision of the procurement documents and thus cannot be implemented in the current state of the tender procedure. Therefore, the Contracting Authority will not alter the procurement documents.

Question No. 4

SPEC_235 - question for Ethernet management interfaces:

In SPEC_235 there is requirement that Ethernet management interfaces of all active network devices and of nodes described in SPEC_279 must be connected to the contracting authority's OOB network implemented by the contracting authority's OOB switch. However, SPEC_279 talks about Scheduler on at least two infrastructure nodes. Does this mean that BMC/IPMI ports of these two servers should be connected to OOB network of contracting authority's OOB switch (not connected to the system LAN)? If yes, could there be an exception (not needed to connect the nodes' BMC/IPMI to OOB switch) if High-availability for other services on these nodes will be used with fencing enabled? So that will prevent the need to have special route from contracting authority's OOB switch into LAN for these two nodes (or in worse case to disable fencing completely)?

Contracting's Authority response to Question. No. 4

In the question, the inquirer assumes that SPEC_279 talks about Scheduler on at least two Infrastructure nodes; however, this assumption is not correct.

SPEC_279 The functionalities described in SPEC_276 to SPEC_278, ***and tools for Scheduler management*** *must be collectively available on at least two Infrastructure nodes (nodes intended for management).*

The requirement SPEC_279 states that some functionalities must be collectively available on at least two Infrastructure nodes and the required functionalities include "tools for Scheduler management". So it is required that "tools for Scheduler management" must be available on the given nodes, but it does not mean that Scheduler itself must be installed and run on the given nodes. For better understanding, by "tools for Scheduler management" commands like qstat, qsub, pbsnodes, qmgr, qdel, qsig, qrls, qhold, etc. are meant; these commands are included in the application package named pbs-client. The package called pbs-server is not required for the purpose.

The requirement SPEC_235 states that *"Ethernet management interfaces of all active network devices and of nodes described in SPEC_279 must be connected to the contracting authority's OOB network implemented by the contracting authority's OOB switch".* For SPEC_235, nodes described in SPEC_279 (nodes intended for management) can be connected to OOB network either using BMC/IPMI Ethernet interface (providing access to baseboard management controller/specialized service processor of node) or using standard Ethernet interface intended for node management (providing access to main processor/operating system of node). Both options are considered applicable for the purpose.

It is also applicable to connect BMC/IPMI interfaces to LAN network and route them to OOB network, as suggested by the inquirer.

Nodes referred in SPEC_279 (*nodes intended for management)* must be connected to OOB network according to SPEC_235, no exception is allowed.

Question No. 5

The contracting authority requires the delivery of a system where the minimum performance Rmax of individual partitions is determined within SPEC_47, SPEC_49, SPEC_57 and SPEC_63. The sum of the minimum required performances of the individual partitions is 6,606 PFLOPS.

At the same time the Contracting Authority requires the total minimum performance defined in SPEC_37 as 8.6 PFLOPS, i.e. higher than the mentioned sum of the minimum performance of the individual partitions.

We see in these two requirements inconsistency and confusion, which makes the input non-transparent, as both require de facto two different values for the same parameter.

In addition, the SPEC_37 requirement is also unsatisfactory in terms of compliance with the price of Work, as the global pandemic COVID-19 has led to significant price increases due to limited producer

capacity and rising prices of primary components (RAM, CPU, etc.) and significant exchange rate movements, when the initial calculations were made many months ago at a significantly lower exchange rate.

Based on the above, the tenderer asks the contracting authority to remove the SPEC_37 condition, which in our opinion is superfluous, confusing, non-transparent and de facto makes the current situation economically unattainable, which significantly threatens the implementation of this public contract as whole.

However, if the contracting authority continues to insist on SPEC_37, it is very likely that the tenderer (and this will probably not only be our case) will not be able to submit a qualified and, above all, feasible proposal.

We are convinced that even without the requirement for SPEC-37, the contracting authority can achieve the purpose and the desired result and benefit of the public contract, as the evaluation of bids is conceived in such a way that the bid with the highest offered overall performance wins and thus ensures for the contracting authority that completely sufficient and above all possible performance exceeding 6,606 PFLOPS.

Another option how to get better performance of the whole system is to allow usage of 4 GPU nodes because the 4GPU node has better HPL efficiency.

Contracting's Authority response to Question. No. 5

The aim of the Contracting Authority's response is, in particular, to dispel any, albeit presumed, ambiguities or misunderstandings which may exist in relation to the requirements concerning the minimum level of computing performance.

