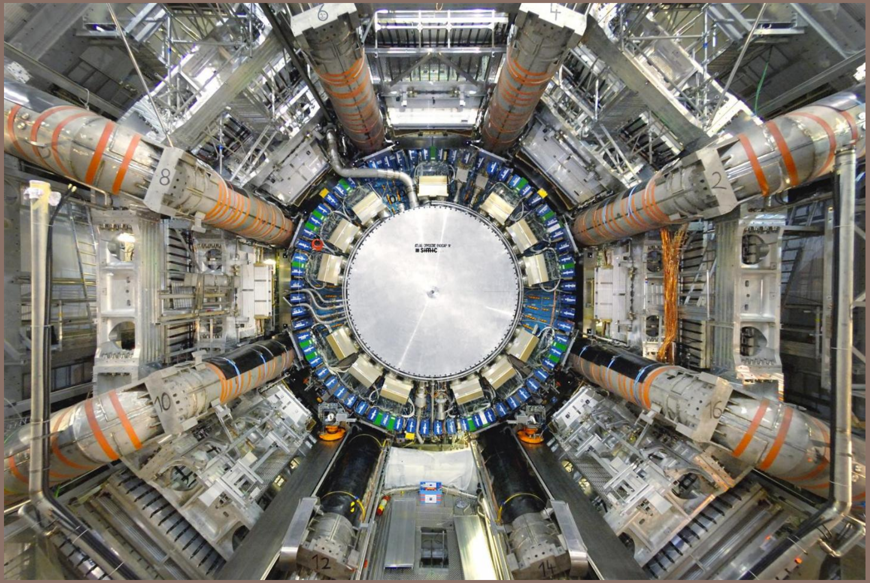


Visual Cluster Analysis for Computing Tasks at Workflow Management System of the ATLAS Experiment at the LHC

