Big Data, Big Transformation: Big Benefits for Large-Scale Engineering Products

Martin McDonald
Andrew Gorrie

October 2019
A top 10 Global Defence & Aerospace Company

Leonardo is a global high-tech company and one of the key players in Aerospace, Defence and Security worldwide.

Divisions

Helicopters  Aircraft  Aerostructures  Electronics  Cyber Security

Subsidiaries/Joint Ventures

<table>
<thead>
<tr>
<th>DRS Technologies</th>
<th>Telespazio</th>
<th>Thales Alenia Space</th>
<th>MBDA</th>
<th>ATR</th>
<th>Vitrociset</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Leonardo</td>
<td>67% Leonardo 33% Thales</td>
<td>67% Thales 33% Leonardo</td>
<td>37.5% BAE Systems 37.5% Airbus Group 25% Leonardo</td>
<td>50% Leonardo 50% Airbus Group</td>
<td>100% Leonardo</td>
</tr>
</tbody>
</table>

2018 Results €bn

Revenue 36.1  Order backlog 15.1  New orders 12.2
AGENDA

01 Background: What and Why Big Data?

02 Our Solution: Technologies and Architecture

03 A Future Towards DataOps...
Key Messages

Transformation | Then and Now - what does good look like?

Example of success | Technologies and Infrastructure

Future Looking | What will we do next?
Why Big Data?

Volume, Velocity, Value...
BIG DATA

Infrastructure / Techniques
Fuzzy transition point after which traditional storage and analysis techniques become inadequate

Investment
Data acquisition, storage, maintenance and exploitation is a business investment and should be treated as such

Value
The goal for Big Data is to extract and leverage the value from data
Backdrop - The Business is Changing

1. Next Generation Products
   - New technologies, new hardware and increased complexity means more data than ever

2. New Development Strategy
   - Data is now more valuable than ever with analysis for Model Driven Engineering favoured over costly experimental aircraft trials.

3. New Customer Environment
   - Modern technologies and products increase the demand for product flexibility and so extensibility.

4. Multi-decade Programmes
   - Long term effective management and utilisation of data is key to unlocking the business investment in data
Project Analysis: what was the Status Quo?

- Manual Processes
- Bespoke Tooling
- Network File Transfers
- Millions of Unmanaged CSV Files
- Data Stored on Network Drives
What does this mean for an engineer?

One year on a single project...

Mouse Clicks

264,160

To process data - before adding value

Equivalent 15 page Word Documents...

23,765,923

This would take over 90 years of continuous effort to read.

CSV Files

769,772

Of human readable radar data - i.e., not including the sensor data.
So what do we want?
Opportunities for improvement against traditional approaches.

Analytics
Make it easy for engineers to find the needle in the haystack...

Customise and Standardise
Make it easy for engineers to perform the modelling tasks they need to.
Keep analytics DRY

Accessibility
Make it easy for engineers to get the data they need.
Our Solution
Use Cases, Data Architecture, Hardware, Software.
KEY USE CASES

**Data Management**

*Data volume* - secure our investment in data for the long term

**Advanced Search**

*Accessibility* - right data for the right problem

**Advanced Modelling**

Deploy our MATLAB and Simulink models on large volumes of data - improved ability to experiment and validate
Data Architecture

Source Data

Big Data Platform

Data Consumers

Raw Data
Data Engineering
Refined Data
Data Science
Data Visualisation
Data Governance
Data Architecture

Existing
- No platform
- Passive
- Disparate
Data Architecture

Source Data

Big Data Platform

Data Consumers

Sensor data, emails, test reports

Raw Data → Data Engineering → Refined Data → Data Science → Data Visualisation → Data Governance

© 2019 Leonardo MW Ltd – All rights reserved
Data Architecture

Automated preparation, format conversions, scaling, standardisation
Data Architecture

Automated validation, testing, labelling
Data Architecture

Existing Models and MDE
Data Architecture

Policies, processes, metadata management
Solution: **BigData Platform as a Service**

**Leonardo Big Data Platform**

- **Number of Server Racks**: 2
  - Space for expansion with COTS hardware.
- **TB of Total Installed Storage**: 900
  - 300TB of usable storage after accounting for distributed file system.
- **Number of Processing Nodes**: 20
  - 3 management nodes, 2 edge, 15 workers.
- **TB Memory**: 6
  - 384GB of memory per processing node.
Technology Stack

Extensible technology stack acting as a data and processing hub.

- **Hadoop**: Distributed file system for efficient storage access and processing.
- **Cloudera**: Managed solution to rapidly introduce the paradigm.
- **Hue**: Graphical interface to create dashboards - edit and search.
- **Solr**: Search and indexing engine for documents and data exploration.
- **MATLAB & SIMULINK**: Core products in our Model Driven Engineering strategy and analytics.
- **Apache Spark**: Processing engine optimised for distributed processing in MATLAB, python, scala.
Processing Architecture...
From this...

How do I load lots of files?

what about different data formats?

how do I process huge volumes?

10,290 items
To this...

Easy to change data location...
Local, HDFS...

Easy to change processing environment
Local, Local-Parallel, Distributed...
What was the experience like with MATLAB?

Use Existing Models and Tools
A lot of our existing models and tools are written in MATLAB

Compile to Linux and Hadoop
Engineers can work on their environment of choice and relatively easily transition

Datstore abstraction
As well as deploying code, we can be flexible with data sources - local, HDFS...

Data Format Support
In 2019a and in 2020a, ability to write to the parquet file format
Key Challenges for the Project

01 Data Structures + Schema Evolution
Traditional development process results in rapidly changing data schemas/definitions

02 Knowledge and Skills
Data is not core competency - domain specialist engineers are extended to data management + manipulation.

03 Data Governance
Data owners are not always formalised, metadata capture is often not part of the workflow so is extra.

04 Encourage Thinking Globally
Traditional usage patterns for data involve massive data reduction at each stage.
Achievements

Benefits so far

Supporting Multiple Programmes with different needs

Processing Times of Minutes, not days

Improved Model Driven Engineering capability with deployment of MATLAB models
3

Future

What's next.
A great model isn't enough...
End to end solution to support a data-driven workflow

Towards DataOps

DataOps is an automation methodology used to improve the quality and reduce the cycle time of data analytics.

- Access and Explore Data
- Clean and Validate
- Experiment and Evaluate
- Visualise and Report
- Integrate, Deploy and Monitor
Future Technologies...

kafka
Streaming
Live Stream testing events from our lab facilities to engineers PCs
Points to take away...

Foundation for model driven engineering workflows.

1. MATLAB
Abstractions are powerful to get us going

2. Engineers can work in their preferred environment and deploy to scale

3. Spark & streaming are the future for interactive engineering development
THANK YOU FOR LISTENING!

Questions?