MARIA GRIGORIEVA (LOMONOSOV MOSCOW STATE UNIVERSITY)

GRAPH NET::WORK::SHOP

INTERACTIVE VISUAL EXPLORER APPLICATION FOR ATLAS COMPUTING METADATA
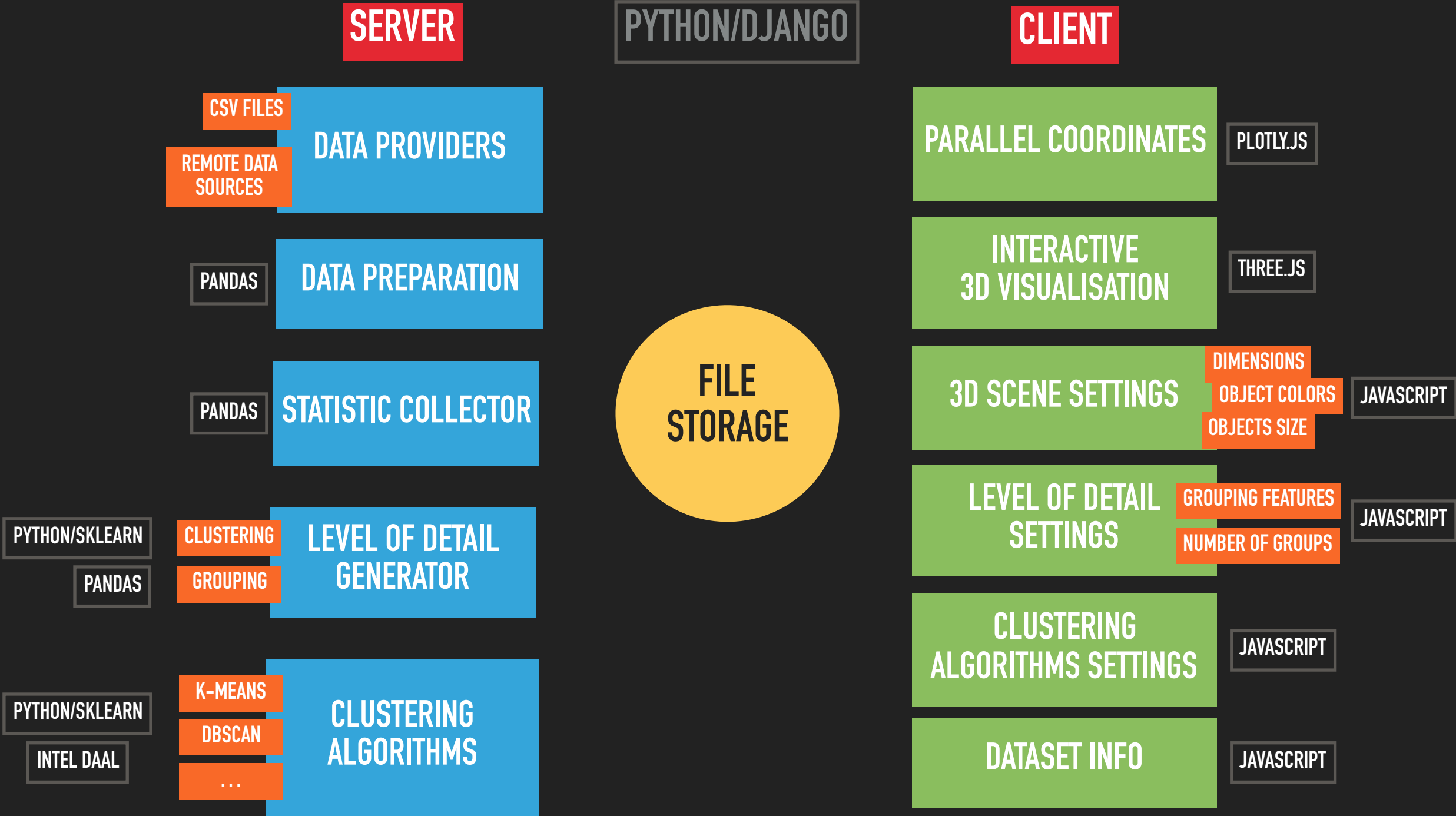
INTERACTIVE VISUAL EXPLORER (INVEX)

▸ InVEx is developing as a generic interactive visual analytics tool for the exploration of big volumes of multidimensional data, based on a combined usage of intellectual data analysis and advanced interactive visualisation techniques.

