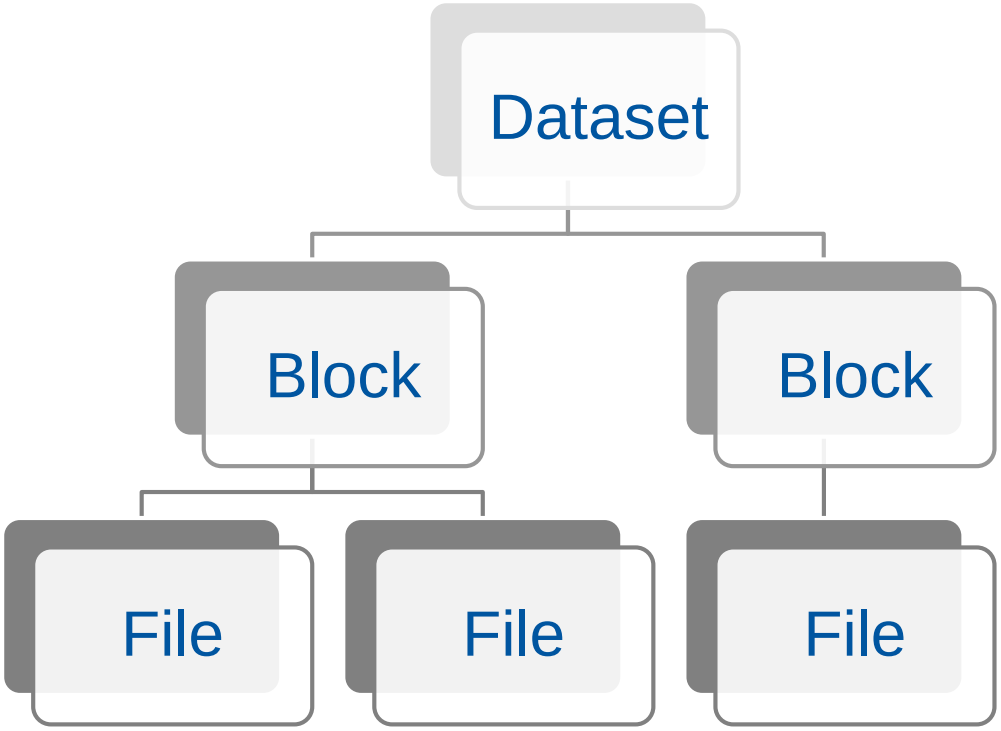
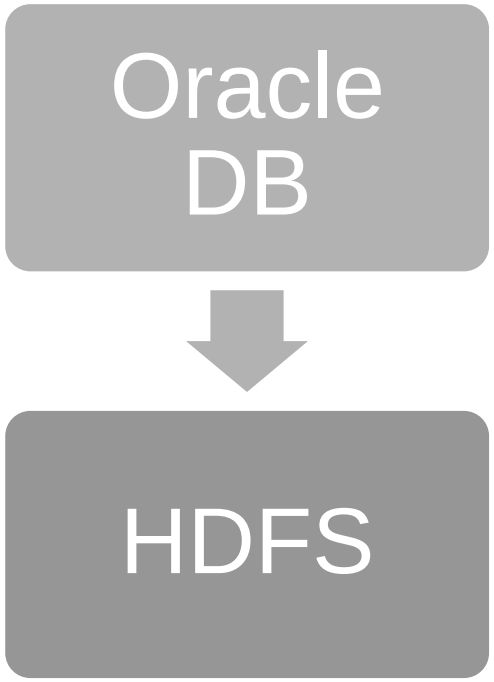