The sum of the minimum required performances of the individual partitions 6.606 PFLOPS determined within SPEC_47, SPEC_49, SPEC_57 and SPEC_63 does not represent the same value as it is set in the SPEC_37, i.e. 8.6 PFLOPS. These are two different values.

While SPEC_37 represents the minimum value for the computing power of the entire cluster as a whole, the aforementioned SPEC_47, SPEC-49, SPEC_57 and SPEC_63 set the minimums for individual system partitions and leave it up to economic operators to decide which "performance mix" they choose to meet the overall computing power requirement (i.e. 8.6 PFLOPS) they offer. The contracting authority does not see anything non-transparent or confusing in this procedure. The aim is again to ensure the price/performance optimization of the computing system, respectively its technical solution, by the economic operators.

It is also worth noting that the minimum overall performance requirement of the computer system as presented in the SPEC_37 has been set by the subsidy provider mandatorily. Hence the Contracting Authority is not allowed to dismiss it during the public procurement procedure. Therefore, in case of SPEC_37, no further alteration of the procurement documents will be made.

Question No. 6

We would like to ask the Contracting Authority a clarification question related to SPEC_68 section D, where the Enhanced Hypercube network topology is specified.

There is stated that the ratio of connectivity-to-the-network to connectivity-to-endpoints is greater than or equal to 2.2. Could you please confirm, that the ratio is intended for Universal Partition compute nodes connectivity only, because otherwise, it could not be possible to meet the number of hypercube dimensions which is required to be less than or equal to six.

Contracting's Authority response to Question. No. 6

No, SPEC_68 section D is **not** intended for Universal Partition compute nodes connectivity only. It applies to all compute partitions that are part of the computing system.

The purpose of SPEC_68 was to determine the conditions for individual types of the network so that the offered solutions would be comparable in technical level and level of performance.

On the basis of Question No. 6, the contracting authority performed a verifying calculation using known data from the preliminary market consultations and did not come to the conclusion that the Enhanced Hypercube (EHC) topology, even with the prescribed constraints, is not applicable to achieve the minimum overall performance of the computer system specified in SPEC_37, i.e. to achieve 8.6 PFLOPS. Given the fact that the economic operator didn't submit any material proving the contracting authority otherwise it is still assumed that the specification is correctly defining and admitting a solution based on the EHC topology.

As to exclude any doubt or misunderstanding of the contracting authority's intent, additional information is provided to the definition of the EHC topology parameters in SPEC_68, section D:

The Compute network topology is Enhanced Hypercube. ***Each switch in the network provides the same connectivity (the number of links and their distribution and throughput) to the network.*** *Switch connectivity is spread to all dimensions of the network. The number of links to each dimension is the same number or differs by one (i.e. for any two dimensions of the network the maximum difference of link count is one); links of the same speed are used; in the case of link count difference, lower dimensions have greater number of links than higher dimensions. The number of hypercube dimensions is less than or equal to six.* ***For each switch, the ratio of connectivity-to-the-network to connectivity-to-endpoints is greater than or equal to 2.2.*** *For the ratio calculation, if provided connectivity of endpoint is higher than connectivity requested in technical specification, the lower value can be used (e.g. if 100Gb/s connection to Compute network is requested and 200Gb/s speed is provided, then 100Gb/s can be used for ratio calculation).*

In both statements (emphasized in bold) the term "to the network" describes the links of each switch to the other switches constituting the compute network. That means links from the switches towards the endpoints are not considered as part of the "to the network" connections. Links between the switches are considered as uplinks (connectivity-to-the-network). Links used to connect endpoints to the switches are considered downlinks (connectivity-to-endpoints).

The statement *"Each switch in the network provides the same connectivity (the number of links and their distribution and throughput) to the network."* sets requirements for the uplinks only, while it does not set any requirements for the downlinks.

The statement "*For each switch, the ratio of connectivity-to-the-network to connectivity-to-endpoints is greater than or equal to 2.2.*" sets the minimum of the ratio between the uplinks and downlinks.

The economic operator is allowed to further optimize (differentiate) the amount of the downlinks used for the endpoints in the different parts of the compute network while having in all cases the ration between the up/down links greater or equal to 2.2.

As an example it's fully acceptable to have different amount of downlinks for the compute nodes in the Universal partition as in the Cloud partition if in both cases the ratio of up/downlinks is greater or equal to 2.2.

Therefore, in case of SPEC_68 section D., no further alteration of the procurement documents will be made.

In Ostrava

prof. RNDr. Václav Snášel, CSc.

Digitally signed by prof. RNDr. Václav Snášel, CSc.
Date: 2020.05.20 17:08:19 +02'00'

---

prof. RNDr. Václav Snášel, CSc.
Rector