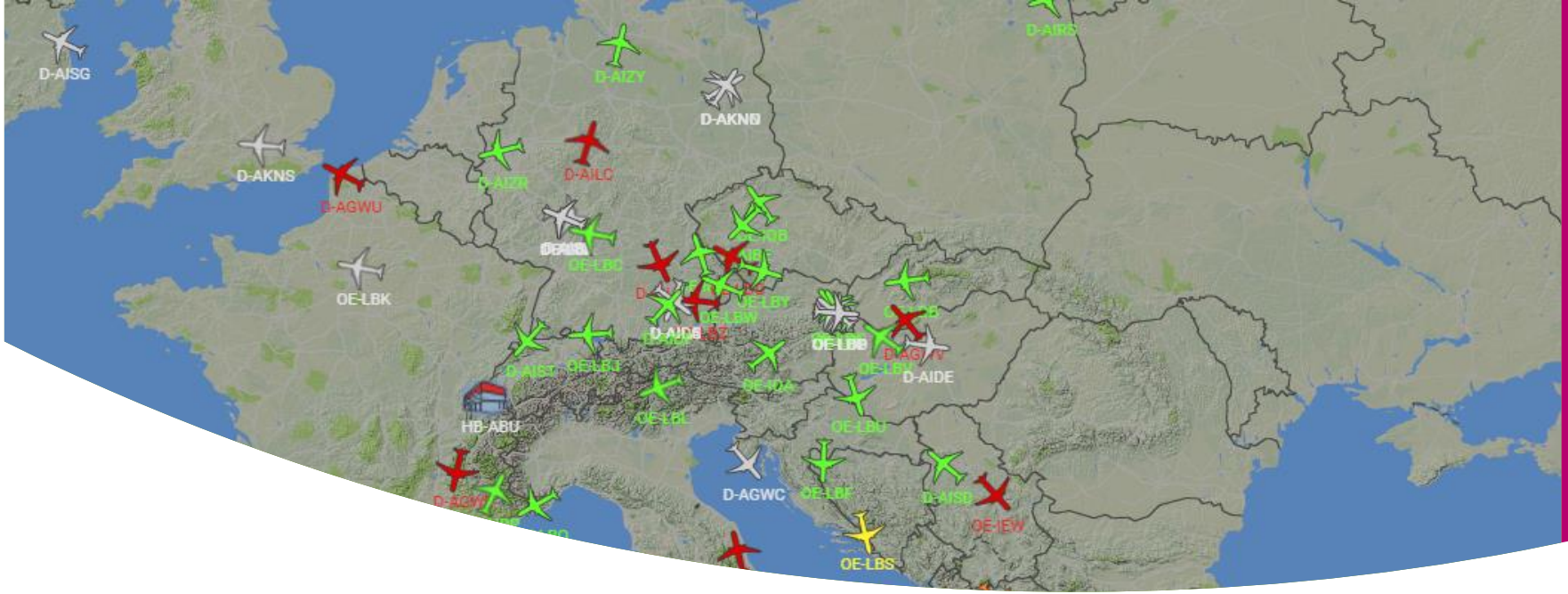


CERN Openlab alumni talk

CERN Openlab: my transition from science to business

Maaike Limper 21/09/2017

inmarsat
The mobile satellite company

Outline

(why I’m not at work right now…)

