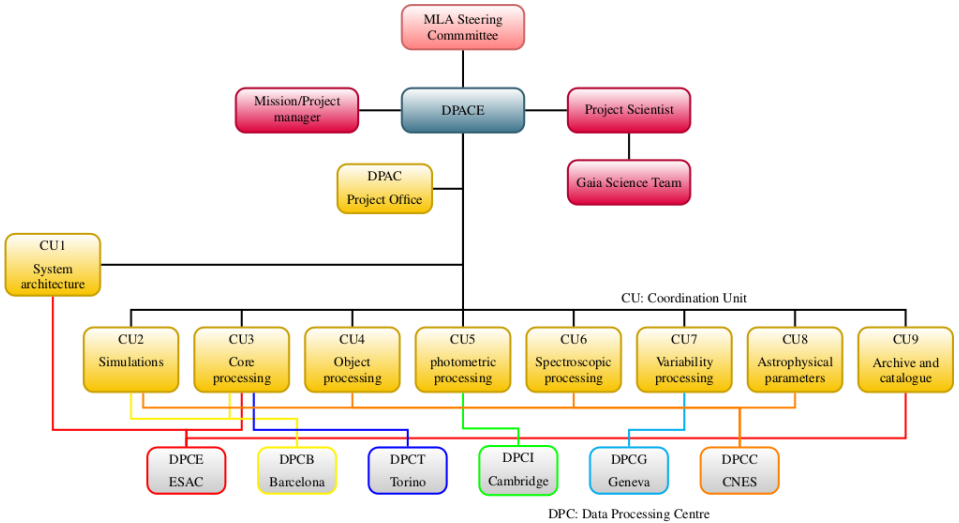
DevOps: the perfect ally for Science Operation

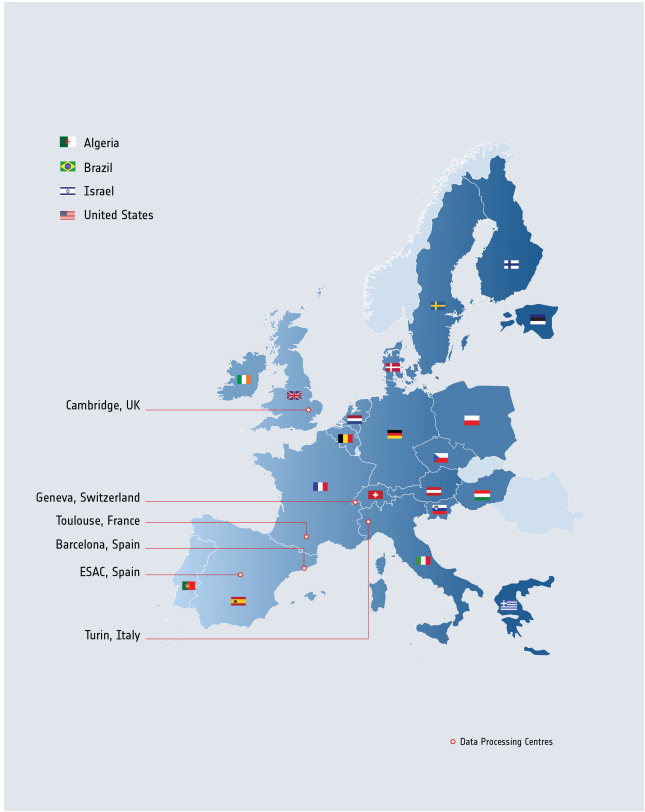
Rocio Guerra - ESA

ADASS XXVVIII, College Park MD, 15th November 2018

Gaia Data Processing - DPAC

- **Large pan-European team**
  - Responsible for the processing of the Gaia data
  - 450 expert scientists, developers and engineers
  - 160 institutes, 24 countries + ESA





Gaia SOC from ESA viewpoint

- **Gaia SOC is an integral part of the Data Processing and Analysis Consortium (DPAC)**

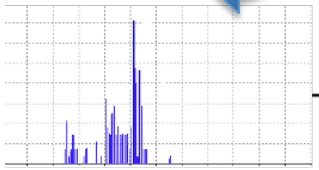
- **A bit unusual for an ESA "classical" roles:**
  - MOC's primary contact for all the payload/science-related mission aspects
  - Mission planning product generation (Scan Law, Science Schedule, etc.)
  - Payload health monitoring
  - First-level product generation and downstream data dissemination
  - Archive development + operations

- **"Special" roles:**
  - CU1: System architecture + central repository (Main Database)
  - CU3: Development + Operations of AGIS
  - CU9: Support for catalogue creation
  - DPCE: Data Processing Centre at ESAC

- **Cyclic and Daily processing**

Daily Pipeline

TM + HK ~ 35GB daily



70M transits/day

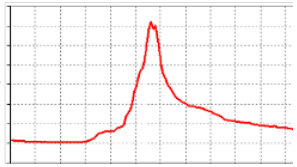
Input

TM files

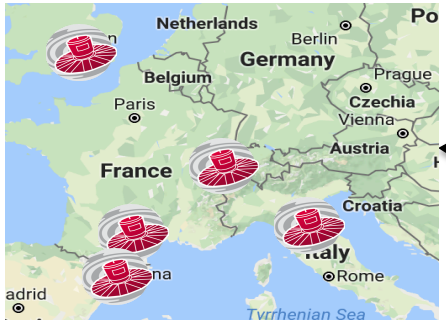
Ops. Pipeline

Output

Gbin files



~ 300-500GB



Internet

1 Gb/s

MDB

Our challenge: dealing with complexity!

- **Operations are very complex**
  - Very ambitious top level requirements
  - Near real-time processing
  - Systems tightly coupled: many dependencies
  - Systems developed by external parties -> hard to integrate

- **Very strict performance requirements**
  - Science alerts can’t wait! + Payload monitoring
- **Data volumes**
  - 35-90GB of TM inputs become 300-500GB to be served to the DPCs daily

CURRENT DATE AND TIME	2018-11-08T18:55:59 (TCB)
MISSION STATUS	
Satellite distance from Earth (in km)	1,569,642
Number of days having passed since 25 July 2014	1567
OPERATIONS DATA (collected since 2014/07/25)	
Volume of science data collected (in GB)	58,834
Number of object transits through the focal plane	112,746,221,016
Number of astrometric CCD measurements	1,111,355,607,154
Number of photometric CCD measurements	224,751,711,050
Number of spectroscopic CCD measurements	21,550,324,020
Number of object transits through the RVS instrument	7,193,871,670

REAL Daily Pipeline

TM & HK
30-90GB

Backup
@ESTEC

MIT/DCS

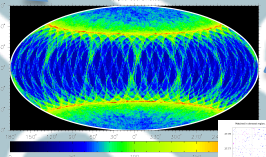
IDTFLDB

~30TB

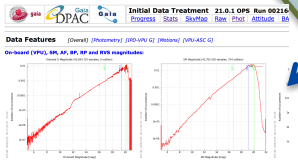
Soc Monitor

POS
Commanding

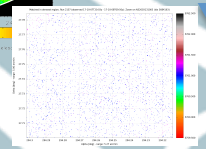
DB Mirror



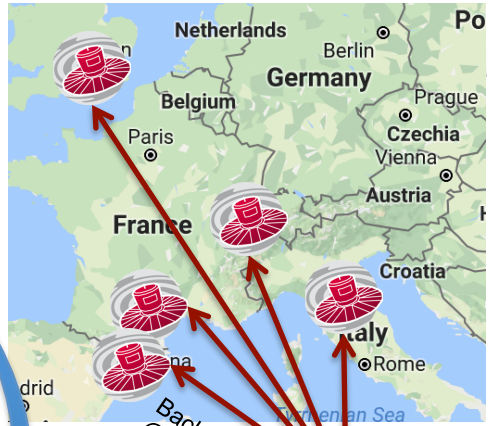
IDT



Monitoring



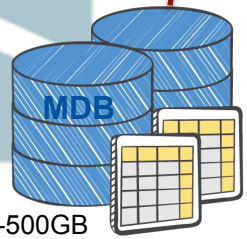
FL



Internet

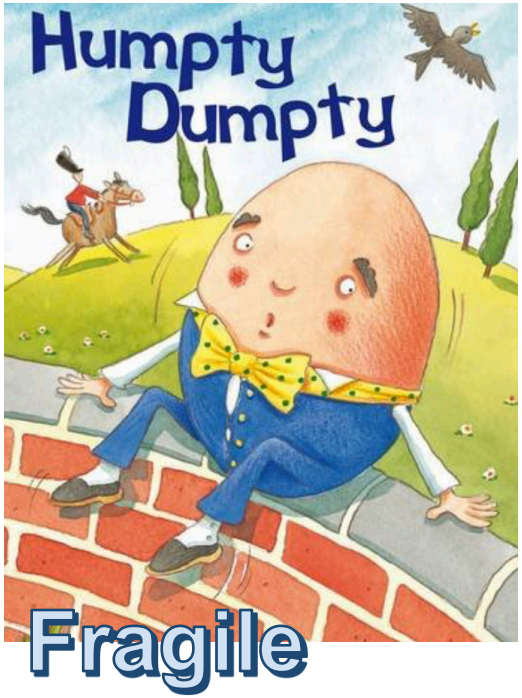
DTS

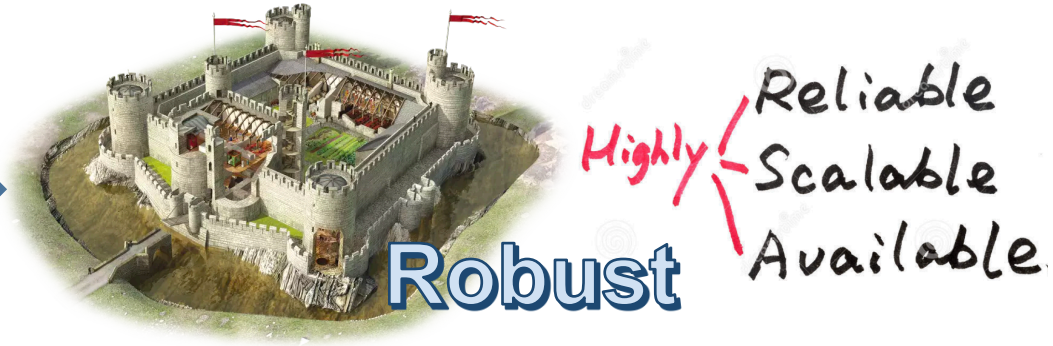
1Gbps



300-500GB

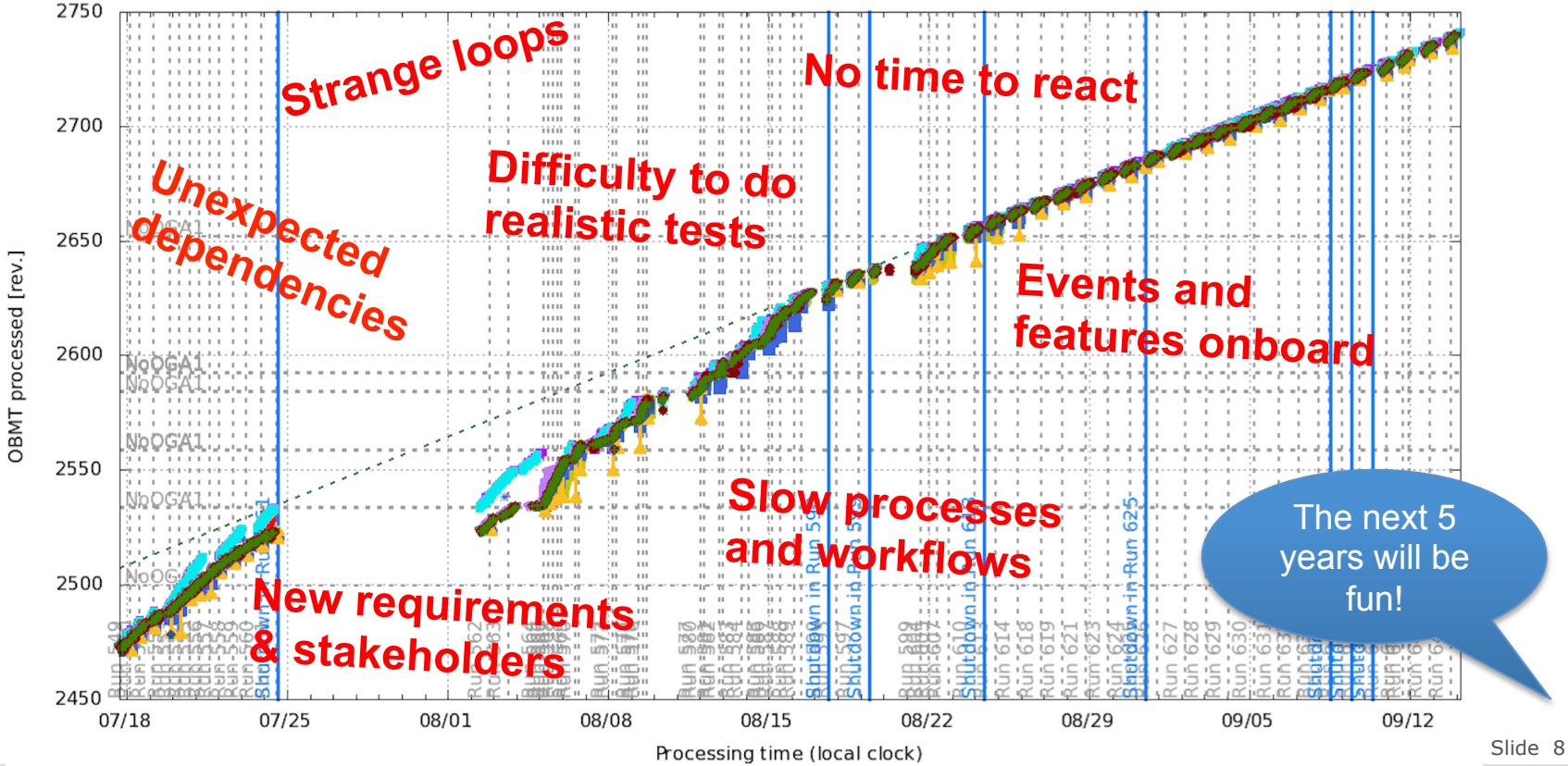
Our mission: creating a robust system





- **Robustness achieved by:**
  - Large simulated datasets provided at different levels
  - Tests campaigns (system, integration) including E2E and Operational Rehearsals – Test Readiness Review and Review Boards
  - Formal CCBs, bug tracking and definition of strict issues workflows and approval cycles
  - Identification of interfaces with the stakeholders
  - Definition of detailed Operational Procedures

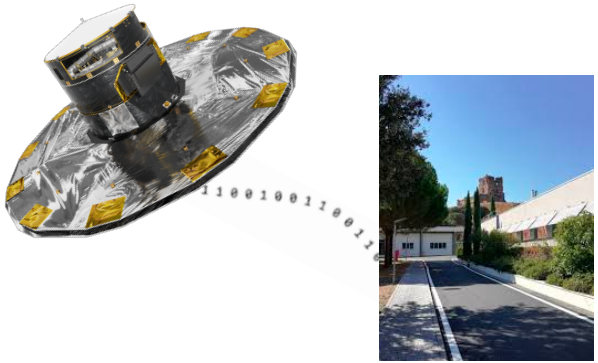
It seems robustness was not enough



Old rules no longer applies

- **Military field: Team of Teams, 2015. General McChrystal**

When General McChrystal took command of the Joint Special Task Force he quickly realised that conventional military tactics were failing. Al Qaeda in Iraq was a decentralised network that could move quickly, strike ruthlessly, then seemingly vanished in the local population. The allied forces had a huge advantage in numbers, equipment and training – but none of that seemed to matter. To defeat them, his Task Force would have to **acquire the enemy's speed and flexibility**.



- **And not for the intensive processing imposed by new astronomical missions either!**

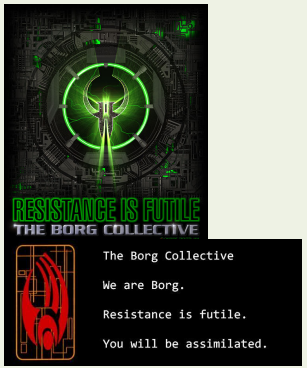
- **We had applied strictly best practices, but:**
  - They were insufficient to solve problems while keeping up with the performance requirements
  - Too many unknown variables, scenarios
  - Too little time to follow the processes as defined (but still needed to conform with quality standards)

Gaia Processing – Our evolution

Fragile

- Work but …don't touch it
- Hidden dependencies
- Not documented
- Tightly coupled

Anti-fragile



- System will fail
- … but it will recover quickly
- … it will learn from it

DevOps

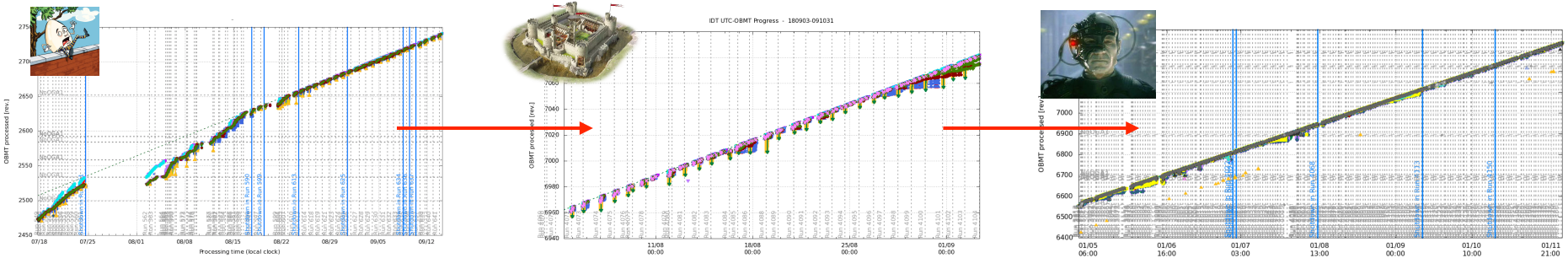
Robust

- Repel all the attacks and remain still
- When it falls it is catastrophic (we can't anticipate every possible attack)

Transformation through DevOps

- **Embracing DevOps practices allows us to:**
  - Velocity: fast time to production
  - Continuous feedback and fast incorporation
  - Stability and quality of operations: optimized and streamlined processes and procedures
  - Reduced risks
  - Empowered teams
  - Saving time and money

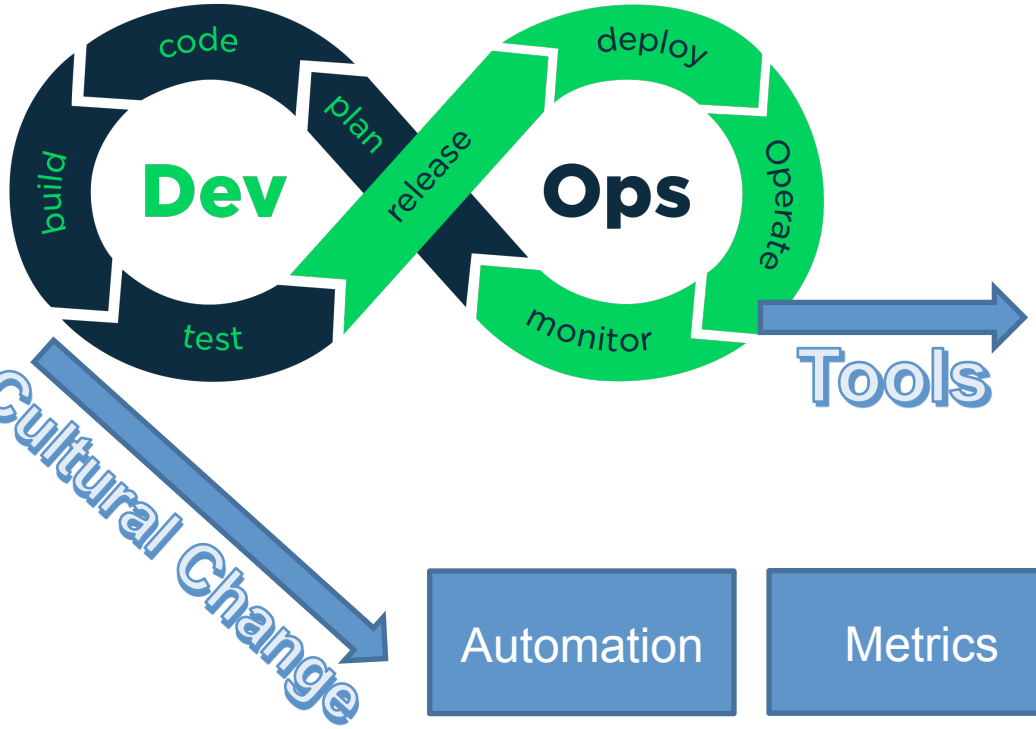
Antifragile system

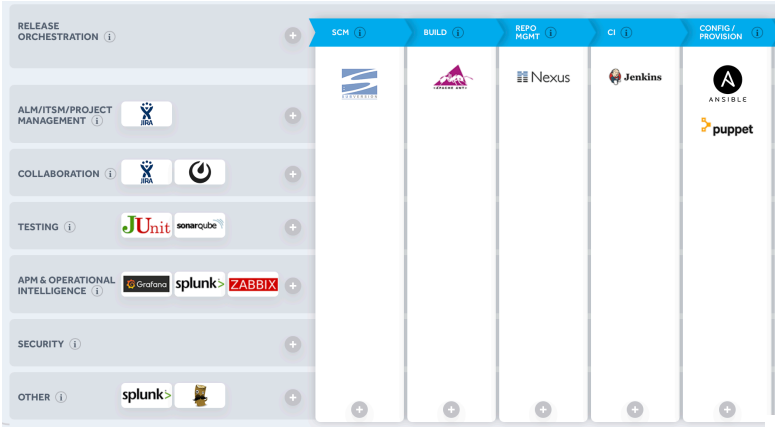


- More stable systems over longer periods
- Downtimes reduced
- Ability to detect and recover from failures faster

- Same with less
- Quality of the data produced
- Team motivated

DevOps is not just about tools



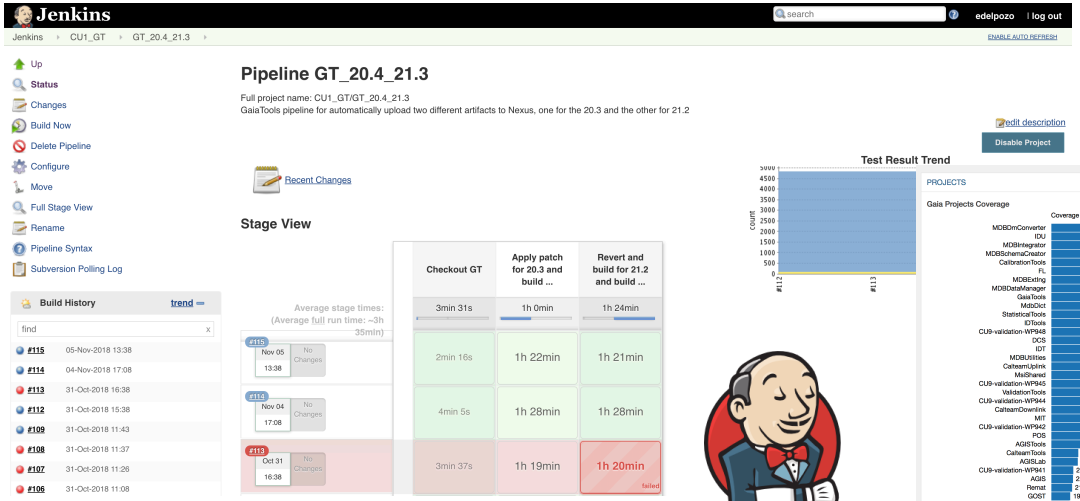


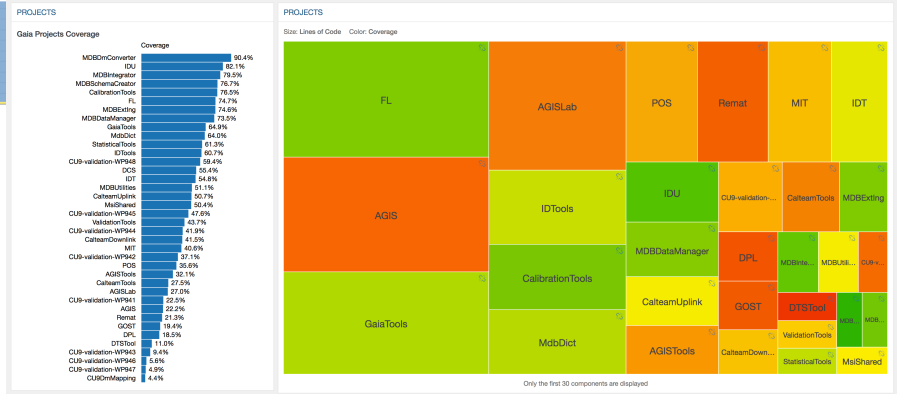
ermmm … we don't use Dockers or Kubernetes (yet) 😳

- Automation
- Metrics
- Lean
- Collaboration

CI / CD

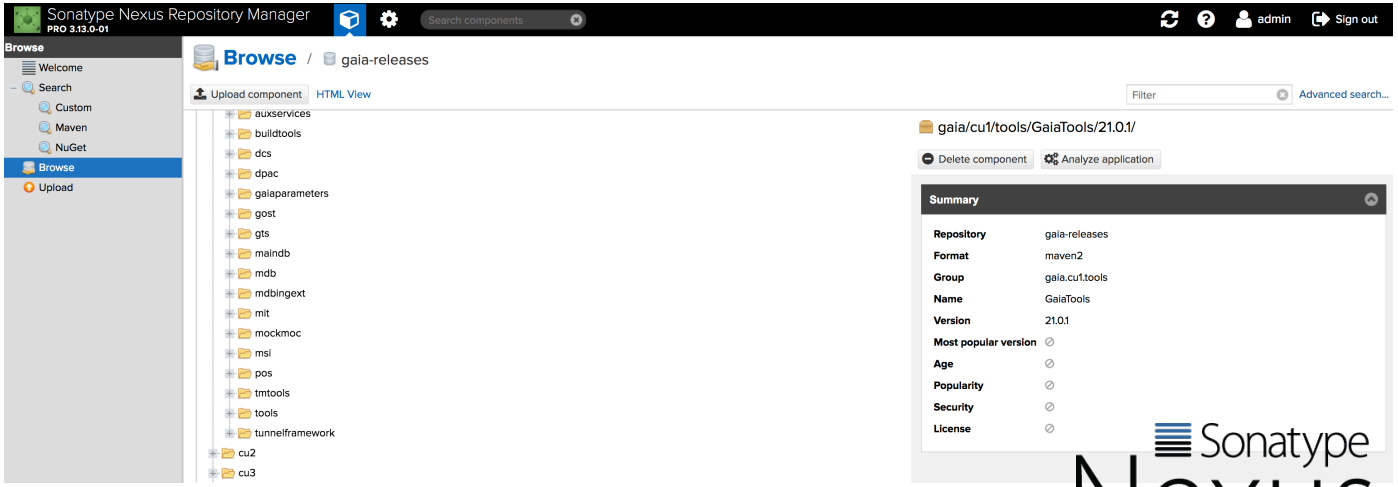
- **Jenkins as the main Continuous Integration tool**
  - Recently introduced the usage of Jenkins Pipelines for implementation and integration of complex jobs: building, testing, QA and deployment: a big step towards automation
  - SonarQube is the source analysis tool for continuous inspection of code quality. Fully integrated with Jenkins





Artifact Repository and Infrastructure

- For storage of SW artifacts Nexus Professional Repository Manager
- IT Infrastructure automates the deployments, OS updates, etc. with Puppet and Ansible



