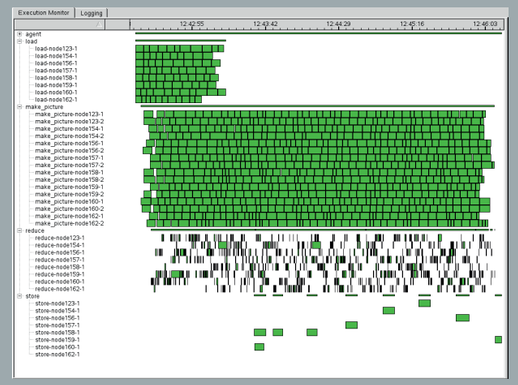
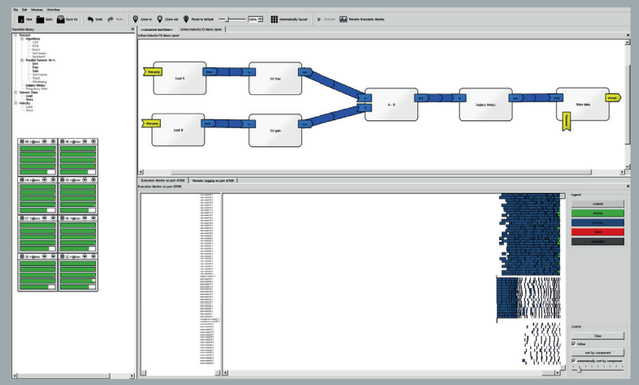


1



2

GPI-Space development:

1 *GPI-Space runtime system monitor*

2 *GPI-Space integrated domain specific graphical development and execution*

Getting Started

As mentioned, GPI-Space separates domain expert and parallelization expert knowledge, allowing each specialist to focus on the field of his expertise.

In a first step, the parallelization experts will – together with the domain expert – build up an understanding of the specific requirements of the given domain and define basic components for the user library, such as parallel I/O. They also configure GPI-Space with the specific parallelization patterns for the data type used, and add data management and improved fault tolerance for ultra-scale machines.

For some domains, like seismic, these steps have already been performed and GPI-Space is ready to be used from day one. Ask us, which domains we have already covered.

For all these steps. Fraunhofer experts will be on your side. We guide you through the process of deploying GPI-Space into your organization and make sure, the transition to the new approach is as smooth as possible.

Once these steps have been executed, the domain user can start expanding the library with his own components.

Users typically deal with very complex algorithms with various dependencies. In a first step, a logical decomposition of this algorithm into independent modules is performed. Those modules will then become part of the user library.

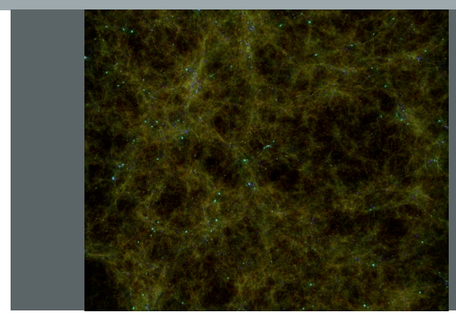
As part of the user library, they can be used to recreate the same algorithm as a workflow which can then be executed in parallel by GPI-Space without further effort.

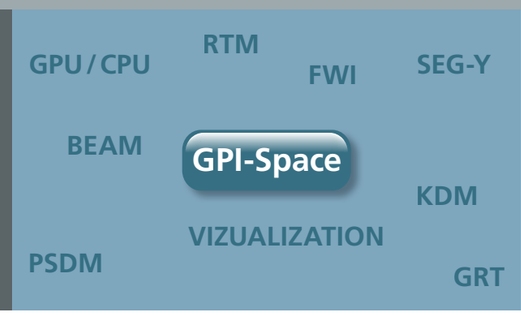
“Software is getting slower more rapidly than hardware becomes faster”
Niklaus Wirth (1995)

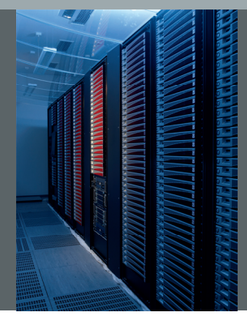
hpc_flyer_GPI-Space-ISC2015_EN

Fraunhofer
ITWM

FRAUNHOFER INSTITUTE FOR INDUSTRIAL MATHEMATICS







GPI-SPACE

GPI-Space is Fraunhofer’s integrated solution to solve big problems on ultra-scale machines. GPI-Space is used for interactive seismic imaging, for reverse time migration (RTM), for financial services, for sorting huge amounts of data, and for live streaming data analysis. GPI-Space is a general purpose system that can be specialized to your domain by adding knowledge about your data, its parallelization patterns, and how to handle it.