Before Openlab

My CERN Openlab fellow

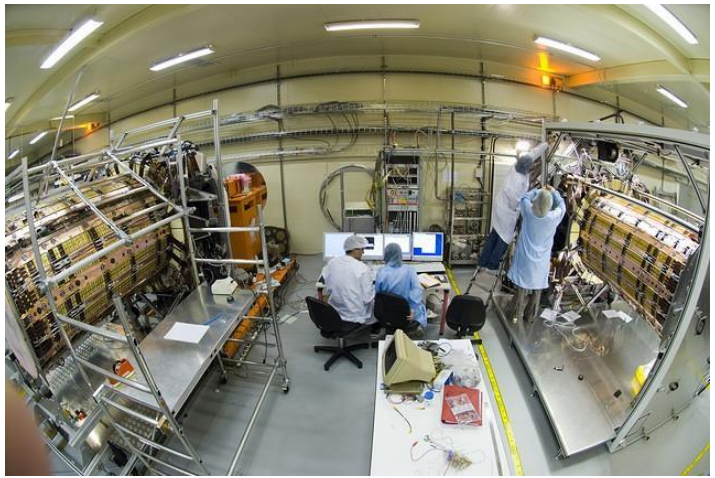
My career since Openlab

CERN Openlab: lessons learnt

Before Openlab

One of the many physicists at CERN…

- Master in Particle Physics at University of Amsterdam 1999-2004
  - First time at CERN as Summer Student in 2003!
- PhD in Particle Physics 2004-2009
  - Building the ATLAS SemiConductor Tracker (SCT)
  - Testing and installation of SCT at CERN
  - Software development for track and vertex reconstruction
- Post-doctoral researcher for University of Iowa
  - based at CERN
  - First physics analysis with LHC data in ATLAS
  - Prompt Reconstruction Coordinator for ATLAS
  - "Co-discovered the Higgs-boson" along with others thousands of physicists…



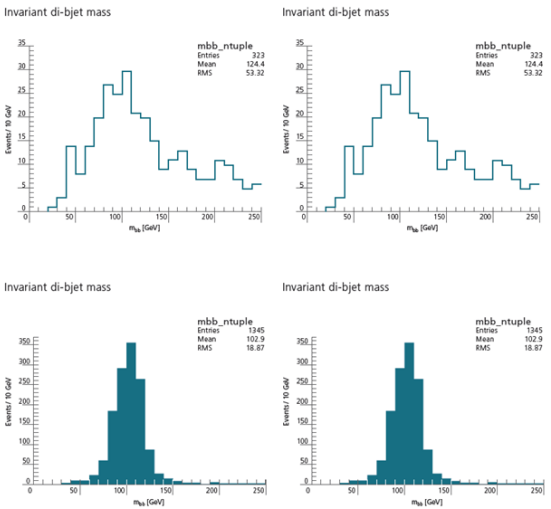
My CERN Openlab fellow

How to discover the Higgs boson in an Oracle database

SQL-based approached to physics analysis:

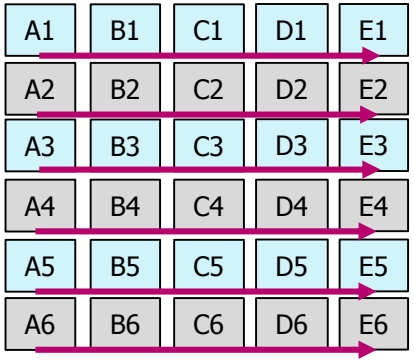
- Databases allow us to do much more than store calibration constants and conditions
- Complex SQL with analytics-functions or precompiled procedures allow to do event filtering and calculations directly in the database
- SQL-analysis of basic physics data potentially faster than ROOT-based analysis
- Physics data has many attributes per row, making it more suitable to be kept in column-storage and partitioned in files across the grid for easy parallel processing

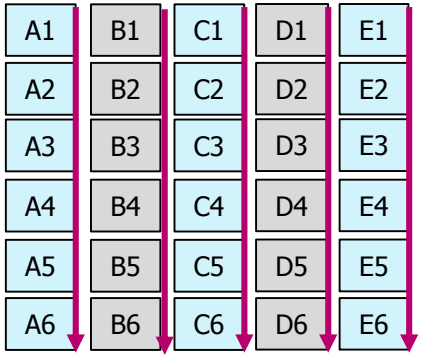
Plots showing results from an Oracle-based analysis (left) identical to those from a ROOT based analysis (right)



My CERN Openlab fellow

What I actually discovered



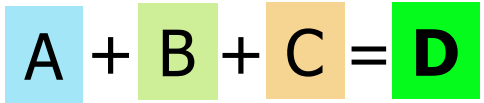


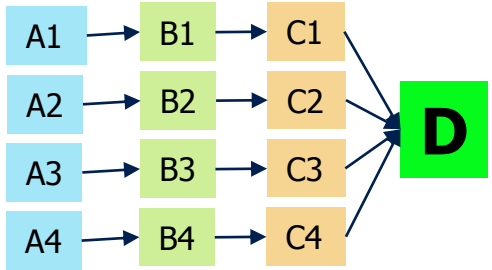
ROOT/PROOF is optimized for physics:

- Hard to improve using database analysis!
- Though users don't always make optimum use of its features

But great test-case for complex DB-analysis:

- Sequential scan vs index look-up
- Column vs row-based storage
- Love-hate relationship with SQL optimizer
- Use of partitions/distributed processing





Hadoop vs PROOF vs Oracle RAC: Any distributed data processing ultimately comes down to the same thing, optimize your partitioning to push down your joins!

My career since Openlab

The final frontier of connectivity

SITA OnAir:

- Business data analyst

Inmarsat:

- Head of Aviation Service Performance

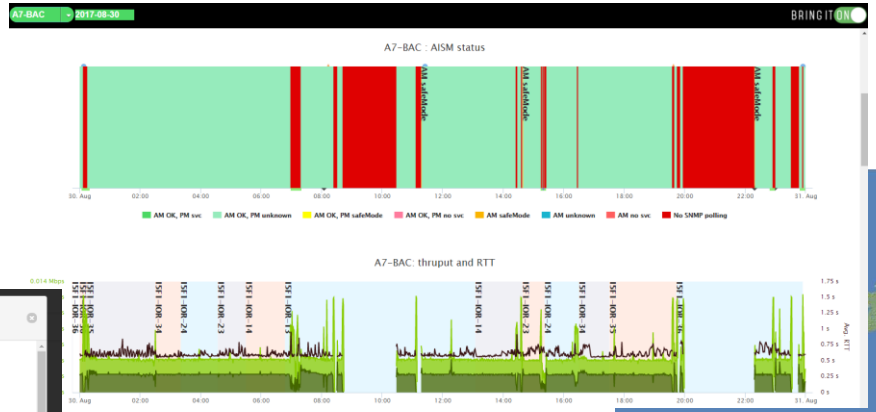
Aviation Service Performance

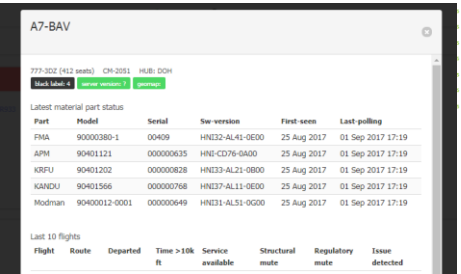
Data-driven approach to performance monitoring

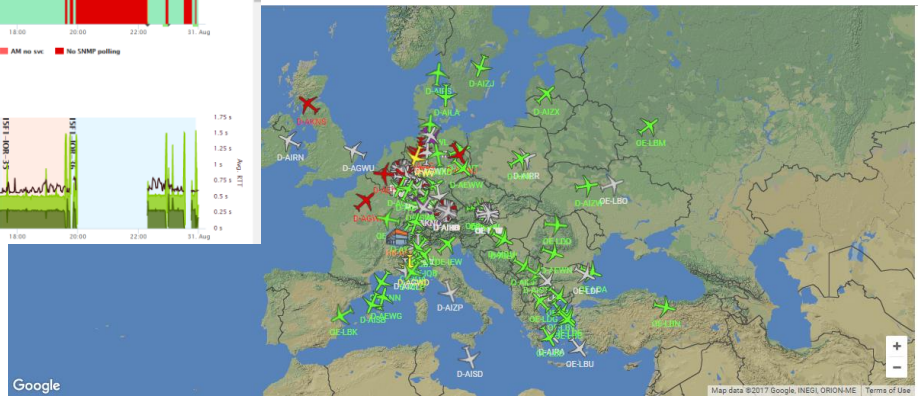
Inmarsat Aviation has a dedicated team for Service Performance

- Our mission: *'to turn data into information'*

Data collected from all IFC systems is analysed using advanced custom data-processing solution and dedicated tools provide monitoring, customer analytics and enable data-sharing with our customer







