

State and evolution of data analytics infrastructure at ESOC

Rui Santos

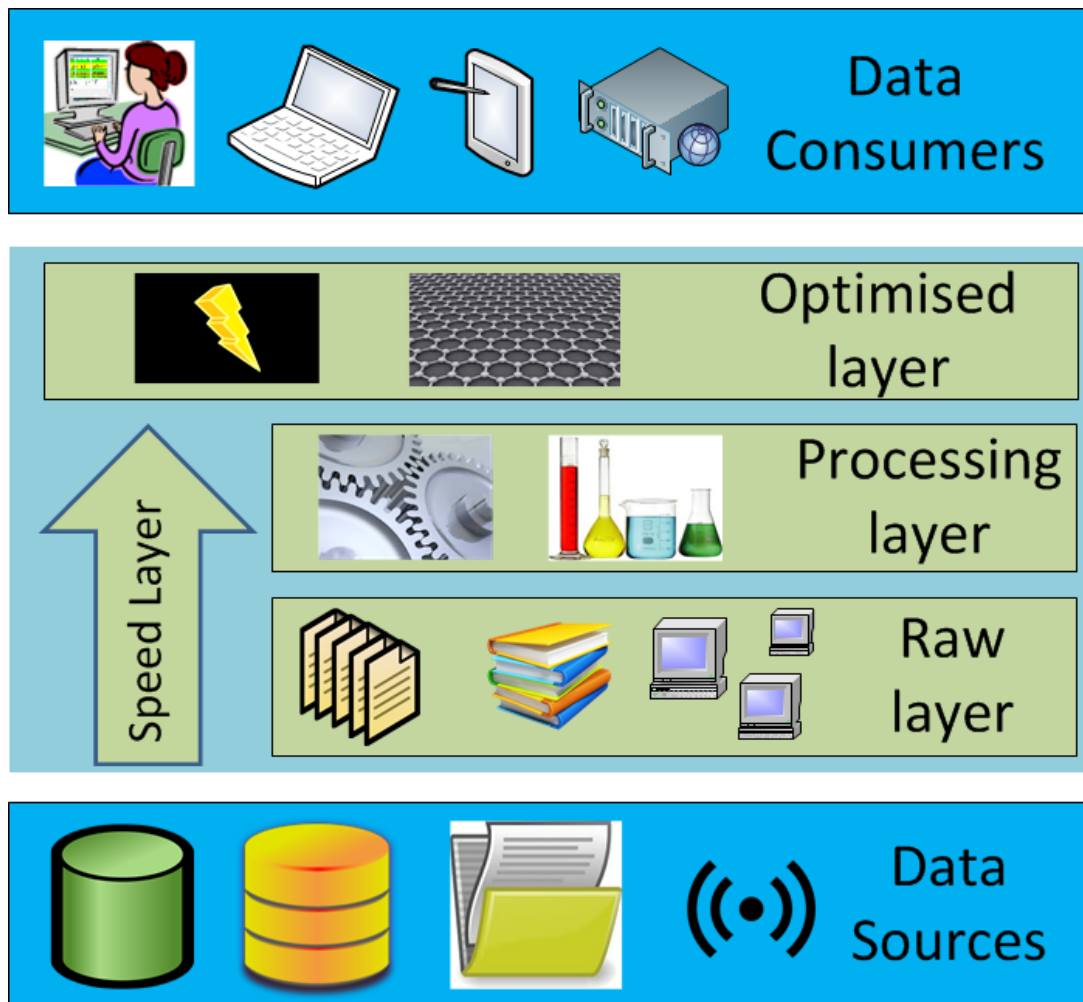
Thanks to James Eggleston & Jose Martinez Heras

03/03/2020

© 2020 by European Space Agency. Published by The Aerospace Corporation with permission



Architecture – Big Data Layers



•Raw layer:



