M.A. Grigorieva, A.A. Alekseev, A.A. Artamonov, T.P. Galkin, T.A. Korchuganova, I.E. Milman,
S.V. Padolski, M.A. Titov, A.A. Klimentov on behalf of the ATLAS Collaboration

CHEP 2019
4-8 November 2019,
Adelaide, Australia

Enhancements in Functionality of the Interactive Visual Explorer for ATLAS Computing Metadata

Interactive Visual Explorer (InVEx) is a web application for the exploration of big volumes of multidimensional data.

ATLAS computing metadata has become the research ground for this application.

Basic features:

- ❑ Machine learning methods for data analysis (clusterization algorithms – Kmeans, DBSCAN)
- ❑ interactive 3D visualization models of data
- ❑ Visualization of the results of clusterization

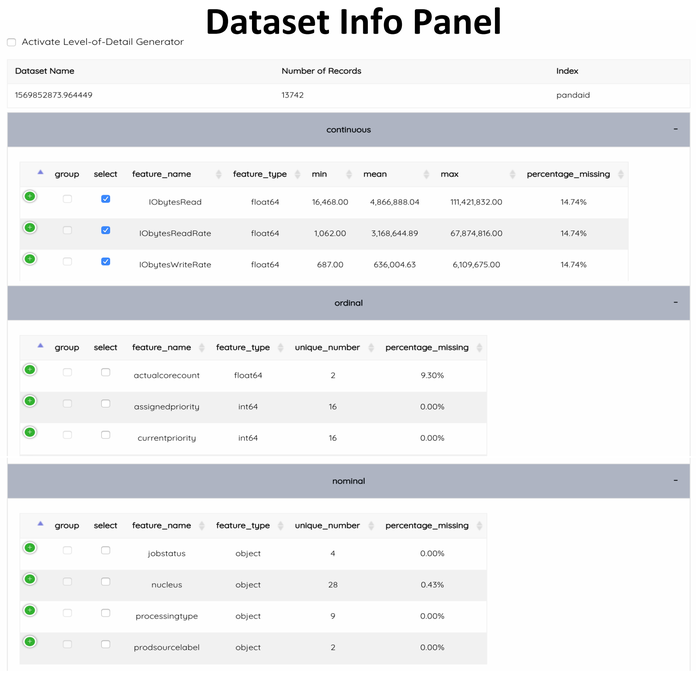
Enhanced functionality:

- ❑ **Dataset Info Panel:** data sample features representation by statistical measurement types: numerical, ordinal, nominal, range non-categorical (string data, that can't be treated as categorical)
- ❑ **Added new clusterization algorithms:**
  - MiniBatchKMeans/sklearn
  - Hierarchical/slkearn
  - K-Prototypes/sklearn
- ❑ **Implemented the Level-of-Detail Generator**, providing grouping of an initial large data sample into clusters/groups to reduce the amount of data presented to the user simultaneously:
  - MiniBatchKMeans Clusterization method
  - Group by nominal/ordinal parameters
  - Group by numerical continuous parameters

The LoD method hides the complexity of the initial data, and allows users to use 3D visual scene to select interesting objects/groups and investigate them separately in the new window.

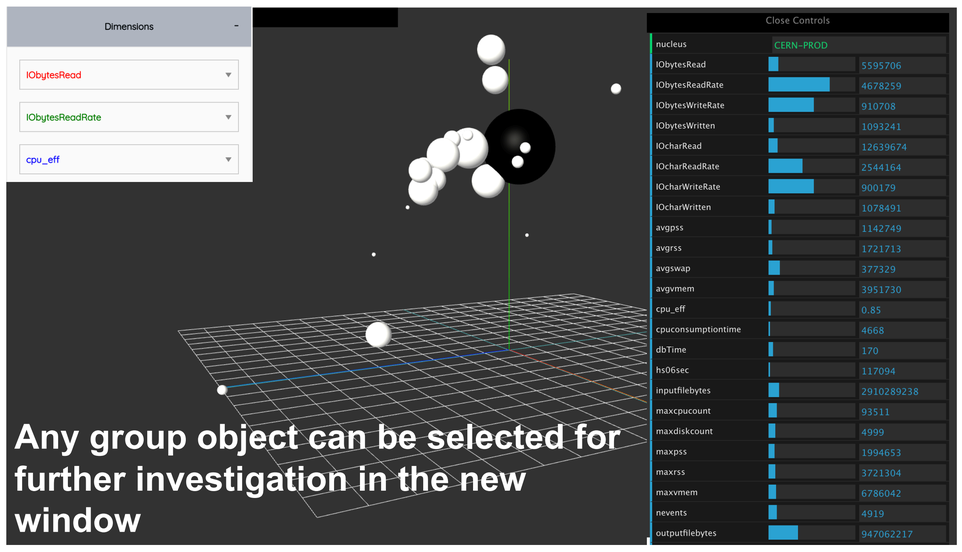
- ❑ **Storage backend** with the ability to store all information about provided operations with data (clusterization, grouping, change of the level of detail) and keep all stages of data derivation sequence.
- ❑ **Interactive parallel coordinates graph** with a linked data table for the exploration of data in all dimensions simultaneously.

