ELECTRONICS DIVISION

LEONARDO

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

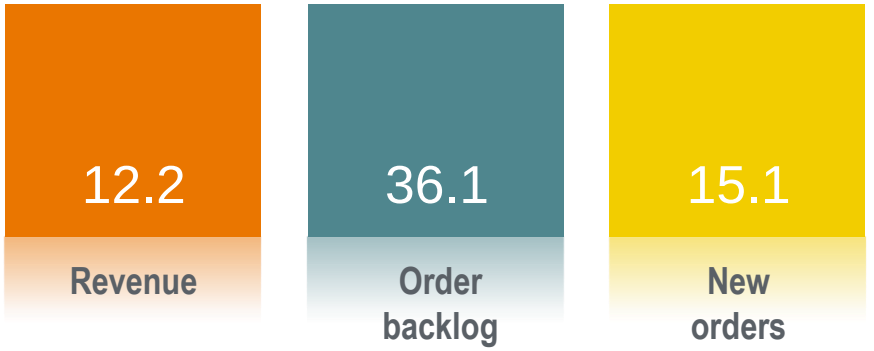
Martin McDonald
Andrew Gorrie

October 2019

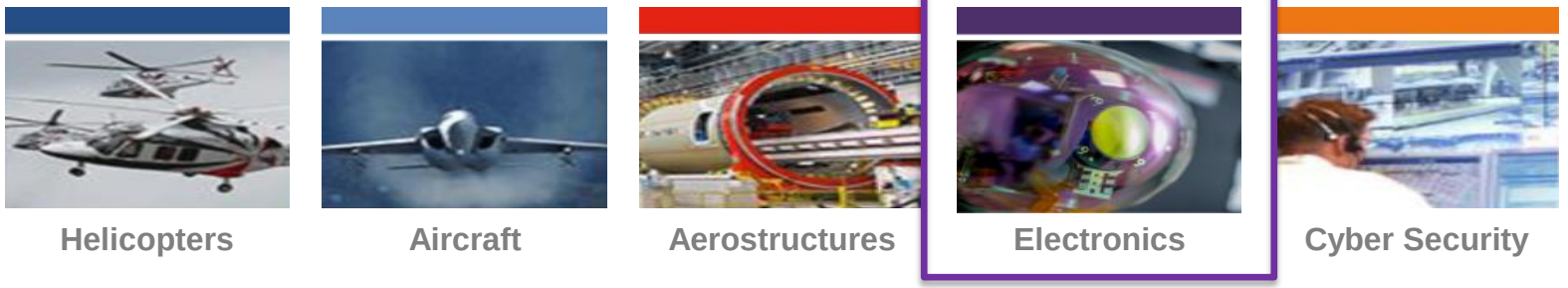
A top 10 Global Defence & Aerospace Company

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

2018 Results €bn



Divisions



Subsidiaries/Joint Ventures

DRS Technologies
100% Leonardo

Telespazio
67% Leonardo
33% Thales

Thales Alenia Space
67% Thales
33% Leonardo

MBDA
37.5% BAE Systems
37.5% Airbus Group
25% