

MARIA GRIGORIEVA (LOMONOSOV MOSCOW STATE UNIVERSITY)

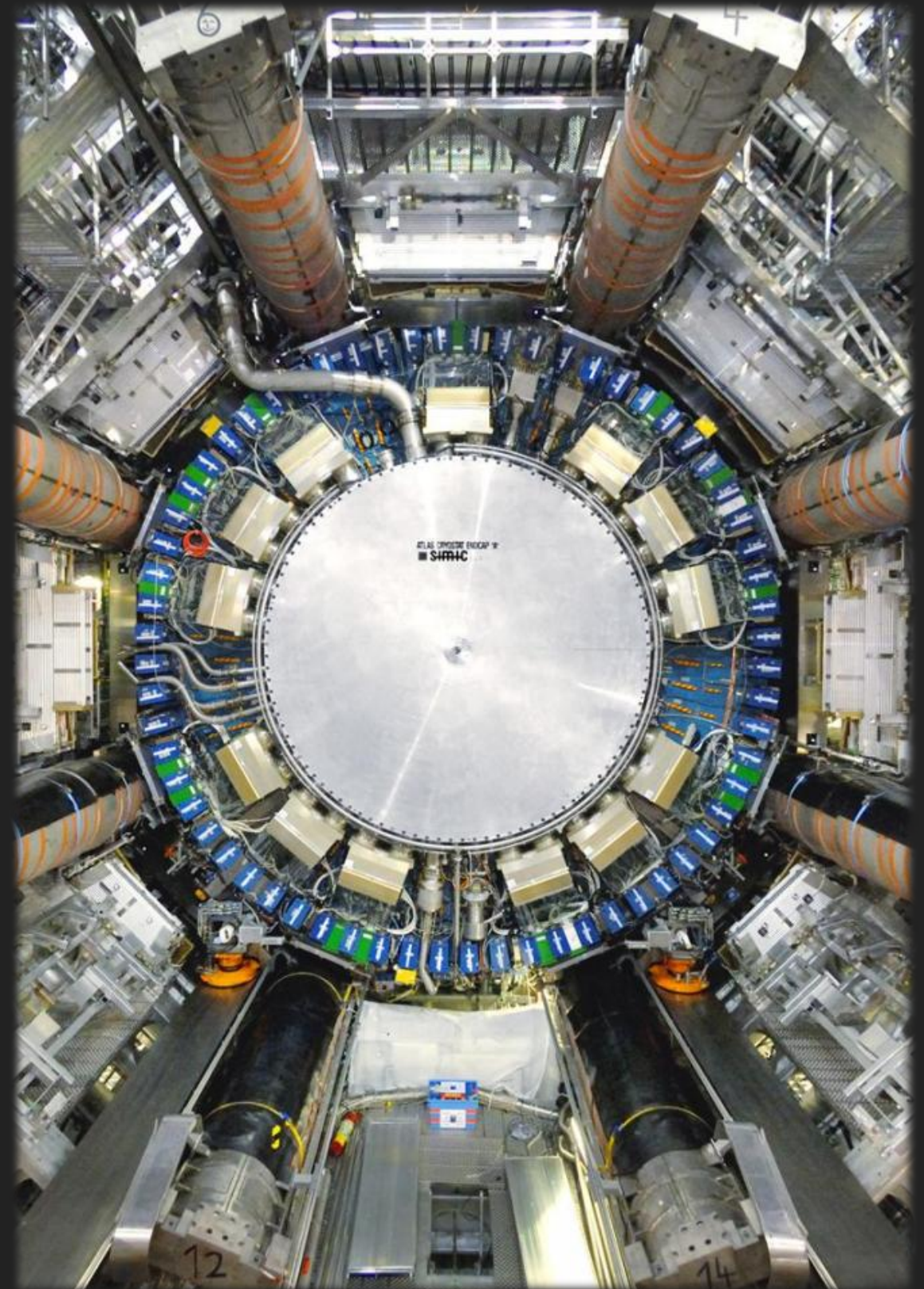
GRAPH NET::WORK::SHOP

---

**INTERACTIVE VISUAL EXPLORER APPLICATION  
FOR ATLAS COMPUTING METADATA**

## INTERACTIVE VISUAL EXPLORER (INVEX)

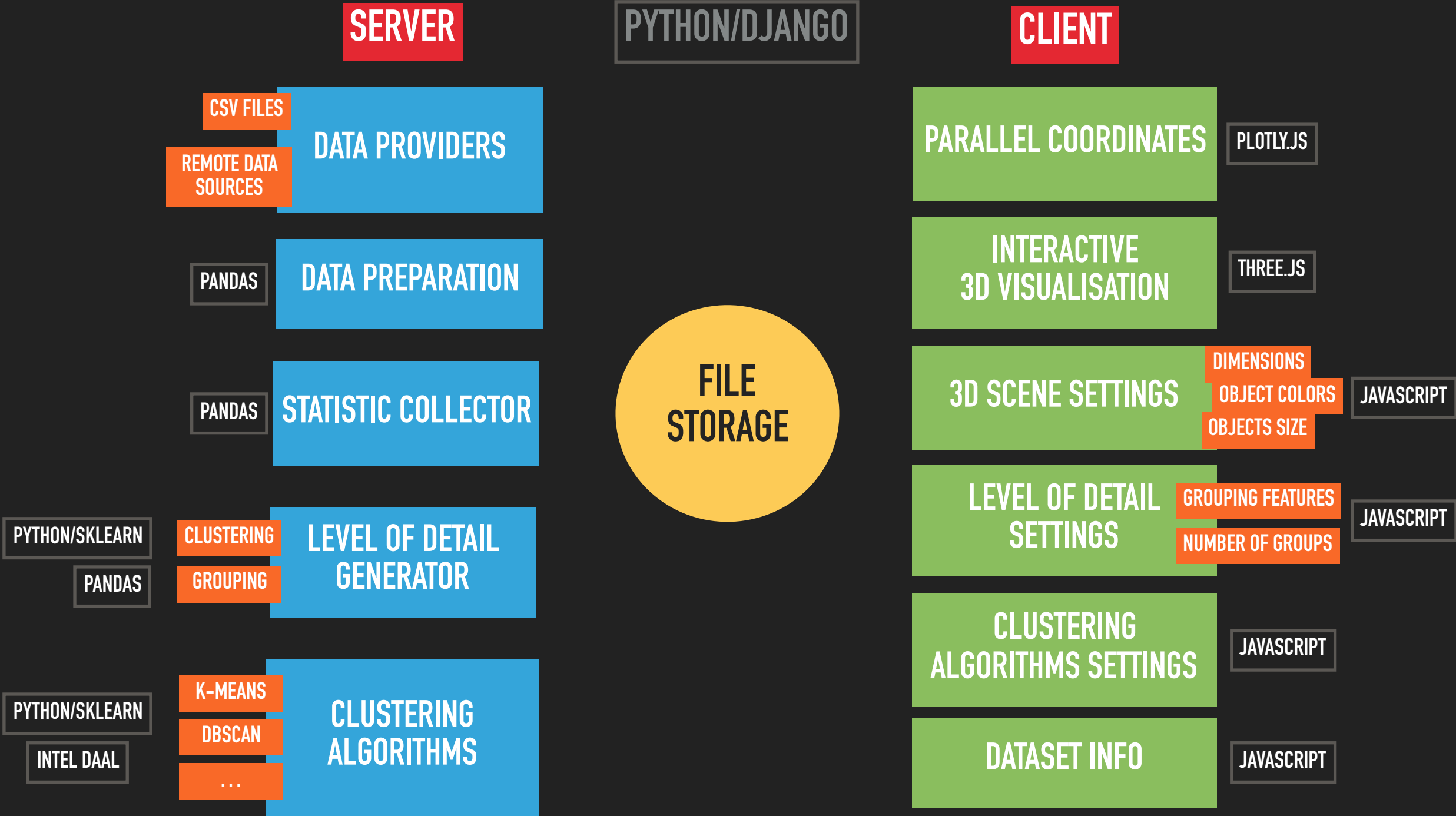
- ▶ InVEx is developing as a generic interactive visual analytics tool for the exploration of big volumes of multidimensional data, based on a **combined** usage of **intellectual data analysis** and **advanced interactive visualisation techniques**.
- ▶ ATLAS Computing metadata is a test polygon
  - ▶ We need to understand how software/hardware is performing and decrease the operational workload.