- Full fidelity
- Organized in file/folder structures

•Processing layer: **MAPR**



- Distributed computation and data transformation

•Optimised layer: **APACHE HBASE**



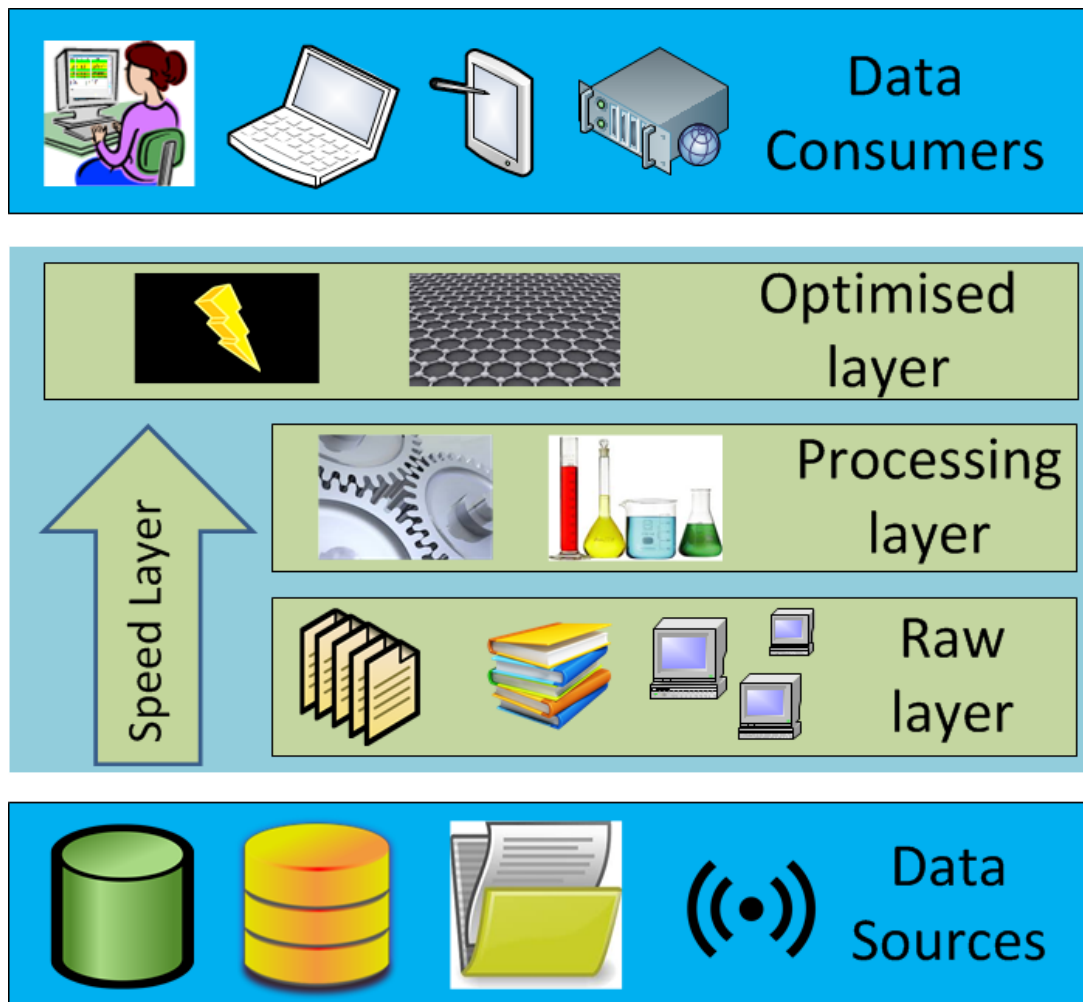
Elasticsearch

- Structured data
- Organised by consumer use cases

•Speed layer:



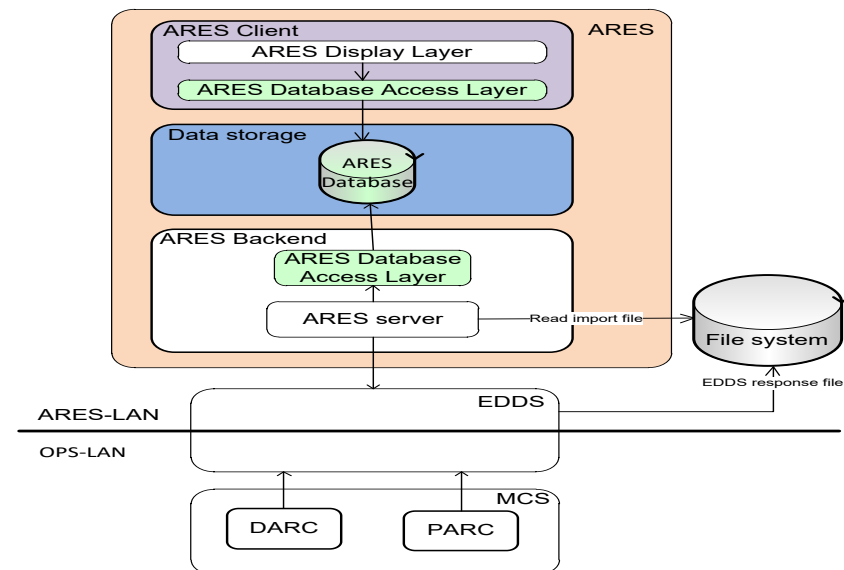
- Low latency
- Consumed from sources



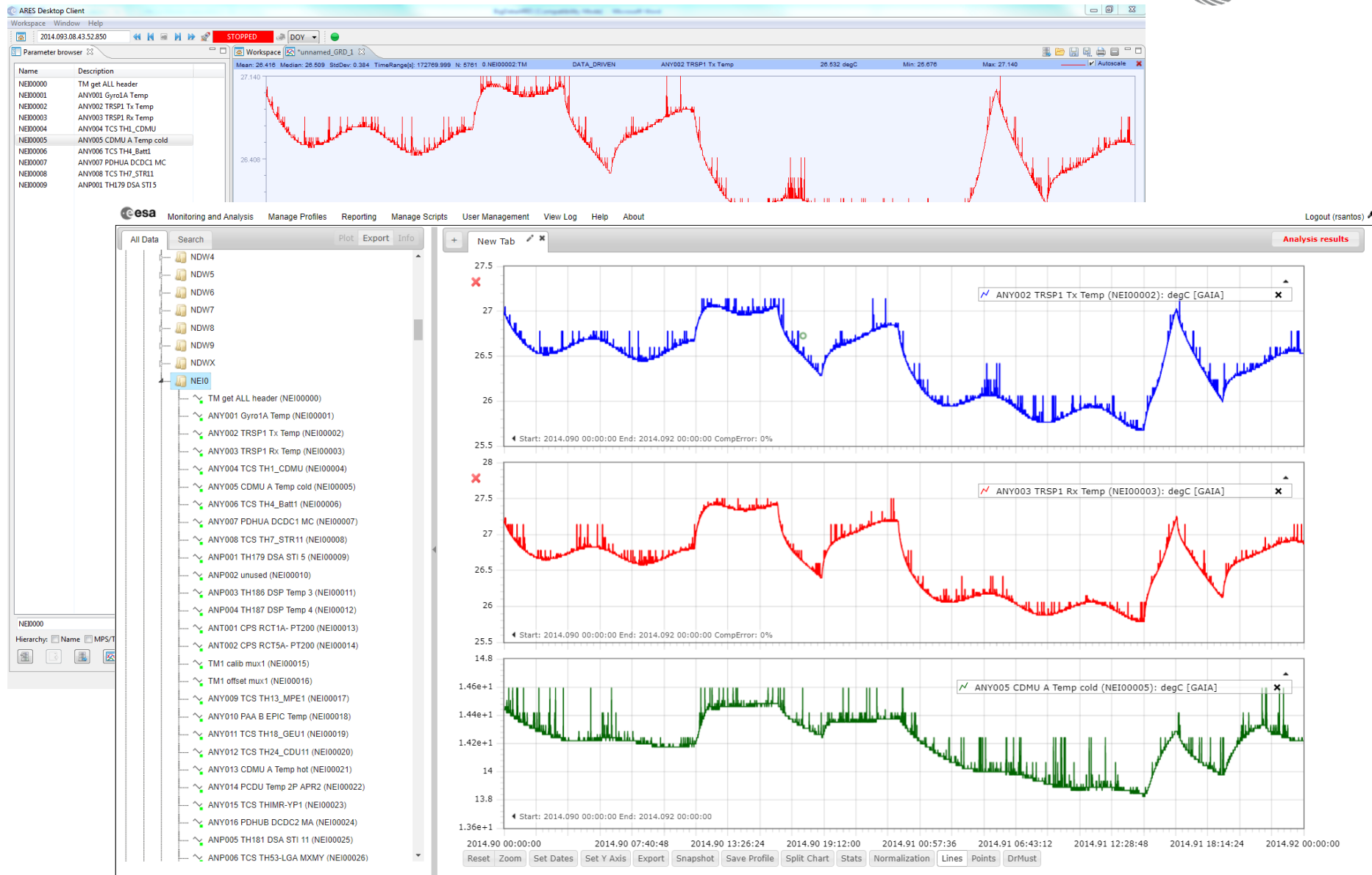
- Storage/Infrastructure:
 - ARES (Analysis and Reporting System)
- Data Accessibility (includes preparation, processing, retrieval and enhancement)
 - PyARES
 - NoveltyDetection system
 - Dependency Finder
 - LTDP products and collections

Infrastructure: Introduction to ARES

- ARES (Analysis and Reporting System) is an off-line data analysis system
- Is an evolution of existing ESA system (Mission Utility Support tools – MUST)
- Build on top of the EGOS User Desktop Framework (EUD)
- Mainly for storage of spacecraft house-keeping data in engineering format
 - Supports generic storage of files, parameters and events
- Optimized for fast data retrieval



Displays: Graphical display



Displays: Alphanumeric display

ARES Desktop Client

Workspace Window Help

2014.093.08.43.53.603 STOPPED DOY

Workspace DSYA0100

Parameter	Description	Value	Unit	Parameter	Description	Value	Unit
0.MF0000...	SPACECRAFT ID	789		0.NSY10...	SYSMGR SGM SC mode	FOM	

esa Monitoring and Analysis Manage Profiles Reporting Manage Scripts User Management View Log Help About Logout (rsantos)

All Data Search Plot Export Info

GAIA

GAIAAND

GAIAGRD

Profiles

Scripts

GAIA params

GAIAAND paramtables

AND System AND (DSYA0001)

AND TC Link Monitoring (DSYA0010)