M. Grigorieva, T. Galkin, A. Klimentov, T. Korchuganova, I. Milman, S. Padolski, V. Piluygin, M. Titov

NRC "Kurchatov Institute"
NRNU "MEPHI"
NR TPU
Brookhaven National Laboratory

GraphiCon 2018, Tomsk

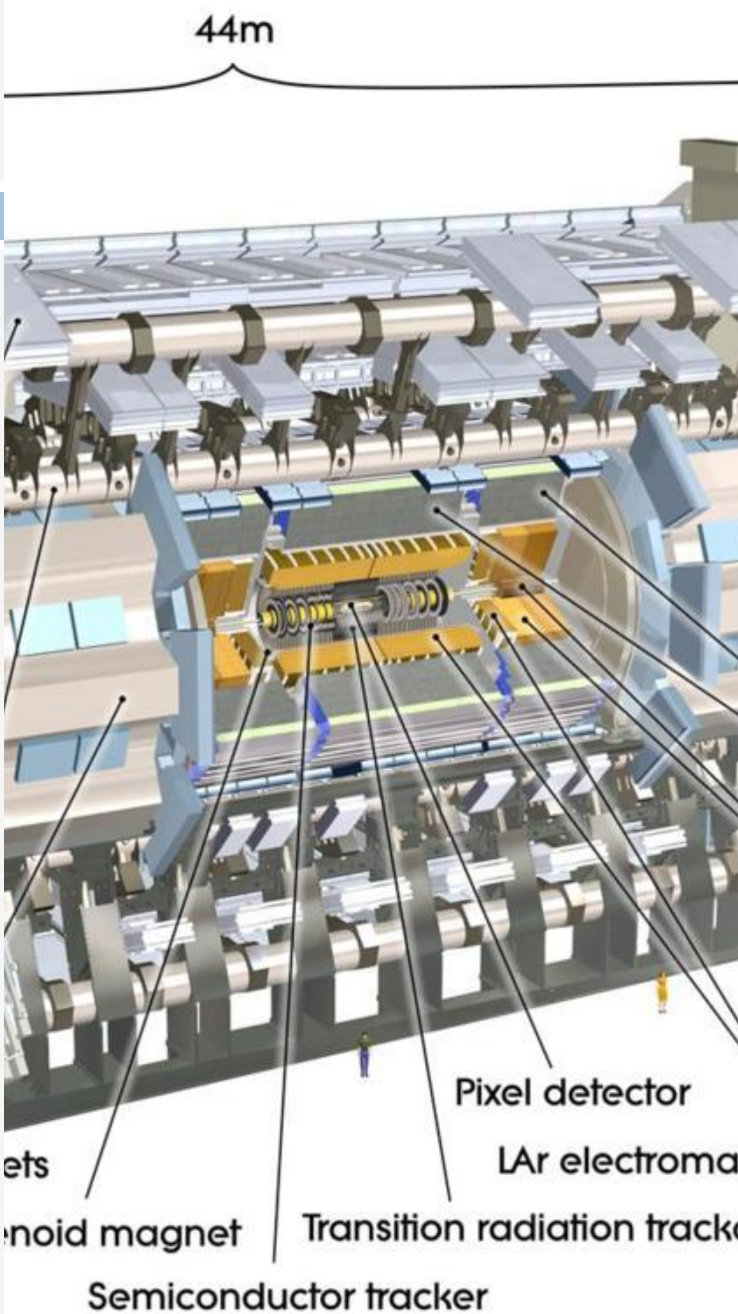
Outline

- Introduction
- The ATLAS experiment computing challenges
- Proposed approach of visual analytics
- BigData visualization challenges
- Analysis of ATLAS tasks execution
- Interpretation of the results
- Conclusions and future work

Introduction

- **LHC physics with ATLAS at CERN**
  - Largest collider detector ever built
  - Higgs discovery, precision measurements of the Standard Model (including Higgs), searching for dark matter, supersymmetry, exotic particles…

- **ATLAS Computing Challenges**
  - Massive computing resources required
  - Huge collaboration – with thousands of data analyzers
  - Ongoing competition for computing resources between different threads of tasks/jobs
  - Complex workloads and workflows
  - Computing needs grow every year with more LHC and Monte-Carlo data
  - Uncountable possible reasons of failures and delays