## DATA ANALYSIS VS DATA EXPLORATION

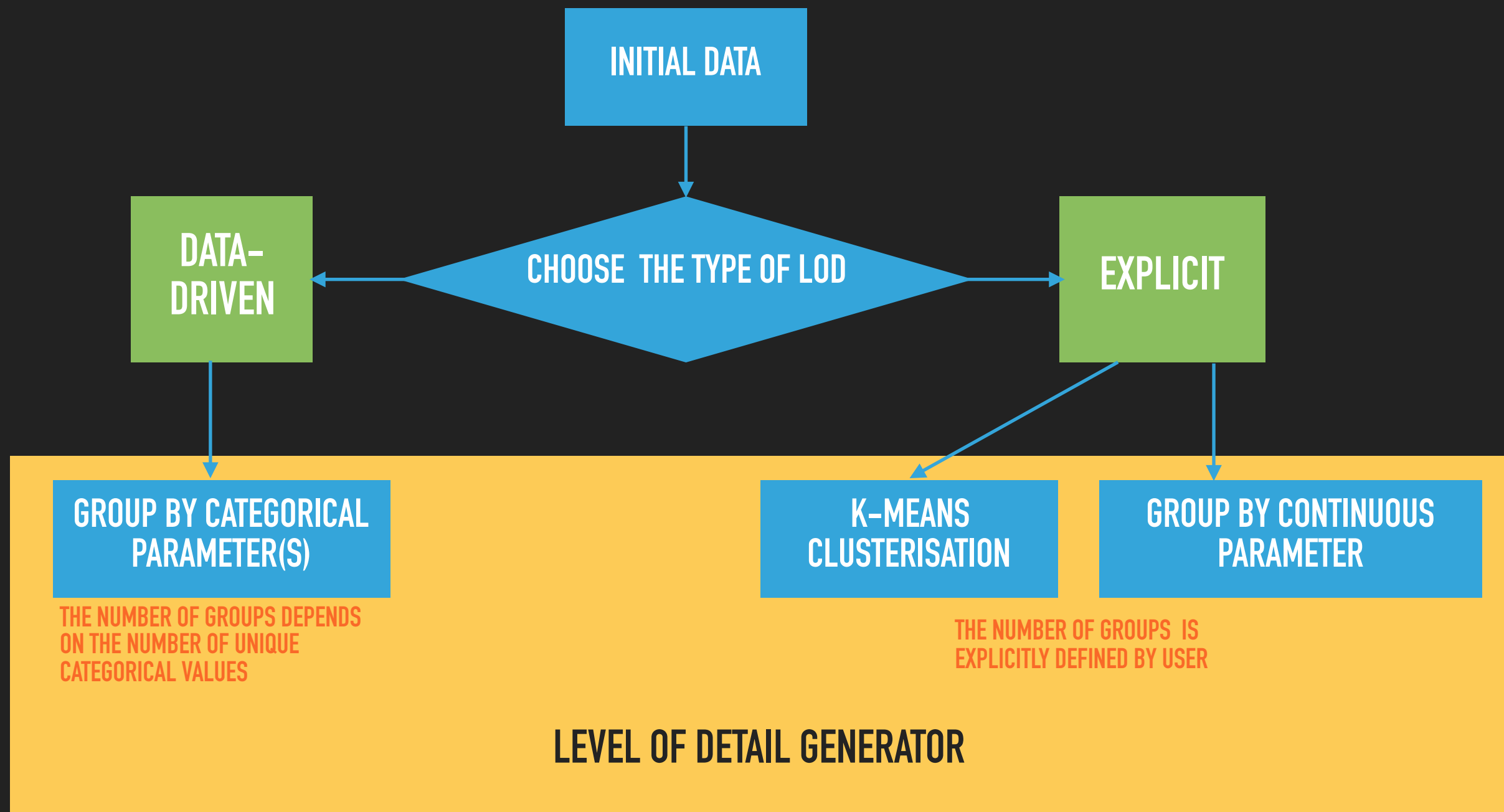
- ▶ In **data analysis** the user knows what he is looking for, in **data exploration**, the user doesn't.
- ▶ **Data analysis** tasks typically investigate specific data instances, and their relation to other instances. The analyst usually has a large understanding of the structure of the data set he 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