AND SVM Systems SOM 1 (DSYA0100)

AND SVT-0 Day 1 Handover Step 2 (DSYA0901)

AND SVT-0 Day 2 Handover Step 2 (DSYA0902)

AND SVT-1A Day 1 Handover Step 2 (DSYA0911)

AND SVT-1B Day 1 Handover Step 2 (DSYA0921)

AND Timing Overview (DSYA1001)

AND LEOP Sep Seq Monitor Part 1 (DSYA5000)

AND LEOP Sep Seq Monitor Part 2 (DSYA5001)

AND LEOP DSA dep monitor Part 3 (DSYA5002)

AND Separation Sequence Overview (DSYA500A)

AND Separation Sequence Step Timings (DSYA500B)

AND SS 2 - TM Downlink Start (DSYA502)

AND SS 3 - Gyros Switch On (DSYA503)

AND SS 4 - PLM Bipods Release (DSYA504)

AND SS 5-1 - CPS PCDU Config (DSYA5051)

AND SS 5-2 - CPS Pressure Xder + TLV (DSYA5052)

AND SS 5-3 - CPS py ma switch off (DSYA5053)

AND SS 5-4 - CPS Prop Release TLV 1 (DSYA5054)

AND SS 5-5 - CPS Prop Release TLV 2 (DSYA5055)

AND SS 5-6 - CPS Thrusters Prime (DSYA5056)

AND SS 5-7 - CPS Raise PR Inlet Pres (DSYA5057)

AND SS 5-8 - CPS Open Low Flow LVs (DSYA5058)

AND SS 5-9 - CPS Prop Tank Pressure (DSYA5059)

AND SS 6 - Start Thermal Control (DSYA506)

AND SS 7 - Start TC Link Monitor (DSYA507)

AND SS 8 - AOCs FM Configuration (DSYA508)

New Tab SVM Sy...

Normal Delta

Timestamp 2014.093 08:43:53

Refresh Export Snapshot <<< << < > >> >>> Auto-refresh

Search:

Parameter	Description	Value	Unit	Parameter	Description	Value	Unit
MF000002	SPACECRAFT ID	789	none	NSY10007	SYSMGR SGM SC mode	FOM	none
NSY10006	SYSMGR objectState	FOM	none	NSY10110	SYSINIT FDIR Level	LEVEL_1	none
NAC19001	AOCsMGR objectState	AOCs_TSM	none	NDW29005	CDMU active PM board	PM_A	none
NPF20001	PF Mgr Object state	ON	none	NDW46001	Active TTR board	TTR_A	none
NPL30001	PLMGR objectState	ON	none	NDW52001	Selected SGM	SGM_A	none
				NDW56005	PM OBT seconds	69065028	none
NAC13002	SAM objectState	SAM_INACTIVE	none	S2KTC013	TCO ONBOARD RAD DELAY	0	none
NAC13003	SAM elapsedTimeRR	0.375	s	NDW61001	MMU set state	OPERATIONAL	none
NAC13004	SAM elapsedTimeSSA	0.625	s	NDW61002	MMU set isFailure	FALSE	none
NAC13005	SAM elapsedTimeSSB	0	s				
NAC13006	SAM elapsedTimeSUN0	8943.6201	s	NPD20001	PDHUSSET maState	MA_ON	none
NAC13007	SAM elapsedTimeSAA45	0	s	NPD20002	PDHUSSET mcState	AVAILABLE	none
NAC13008	SAM elapsedTimeSLEW	0	s	NPD20014	PDHUSSET isMcFailed	FALSE	none
NAC13009	SAM elapsedTimeSUN45	0	s	NPD20015	PDHUSSET isMaDcDcFailed	FALSE	none
NAC13010	SAM isCompleted	TRUE	none	NPD20016	PDHUSSET isMaDcDcBFailed	FALSE	none
NTR81001	TRSP SET Tx state	ON	none	NST80001	STRSET objectState	ON	none
NTR81002	TRSPSET is Failed	FALSE	none	NST80002	STRSET isFailed	FALSE	none
NTR81006	TRSPSET nominal	TRSP1	none	NST80003	STRSET isValid	TRUE	none
NTR80125	TRSP1 non-coherent AGC	-141.105	dBm				
NTR80005	TRSP1 coherent AGC	-139.199	dBm	NMP30001	MPSSET objectState	READY	none
NTR80121	TRSP1 carrier lock	Not Locked	none	NMP30003	MPSSET isFailed	FALSE	none
NTR80122	TRSP1 subcarrier lock	Not Locked	none				
NSY37007	Nb TC packets routed	14549	none	NGY20011	GYROSET objectState	ON	none
NPA00001	PAA SET object state	ON	none	NGY20009	GYROSET isValid	TRUE	none
NPA00002	PAA SET is failed	FALSE	none				
NPC20001	PCDUSSET object State	ON	none	NFS20001	FSSSET objectState	ON	none
NPC20002	PCDUSSET isFailed	FALSE	none	NFS20002	FSSSET isFailed	FALSE	none
NPC10205	RTA MB Voltage	27.9111	V	NCP20001	CPSSSET objectState	OFF	none
NPC10202	RTA Ve MEA	16.7368	V	NCP20002	CPSSSET isValid	FALSE	none