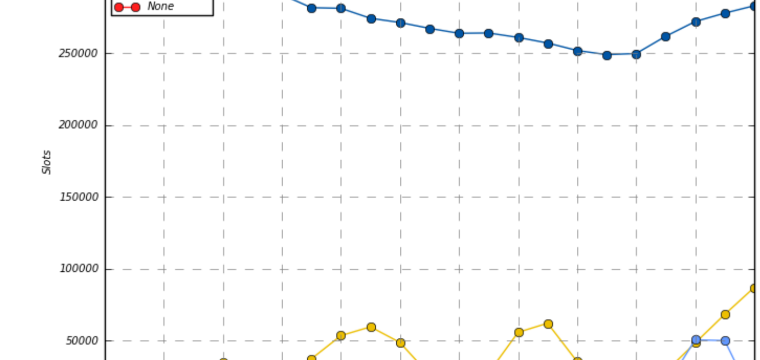


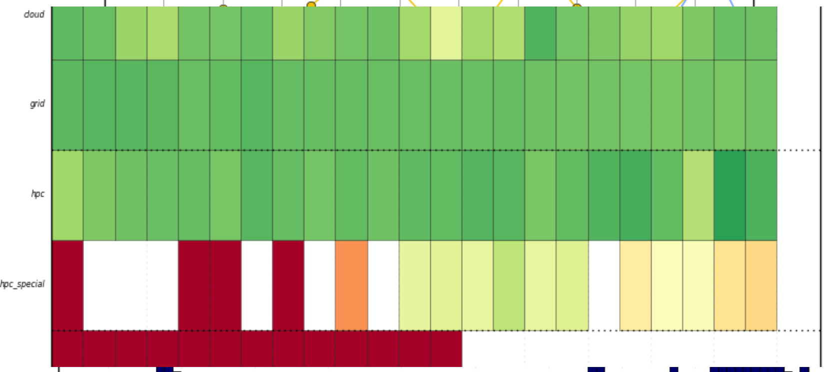
Distributed Software Pillars

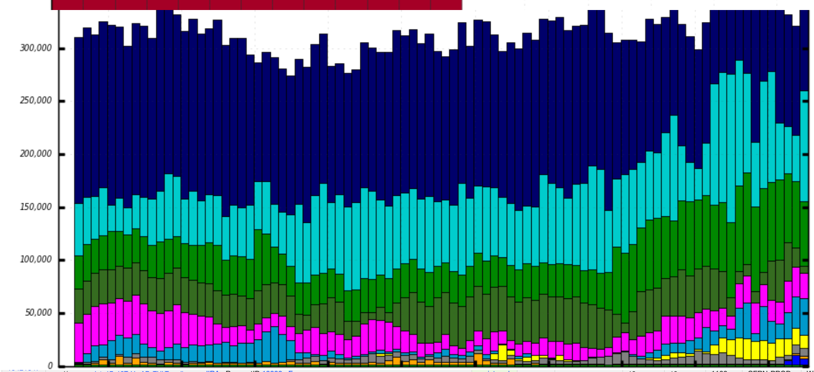
1. **Workflow Management System** provides an infrastructure for the set-up of a defined sequence of tasks, arranged as a workflow application
   - Physics Analysis **tasks** contain execution code and input/output files, corresponding to underlying physics process and initial conditions.
   - Each task is fragmented into **jobs** which correspond to a fixed number of particle collision events.
   - Jobs are executed on **computing sites**

For the last decades ATLAS has processed:
10 millions of tasks and
3 billions of jobs

2. **Workload Management System** realizes the task scheduling. It solves many problems of using unstable distributed computing resources.
3. **Data Management System** is software that takes in data files and converts or aggregates various kinds of data into a single storage datasets or containers.
4. **Grid Information System** provides the information about computing sites and queues.
5. **Accounting** tools used to monitor the utilization of the available computational and storage resources.
6. **Monitoring** provides interface to the system, containing overviews of its operation.







15421845	prod/merge data17_13TeV AP_THLT sdjones JIRA RequestID:19082 Errors pending until parent produces input	registered			18 15:01:03	18 15:01:03	1100	CERN-PROD	WORLD
15421845 15392178	data18_13TeV.00360026.physics_EnhancedBias.recon.r10827_r10841 prod/recon data17_13TeV AP_THLT sdjones JIRA RequestID:19082 Errors	ready 2664		1107755 / 0	2018-09- 18 14:01:17	2018-09- 18 15:01:17	1100	CERN-PROD	WORLD
