

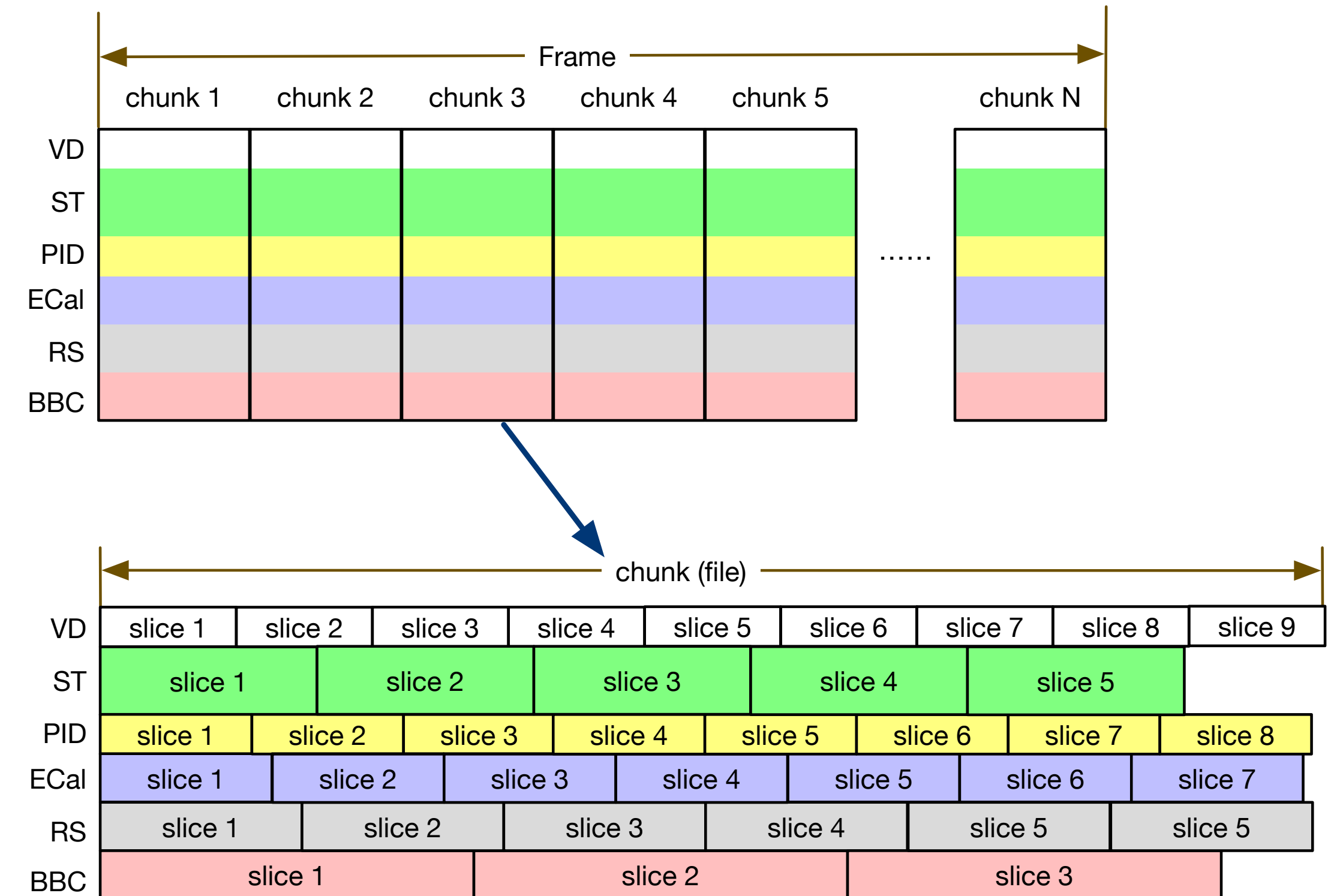
SPD OnLine Filter. Status update.