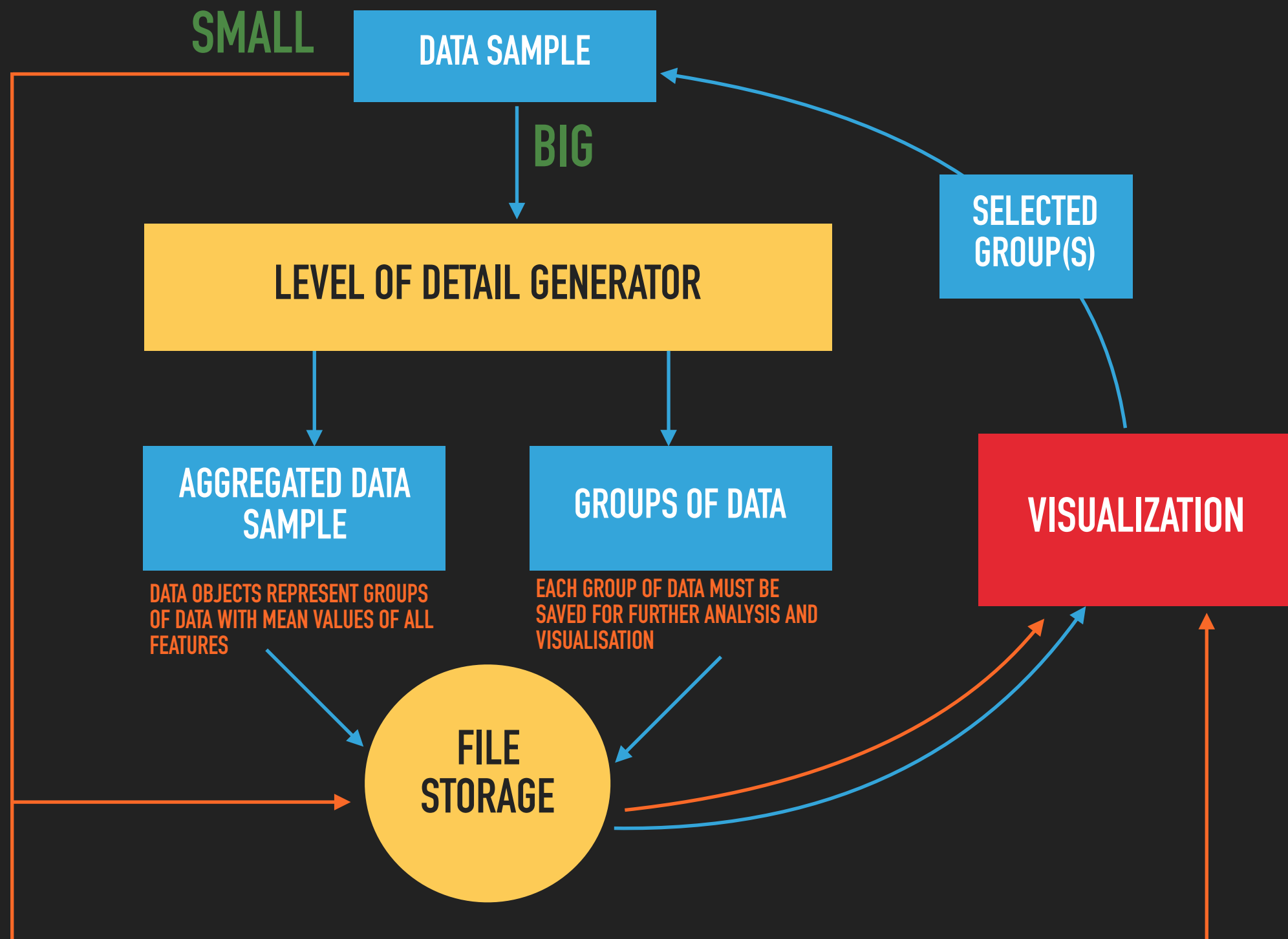




# LEVEL OF DETAIL GENERATOR IN INVEX



# LEVEL OF DETAIL IN INVEX





LEVEL OF DETAIL IN INVEX

Activate Level-of-Detail Generator

Choose LoD Type:

MiniBatchKMeans(sklearn/python) clusterizator

Set the number of clusters and select numerical continuous features using "group" selector

3000

K-MEANS  
CLUSTERISATION

EXPLICIT

Dataset Name		Number of Records		Index	
1559332254.224132		158820		pandaid	

group	select	feature_name	feature_type	measure_type	min	mean	max	std	percentage_missing
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesRead	float64	continuous	0.00	879,301.71	229,699,940.00	3,334,701.14	9.79%
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesReadRate	float64	continuous	0.00	26,191.36	731,891,712.00	1,950,195.22	9.79%
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesWritten	float64	continuous	0.00	224,716.82	4,834,692.00	65,903.03	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgpss	float64	continuous	0.00	1,426,694.52	1,683,919.00	139,078.84	1.45%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgrss	float64	continuous	0.00	1,444,575.11	1,689,116.00	140,529.46	1.45%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgvmem	float64	continuous	0.00	3,054,447.60	3,151,974.00	166,782.28	1.45%
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	cpu_eff						8	0.00%
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	cpuconsumpt						013.86	0.00%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	hs06sec						5,954.66	0.00%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	maxpss						2,890.70	1.45%

Activate Level-of-Detail Generator

Choose LoD Type:

Group by numerical continuous parameter

Select single numerical continuous parameter for grouping using "group" selector and set the number of groups

