Implementation of ATLAS Distributed Computing monitoring dashboards using InfluxDB and Grafana

Thomas Beermann, Aleksandr Alekseev, Dario Barberis, Sabine Crépé-Renaudin, Johannes Elmsheuser, Ivan Glushkov, Michal Svatos, Armen Vartapetian, Petr Vokac, Helmut Wolters

CHEP 2019, Adelaide

ATLAS EXPERIMENT

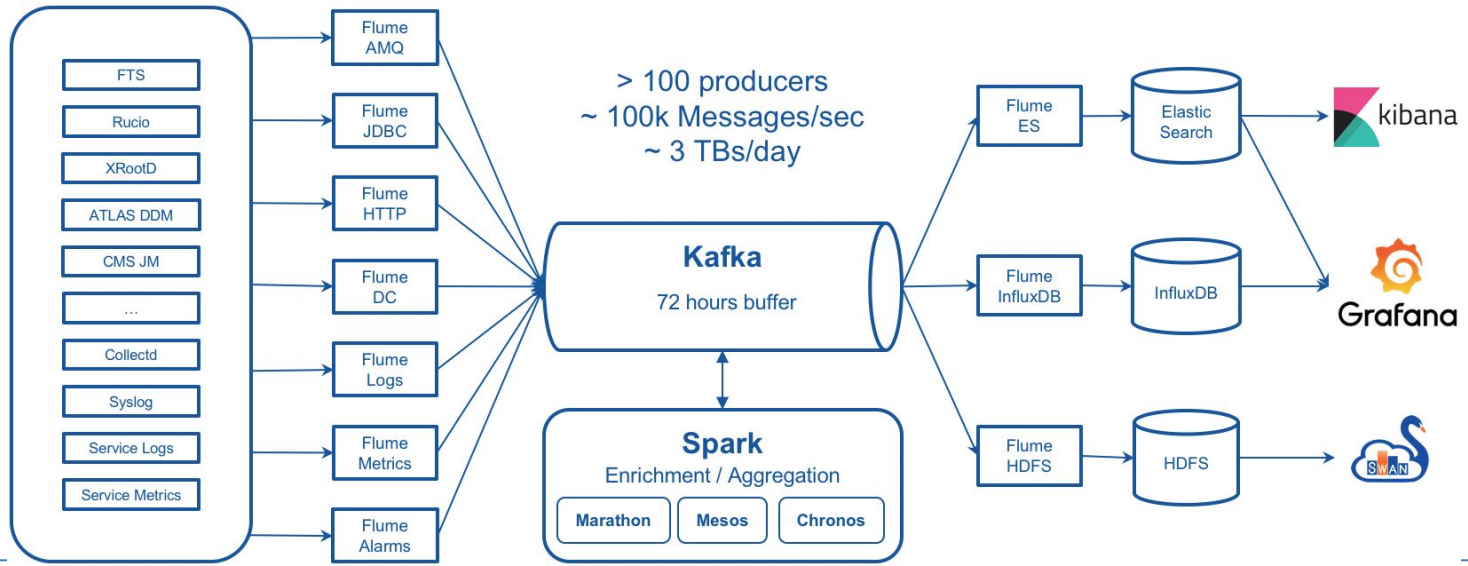
Overview

- The monitoring and accounting infrastructure used by ATLAS during Run-1 and Run-2 was based on custom applications developed ~10 years ago by CERN IT and ATLAS members
  - After all these years the old dashboards showed some aging effects:
    - Data retrieval was slow due to huge amounts of data stored in Oracle databases
    - The main developers left and with them in-depth maintenance knowledge
    - It was difficult to add more features
  - The dashboards worked well enough for the old requirements but were in need of an overhaul
- In 2016 the CERN IT monitoring group started to build a new Unified Monitoring Architecture based on modern open source components
    - The new ATLAS jobs accounting and DDM dashboards presented in here have been developed on top of this infrastructure

CERN MonIT infrastructure

- Provides Monitoring as a Service for the CERN Data Centre, IT Services and the WLCG collaborations
- Collect, transport, store and process metrics and logs for applications and infrastructure

Sources > Transport > (Processing) > Storage > Access



Data Sources for ADC Dashboards

- Transfer Monitoring (from ATLAS DDM system Rucio):
  - Rucio produces events for each successful or failed file transfers between sites and for file deletions
  - Pilots and Rucio clients produce traces for grid file accesses or user downloads
  - Both the events and traces are sent to a central ActiveMQ instance
  - Collected from there by Monit to inject into the Kafka pipeline
- DDM / Storage Accounting (Rucio):
  - Daily Oracle procedures running on the Rucio database to create a dump of the data registered in Rucio per storage / project / datatype / etc.
  - Aggregated data is sent again to ActiveMQ and forwarded to Kafka
- Jobs Accounting (from ATLAS WFMS PanDA):
  - Collector from Monit runs every 10 minutes to fetch submitted, pending, running, finalising and completed jobs directly from the PanDA database
  - Directly injected into Kafka

Processing

- All the processing / aggregations are done using the streaming platform Kafka together with Spark for the stream-processing
- Transfer monitoring:
  - Events and Traces are unified to a common format and then enriched with topology information (Country, Cloud, Tier, etc.) from the ATLAS Grid Information System (AGIS)
  - Then the data is aggregated in 1 minute bins grouped by activity, destination/source, protocol, etc.
- DDM / Storage accounting
  - Processing is only used to enrich the data with same topology info as for transfers
- Job accounting:
  - Again topology information added from AGIS
  - Then data is aggregated in hourly bins
  - Bins are continuously updated as data is flowing in

Backends

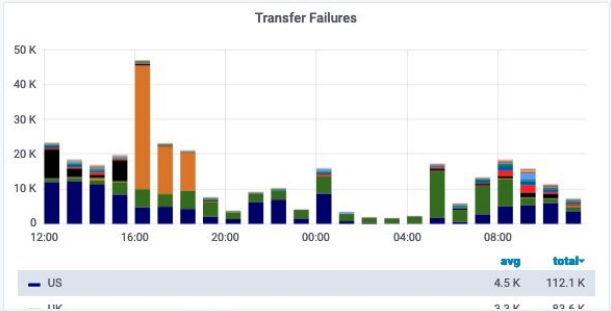
- InfluxDB
  - Used to keep the time series data for the transfer monitoring and jobs accounting
  - Kept for 30 days in original binning
  - Then auto down-sampling to larger bins (1h / 1d / 7d / 30d) for up to 5 years
- Elasticsearch
  - Used to store transfer details and the storage accounting data
  - Data retention of 30 days
- HDFS
  - Used for long term storage of all the data
  - Data stored in compressed JSON
  - Data kept forever
  - Can be used in batch and data-intensive analytics workflows

Dashboards

- All dashboards are available in a central Grafana instance
- The dashboards are split in three different categories:
  - Production: Stable / official dashboards that are used by experts / shifters. Can only be edited by admins.
  - Development: Copies of Production dashboards to test new features as well as completely new dashboards that are still in testing. Can be edited by users having the editor role.
  - Playground: Open to anybody having access to Grafana. Can be used to build your own specific dashboards that might not be interesting to other people, e.g. specific site view.

Transfer Dashboard

- The transfer dashboard provides a lot of useful information for Shifters, Experts and Management
- Possible to drill down from a global view to single file transfers / deletions
- Main entrypoint is the transfer efficiency matrix to quickly spot problems
- Histograms for transfer / deletion volume, successes, failures, throughput
- Error messages are classified and grouped to identify common problems