Danila Oleynik 26.04.2023

SPD OnLine filter

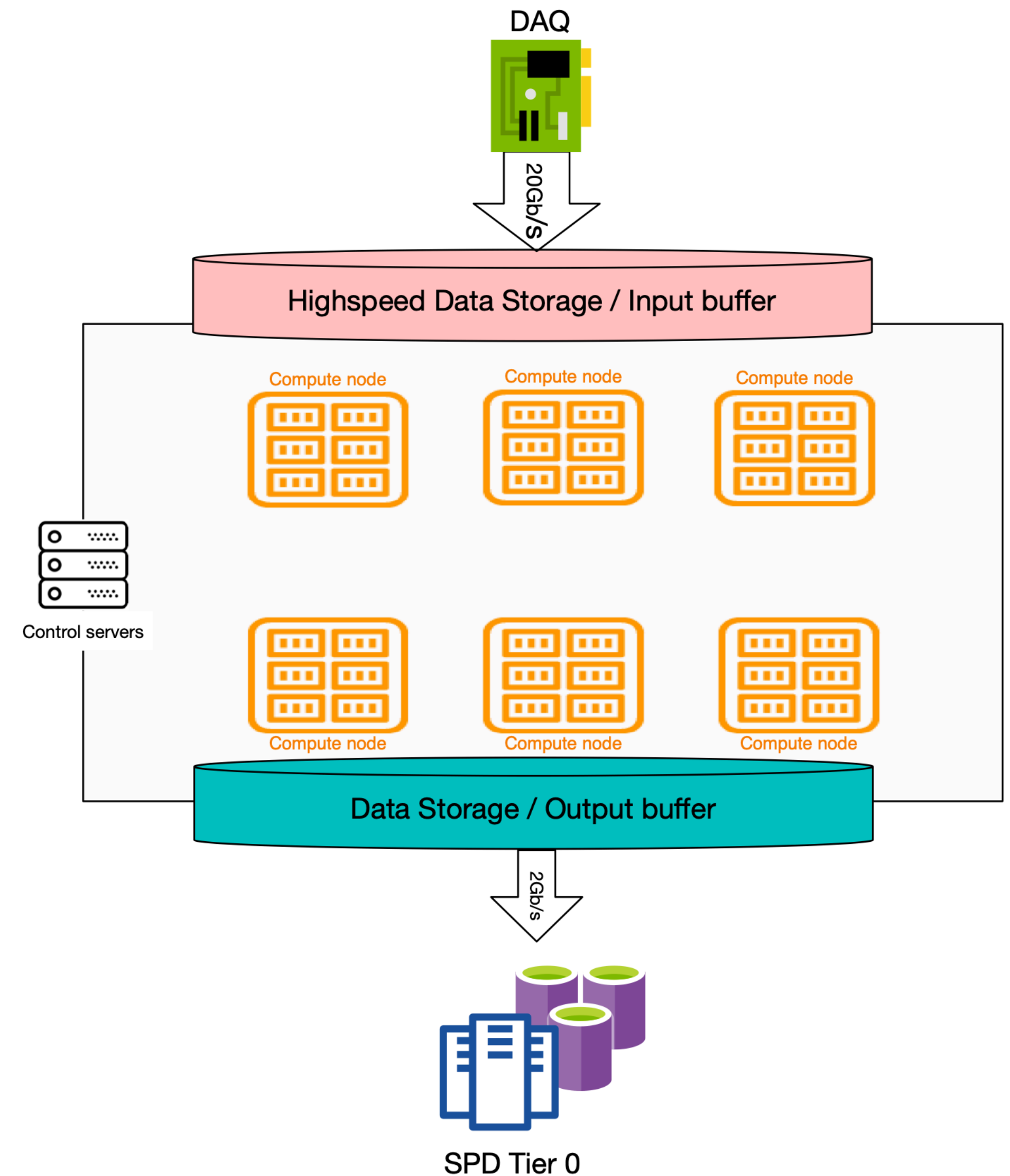
Remainder

- SPD Online Filter is a high performance computing system for high throughput processing of data
- This computing system should carry out next transformations of collected data:
 - reorganize data into event's oriented format – Event unscrambling;
 - partial preliminary reconstruction;
 - filter 'boring' events and leave only 'hot';
 - settle output data, merge events into files and files in datasets for future processing.
 - Prepare data for on line DQM



Online filter infrastructure

- High speed (parallel) storage system for input data written by DAQ.
- Compute cluster with two types of units: multi-CPU and hybrid multi CPU + accelerator(s) (GPU, FPGA etc.).
- A set of dedicated servers for managing of processing workflow, monitoring and other service needs.
- Buffer for intermediate output and for data prepared for transfer to long-term storage and future processing.