50

GROUP BY CONTINUOUS  
PARAMETER

EXPLICIT

Dataset Name		Number of Records		Index	
1559332254.224132		158820		pandaid	

group	select	feature_name	feature_type	measure_type	min	mean	max	std	percentage_missing
<input type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesRead	float64	continuous	0.00	879,301.71	229,699,940.00	3,334,701.14	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesReadRate	float64	continuous	0.00	26,191.36	731,891,712.00	1,950,195.22	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesWritten	float64	continuous	0.00	224,716.82	4,834,692.00	65,903.03	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgpss	float64	continuous	0.00	1,426,694.52	1,683,919.00	139,078.84	1.45%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgrss	float64	continuous	0.00	1,444,575.11	1,689,116.00	140,529.46	1.45%

Activate Level-of-Detail Generator

Choose LoD Type:

Group by nominal/ordinal parameter(s)

Select from one to several categorical features for grouping using "group" selector. The sequence of checking features = the sequence of grouping

GROUP BY CATEGORICAL  
PARAMETER(S)

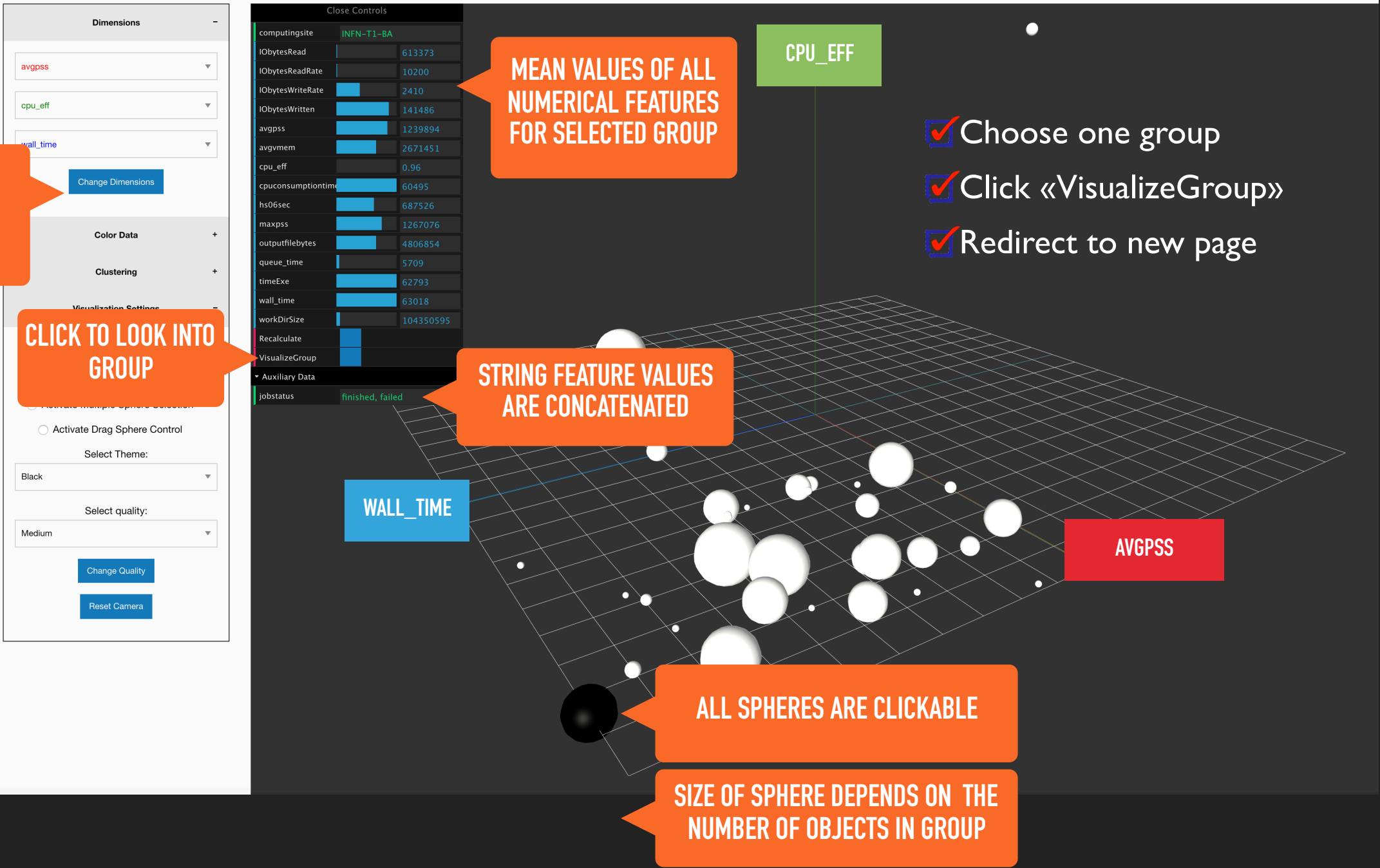
DATA-DRIVEN

Dataset Name		Number of Records		Index	
1559332254.224132		158820		pandaid	

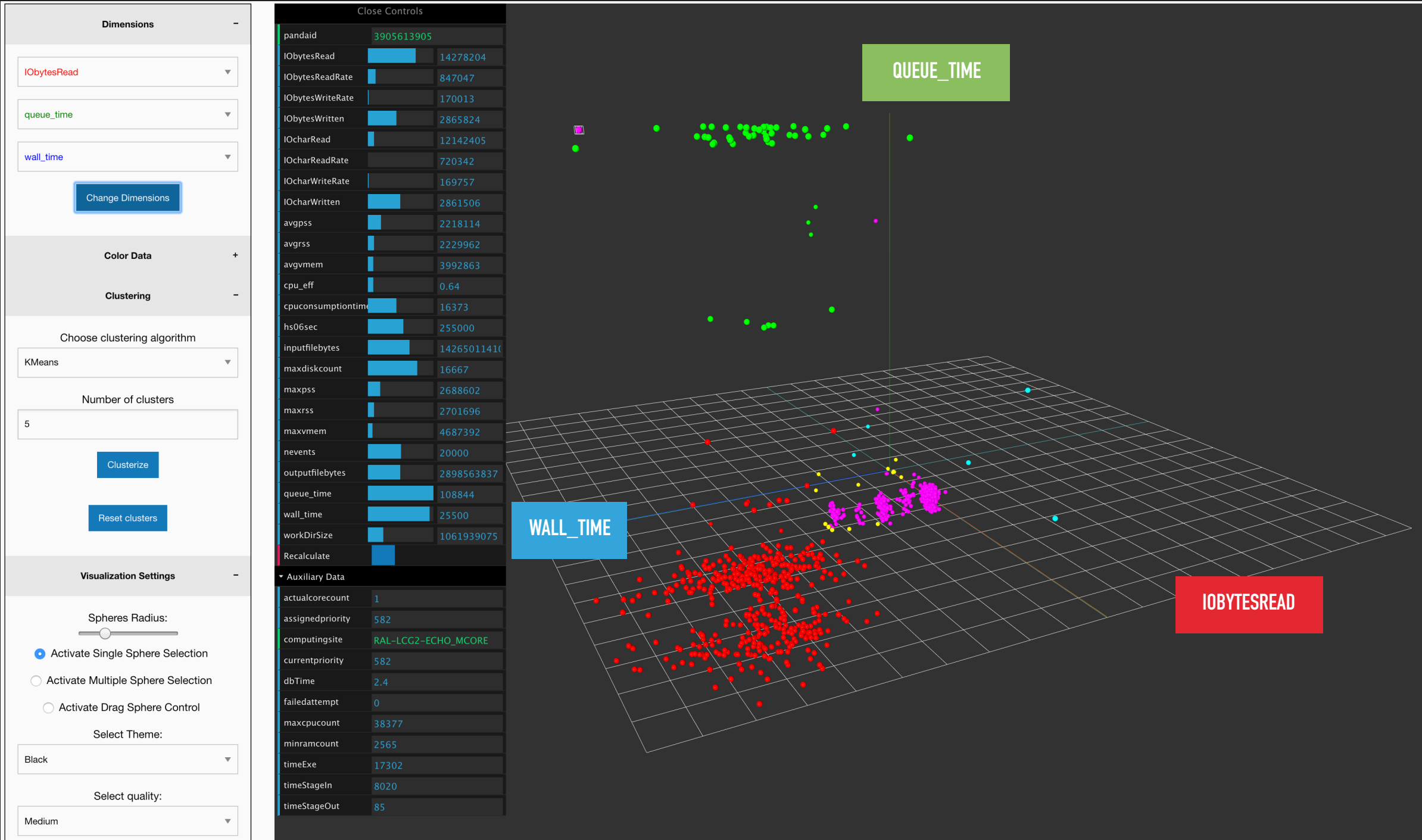
group	select	feature_name	feature_type	measure_type	min	mean	max	std	percentage_missing
<input type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesRead	float64	continuous	0.00	879,301.71	229,699,940.00	3,334,701.14	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesReadRate	float64	continuous	0.00	26,191.36	731,891,712.00	1,950,195.22	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	IBytesWritten	float64	continuous	0.00	224,716.82	4,834,692.00	65,903.03	9.79%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgpss	float64	continuous	0.00	1,426,694.52	1,683,919.00	139,078.84	1.45%
<input type="checkbox"/>	<input checked="" type="checkbox"/>	avgrss	float64	continuous	0.00	1,444,575.11	1,689,116.00	140,529.46	1.45%



