# Příloha č. 3 ZD – Annex 3 of the Tender Documentation (courtesy translation – for information only)

# Specification of technical and other requirements

Within this Specification of Requirements, points of specification are categorised by the notations M or I on the beginning of the pertinent paragraph:

**M** denotes MANDATORY requirements which must be met in order that the tendered solution can be accepted by CHMI. CHMI will not consider a tendered solution that fails to meet a mandatory specification requirement (marked **M**) unless the Tenderer offers valid reasons why the feature in question is either unnecessary or irrelevant for the tendered solution.

**I** denotes requests for INFORMATION. A response must be given to all such requests. Requests for information (marked **I**) are intended to prompt a description of the construction, philosophy or operation of tendered hardware, software, or services in areas which are regarded as being of particular importance.

Tenderer documentation, manuals and diagrams must be written in English or Czech. In case both language versions of the same text exist in the tender it must be defined and clearly marked which version is considered official and which is just for information.

*Tenderers should note that, where relevant, when responding to points of specification that are met, tenderers must give sufficient detail to explain the way in which requirement is met – a simple expression, such as “compliant” or “agreed” , will not normally suffice.*

**Overall requirements**

1. (M) The overall System must serve CHMI’s needs for at least 5 years from its initial installation. The Tenderer must commit to deliver, maintain and support the tendered system for the whole period of the Agreement. The subject of the Agreement will be the hardware and software licence rights ownership conversion to CHMI, their delivery and installation and their support as required by this Call for Tenders. The software will include both operation systems for all System components and other software tools explicitly listed in the tender requirements.
2. (M) The System will consist off
- hardware and software of the high-performance computing server (HPCS),
- hardware and software of two access (login) and two auxiliary servers,
- hardware and software of the shared disk file system (SDFS),
- uninterruptible power supply devices (UPS),

- components of the cooling system (see „Integration to CHMI infrastructure“)
- other auxiliary facilities and items (network switches, cables, racks etc.) necessary for proper installation and operation of all System components.

1. (I) The Tenderer must give a full and detailed description of the architecture of all components of the tendered System as well as its technical characteristics.
2. (M) The Tenderer must adjust to the evolving computing performance requirements and infrastructural limitations described in this documentation. In order to meet the requirements some components of the System may be delivered in two phases, the Phase A and the Phase B.
3. (M) The System Phase A must be installed in time to facilitate the start of the acceptance tests for the System at latest until four months after signature of the Contract. The installation and operation of the System Phase A must not affect parallel operations of the current CHMI computing system.
4. (I) The Tenderer must describe in detail how the upgrade from the Phase A to the Phase B will be performed, including the time schedule.
5. (M) The Tenderer must perform the upgrade of the System from the Phase A to the Phase B in such a way that the operational Works of CHMI running on the System are not interrupted for more than 6 continuous hours.
6. (M)Any other repairs and scheduled replacements of the System must be performed in such a way as to minimise the effect on the normal continuity of computer service at CHMI. If downtime is unavoidable it must be scheduled in such a way that the System service is not interrupted by more than 6 continuous hours during any 48 hour period