Dataset Info Panel

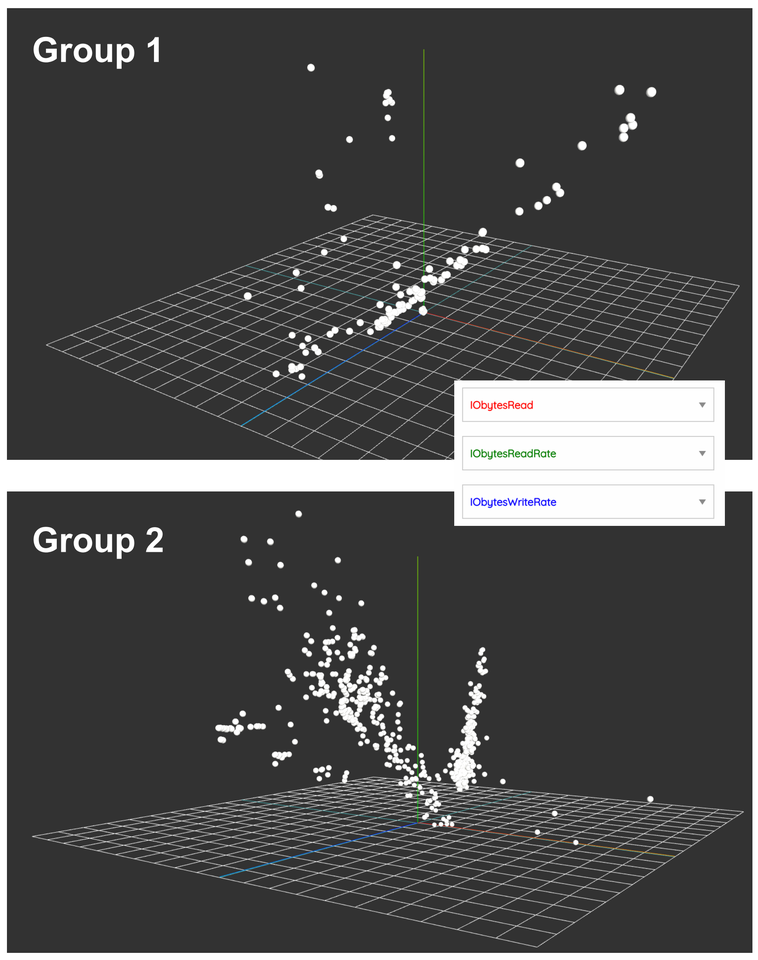


3D visualization of the grouped data sample (Level-of-Detail Generator)

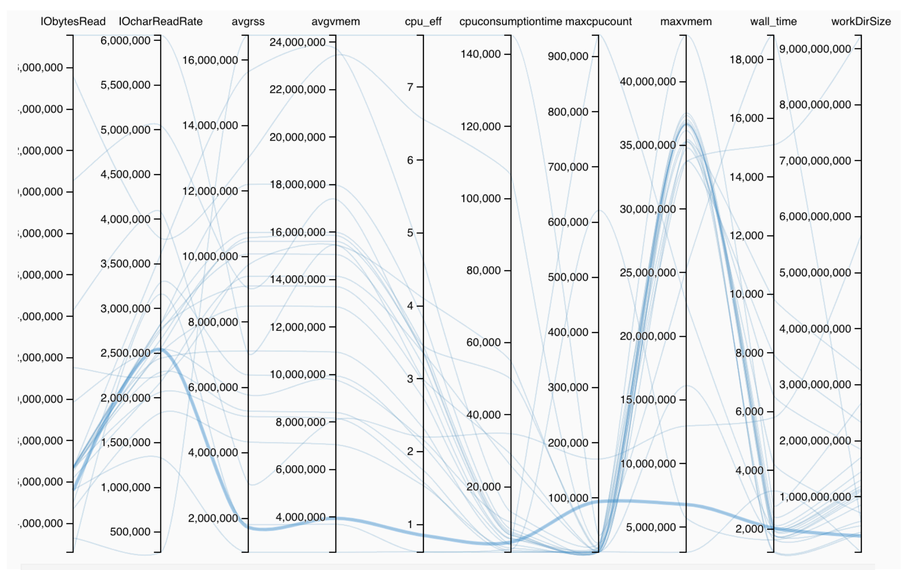
Figure shows ATLAS BigPanDA jobs metadata for some period of time grouped by nucleus