15421471 15307059	mc15_14TeV.346217.PowhegPythia8EvtGen_Ztautau_tautauhh_1pXn.deriv.e6905_p3401 prod/deriv mc15_14TeV GP_PHYS atlas-dpd-production JIRA RequestID:18950 Errors	submitting 3986		39860000 / 39860000	2018-09- 18 14:16:13	2018-09- 18 14:16:13	560	wuppertalprod	WORLD
15421465 15307056	mc15_14TeV.346216.PowhegPythia8EvtGen_VBFH125_tautauhh_1p1n_unpol.deriv.e6905_p3401 prod/deriv mc15_14TeV GP_PHYS atlas-dpd-production JIRA RequestID:18950 Errors	submitting 994		4970000 / 4970000	2018-09- 18 14:16:14	2018-09- 18 14:16:14	560	INFN- NAPOLI- ATLAS	WORLD
15421358 14659527	mc15_14TeV.411045.PowhegPythia8EvtGen_ttbar_171p00_SingleLep.deriv.e6812_p3401 prod/deriv mc15_14TeV GP_PHYS ndang JIRA RequestID:18364 Errors	submitting 2000		10000000 / 10000000	2018-09- 18 13:43:29	2018-09- 18 13:43:29	560	TOKYO-LCG2	WORLD
15421207 15421200	mc16_13TeV.399005.MGPy8EG_A14NNPDF23LO_StauStau_directLLP_100_0_0p01ns.deriv.e6633_e5984_s3307_r10740_r10706_p3384 prod/deriv MC16:MC16a AP_SUSY wanghill JIRA RequestID:18744 Errors insufficient inputs are ready. 0 events (0 files) available, 500000*10 events required for scout	registered			2018-09- 18 14:33:00	2018-09- 18 14:33:00	885		WORLD
15421200 15305786	mc16_13TeV.399005.MGPy8EG_A14NNPDF23LO_StauStau_directLLP_100_0_0p01ns.merge.e6633_e5984_s3307_r10740_r10706 prod/merge MC16:MC16a AP_SUSY wanghill JIRA RequestID:18744 Errors	running 5		10000 / 10000	2018-09- 18 13:21:54	2018-09- 18 13:21:54	885	CERN-PROD	WORLD
15421057	data18_13TeV.00349014.physics_BphysLS.deriv.f969_m2020_p3628 prod/deriv data18_13TeV:openended18 GP_PHYS mdobre JIRA RequestID:19064 Errors	ready 459		3055968 / 0	2018-09- 18 11:51:22	2018-09- 18 12:51:22	880	UNI- FREIBURG	WORLD
15421052	data18_13TeV.00349014.physics_BphysLS.deriv.f969_m2020_p3628 prod/deriv data18_13TeV:openended18 GP_PHYS mdobre JIRA RequestID:19064 Errors	ready 459		3055968 / 0	2018-09- 18 11:51:21	2018-09- 18 12:51:21	880	MPPMU	WORLD