CERN

CMS Dataset Replica Monitoring

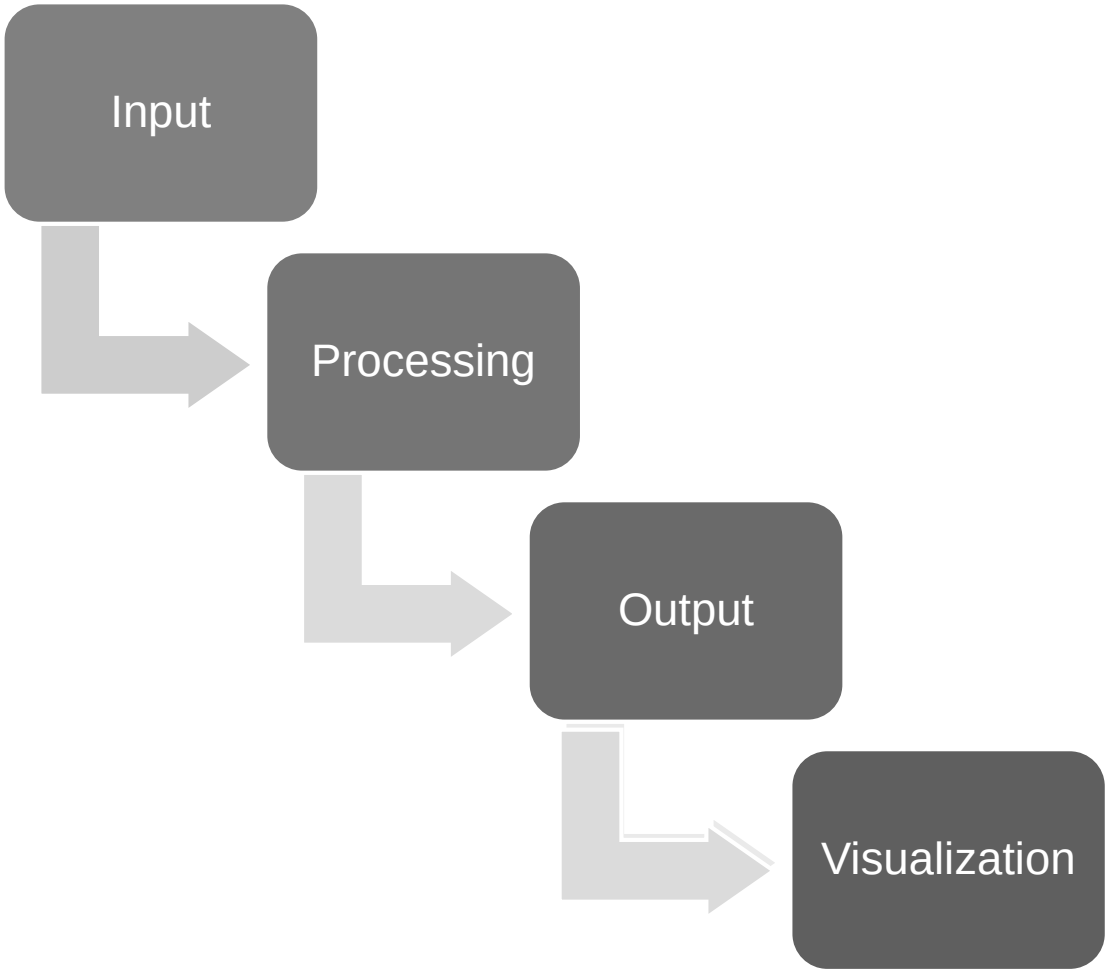
Presenter: Aurimas Repečka
Supervisors: Nicolo Magini, Daniele Bonacorsi
Technical supervisors: Valentin Kuznetsov, Luca Menichetti