**High-Performance Computing Server (HPCS)**

1. (M) A set of benchmarks is provided with this Call for tenders, the specifications of which are given in Annex 5 of the tender documentation. The results of the benchmark runs must be supplied in electronic form to CHMI with the tender together with all changes made to the code, used scripts and also a comprehensive description of the system on which the benchmarks were run. The results must include listings of the tests and other outputs (stdout, stderr, logging information) produced during the running of the benchmarks, except binary files, see Annex 5. Only the set of runs “ASIS” results shall be used to judge the compliance to the mandatory requirements SPEC\_12, SPEC\_16, SPEC\_39 and SPEC\_40.
2. (M) The benchmarks should be run on one system only which is as close as possible to the tendered System.
3. (M) The Tenderer must include with the tender full details of the analytical reasoning that leads from benchmark performance measurements to the performance commitments provided in response to the requirement SPEC\_12 and SPEC\_16. Absolute binary code used in the tests must be generated by compilers and libraries provided with the tendered System.
4. (M) The Tenderer must commit to benchmark execution times and Acceptance dates for Phase A and Phase B. Tests involve running the forecast task ALADIN (performance test MORGANE), ALADIN post-processing task (test FULLPOS), ALADIN Memory test and the test of the operational priority (SWITCHOVER test) on the High-Performance Computing Server (HPCS). To meet the required performances only the Computing nodes of the System (see SPEC\_17) may be used.
5. The System Phase A must be able to run one copy of the test MORGANE within less than 1000 seconds of the wall-clock time not later than 1 month after the start of the acceptance tests.
6. The System Phase A must be able to run 4 concurrent copies of the test MORGANE within less than 3300 seconds of the wall-clock time not later than 1 month after the start of the acceptance tests.
7. The System Phase B must be able to run one copy of the test MORGANE within less than 500 seconds of the wall-clock time not later than 1 month after the start of the acceptance tests.
8. The Tenderer reports the wall-clock time needed to execute 8 concurrent copies of the test MORGANE on the HPCS of Phase B; this performance test should be completed in as short as possible wall-clock time.
9. (M) The Tenderer must successfully complete all benchmark tests as they are described in Annex 5. CHMI will not consider a tendered solution which fails to meet any of these benchmark tests.
10. (M) After the Acceptance Tests are passed the HPCS must maintain such a stability that the mean time between two HPCS major failures within 30 consecutive days exceeds 170 hours. Major failure is considered to be such one that has a serious impact on the operations and/or on the use of the System. The impact on the operations and/or on the use is evaluated by CHMI in function of serious technical and operational production reasons. In all circumstances the System performance below 95% is considered as major failure.
11. (M) HPCS should be of a sufficient quality of design and built in order to achieve the total operational availability of the whole system of at least 99.5% computed as MTBF/(MTBF+MTR)\*100 where MTBF is the mean time between failures (in hours) and MTR is the mean time of repair (in hours). Therefore the total accumulated unavailability time of HPCS should not exceed 44 hours during any period of 365 days including hardware and software maintenance sessions. The Tenderer will describe which hardware failures are critical for the HPCS operations and if and how in case of a critical failure it would be possible to operate HPCS in at least reduced mode (decreased number of processors, limited memory etc.).
12. (M) The System Phase B must have sufficient internal memory to execute concurrently 20 copies of the test MORGANE. The wall-clock time taken to run memory verification test must be less or equal to 20/8 times the one needed to run concurrently 8 copies of the test MORGANE. At the same time the wall-clock time must be less than 3300 seconds. The Tenderer reports the maximal number of copies of the MORGANE test that can concurrently run on the Computing nodes of the HPCS (see SPEC\_17) without error messages linked to the memory allocation.
13. (M) The HPCS nodes are divided to two categories:
	1. Computing Nodes,
	2. Service Nodes which are other than Computing Nodes.

It is required that all Computing Nodes are exactly identical and that they must use the same operational system. It is required that a Computing Node should have at least 64 GB RAM, directly accessible by one copy of the operational system. It is required that one application task can be executed on all Computational Nodes at the same time. HPCS Phase A and B must have each at least two login (access) nodes which would belong to the category of the Service Nodes.