High-throughput data processing

On SPD OnLine filter

- Expected dataflow after DAQ system up to 20 Gb per second. A lot of data.
- If this flow divided to some homogenous but independent pellets, a lot of them can be processed concurrently. That's allow simplify applied application architecture and put less requirements for hardware part of cluster.
- To be able to manage high-throughput data of data on the facility few middleware components should be developed

Online filter middleware functionality

Data management;

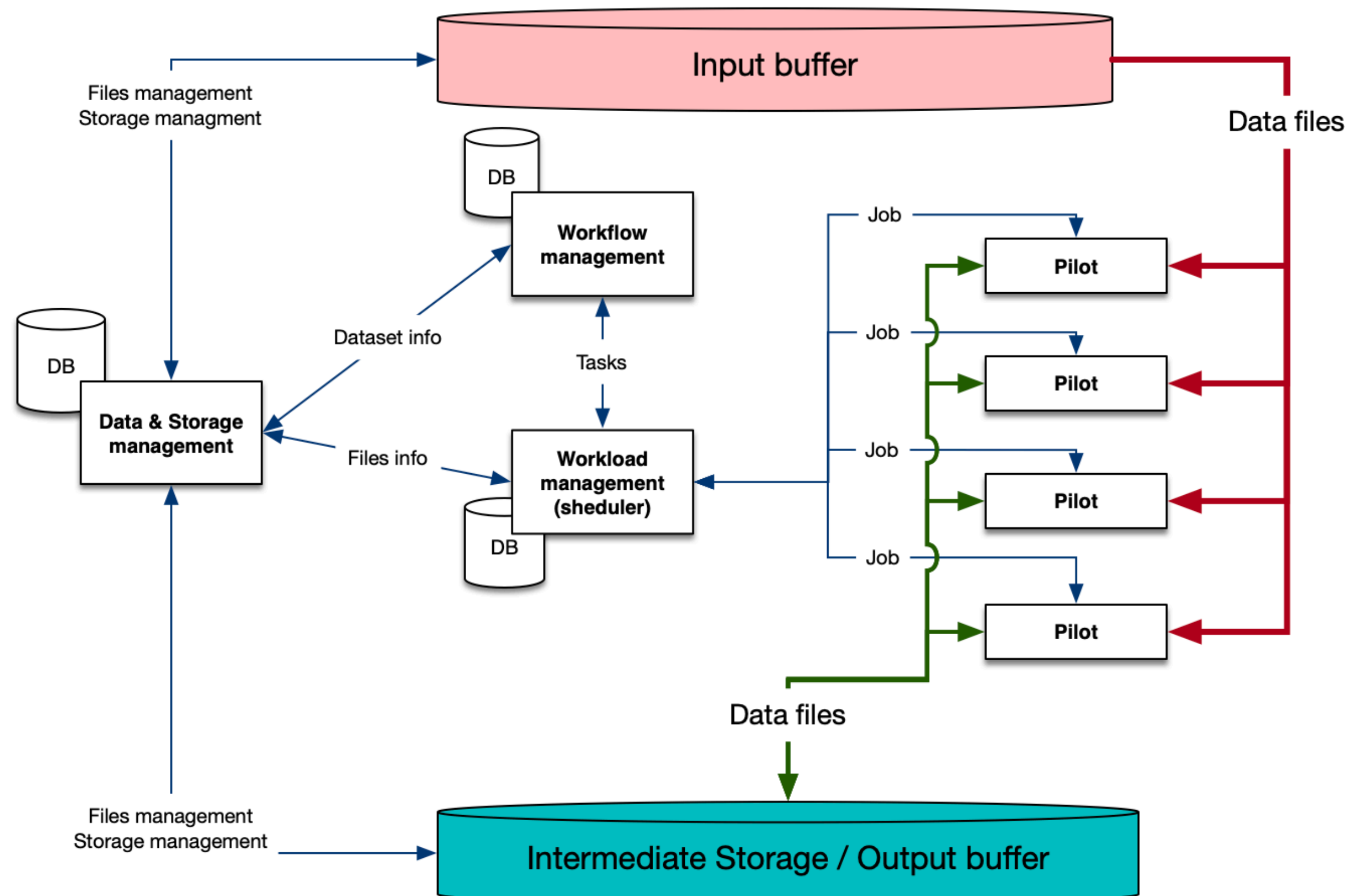
- Support of data lifetime (registering, catalog, global transfer, cleanup);

