ESOC's Successes, Complications and Opportunities in using Cloud Computing and Big Data Technology

James Eggleston

Head of Data Systems Infrastructure Section (OPS-GDI)
European Space Operations Centre

Thanks to Rui Santos & Alex Donati

27/02/2019

Context: ESOC and OPS within ESA Organisation

ESA has multiple sites, multiple directorates…

Focus of today's presentation:

- D/OPS - Operations directorate and ESOC site that focuses on ESA missions
- **Data Systems Engineering:** software system architecture & engineering for our missions operations
- Neither ESA wide nor IT layer implementation nor missions operations

Context: ESOC and OPS within ESA Organisation

Traditionally: mission built bespoken infrastructure…

- Designed for mission needs, paid for by missions, operated by the ESOC IT teams
- Problem: these technologies don't make sense at small scales…

ESA IT now provide standard services – but not targeted at mission operations

- Virtualisation via in-house vCloud platform
- Containerisation via Docker/Kubernetes platform (solution being rolled out)
- (no big data solution yet provided…)

ESA wide initiatives provide overarching control across the directorates and sites of e.g. Studies and Long Term Data Preservation

Cloud Computing at ESOC Situation Report

Historically - initial distrust of cloud computing…

- Yes, use it! But not for online operational use
- Yes, use it! But keep the old architecutre
- How reliable is it?
- What are the security aspects?

Leads to Current Solution:

→ Cross-site ESA private cloud, provided by corporate services

→ Used for mission business uncritical auxiliary systems

→ Used for development and test facilities

→ Only certified public commercial cloud allowed

Cloud Computing Complications

1) Demonstrability

- How do we prove reliability and availability for long time periods?
- How do we guarantee redundancy and security?
- Culture change to be completely reliant on the technology and support of the provider, versus in house engineering and teams

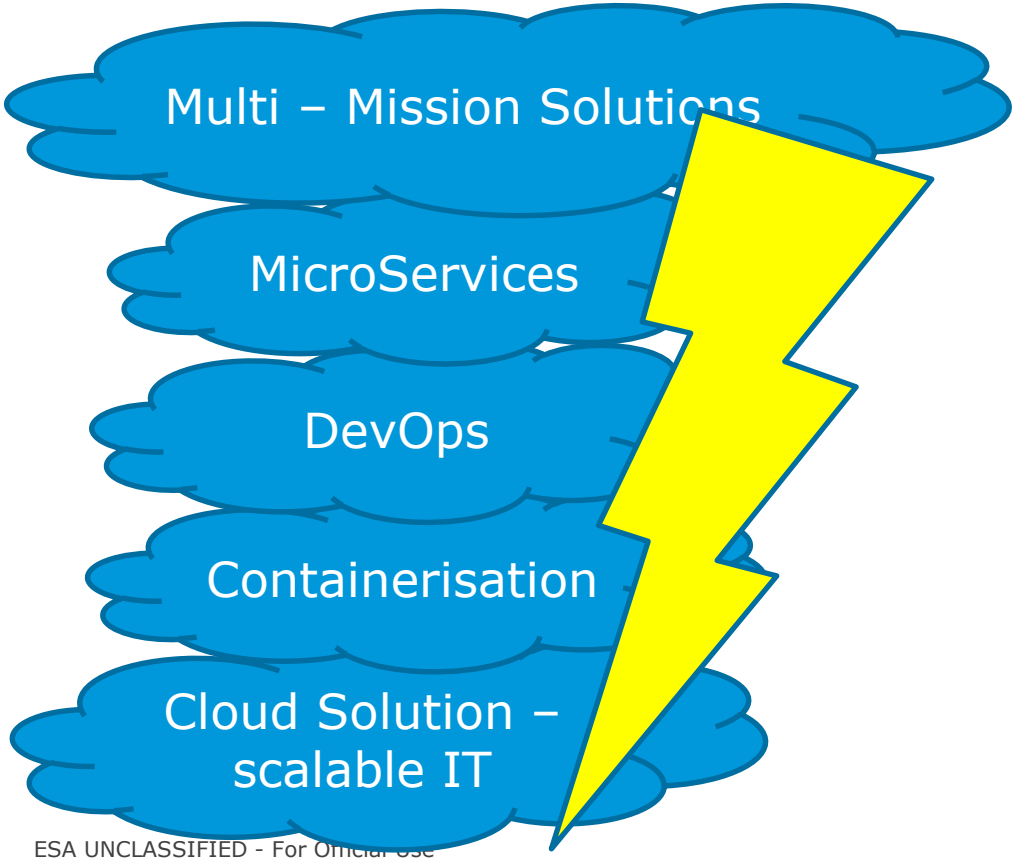
2) Capacity management

- Who is responsible? The mission? The IT layers? The Data Systems Managers?
- Answer seems to present a power shift in the organisation, creating challenges

3) Resource Usage of Legacy Systems

- E.g. if each physical machine needs 32 Gb, only 8 machines fit into 256 Gb
- How restrictive can we be with sharing resources? Shared resources are problematic

New Opportunities, Possibilities, Architectures



Providing reliably scalable mission operations

Providing easily scalable software

Providing reliably configured operational installations

Providing reliably installable software

Providing reliably scalable hardware

The Way Forward

Demonstrability of Software Processes Maturity

- Start with the auxillary systems, multi-mission traditionally designed solutions
- Centrally managed configuration management and deployment orchestration
- Following container & DevOps principles where we can

Demonstration of Capacity management maturity

- Planning and governance that is trusted

Better Resources Usage

- Migration of systems toward microservice architecture & multi mission solution

Create Knowledge Pool

- Education & Training mission staff, developers, gaining experience in house

© Paramount Pictures

Big Data at ESOC Situation Report

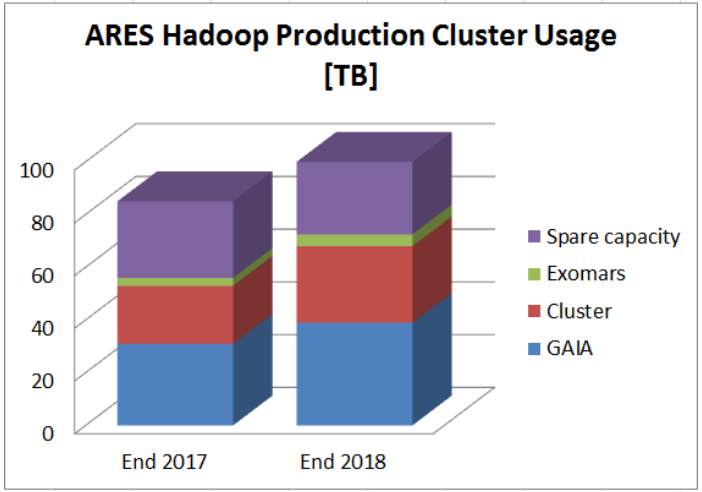
Brief History

- Housekeeping archive ('ARES') service since 2015
- Now 5 missions, 80 Tb HDFS storage for mission Housekeeping data
- Storage needs varies widely (factor of 10) per mission based per MIB

Priority of 5Vs of Big Data at ESOC, in order…

- Velocity (many small rapid parameters)
- Volume (Tbs per year per mission)
- Veracity (completeness, correctness)
- Value (we never had these abilities)
- Variety (typically fixed forever per mission)

Offline data analytics, and not mission control - yet



ESOC Current Solution Keypoints

Open Source solutions based Hadoop ecosytem…

- Long term availability of data – and hence availability of solution - is key driver
- Tech stack with HDFS, Hbase, Yarn/MR, Spark, Kafka, Zookeeper, …
- Highly complex installation and configuration to create aggregate solutions

→ Use Cloudera CDH

Current cluster: 10 machines / 80 vCPUs / 640 Gb RAM / 120Tbs storage

- Growing rapidly with new missions with higher data rates
- Attracting existing missions … with large existing archives
- Expected 10 new missions by 2020

New Opportunities

Original Hadoop justification was to address data volume / velocity issues with RDMS…

No perceived driver for using Big Data techniques…

- Why? Missions have a tradition of creating Limits, synthetic parameters, etc with similar benefits
- Is doing the same thing enough to justify added complexity?