Introduction





Monitoring Process Structure



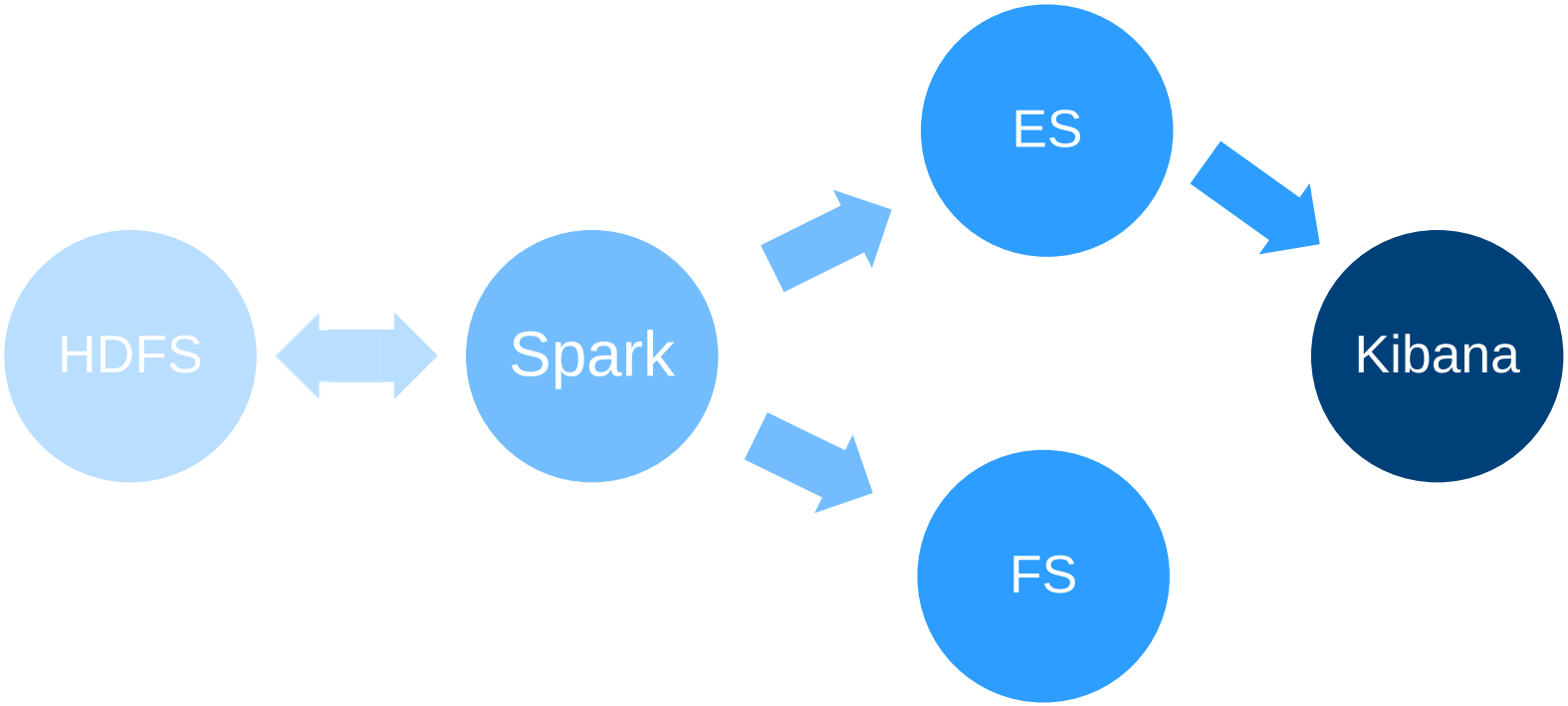
Input – Snapshots

- Daily snapshots in CSV format
- Size for snapshot: 2-3.5GB
- Fields
  - Now
  - Node name (Node Tier)
  - Dataset name (Acquisition Era, Data Tier)
  - Block name
  - Block replicas node bytes
  - Block replicas node files
  - Etc…

Processing – Spark Job

- Group keys
- Result values
- Aggregations
  - Sum, count, first, last, min, max, mean, delta, avg-day
- Data ordering
- Data filtering
- Local/Yarn modes

Output



Hadoop-Elasticsearch connection

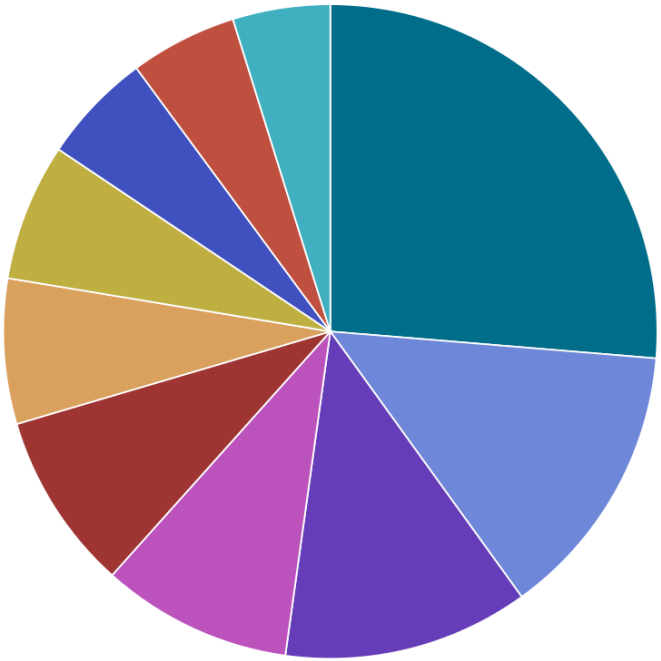
- Allows HDFS files transferring directly to Elasticsearch
- Reading one or multiple HDFS files with any number of partitions
- Configurable Elasticsearch parameters: node, port, resource(index/type)
- Data schema applied dynamically from user specified json file

Performance

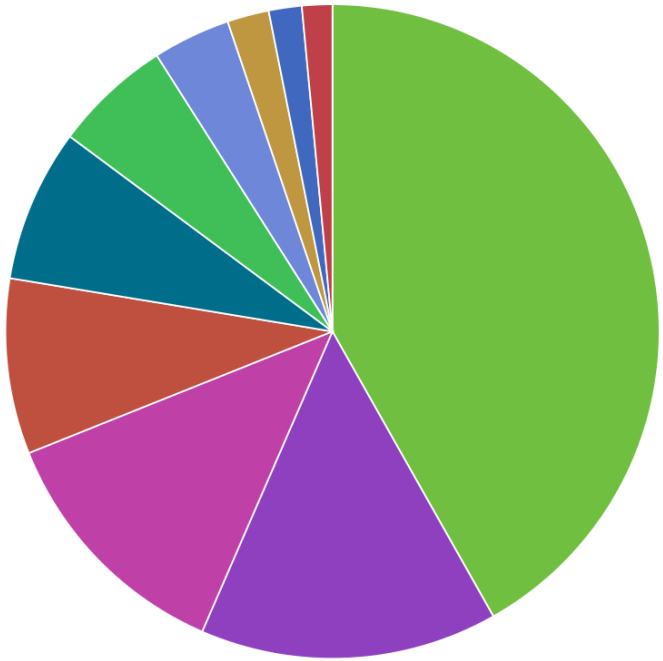
- Group keys: now, user group, acquisition era, data tier, node kind
- Results: node bytes, destination bytes
- Aggregations: sum, sum

Interval	Input	Cores	Memory	Output	Duration
1 day	~3GB	65	361472MB	~600KB	~1.6min
1 month	~100GB	65	361472MB	~18MB	~4.3min
3 months	~310GB	65	361472MB	~52MB	~9.7min
1 year	~1.1TB	65	361472MB	~186MB	~28min