Efficiency

	CA	CERN	DE	ES	FR	IT	ND	NL	RU	TW	UK	US
CA	92%	97%	97%	65%	94%	93%	89%	91%	82%	72%	95%	84%
CERN	86%	96%	97%	96%	97%	96%	96%	93%	90%	85%	96%	87%
DE	76%	98%	91%	79%	87%	89%	96%	85%	82%	94%	93%	88%
ES	96%	98%	94%	99%	97%	93%	97%	89%	84%	99%	98%	90%
FR	95%	95%	98%	97%	96%	96%	96%	88%	92%	91%	95%	90%
IT	81%	93%	92%	85%	85%	97%	97%	86%	95%	97%	96%	83%
ND	89%	93%	95%	87%	74%	97%	98%	89%	92%	99%	98%	90%
NL	93%	43%	98%	98%	97%	97%	97%	97%	99%	99%	95%	90%
RU	85%	99%	99%	98%	95%	99%	98%	83%	100%	99%	98%	93%
TW	84%	100%	98%	98%	98%	95%	98%	71%	100%	-	97%	94%
UK	44%	74%	63%	28%	33%	30%	29%	19%	81%	77%	71%	74%
US	52%	88%	70%	53%	54%	62%	73%	77%	63%	88%	76%	80%

DDM transfers

Destination cloud	Source cloud	efficiency	volume	successes	failures
NL	CERN	42.54	91.29 TB	44863	60592
US	US	79.67	138.07 TB	98650	25178
UK	FR	32.72	22.95 TB	9775	20102
US	FR	54.46	40.12 TB	22606	18906
US	CA	51.66	35.55 TB	17157	16056
UK	US	74.26	72.24 TB	41505	14384
US	UK	75.80	55.64 TB	37410	11943
US	DE	70.29	53.26 TB	26176	11063
UK	DE	62.91	27.63 TB	14570	8589
US	ES	52.61	11.82 TB	9325	8401
UK	NL	18.75	1.03 TB	1609	6972
US	IT	61.93	11.85 TB	10099	6207
UK	ES	28.13	2.33 TB	2300	5875

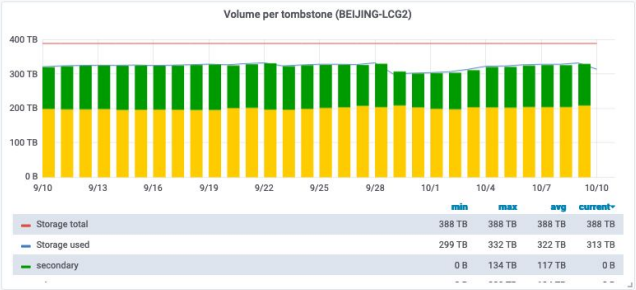
1 2 3 4 5 6 7 8 9

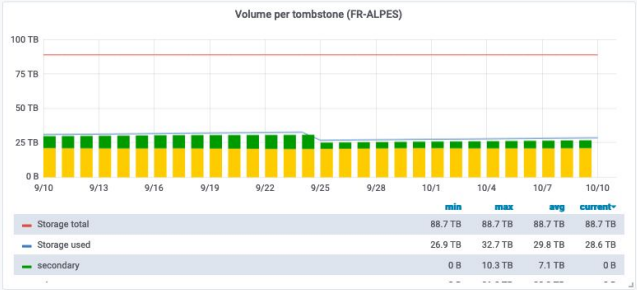
Site Accounting

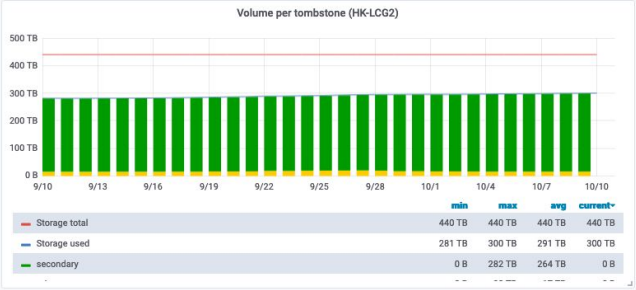
- Useful to show current occupation of sites / storage endpoints
- A lot of different groupings available to identify the data type stored at site:
  - datatype, physics_stream, prod_step, etc.
- Data can be filtered by tier, country, cloud, site, token, etc.
- Additionally, storage provided numbers are added to spot possible inconsistencies and to show free space a site