# VISUALIZATION OF GROUPS – COMPUTING SITES



LOOKING INTO CHOSEN GROUP – CLUSTER ANALYSIS WITH 3D VISUALIZATION



Close Controls

pandaId

3905613905

IBytesRead

14278204

IBytesReadRate

847047

IBytesWriteRate

170013

IBytesWritten

2865824

IOcharRead

12142405

IOcharReadRate

720342

IOcharWriteRate

169757

IOcharWritten

2861506

avgpss

2218114

avgrss

2229962

avgmem

3992863

cpu\_eff

0.64

cpuconsumptiontime

16373

hs06sec

255000

inputfilebytes

14265011410

maxdiskcount

16667

maxpss

2688602

maxrss

2701696

maxvmem

4687392

nevents

20000

outputfilebytes

2898563837

queue\_time

108844

wall\_time

25500

workDirSize

1061939075

Recalculate

Auxiliary Data

actualcorecount

1

assignedpriority

582

computingsite

RAL-LCG2-ECHO\_MCORE

currentpriority

582

dbTime

2.4

failedattempt

0

maxcpucount

38377

minramcount

2565

timeExe

17302

timeStageIn

8020

timeStageOut

85

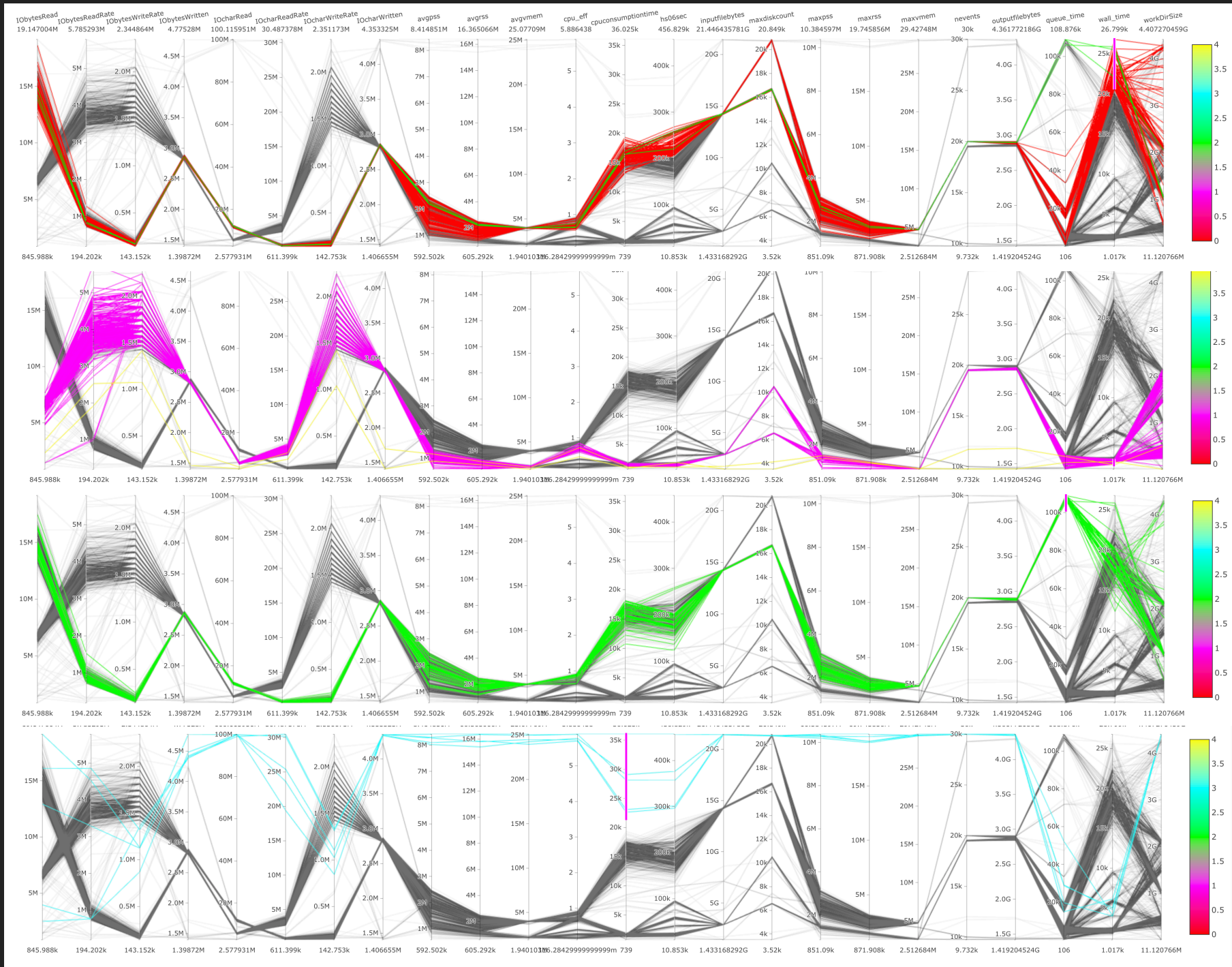
3D Visualization

QUEUE\_TIME

WALL\_TIME

IBytesRead

# 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 ▼

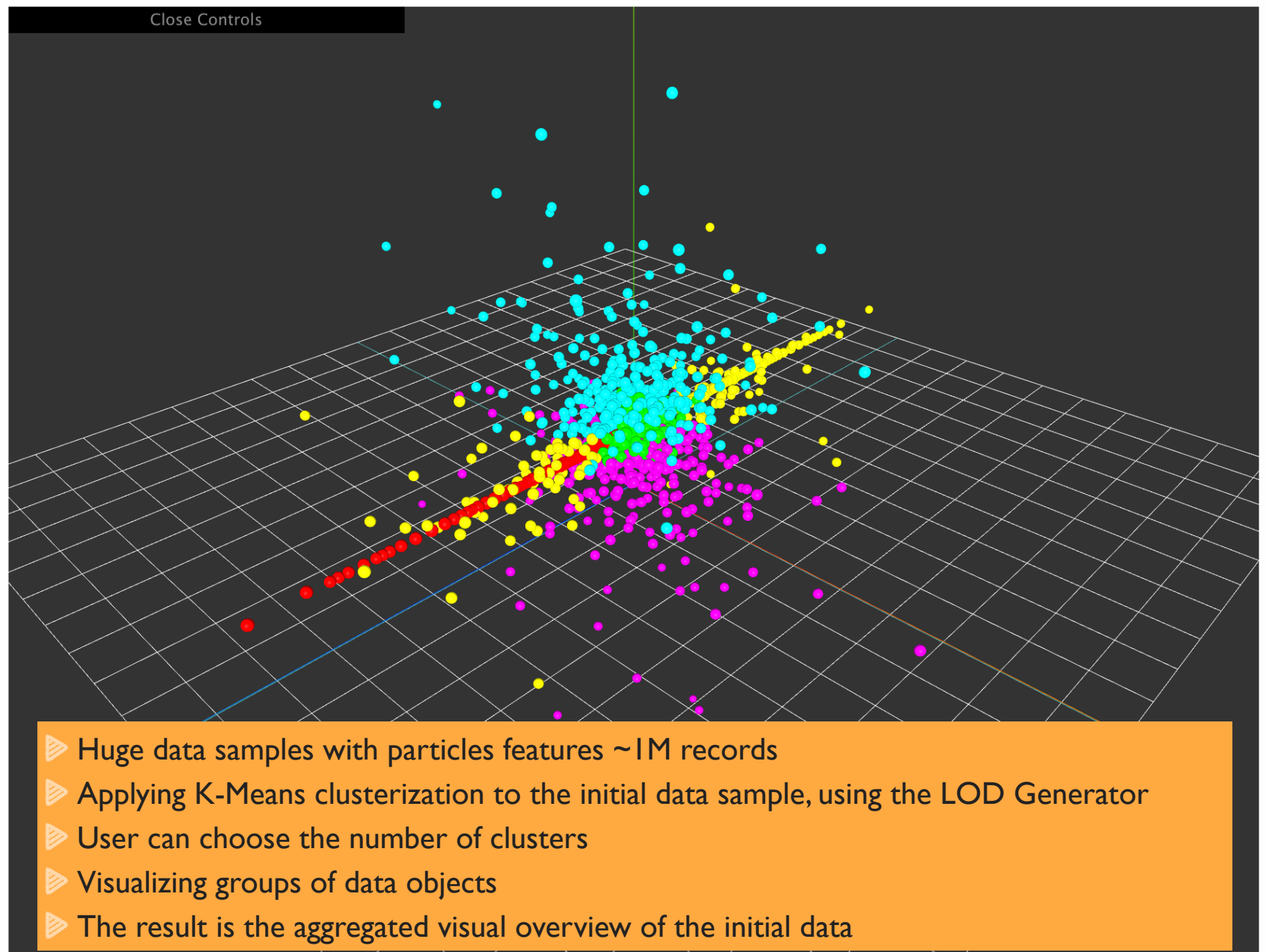
Number of clusters

5

Clusterize

Reset clusters

Visualization Settings +





## 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/>