▸ ATLAS Computing metadata is a test polygon

    ▸ We need to understand how software/hardware is performing and decrease the operational workload.
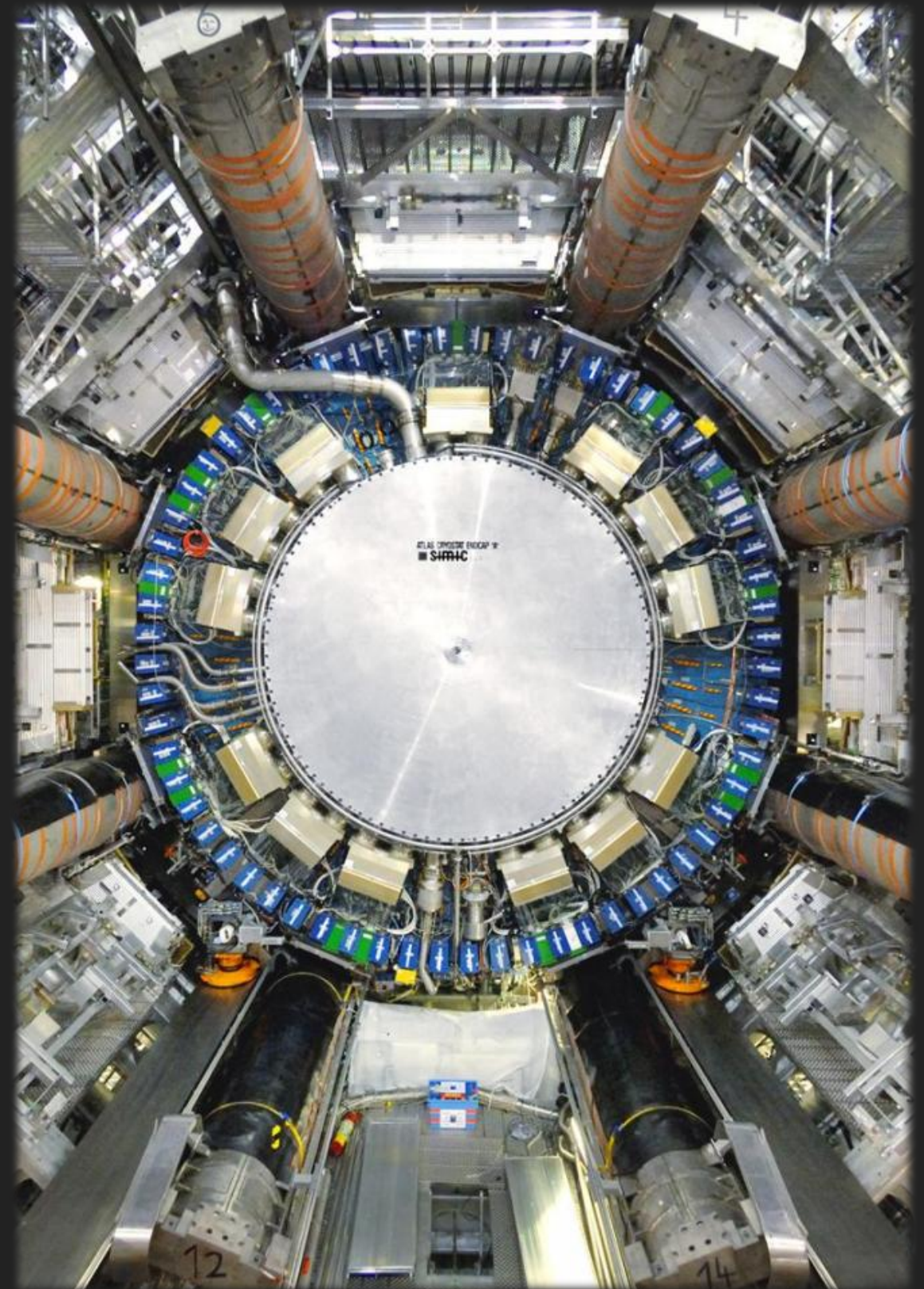