GPI-Space makes your ideas work

In GPI-Space, the user can easily express arbitrary parallel workflow patterns, including cyclic dependencies. Workflows are formally verified and targets for a large number of well-founded optimizations. It’s way more than just Map & Reduce!

Coupling never was easier

The application independent virtual memory layer of GPI-Space is based on GPI, Fraunhofer’s scalable, fast and efficient PGAS layer. Data stored in the virtual memory can be accessed by arbitrary applications and hardware resources like CPUs, NVIDIA GPUs, and Intel XEON Phi. It’s much more than just Spark!

GPI-Space will do the job for you

Besides dynamic load balancing and fault tolerant execution on ultra-scale machines, GPI-Space allows to execute arbitrary legacy codes. There is no need, to learn parallelization, the automatic parallelization in GPI will do it.

Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM

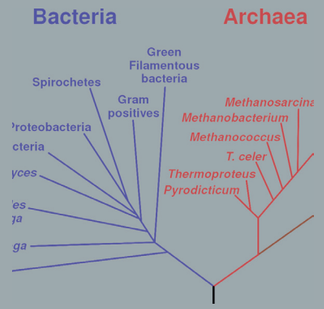
Fraunhofer-Platz 1
67663 Kaiserslautern
Germany

Contact
Dr. Franz-Josef Pfreundt
Phone +49 631 31600-4459
pfreundt@itwm.fraunhofer.de

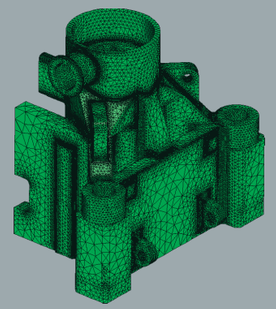
Dr. Mirko Rahn
Phone +49 631 31600-4553
rahn@itwm.fraunhofer.de

www.itwm.fraunhofer.de
www.gpi-space.com

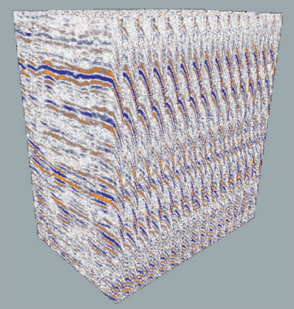
1



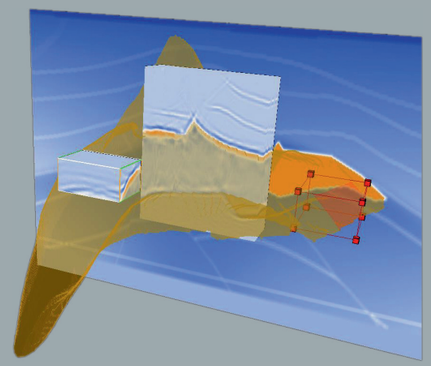
2



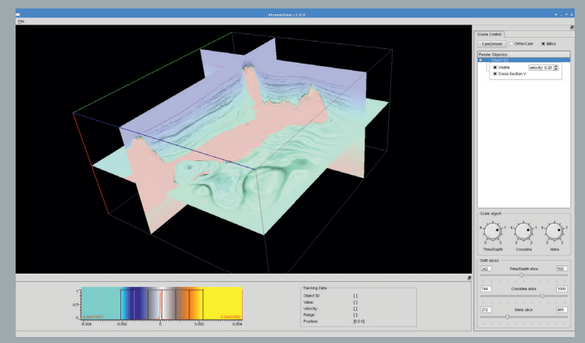
3



4



1



2

GPI-Space use cases:

1 *Financial service: portfolio evaluation*

2 *Life science: phylogenetic tree analysis*

3 *Engineering: structural optimization*

4 *Seismic imaging workflows and migration*

Bridge the gap!