Processing management;

- Generate tasks for each type of processing:
 - Events identification (building);
 - Verifying of processing results (AI vs traditional processing);
 - Select (Filter) events;
 - Pack (merge) output data for transferring to "offline";

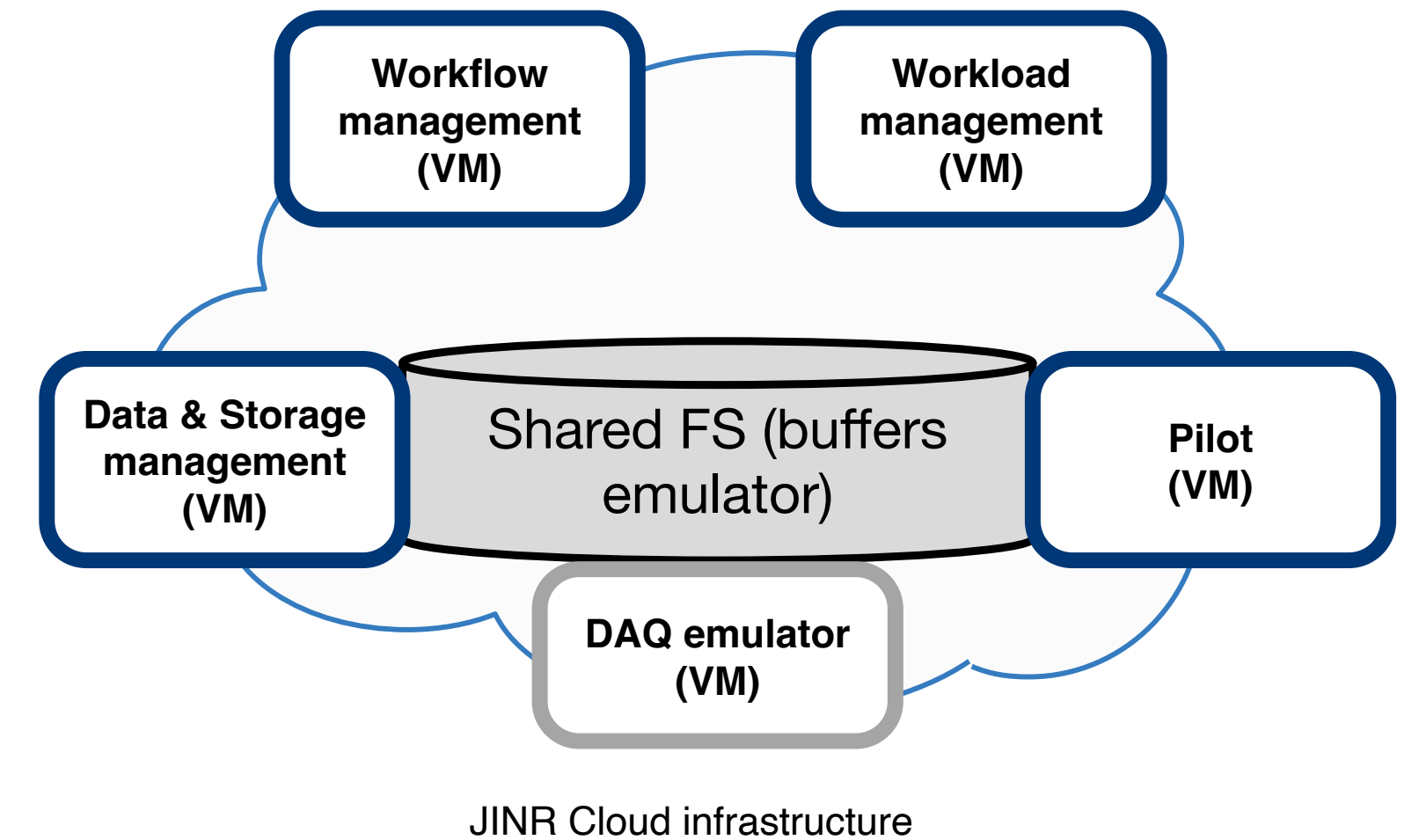
Workload management:

- Generate required number of jobs to perform task;
- Dispatch jobs to pilots;
- Control of jobs executions;
- Control of pilots (identifying of "dead" pilots)
- Manage efficient resources usage



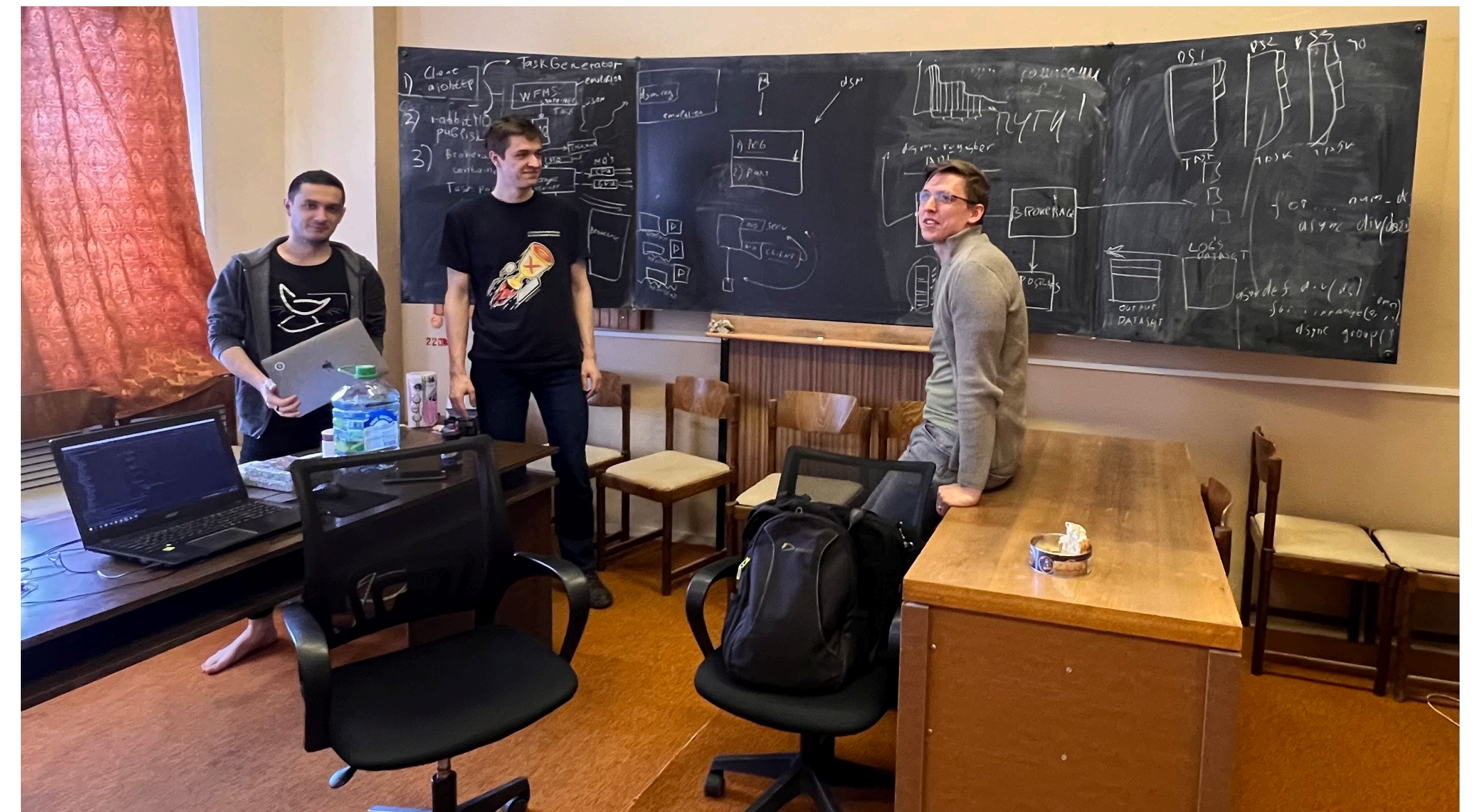
Prototyping

- At least four services should be realised for initial prototype of Online Filter middleware complex
- Prototype will demonstrate proof of concept and will provide initial code base
- Compute resources for development and testing are provided by JINR Cloud service
- DAQ team provides first (simplified) version of emulator
- Basics of data organisation on input buffer also agreed with DAQ team



Manpower

- Nice outcome from first JINR IT School. A few master students were interested to participate in SPD computing and involved in software design and development process.
 - Mostly remote work started in December - January
 - Weekly meetings + personal sprints
 - 2 weeks of practice in JINR (for the moment)



Workflow management

Current status

- Workflow management system is the high level system which interacts with data management system and workload management system for realising multistep processing of blocks of data (datasets)
- The functional decomposition of the subsystem was carried out and the initial set of microservices was proposed
 - “Chain definer” - user oriented application which allow define sequences of processing steps
 - “Processing starter” - microservice responsible for triggering of processing chains
 - “Chain executor” - microservice responsible for control of execution of processing chain.
- In progress, coding of interfaces with external systems, tuning of state model.

Data management

Current status

- A lot of work already done for data management system starting from decomposition of functionality to microservices, definition of set of tools for storage management, realisation of DB design for data catalogue (datasets and files) up to definition and realisation of the set of internal and external interfaces
- Microservices:
 - dsm-register – responsible for registration of input data from DAQ in the catalogue
 - dsm-manager – realise interfaces to the catalogue for subsystems