Can we use the data processing features of Hadoop?

- At least do the same things… but faster!
- Now what about the extra value of data?
- Can we apply Machine Learning techniques?

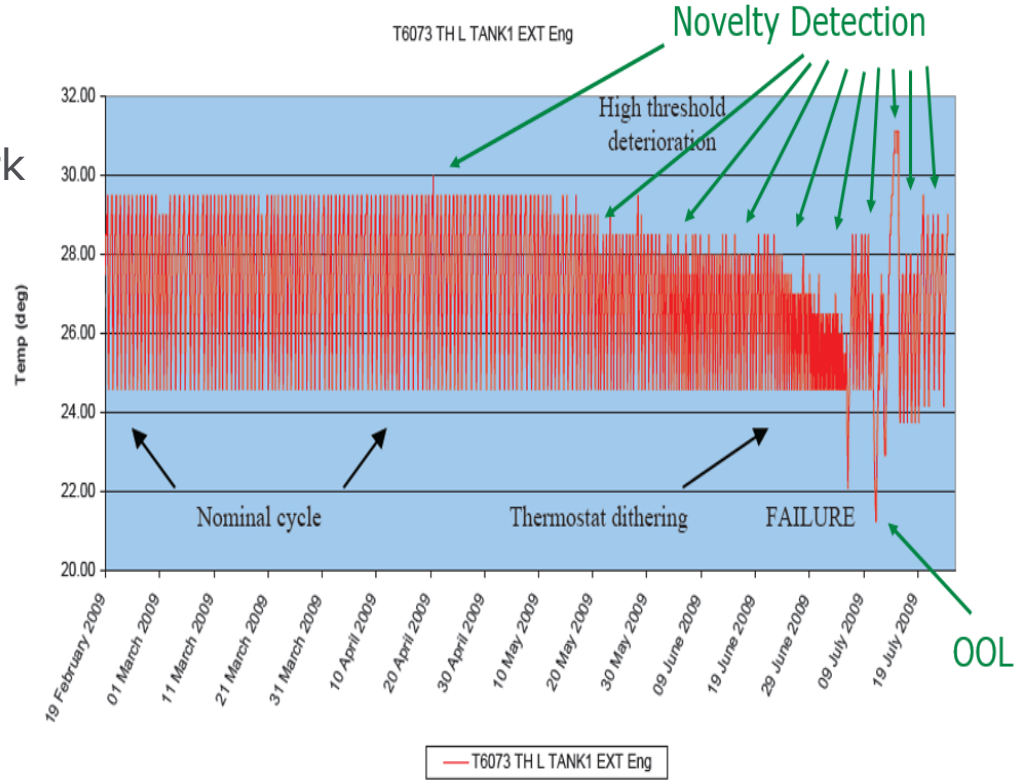
Same things, but faster…

Application to existing ESOC Analytic Tools, e.g.

- "DrMUST" application ported to Spark – x10-50 performance improvement
- Novelty Detection application is expected to deliver large performance benefits allowing more analysis