Visualization



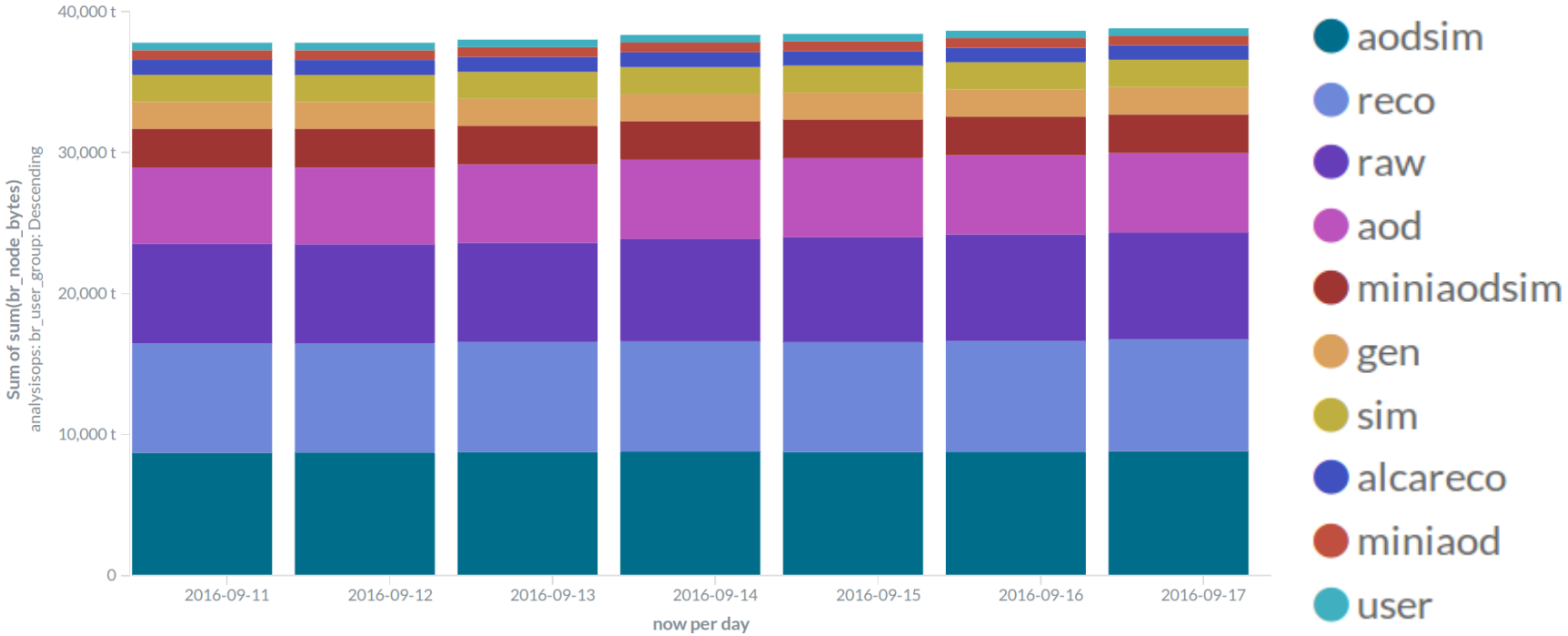
analysisops: br_user_group



dataops: br_user_group

- run2016b
- runiispring16dr80
- run2016g
- runiifall15dr76
- runiispring15dr74
- run2016d
- run2016e
- run2016c
- run2015d
- run2016f
- runiisummer15gs
- runiispring15prepremix
- runiisummer15wmlhegs
- run2015e
- runiiwinter15gs
- runiisummer16dr80ba...
- runiispring16fspremix

Visualization



Deployment

- Development
  - Spark job submission – analytix cluster
  - Elasticsearch/kibana – personal virtual environment
- Pre-production
  - WMArchive node
- Production (future plan)
  - DMWM umbrella
  - Elasticsearch/kibana – central CERN IT service

Conclusion

- Enables CMS block replica analysis by aggregating and visualizing data snapshots
- Efficient
- Fully-covered
- Highly configurable

https://github.com/aurimasrep/PhedexReplicaMonitoring


CERN