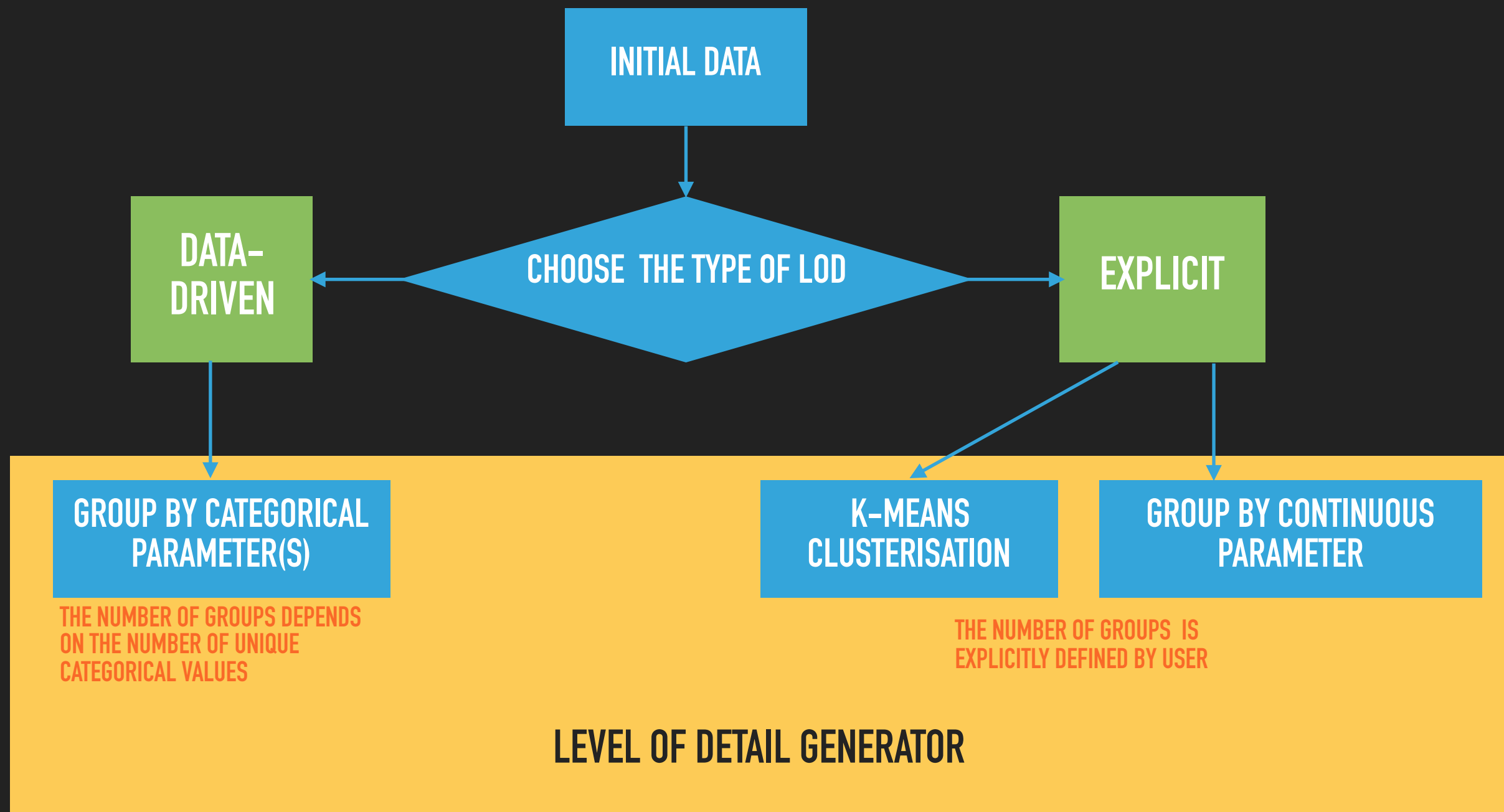
DATA ANALYSIS VS DATA EXPLORATION

▸ In data analysis the user knows what he is looking for, in data exploration, the user doesn't.