1. (M) The operating system of HPCS must be based on UNIX or Linux. Operating system interfaces and file-systems must be generally compliant with relevant POSIX and/or X/OPEN standards.
2. (M) HPCS shall provide the IEEE arithmetics.
3. (M) HPCS operational system must support AT&T Korn shell.
4. (M) An accounting system enabling CHMI to account accurately for the use of main system resources of the HPCS (processors, memory, etc.) on a per user basis is required.
5. (M) HPCS operation system should maintain a comprehensive set of log files. Items to be logged should include:
- hardware malfunctions;
- system demon exception conditions;
- all events that could have a bearing on system security;
- use of secure accounts (root, system administrator, etc.);
- history of all user batch jobs;
- operator actions.
6. (M) HPCS must support the following TCP/IP services: ssh­­2 and scp/sftp (client and server), ftp (client and server), rpc and smtp.
7. (M) It is required that the clock of all the System components can be synchronized with the reference time-server outside the System using the NTP protocol.
8. (M) Diagnostic and analytical tools such as ping, traceroute and tcpdump must be at disposal on the System. The Tenderer reports whether other tools of this type are part of the software delivery.
9. (M) An alarm facility that automatically alerts system operators about potential intruders is highly desirable. A facility to restrict access from remote hosts to individual services as a function of the originating host address is highly desirable. The Tenderer describes the way in which CHMI shall be informed on the system software security risks such as mail services.
10. (M) The System must support a control mechanism of the access to files in function of the ownership. The standard Unix scheme „owner/group/world“ is considered as adequate. The System must use NIS mapping and it must support LDAP to verify and control the access.
11. (M) HPCS has to be seen as one entity by the end user. To meet this requirement, the Tenderer has to provide a tool for synchronizing user accounts and passwords of all the HPCS components for which it is needed.
12. (M) The tendered HPCS must be capable of receiving and executing batch jobs submitted from other hosts, at least from the Access and Auxiliary Servers (SPEC\_50 a SPEC\_51). HPCS must be capable of providing real time status information about these jobs to user, and of returning the output from the jobs to the machines where the jobs were originated from.
Necessary software licences must be the part of the System and the software must be installed and maintained on the Access and Auxiliary Servers. The Tenderer must provide a list of operating systems and hardware systems which client software capable of submitting batch jobs to HPCS is available for.
13. (M) The tendered System must be capable of supporting the shared operational and research use of HPCS. It is required that operational jobs whenever they arrive be given immediate and unhindered access to all resources needed to complete these jobs in minimal time, while still allowing research workload access to any residual capacity. It is requested that this requirement is not met by aborting work in progress. The Tenderer must describe how this requirement can be met on the tendered HPCS.
14. (M) The following job scheduling facilities should be provided:
- a site-definable priority scheme, whereby the process responsible for selecting jobs for execution gives higher priority jobs more favourable treatment;
- a site-definable job class system, allowing the workload to be controlled. Such control should be capable of preventing a “flooding” of the System by a single user or user group and ensuring a fair access to resources for all users;
- a site-tunable resource scheduling facility, capable of ensuring efficient and full use of the system;
- a resource control facility, enabling system administrators to allocate to users shares of the computer resources;
- facilities for tuning of all scheduling parameters in real time without interruption of normal service.
15. (I) The Tenderer must describe what algorithms are used for the allocation of system resources in a situation where jobs waiting for execution and/or jobs already in execution are requesting more resources than are available.
16. (M) The batch processing system should be designed so that as far as possible, no intervention by operations staff is required to keep it operational at a high level of efficiency in normal circumstances.
17. (M) Job visibility and control must be available to allow the system operator to perform the following tasks:
- determine the status of all jobs in the System;
- determine the status of the batch queuing system;
- determine the detailed status of an individual job;
- terminate a job, rerun, execute and divert jobs.
18. (M) System performance visibility and control must allow the system operator to:
- determine in a simple and straightforward manner that the system is functioning correctly;
- determine the various aspects of performance of the System such as CPU utilisation, memory utilisation, I/O activity, disk space status etc.;
- identify bottleneck in the system;

- monitor hardware performance.