Monitoring

- Monitoring includes many views (**operational and historical**)
  - Tables
  - Plots
  - Bar Charts
  - Pie Charts
  - Gantt Diagrams
  - HeatMaps

Monitoring Challenges

- Until now, the requirements for the monitoring system have been limited to the use of basic visual analysis methods for a limited class of tasks, and data dimensionality limited to 3 dimensions.
- Current monitoring infrastructure **does not provide**:
  - analytical studies for estimation of correlations between the numerous properties of objects over time
  - the analysis of the causes of failures or time delays of computing tasks in a distributed computing environment.
- The constant increase in the amount of processed data and the complexity of ATLAS computing infrastructure produces new challenges related to the visual analysis of large volumes of multidimensional data.

Machine Learning ”BLACK BOX” Explanations

Solving the problem of optimal and effective usage of the computing infrastructure requires the use of **ML methods to predict the progress of the tasks/jobs execution.**

ML methods are often treated as a **”black box”** since there are no effective techniques for understanding the internal mechanisms of these complex systems and interpretation of their results.

Relative **lack of human supervision** makes it hard to fully understand the inner workings of trained models and limits the ability to verify that the models work correctly.

Visual analytics plays the most important role in data interpretation, it uses **unique human abilities of understanding the complex graphic images.**

By designing effective interactive visual images, it is possible to provide a fast, accurate and reliable **interpretation of the ML** models.

Visual analytics of multidimensional data

- **Visual Analytics** uses the power of visual perception as a way to involve humans in solving complex problems with data.
- Visual Analytics is the area of studying **interactive** graphical representations for analyzing complex data. This is often done in addition to and with the help of computational and statistical procedures.
- Visual Analytics can help domain experts with a problem that is to be solved using ML, to verify the **semantic accuracy** of a model, whether decisions made by model **"make sense".**

An ultimate goal of this project is use a modern visual analytics methods to increase the stability and efficiency of the distributed data processing and analysis systems.

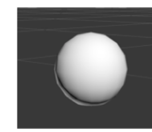
The method of multidimensional data visual representation

coordinates

	Parameter 1	Parameter 2	…	Parameter n
Job 1	x_{11}^j	x_{12}^j	…	x_{1n}^j
…	…	…	…	
Job i	x_{i1}^j	x_{i2}^j	…	x_{in}^j
…	…	…	…	…
Job m	x_{m1}^j	x_{m2}^j	…	x_{mn}^j

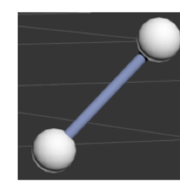
multidimensional points

$$p_i = (p_i^1, p_i^2, \dots, p_i^n) \xrightarrow[SphC]{p_i^{l_1}, p_i^{l_2}, p_i^{l_3}, SphR}$$