Comparative analysis of the selected data groups in different dimensions

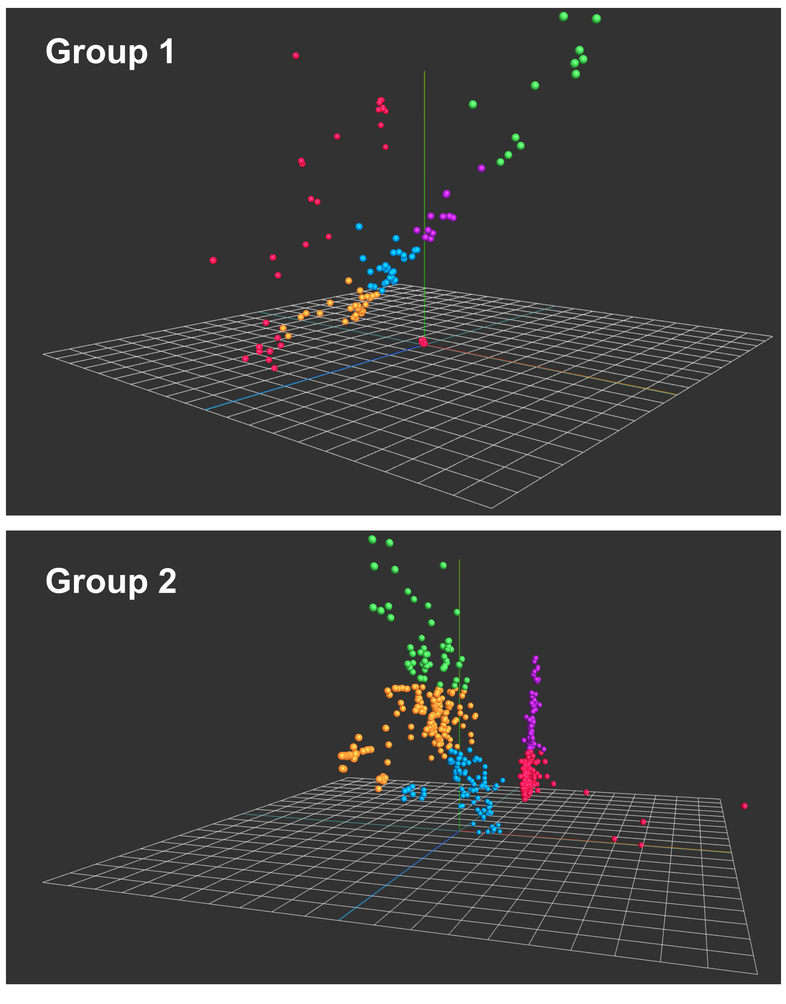


Parallel Coordinates graph with linked table (exploration of aggregated groups parameters)