  ▸ Data analysis tasks typically investigate specific data instances, and their relation to other instances. The analyst usually has a large understanding of the structure of the data set he is working with. The relations between attributes are typically well understood, or at least the characteristics of a particular attribute are well known.

  ▸ Data exploration tasks are those which attempt to uncover the general structure of the data. Here, the user may not know which attributes best separate or explain the data, may not know the relationships between attributes, and may not even know which attributes are useful.

▸ InVEx is a system for data exploration!
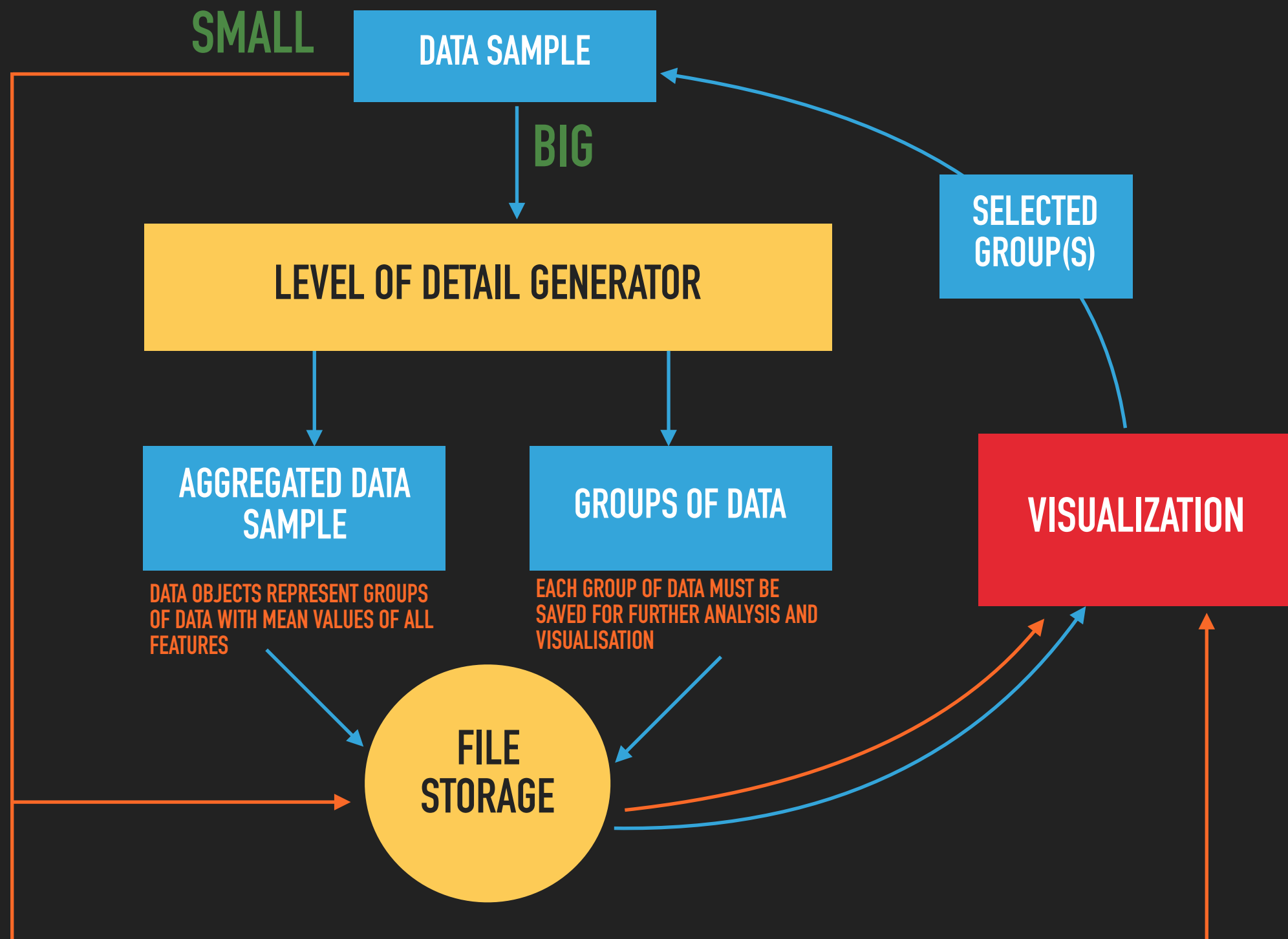
INVEX ARCHITECTURE AND TECHNOLOGY STACK