1. (M) HPCS must inform the operator of error conditions in order that the operator can take appropriate action. At the same time the System must warn the operator of any critical shortage of system resources.
2. (M) An effective implementation of the FORTRAN 77, FORTRAN 90, FORTRAN2003, C and C++ programming languages on the tendered HPCS is required. The tender must include licence for unlimited number of compiler users on the HPCS. A complete installation of the so-called profiler facility must be available on the HPCS so that parts of the Fortran and/or C code prone to an optimization may be detected. The so-called “trace-back” information when a process gets abnormally terminated must be supported by the System.
3. (I) The Tenderer shall provide a description of the main compiler flags which are available for the compilers provided with the tendered System and in particular those that have been used when compiling the benchmark code to complete the performance tests.
4. (M) Bit reproducible results must be obtained when the blocking length of the inner loops of the model code is changed (NPROMA variable).
5. (M) Bit reproducible results must be obtained when the number of processors used by the benchmark tests MORGANE is changed.
6. (M) If the processors of the tendered HPCS are of a vector type it is required that a cross-compilation software for the languages listed in the requirement SPEC\_37 is installed and maintained on the Auxiliary Servers. The licence for unlimited number of users of the cross-compilation system is then requested.
7. (M) Tendered HPCS must support MPI (*message passing*) libraries, namely MPI-1, and MPI-2 with a full support of MPI-IO and of a dynamical control of processes.
8. (M) Tendered HPCS must support current OpenMP standard, at least OpenMP 2.5, for Fortran, C and C++ programming languages listed in the requirement SPEC\_37.
9. (M) An interactive source level debugger must be fully installed on tendered HPCS that is capable of giving application developers access to information necessary for quick and effective debugging of a single application written in Fortran or C and executing over multiple nodes, launched either from a command line and/or as a batch task.
10. (M) The Tenderer must specify a list of manuals and documentation to the System. These should also be available in English.
11. (M) An online set of manuals in HTML or PDF formats is required including the permission to publish it on CHMI internal web server.
12. (M) The Tenderer will commit to provision of the support for migration of applications, ALADIN operational suite and other utilities from the current system to the tendered System.
After the System Acceptance Tenderer’s specialists will be ready to work with CHMI on:
- installation and tuning of the main ALADIN applications for HPCS (current version of ALADIN code for FULLPOS and MORGANE configurations, ODB library and corresponding data assimilation tools, important libraries and further applications written in Fortran and C specified by CHMI;
- installation of the client programs of Scheduler Monitor Supervisor and/or ecFLOW (©ECMWF);
- configuration of the batch processing system on HPCS and its tuning for the shared operations of development and production (high-priority) tasks;
- demonstration of system tools for monitoring of system resources on HPCS;
- providing assistance to the building of new operational suite (spanning HPCS and the Auxiliary Servers), in particular the optimisation of job submission.
Phase B migration support will start by the analysis of the ALADIN operational suite running on the System Phase A with the aim to identify tasks using the System inefficiently and to propose optimisation of those tasks.
At present, the production software of the model ALADIN is run on the computer system composed by the computing server NEC SX-9 and the linux server Express5800/140Rf-4.
The migration support is requested in a necessary range, at most 3 person-months of labour of experienced specialists skilled in Fortran and C applications and system analysts, preferably in the CHMI headquarters. The Tenderer will propose the organisation of works for both Phase A and B in order to complete the migration within 2 months After the acceptance of the Phase A and 1 month after the Acceptance of the Phase B.
13. (M) In order to facilitate migration and optimization of ALADIN production software to the tendered HPCS, the Tenderer is requested to offer at least 70 hours of support of a specialist in Fortran and C applications. This support might be used by CHMI for porting and optimisation of new versions of the model and auxiliary libraries in connection with the migration on the System Phase A and B.
14. (M) The Tenderer must provide a demonstration programme for HPCS application programmers stating the detailed content of this programme as a part of the Acceptance tests.

**Access (login) Servers**

1. (M) The System will include two Access (login) linux Servers (nodes), each of them having the following minimal parameters:
* 2x CPU 20 cores, frequency 2.20 GHz or better;
* 64 GB RAM;
* internal SSD RAID, 2 TB usable capacity;
* Dual Gigabit Ethernet card;
* adapter and software to connect to the Shared disk file system (SPEC\_53)
* Linux operational system (Red Hat, Suse, …);
* software for job submission to the HPCS batch queuing system;
* Compiler and corresponding run time libraries for the FORTRAN, C a C++

**Auxiliary Servers**

1. (M) The System will include two Auxiliary Servers. Each of them shall have the same parameters like the Access Servers except:
* 256 GB RAM;
1. (M) Auxiliary Servers shall be included in the System running the batch tasks, in the sense of a client, i.e. it shall be possible to submit from them batch tasks to all HPCS queues, and also in the sense of the server, i.e. there shall be a queue (or queue group) to process batch tasks on the Auxiliary Servers.

**Shared Disk File System (SDFS)**

1. (M) The System shall include necessary hardware and software for providing storage capacity in the Shared Disk File System (SDFS). The total usable capacity of SDFS shall be at least 200 TB in RAID-5 or RAID-6 on the System Phase A and 800 TB in RAID-5 or RAID-6 on the System Phase B. RAID-5 variant must have minimal one spare disk. SDFS must be fully accessible from HPCS and from all Access and Auxiliary Servers.
2. (M) The data stored on SDFS shall be backed up to the CHMI Legato backup server. The Tenderer is requested to indicate which part of the System will be suitable for running the Legato Networker back-up client and to deliver and install the necessary client software.