The speed-up of today's compute systems is mostly generated by the massive use of parallelism which is also essential for the ability to solve large scale problems in the first place. This parallelism and factors that come along with it, increase the complexity of creating efficient applications for those systems to a level, a domain user can no longer master. Thus Wirth's Law will prevail if an approach does not bridge this gap.

Programmers are facing an ever increasing gap between the need to focus on their domain and create solutions, and the increasing challenge to make this solution efficient in a parallel environment.

GPI-Space bridges this gap between domain user and HPC expert and takes the parallelism burden off the domain user.

Sources of complexity are manifold. Some factors that contribute to a high level of complexity are:

- parallel systems with a large number of cores per compute node,
- distributed systems with large numbers of nodes,
- interconnectivity between the nodes grows with $O(n^2)$ in worst cases,
- uncertainty, i. e. the more components in a system, the more likely resource failures become,
- heterogeneity of components, such as CPUs, GPUs, hardware accelerators etc., and reliability requires redundancy and separate levels of logic.

Application fields

GPI-Space is a great solution for all classic large scale HPC problems and is already widely used in various industries to speed up calculation time by efficiently using the available resources, and to make the process more failure tolerant so a system error does not lead to a complete loss of weeks or months worth of processing time.

Companies from the oil and gas industry are using GPI-Space for speeding up seismic imaging algorithms, and scaling them to higher resolutions in order to find oil and gas deposits to ensure our energy supply for the future.

In life sciences, GPI-Space is used to analyze the human genome in order to find cures to diseases like Alzheimer's, cancer, MS and many others, millions of people suffer from.

Taking GPI-Space to the next level of possibilities, batch and speed layer of the Lambda Architecture have been unified within a single framework, allowing new approaches to existing big data challenges. A data stream from any source can be processed in memory with any kind of topology and written to a fast parallel file system. Batch data from a file system can be loaded simultaneously and combined with the streaming data.

With this functionality, real time data analysis projects can be realized without leaving the familiar development environment of GPI-Space, which can also be used to feed data into today's big data systems.

Rapid and Modular Development

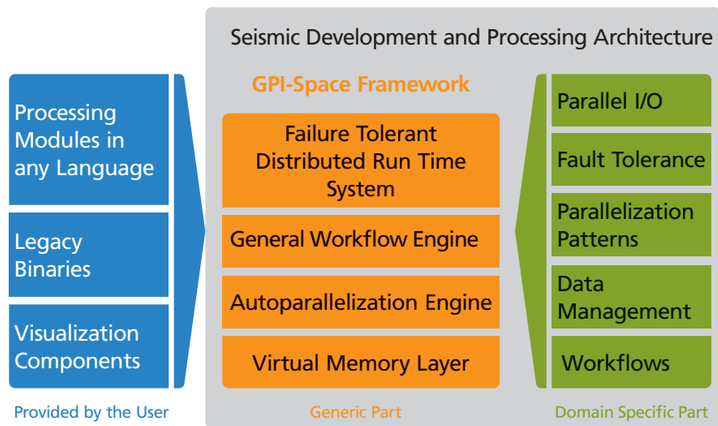
GPI-Space uses modular workflows to execute each task from simple to most complex. Each such workflow is comprised of independent components and can again be used as such a component itself.

The foundation for this workflow approach is a library of such components which comes with the domain specific configuration of GPI-Space. Over time, simply by using it, users will expand this library and it will contain more and more reusable components, each solving one individual problem. This means, solutions only have to be implemented once and not over and over again.

Examples for such individual components are: loading/storing processes for given data patterns, pre-processing and processing components, but also legacy modules either as source code or in form or existing binaries.

Internally, GPI-Space uses a local, concurrent and reversible theoretical modeling language (Petri Nets) to describe workflows. But there are domain specific graphical workflow editors to compose new workflows by adding data dependencies between existing components.

The more GPI-Space is used, the faster the library will be filled with reusable components and the faster users are able to create new solutions just by plugging a new piece of code in between existing components. This way, rapid parallel prototyping is possible because the user can focus on his problem and doesn't have to worry about the challenges and traps of parallelism, because that part has already been solved.



GPI-Space tools:

1 *Interactive salt dome modification*

2 *Direct connection with XTreemView*