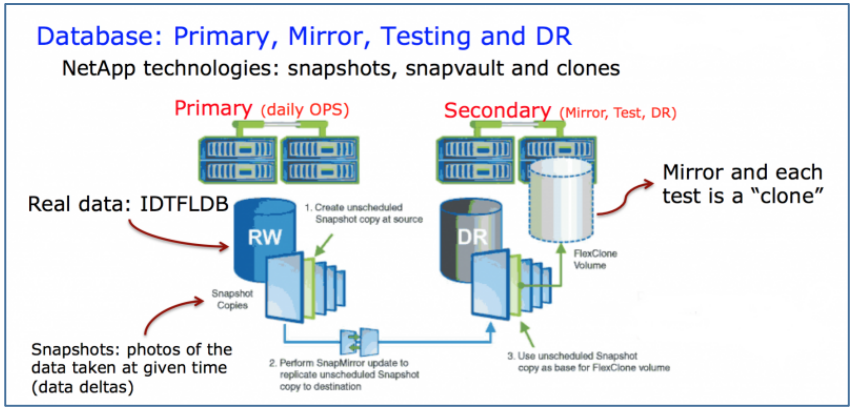
Sonatype Nexus

ANSIBLE

puppet labs

Testing

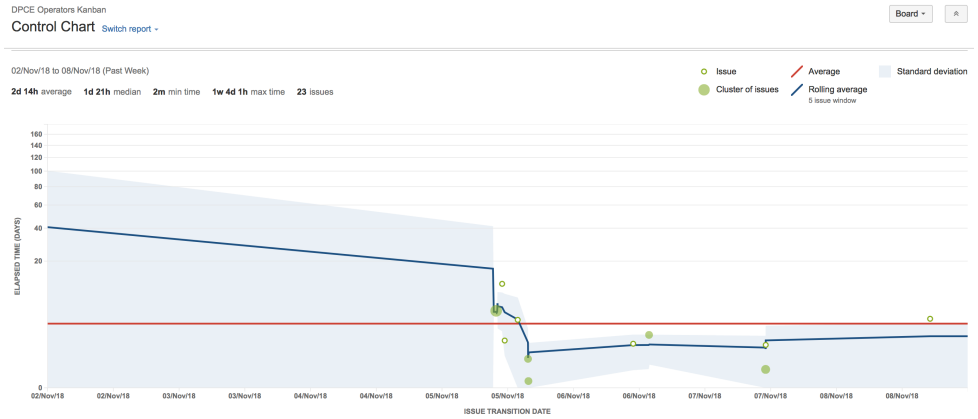
- **Constant testing and re-evaluation**
- Main challenge: **reproduce DB** state, volume and load
- **Representative tests in ”real time”**
  - Past: very costly – time and resources – to conduct one, let alone repeated test
  - Now we can do small, simple, fast experiments, trials
  - Clone of the working database (˜30TB), same environment and receiving real time TM
  - -> Direct comparison with OPS
  - + backup!



- **Future plans:**
  - Testing management in Jira
  - Cucumber for automating system tests

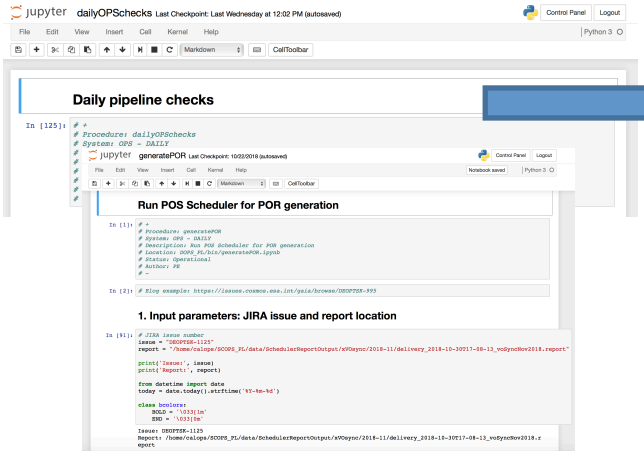
Agility - Kanban

- **Kanban to visualize the work**
- Based on:
  - Making the policies and procedures explicit
  - Work in progress (WIP) limited
  - Manage the workflows
- Whiteboard with sticky notes (moving to **digital board in Jira** – useful for collecting metrics)
- **"Person of Interest"** role trains the team members and enhance the procedures

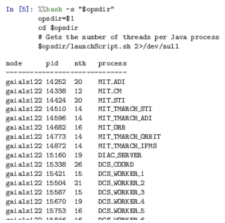


Procedures – Jupyter Notebooks

- **Jupyter Notebooks** for procedures and runbooks
  - Markdown text, executable code and output all inside a single document -> they all fit the procedures and runbooks need very well
- **Connection to operational logbooks**
- **Benefits:**
  - Up-to-date
  - Often reviewed, collaboration
  - Automation, reduce risks for manual intervention



Message ID: **250**	Entry time: **Tue Nov 6 06:02:23 2018**
System:	IDT
	Accountability
	JN
Subject:	Daily OPS Checks (JN)
Status:	Pass
test.txt	
Attachment 1:	dailyOPSchecks.pdf 455 kB | Hide | Hide all



Takeaways

- **By embracing some DevOps practices the Gaia SOC enhanced the data processing and the effectiveness of the team**
- **Fundamental drivers: automation, lean, metrics and collaboration**
- **Continuous learning (it's a culture)**
- **Goal: collecting feedback from the systems and the team so that they can be used for improving them and achieving antifragile systems**

Thanks
QA