Showing 1 to 32 of 32 entries

Displays: Table displays

ARES Desktop Client

Workspace Window Help

2014.097.16.46.09.491 STOPPED DOY

Workspace Spacecraft Event Display(6259) Full Mode

Event ID	Mnemonic	Generation time	Storage time	Time quality	SSC	SPID	APID	Severity	Description
0	YPLX8076	2014.076.11.53.14.721	2014.077.11.37.47.689	GOOD	4319	661232886	199	NORMAL	5_1 INFO YPLX8076 VPU Sequencer component
0	YPLX8076	2014.076.11.53.14.721	2014.077.11.37.47.689	GOOD	4320	661232886	199	NORMAL	5_1 INFO YPLX8076 VPU Sequencer component
0	YPLX8001	2014.076.12.00.08.971	2014.077.11.37.47.690	GOOD	4321	661232977	199	LOW	5_2 LOW YPLX8001 VPU Health Monitoring comp...

Monitoring and Analysis Manage Profiles Reporting Manage Scripts User Management View Log Help About Logout (rsantos)

New Tab GAIA S...

Normal DOY Last N Delta

Start 2014.076 11:53:14

End 2014.097 00:00:00

Refresh Export Auto-refresh

Nominal Warning Error Alarm

Mnemonic Description APID SPID

Event ID

Search:

DateTime	Time Quality	SSC	Event ID	Mnemonic	Description	Severity	APID	SPID	Parameters
2014-03-17 11:53:14.721	GOOD	4319	0	YPLX8076	5_1 INFO YPLX8076 VPU Sequencer component	Nominal	199	661232886	View...
2014-03-17 11:53:14.721	GOOD	4320	0	YPLX8076	5_1 INFO YPLX8076 VPU Sequencer component	Nominal	199	661232886	View...
2014-03-17 12:00:08.971	GOOD	4321	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 12:01:08.096	GOOD	4322	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 12:02:05.221	GOOD	4323	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 12:03:05.346	GOOD	4324	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 12:04:08.471	GOOD	4325	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 12:05:05.596	GOOD	4326	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 12:06:05.721	GOOD	4327	0	YPLX8001	5_2 LOW YPLX8001 VPU Health Monitoring component	Warning	199	661232977	View...
2014-03-17 13:30:31.722	GOOD	4328	0	YPLX3264	5_4 HIGH YPLX3264 PAYLOAD service 12 monitoring 1	Alarm	199	661212900	View...
2014-03-17 15:03:09.800	GOOD	46	54903	YACX2B00	5_1 INFO YACX2B00 AOCS entered IGM	View...			
2014-03-17 15:03:09.800	GOOD	46	54903	YACX2B00	5_1 INFO YACX2B00 AOCS entered IGM	View...			
2014-03-17 16:49:54.971	GOOD	4329	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-17 17:32:36.546	GOOD	175	0	YPFX32CB	5_3 MED YPFX32CB TCS control line failure	View...			
2014-03-17 17:45:24.351	GOOD	176	0	YPFX32CB	5_3 MED YPFX32CB TCS control line failure	View...			
2014-03-17 18:20:34.971	GOOD	4330	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-17 22:23:18.971	GOOD	4331	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-18 11:36:28.971	GOOD	4332	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-18 18:03:42.971	GOOD	4333	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-18 18:22:47.971	GOOD	4334	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-18 21:15:42.971	GOOD	4335	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-19 08:39:16.547	GOOD	177	0	YPFX32CB	5_3 MED YPFX32CB TCS control line failure	View...			
2014-03-19 08:52:04.351	GOOD	178	0	YPFX32CB	5_3 MED YPFX32CB TCS control line failure	View...			
2014-03-19 11:37:07.971	GOOD	4336	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-19 13:08:29.971	GOOD	4337	0	YPLX2F07	5_1 INFO YPLX2F07 PDHU event DOWNLINK_MODE_TR	View...			
2014-03-19 13:09:21.971	GOOD	4338	0	YPLX2F08	5_1 INFO YPLX2F08 PDHU event STORAGE_MODE_TRA	View...			
2014-03-19 13:09:21.971	GOOD	4339	0	YPLX2F11	5_1 INFO YPLX2F11 PDHU event TM_MODE_TRANSITIO	View...			

Showing 1 to 500 of 2,190 entries

YPLX3264 - 5_4 HIGH YPLX3264 PAYLOAD service 12 monitoring 1

Name	Description	Value
NDW00130	Parameter ID	NOS12059
NDW00138	RID	12900.0
NDW00202	Received value of param	30375.0
NDW00288	Monitoring ID (4bytes)	2.0
NDW00289	Actual Check Status	OVER_LIMIT
NDW00193	Violated value	30373.0

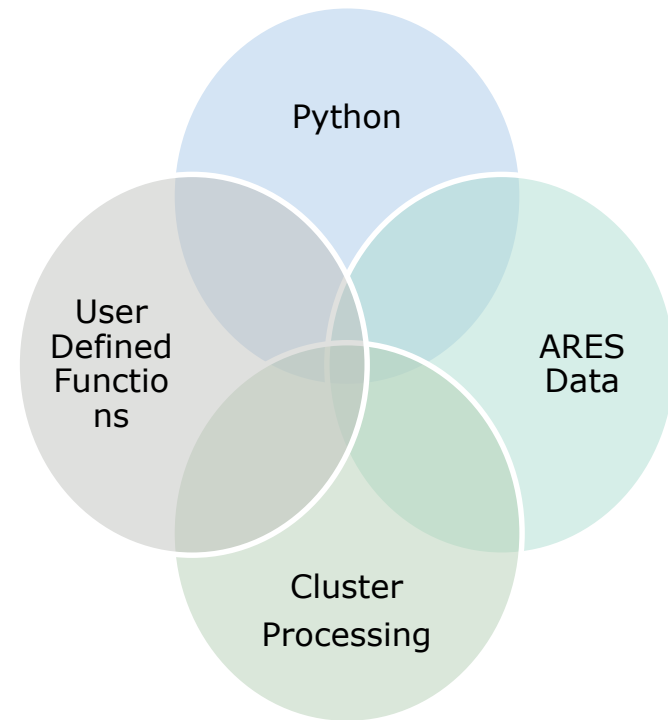
Close

Infrastructure: (Big Data) Multi mission shared storage

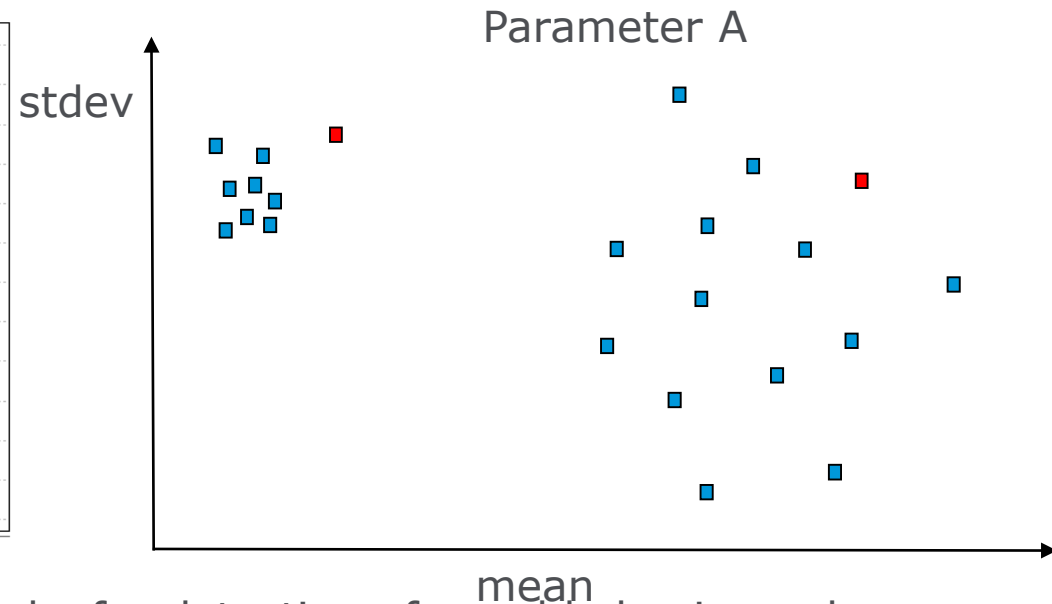
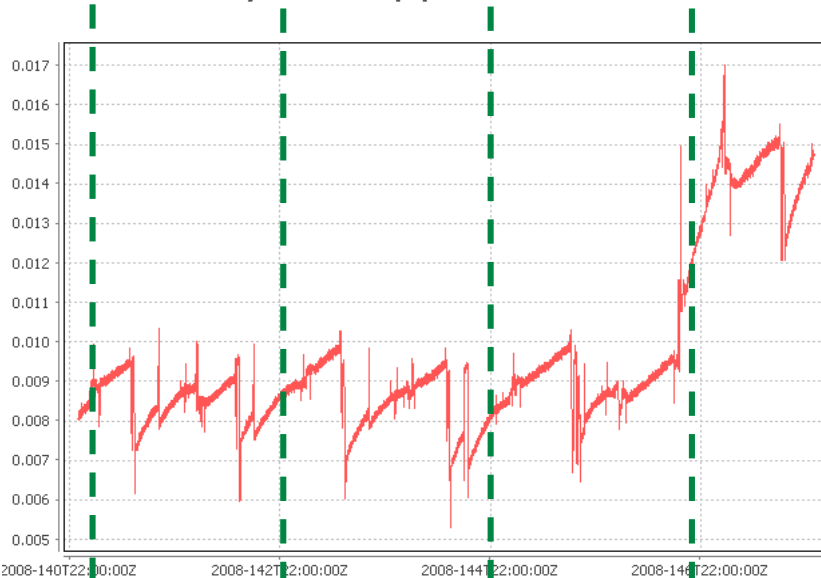
- Ever increasing volume of data (e.g. GAIA receives around 150 million unique HKTM samples per day)
- Remove data storage limitations on type and content
 - Store all available data instead of a selected sub-set
 - Data analysis and mining across different missions
- Scale without impacting performance with stable (and good) retrieval performance
- High availability and data redundancy
- Enable effective Long Term Data Preservation
- Homogenous service and system evolution for all missions



- PyARES is a python library to interface ARES for:
 - Metadata and data retrieval
 - Processing based on user defined functions
 - Machine learning jobs (Random Forest)
- Aims at bridging the gap between data analytics and software development
- Bring the processing to the data (with PySpark)
- Abstraction with underlying technologies such as Spark, Hbase, Protobuf



- Novelty Detection: **Unusual behaviour** is often the signature of an **anomaly** in the way to happen

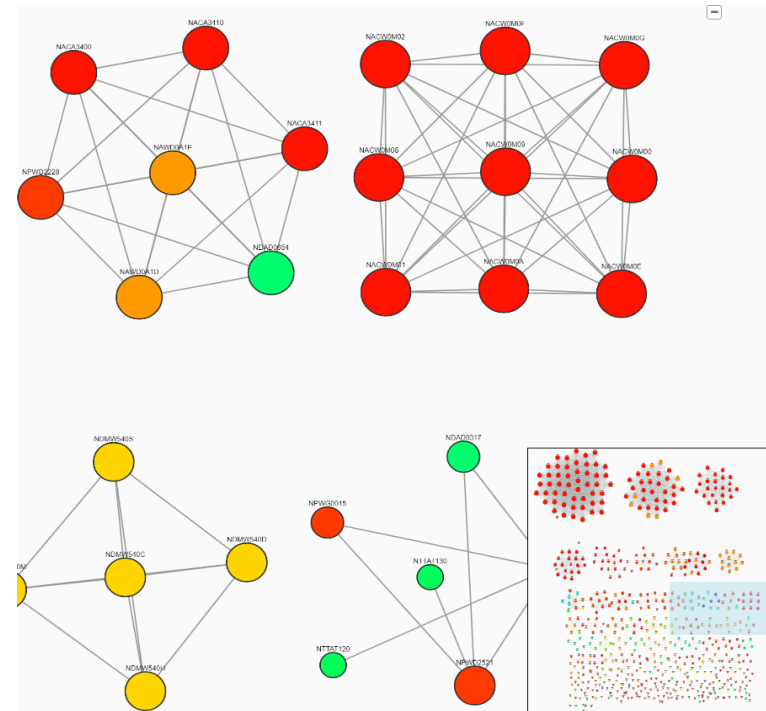


- Deep Novelty: using neural networks for detection of novel behaviour when parameter behaviour could depend on other parameters
- The field of Deep Learning looks promising for anomaly detection but more effort needs to be devoted to it

SpaceOps2018 paper "Novelty Detection with Deep Learning" by Jose Martinez and Alessandro Donati