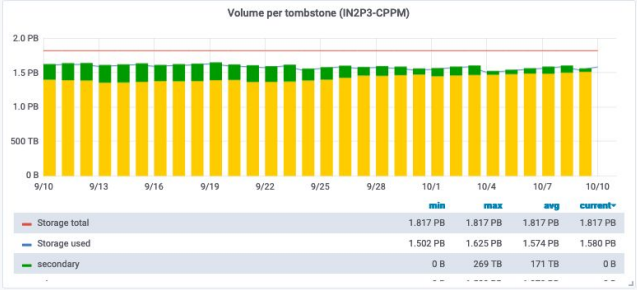
Snapshot — Last 24 hours

Endpoint	Files	Bytes	Storage used	Storage total
BEIJING-LCG2_DATADISK	712936	275.98 TB	278.32 TB	310.00 TB
BEIJING-LCG2_LOCALGROUPDISK	140386	8.77 TB	8.77 TB	18.00 TB
BEIJING-LCG2_SCRATCHDISK	232592	44.19 TB	44.23 TB	60.00 TB
FR-ALPES_DATADISK_DATALAKES	251082	26.78 TB	28.42 TB	60.01 TB
FR-ALPES_SCRATCHDISK_DATALAKES	56082	101.69 GB	129.83 GB	28.73 TB
GRIF-IRFU_DATADISK	3714862	1.49 PB	1.47 PB	1.72 PB
GRIF-IRFU_LOCALGROUPDISK	597888	370.36 TB	370.82 TB	395.82 TB
GRIF-IRFU_SCRATCHDISK	741428	74.37 TB	75.43 TB	101.16 TB
GRIF-LAL_DATADISK	2897600	901.38 TB	900.67 TB	1.00 PB
GRIF-LAL_LOCALGROUPDISK	318304	14.43 TB	16.29 TB	26.39 TB
GRIF-LAL_SCRATCHDISK	339702	29.80 TB	32.97 TB	43.98 TB



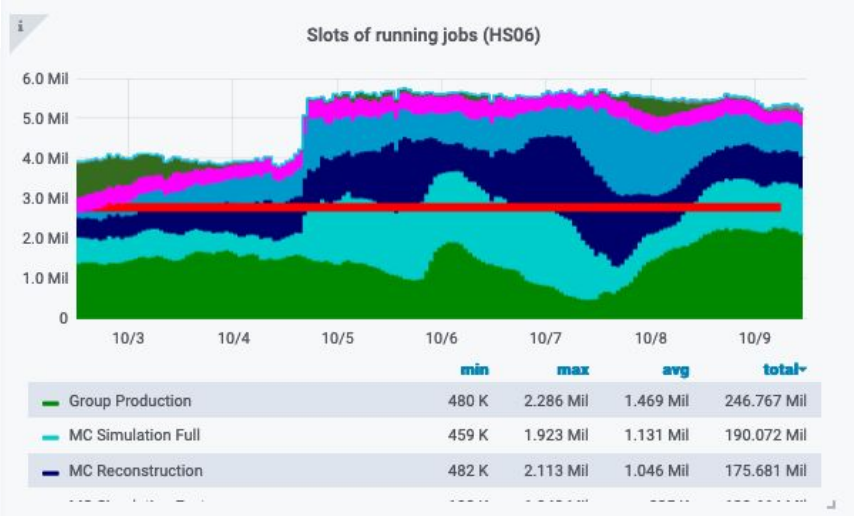






Jobs Accounting

- Useful overview to find potential problems with the WFMS and jobs on the grid
- Plots available for transient and permanent job states
- Accounting of number of job slots, wall clock time, HS06, etc.
- Plots can be grouped / filtered by a lot of different parameters like site / job type / processing type / queue



Conclusion

- ATLAS Distributed Computing has a coherent set of monitoring and accounting dashboards
- Removing the custom, in-house developed tools and using instead widely available open-source technologies brings a lot of benefits:
  - Better scalability
  - Better maintainability
  - Easier development of new features
- But there are also some minor disadvantages like the reduced number of visualization options in Grafana
- Overall, we are well set for LHC Run-3

Questions?