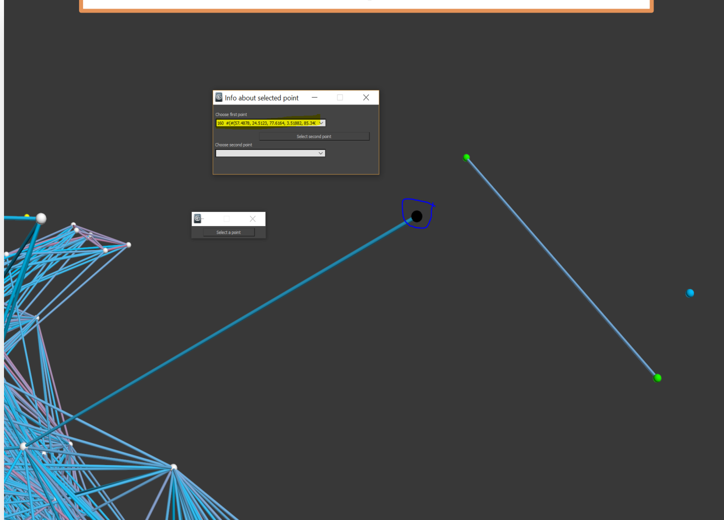


Euclid/Mahalanobis distances between all points

$$p_{i_1}, p_{i_2}, d: \rho(p_{i_1}, p_{i_2}) \leq d \xrightarrow[CylC]{CylR}$$

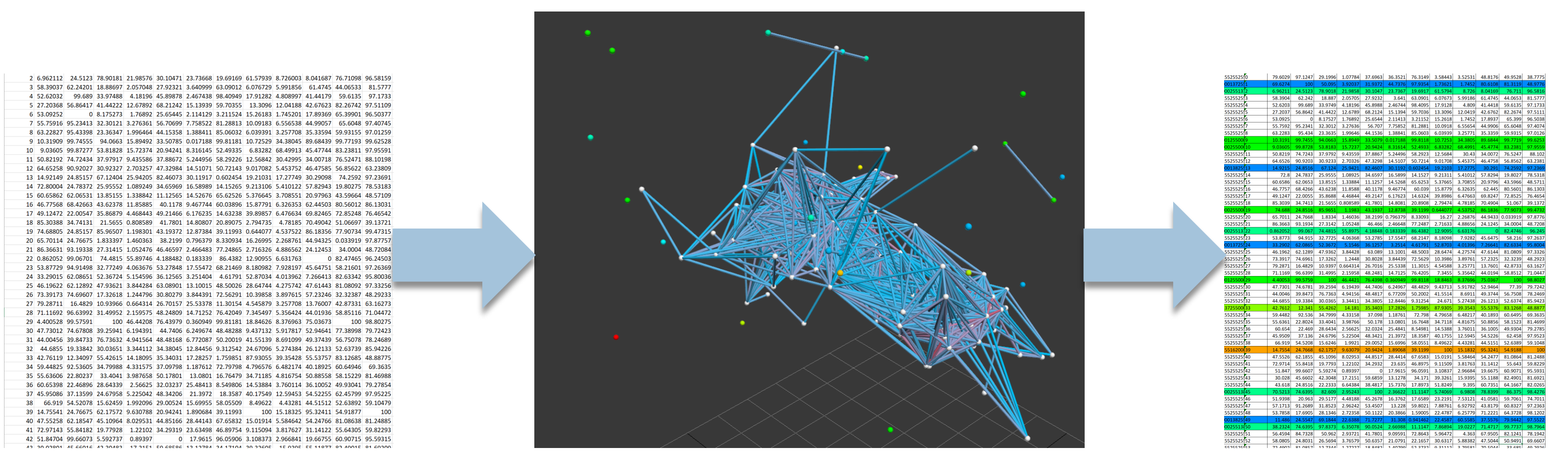


Interactive spatial scene



Multidimensional visual analysis prototype IVAMD

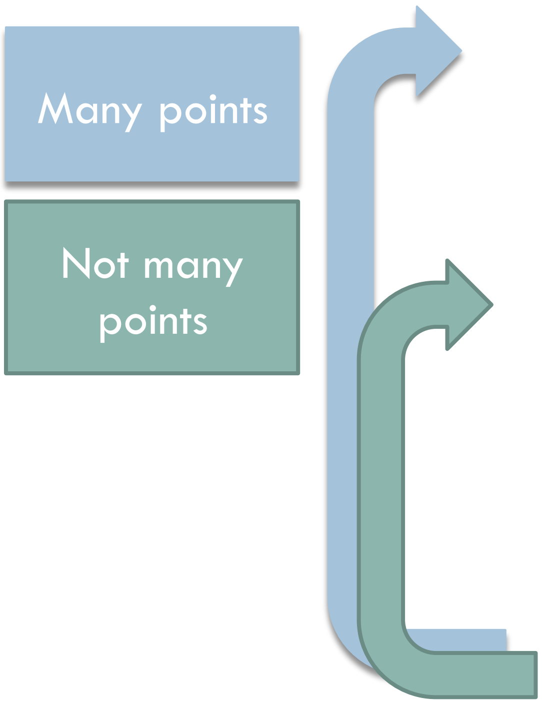
- First prototype is created based on **3ds Max**
- A combination of MAXScript scripts and C# module was used
- Depending on the amount of RAM, the software can handle up to a couple of hundreds of objects
- Points (Spheres) are coded with colors, depending on the clusters
- Results can be exported to excel (xlsx) files



Big Data Visualization Challenges

- When data volume exceeds tens of thousands/millions of points:
  - Some information may be lost or suffer from overplotting
  - Require considerable amount of computational resources (storage, memory)