 - dsm-inspector – realise auxiliary tools for storage management (consistency check, cleanup, dark data identification)

Workload management

Current status

- Realize a task execution process by shredding a required number of jobs to provide controlled loading to compute facility, tacking into account priority of tasks and associated jobs. A task is one step in a processing chain of a block of data. Job is a processing of a single piece of data (file or few files).
- Microservices: task manager, task executor, job manager, job executor
- Workload management system intensively interacts with Pilots, Data Management system, accept tasks from Workload management system and reports progress of execution back
- Base architecture in place, coding of internal and external interfaces.

Pilot

Current status

- Pilot is the application to manage the execution of a job, produced by WMS, on a compute node. It is responsible for setting up of environment, stage-in/out of data from storage to compute node, execution of payload and monitoring of different conditions during execution. Intensively interacts with workload management system
- Base architecture and initial functionality of pilot application is defined. It is a multithread application with interactions between threads through queues
- In progress intensive development of interfaces and initial definition of internal entities

Nearest plans

- We have ambitious plan to deliver prototype before July.
 - The team is in progress of rapid coding of each component
 - We would like to start integration testing before summer, and collect some very initial info about system behaviour during summer
 - It will be functional prototype but for emulation of realistic behaviour we need agreed initial processing chains and got some solutions for applied software.

The team

- Dmitry, Evgeny, Leonid and Nikita shows great progress in quite short period of time.
- I hope, that their enthusiasm will not significantly decrease after magistrate theses defence and we will continue this (and other works) for SPD (NICA) Computing.
- Project is mature enough to get a name :-). A few variants were formulated and team chose:
- **VISOR** – **VISOR Is SPD Online Filter** [reverse acronym like GNU, PNG, YAML]