SERVER   PYTHON/DJANGO   **CLIENT**

CSV FILES

DATA PROVIDERS

REMOTE DATA SOURCES

PANDAS

DATA PREPARATION

PANDAS

STATISTIC COLLECTOR

FILE STORAGE

PYTHON/SKLEARN

CLUSTERING

LEVEL OF DETAIL GENERATOR

PANDAS

GROUPING

K-MEANS

PYTHON/SKLEARN

CLUSTERING ALGORITHMS

DBSCAN

INTEL DAAL

. . .

PARALLEL COORDINATES   PLOTLY.JS

INTERACTIVE 3D VISUALISATION   THREE.JS

DIMENSIONS

3D SCENE SETTINGS   OBJECT COLORS   JAVASCRIPT

OBJECTS SIZE

GROUPING FEATURES

LEVEL OF DETAIL SETTINGS   JAVASCRIPT

NUMBER OF GROUPS

CLUSTERING ALGORITHMS SETTINGS   JAVASCRIPT

DATASET INFO   JAVASCRIPT

LEVEL OF DETAIL GENERATOR IN INVEX

INITIAL DATA

CHOOSE THE TYPE OF LOD

DATA-DRIVEN

EXPLICIT

GROUP BY CATEGORICAL PARAMETER(S)

THE NUMBER OF GROUPS DEPENDS ON THE NUMBER OF UNIQUE CATEGORICAL VALUES

K-MEANS CLUSTERISATION