- One general approach is **to reduce the amount of data** presented to the user in one go, first presenting an overview of the data to the user, then allow him/her to zoom in/filter out the relevant area/volume and finally provide details when needed.
- **Our approach:**
  - Multi-Layered Clustering

Multi-Layered Clustering



- Feature Selection
- Normalization
- **1st level clustering:** **K-Means** (to split the initial data sample into several hundreds of data clusters)
- Grouping the initial data sample by clusters with the mean values of all features
- **2nd level clustering:** **calculating distances between multidimensional points**
  - Euclid, Mahalanobis, etc.
- Building 3D spatial scene with spheres and cylinders
- Using interactive interface to tune the distance threshold
- Choosing clusters or anomalous points for the detailed analysis

ATLAS Jobs Execution Data Analysis

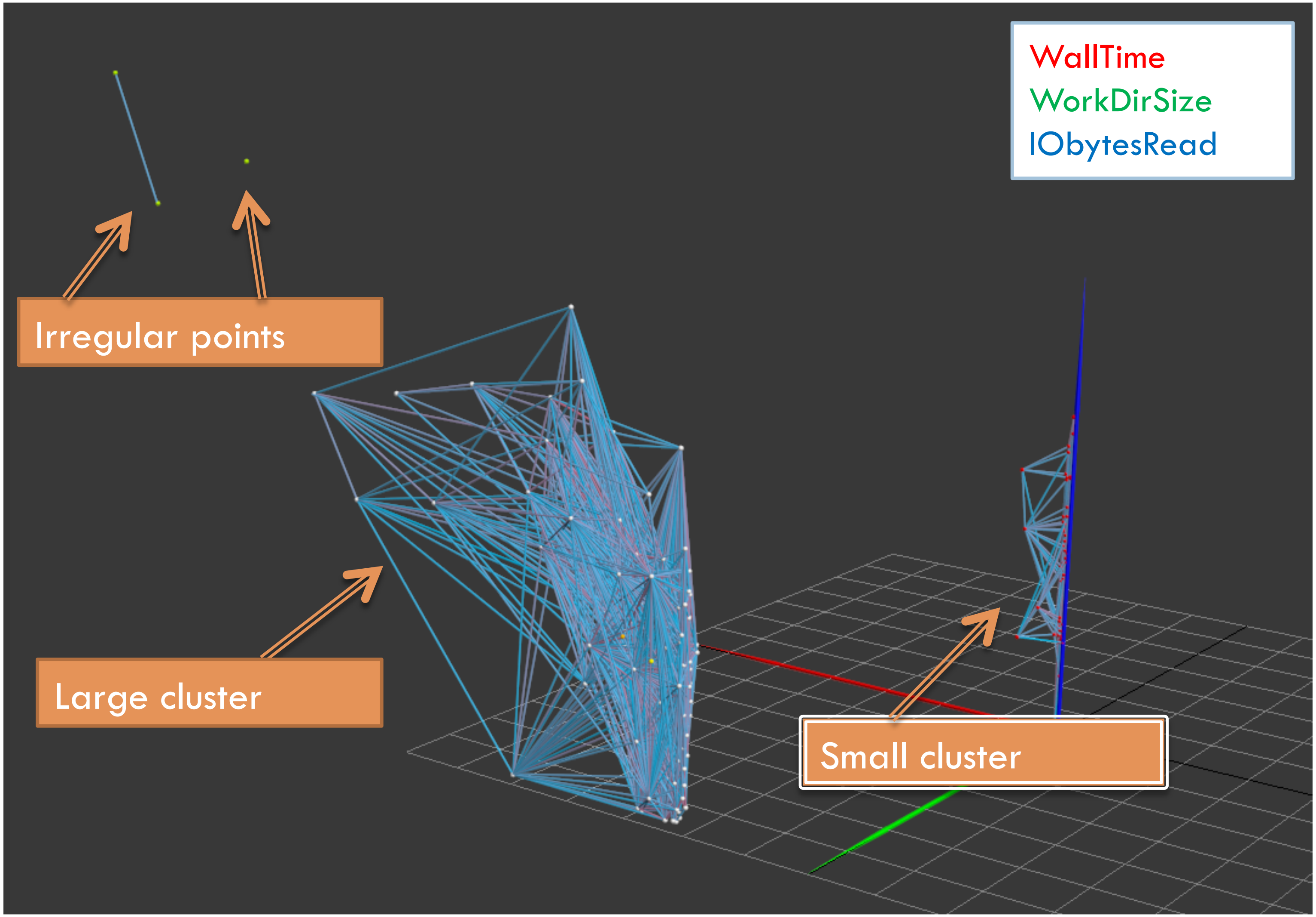
Demonstrated on the proposed visual analytics prototype IVAMD