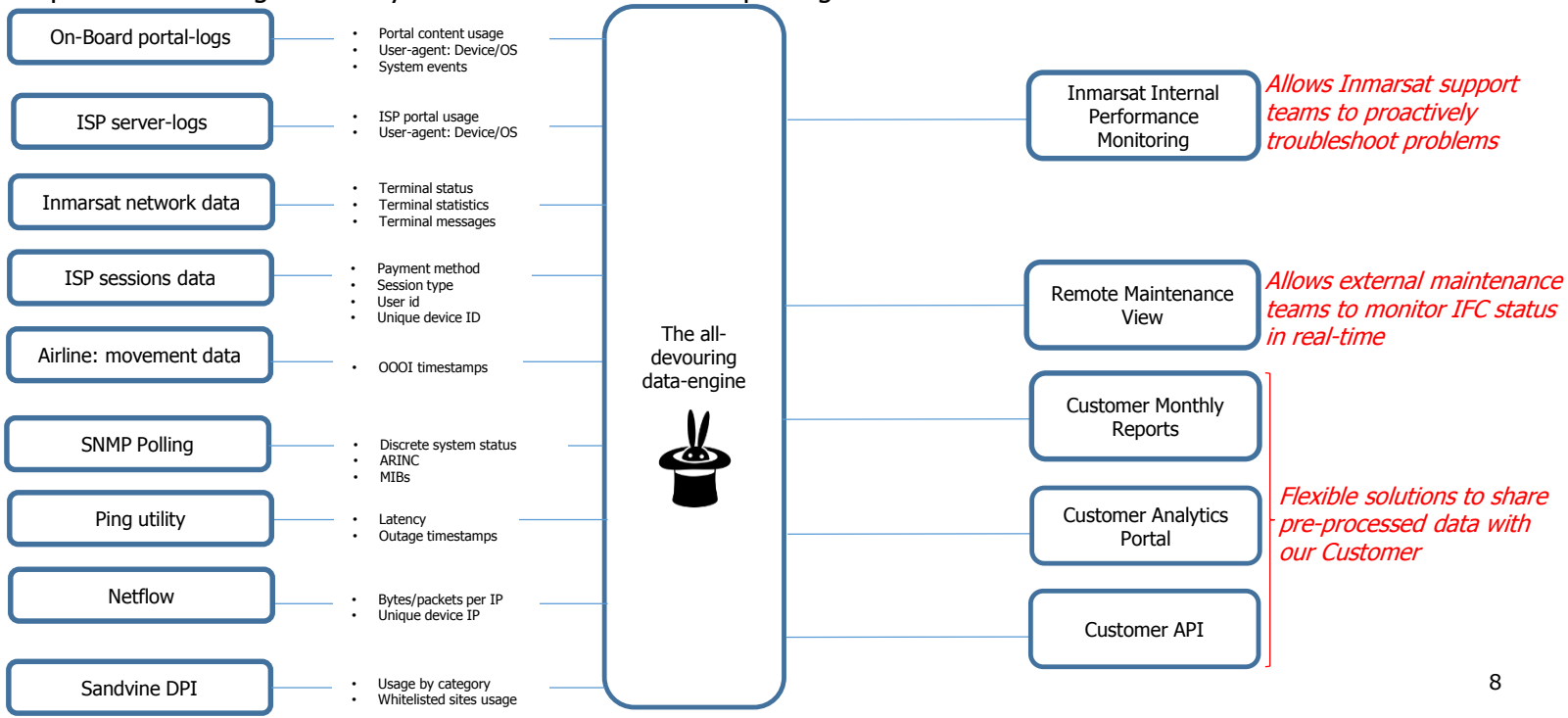
The all devouring data-engine

"Any sufficiently advanced technology is indistinguishable from magic"

Automatic processes *extract, consolidate and aggregate* data from every possible data source

Correlating data from different sources allow unique opportunities to study performance dependencies between our systems

… pre-processed data is presented through a variety of internal and external reporting tools.



CERN Openlab: lessons learnt

What working at CERN has taught me

- No Higgs boson without IT
- Multi-culturalism
- Lateral thinking
- DIY
- Don’t trust technology
- Love, trust & respect your data