**System Service Requirements**

1. (I) The Tenderer must state his release policy for system level software. In particular, the Tenderer must state his commitment to support a previous software version once a new version of that software has been released.
2. (M) The Tenderer must commit to provide the software and hardware service to all components of the tendered System Phase A and Phase B for the period of at least 5 years from the acceptance of the System Phase A. All costs related to the service for the period of 5 years must be included in the bid price.
3. (M) During the hardware and software service period an on-call service support must be available 24 hours per day, 7 days per week. The maximum response time may be 4 hours on-call, next business day on-site (*4 hours on-call, next business day on-site*).
4. (I) The Tenderer must describe in detail how the repairs will be organized as well as the standard response time and remedy time. Further, he will state which experience the subjects of the proposed organization have with repairs of hardware and software of the type and use similar to the proposed System, including the support provision to the programmers of demanding scientific applications written in the Fortran and C languages.
5. (M) The Tenderer must undertake to provide, or to arrange provision of, repairs, spare parts and support for the System hardware and software for the whole duration of the Agreement.
6. (I) The Tenderer is requested to give details of the amount of time required for scheduled preventive sessions to check the correct functionality of the System, and the frequency of such sessions required to maintain correct functioning of the hardware. The impact upon on normal functioning of the System of such sessions should also be clearly stated.

**Integration into CHMI infrastructure**

1. (I) The Tenderer must provide firm details of physical configuration of all items of the System, size and weight of its physical subcomponents, including average and point floor loads.
2. (I) The Tenderer must provide firm details of HPCS electric supply requirements in the full operations when running the most demanding ALADIN performance tests described in SPEC\_12 and SPEC\_16 (including the magnitude of oscillations), power-up and shut-down characteristics as well as the distribution over phases.
3. (I) The Tenderer must provide details of the environmental requirements of HPCS as well as cooling requirements with the Tender.
4. (M) The total electric consumption of all components of the System must not exceed 160 kVA in either Phase A or B. This consumption will be measured for the System running the most demanding ALADIN performance tests described in SPEC\_12 and SPEC\_16. The Tenderer shall provide the breakdown of this total electric consumption for all sub-components of the System for Phase A and Phase B in kVA. The total electric incoming power measured on the input of the System UPS during the Acceptance tests must not exceed the values provided in the Tender.
5. (M) The System must be equipped by Uninterruptible Power Supply (UPS) devices which allow the System to maintain normal operations in case of a loss of external electric energy supply. UPS in Phases A and B must supply the whole System normal operations for at least 10 minutes.
6. (M) The spare capacity of the CHMI Diesel Electrical Aggregate (DEA) available for the System Phase A and B is 160 kVA.
7. (M) Due to limitations of the computer room regarding the possibility of cooling the System by air, the use of water cooling technologies may become necessary. The System Phase A may use the air cooling up to 30kW of released heat only – for that the Tenderer is allowed to install a temporary cooling device up to the time of installing the System Phase B (this concerns the room A of the computer hall, see Annex 4 of this Tender Documentation). The System Phase B may use at maximum the current air cooling capacity of the computer hall (room B; see Annex 4 of this Tender Documentation).
8. (M) If needed to cool the System by a water based technology, the Tenderer delivers and installs the necessary components of that technology. Preparations made by CHMI are described in Annex 4 of this Tender Documentation.
9. (M) Access and Auxiliary Servers will be connected to CHMI computer network via Gigabit Ethernet.
10. (M) CHMI requires that HPCS and Auxiliary Servers provide network card fault-tolerance facility like IP Network Multi-pathing (IPMP) or similar facility.
11. (M) Two network switches Catalyst WS-C3560G-24TS-E will be delivered with the System.
12. (M) All network cables will be delivered with the System. The length of cables between the System and the switch will be 15 meters.
13. (M) The System components must fit into available space in the Computer Room indicated in Annex 4 of this Tender Documentation.