Jobs execution metrics chosen for data analysis

- **WallTime** – the time it takes for the jobs to run from start to finish
- **timeStageIn** - time for job wrapper (pilot) to stage the input files to the worker node
- **timeStageOut** - time for pilot to stage the output files to local storage element
- **TimeEXE** – WallTime without staging
- **CPUConsumptionTime** – the time that the CPU is actually working on task (without memory swap)
- **inputFileBytes** - The total size of input files
- **maxRSS** – max available RAM
- **maxVMem** – max available virtual memory
- **IObytesRead** – data read during job execution
- **IObytesReadRate** – average rate of data read
- **workDirSize** – available space of the working directory

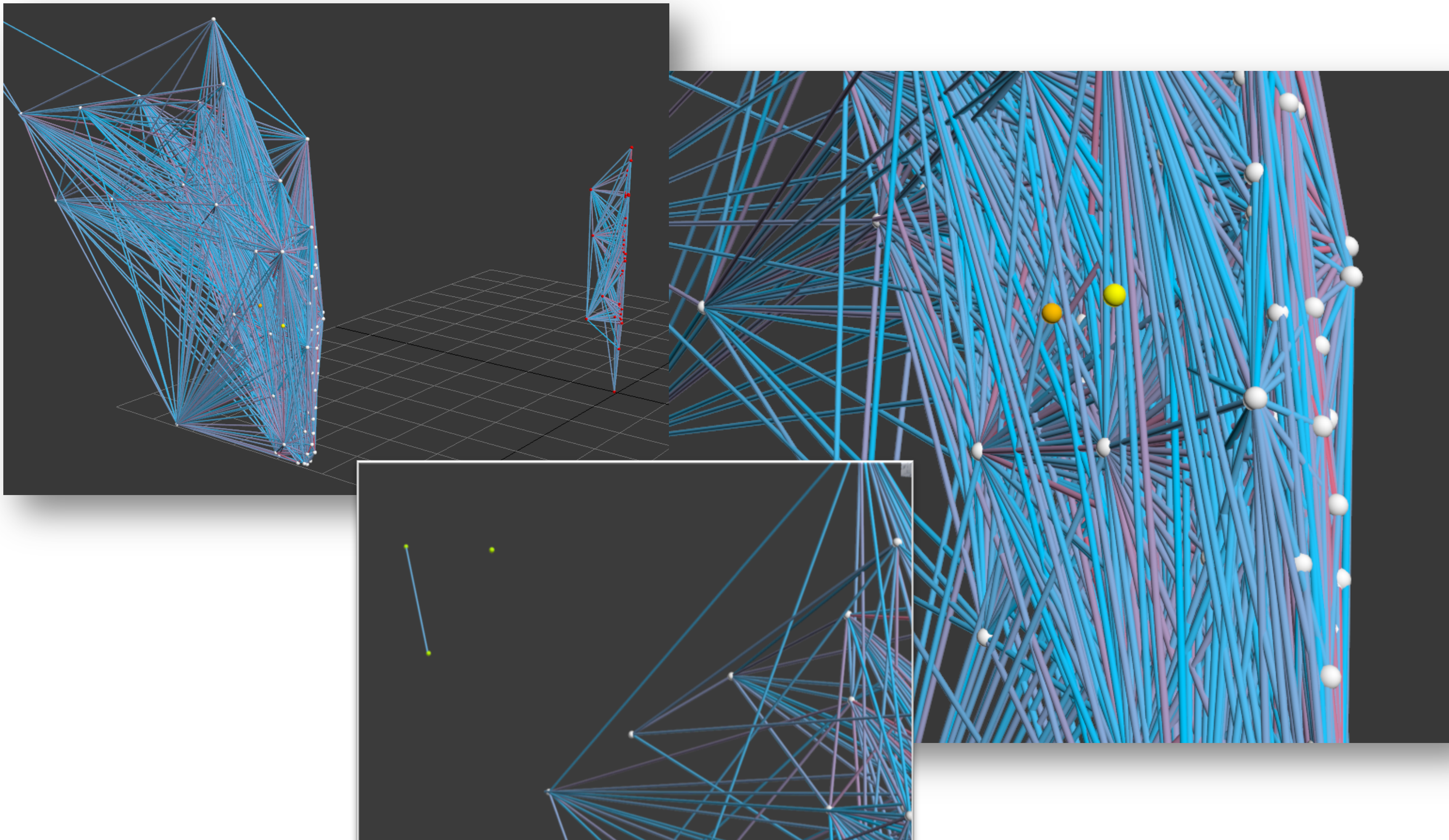
Clusters and Anomalies

Task №14296407



Clusters and Anomalies

Task №14296407



Features	Large cluster	Small cluster	Irregular points
WallTime	25 min	**10 min**	**227 minutes**
CPUTime	3,8 min	3,2 min	3,5 min
TimeStageIn	370 sec	**110 sec**	356 sec
TimeStageOut	59 sec	**33 sec**	**680 sec**
InputFileBytes	300 MB	300 MB	300 MB
IObytesRead	1 957 MB	1 704 MB	**3 029 MB**
IObytesReadRate	3,868 MB/sec	**6,716 MB/sec**	**0,690 MB/sec**
WorkDirSize	600 MB	**8 MB**	600 MB
MaxRSS	825 MB	817 MB	814 MB
MaxVmem	3 041 MB	2 767 MB	3 056 MB

Results of cluster analysis

Task №.14296407

Results of cluster analysis

Task №14296407

- The **CPU Time** is in the expected range in all clusters and irregular points
- **Input files sizes** are the same for all clusters and points
- But observed **input data (IObytesRead)** much larger than input file sizes
  - The amount of read information is a subject to further investigation and analysis. Hypothesis of the reasons could be:
    - Failed jobs on the same site lead to overload of the data streams
    - Input of the data failed and had to start from the beginning
- **Read rate** of irregular points is 5 times slower than in large cluster
- The small cluster has the **highest rates** of data read/write and the **shortest wall time**

Conclusions and Future Plans

- The methodology of data analysis with the combined usage of machine learning and visual analytics methods has been proposed
- The first prototype of interactive visual analytics platform was developed on the basis of 3dsMax
- The developed methodology and visual prototype have been applied to analyze ATLAS jobs execution
- The development of the next phase of visual analytics model:
  - Moving to open-source platform
  - Using web-interface for the integration with BigPanDA monitoring in ATLAS