→ Big Data technology enabling faster and more detailed analysis

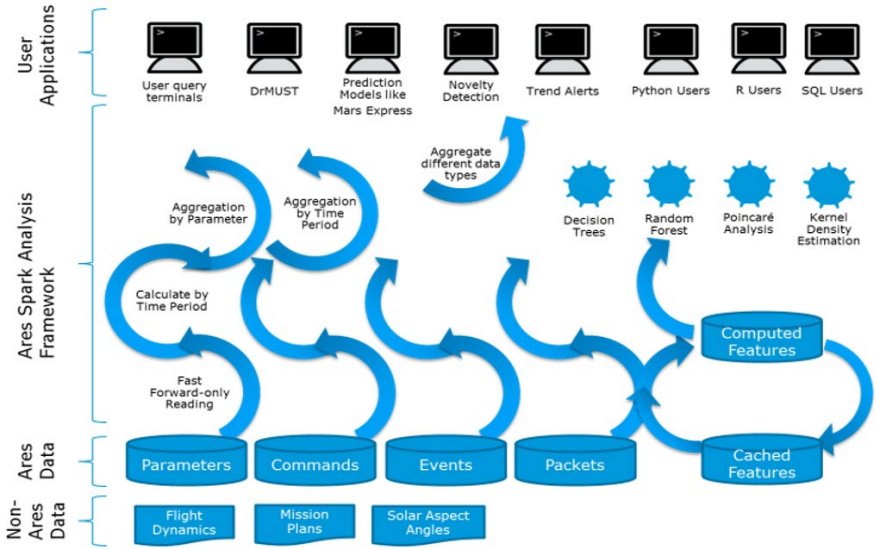
→ **Why do missions trust humans? Why don't they trust AI?**



Machine Learning (not Big Data) adds Value

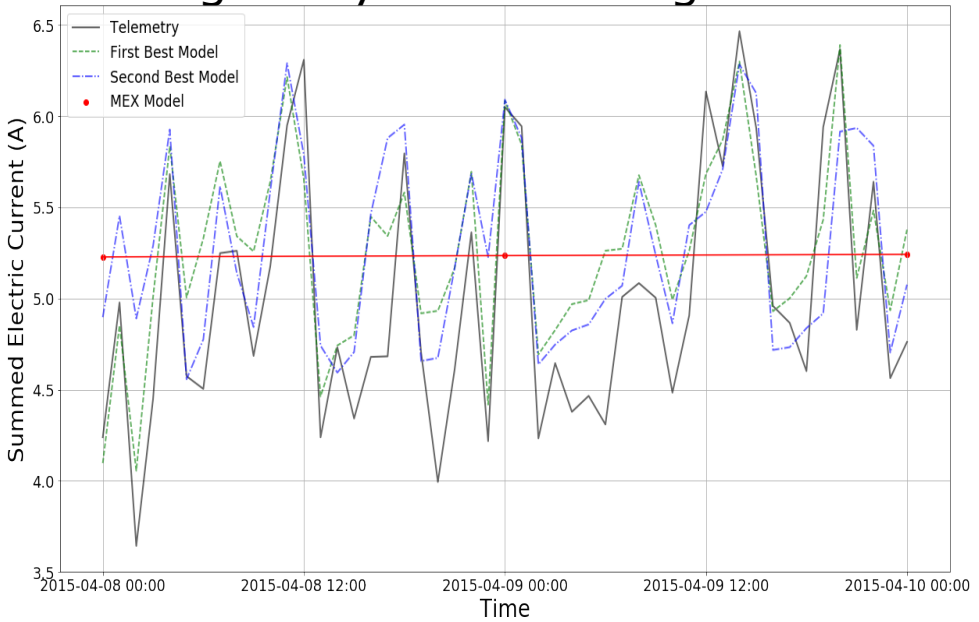
Believe Machine Learning **Support** is key…

- Infrastructure access layer to python scripting for e.g. ML kits, Jupyter, etc



Will missions take advantage?
Why do they need to change?

→ e.g. They can live longer



Current Problems for ESOC

Basic entry point for usage is massively more complex…

- SQL server v Hadoop cluster
- Creates issues in testing, understanding and experience

Big Data Exploitation - bringing ideas and opportunities into missions…

- We have successes… but only with FCT / mission specific issues
- Convincing the missions to prepare & work differently is a massive challenge!

Non-Provable ML is not trusted in current culture and mission profile

- Would a Spacecraft Manger trust Alpha Zero?
- When AIs be allowed to send TCs? What about on-board autonomy?

Mixes the world of hardware infrastructure, software infrastructure, mission solutions and data analytics

Recent reduction of competition in solution supplier space