GROUP BY CONTINUOUS PARAMETER

THE NUMBER OF GROUPS IS EXPLICITLY DEFINED BY USER

LEVEL OF DETAIL GENERATOR

LEVEL OF DETAIL IN INVEX

INVEX DATA SAMPLE INFO PAGE

The exploration of ATLAS PanDA jobs

LEVEL OF DETAIL IN INVEX

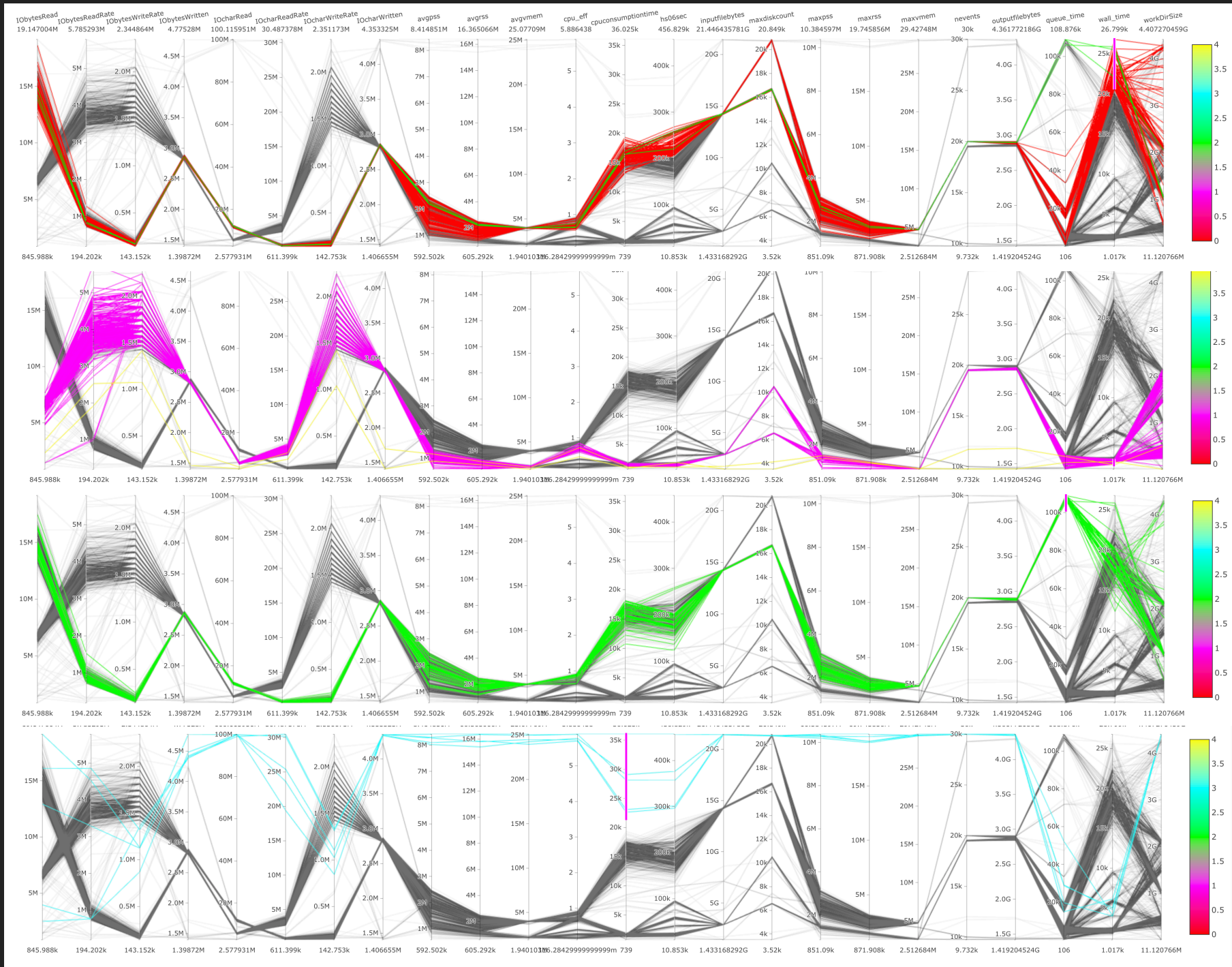# VISUALIZATION OF GROUPS – COMPUTING SITES
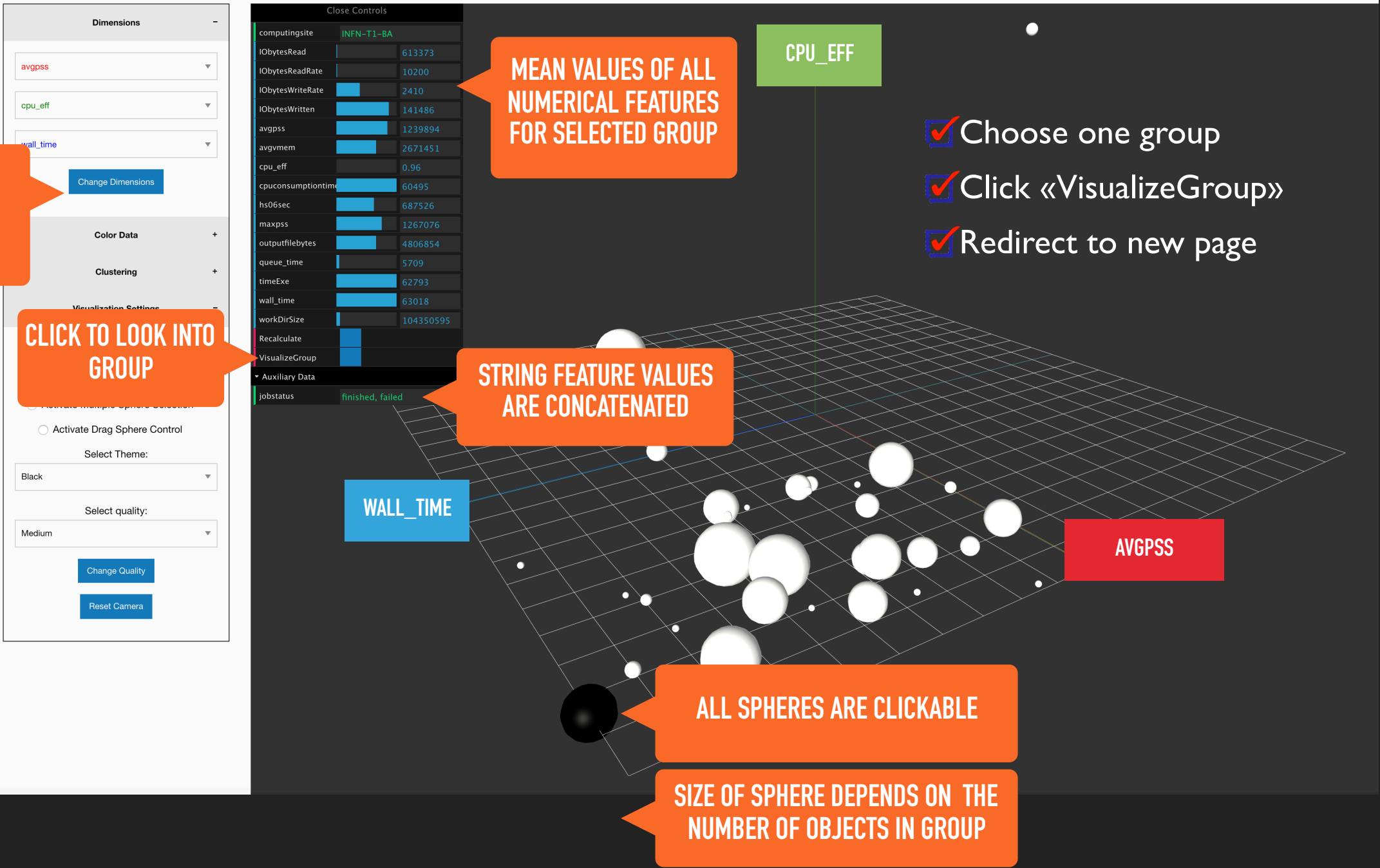
LOOKING INTO CHOSEN GROUP – CLUSTER ANALYSIS WITH 3D VISUALIZATION