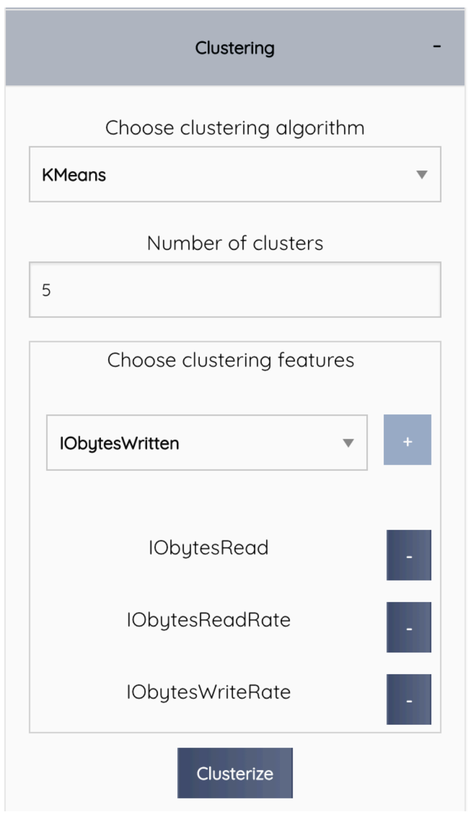


ID	IObytesRead	IObytesReadRate	IObytesWriteRate	IObytesWritten	IOcharRead
BNL-ATLAS	2 607 134.14	4 963 861.61	1 688 020.86	1 065 786.35	2 415 999.1
BU_ATLAS_Tier2	3 285 802	183 866.71	50 896.87	910 172.12	4 950 514.26
CERN-PROD	5 595 705.96	4 678 258.88	910 708.39	1 093 240.55	12 639 673.86
DESY-HH	6 183 737.04	4 239 618.84	542 276.86	884 639.51	3 724 939.16
DESY-ZN	6 654 312	3 741 940.13	731 905	1 301 324.5	4 551 085.75
FZK-LCG2	6 722 965.08	3 820 331.25	751 962.32	1 479 362.84	5 499 766.99

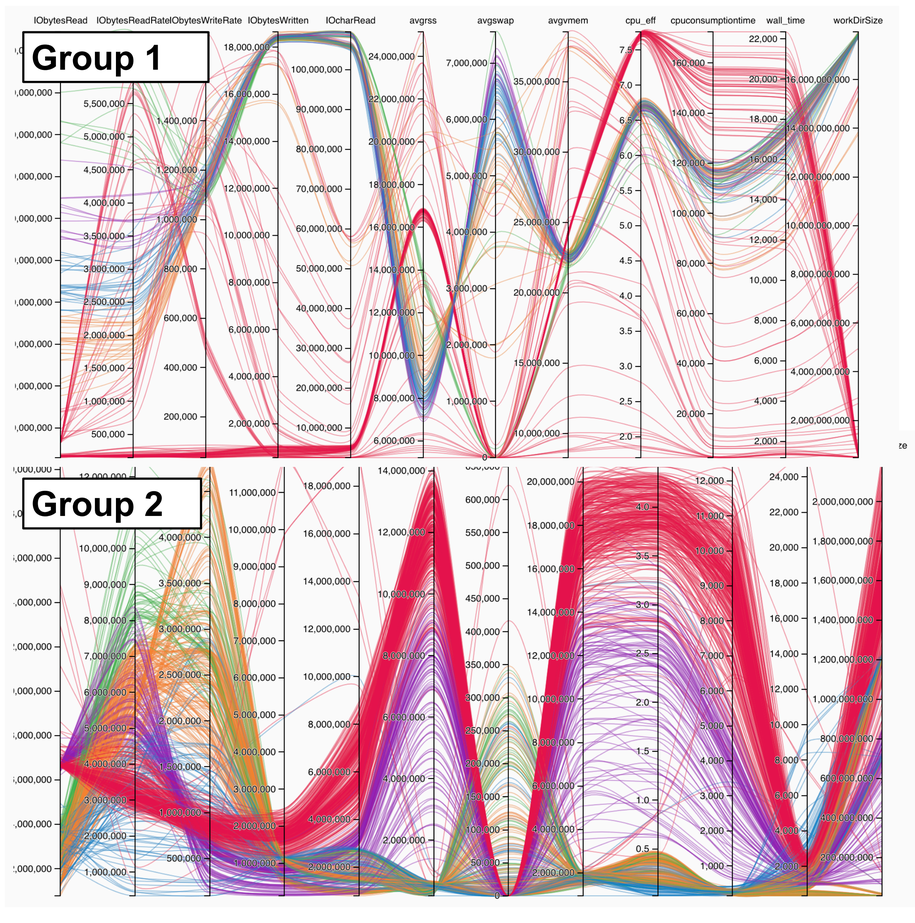
Visualization of the clusterization results



Clusterization with the selected parameters and features



Parallel Coordinates allows to explore trends of parameters for each cluster



The project **"Visual analytics framework to monitor workflow management systems behavior in exascale era"** is supported by **RSCF grant №18-71-10003 for 2018-2021 years**.