Processing: Dependency Finder

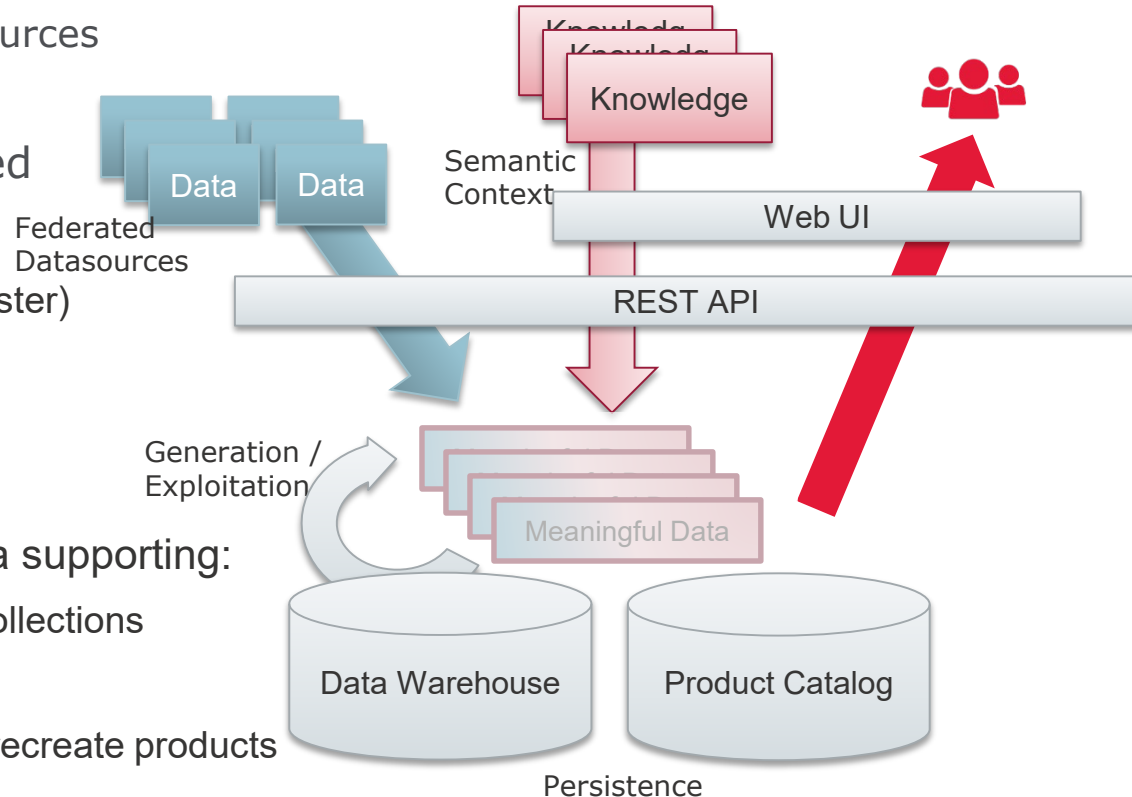
- Motivation:
 - Better understanding (training, confirm operational experience knowledge)
 - Support anomaly investigation
 - Support operations preparation (anticipate which other parameters would be affected by the new operation)
- We look for cases where the joint distribution is very different from the theoretical joint distribution if it would have happened by chance
- Prototype tool has been developed by the Artificial Intelligence & Operations Innovation Team and validated by the Mars Express Flight Control Team



SpaceOps2018 paper "Dependency Finder: Surprising Relationships in Telemetry" by Jose Martinez, Luke Lucas and Alessandro Donati

Enhancement: LTDP products and collections

- Vision:
 - Combine diverse data types
 - Preserve and enhancement data with semantic context and metadata
 - Federate multiple data sources
- A platform has been prototyped using:
 - ARES (Hadoop database cluster)
 - Elastic Search
 - RESTful API
- A generic and flexible data schema supporting:
 - Multi mission products and collections
 - Different data types
 - Instructions to preserve and recreate products

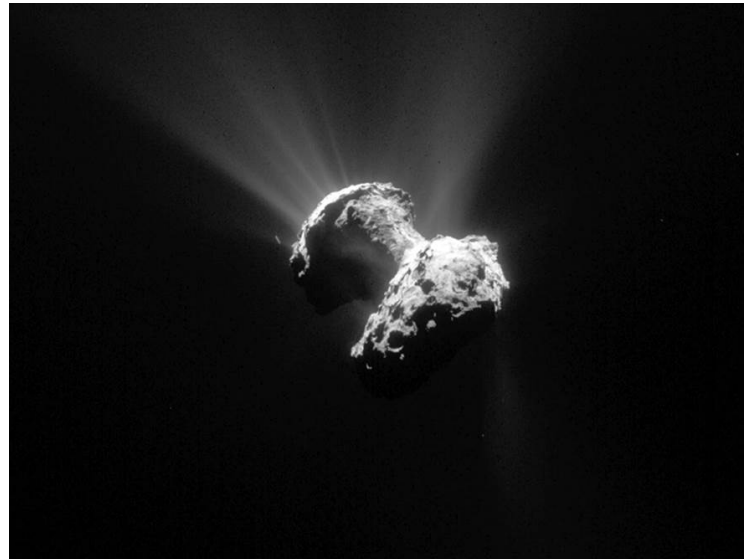


Examples of potential future work

- Processing and storage of live stream of TM data
- Interface/Federation with central data portal for enhanced data access and correlation with Science data
- Augment capabilities of data processing layer (including machine learning techniques)
- Preserve legacy missions by migrating them to the Hadoop infrastructure
- Introduce operationally validated AI and machine learning tools to operator teams
- Create a curated “Anomaly Detection ESA Dataset” to allow to benchmark different anomaly detection approaches



- Data analytics activities need a stable and scalable support infrastructure
- More data = more value = better data analytics = new opportunities
- Multi-mission service allows for optimization and better inter mission cooperation activities
- Sustainable and contextualised long term data preservation from day one



Thank you for your attention!