FEATURES TRENDS IN PARALLEL COORDINATES

MAX WALL_TIME

MIN WALL_TIME

MAX QUEUE_TIME

MAX CPUCONSUMPTIONTIME

VISUALIZATION OF PARTICLES DATA

Dimensions −

ParticlePx ▼

ParticlePy ▼

ParticlePz ▼

Change Dimensions

Color Data +

Clustering −

Choose clustering algorithm

KMeans ▼

Number of clusters

5

Clusterize

Reset clusters

Visualization Settings +

Close Controls

▷ Huge data samples with particles features ~1M records
▷ Applying K-Means clusterization to the initial data sample, using the LOD Generator
▷ User can choose the number of clusters
▷ Visualizing groups of data objects
▷ The result is the aggregated visual overview of the initial data
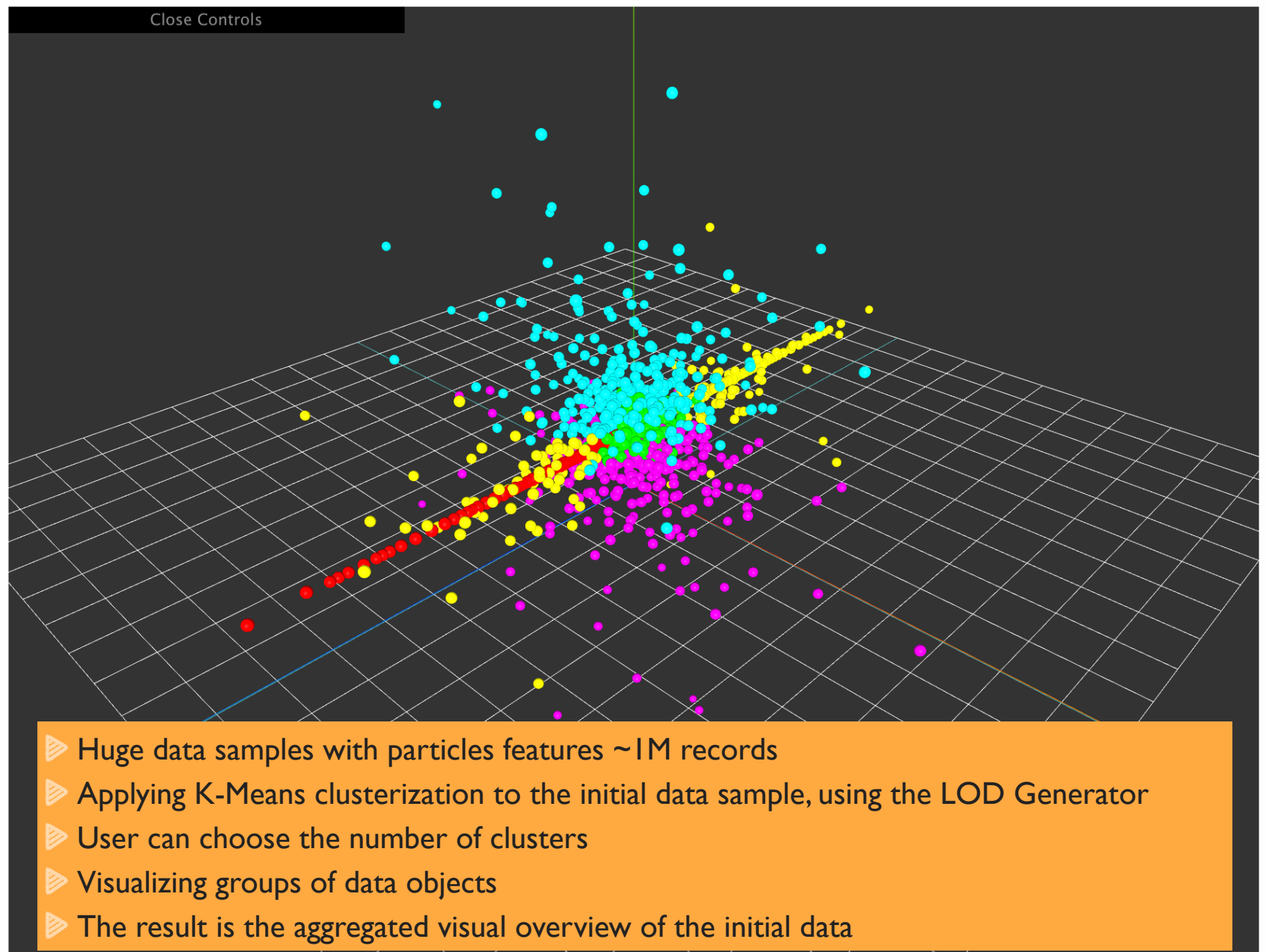
FUTURE

- ▸ More clustering algorithms

    - ▸ text

    - ▸ mixed features (continuous with categorical)

- ▸ Auxiliary data for clusterization

- ▸ Interactive parallel coordinates with linked tables

- ▸ Dimensionality reduction algorithms

- ▸ Replace file storage with database

- ▸ **Search for new use-cases!**

RESOURCES

- ▸ http://vap-dev.tpu.ru/

- ▸ https://github.com/PanDAWMS/InVEx/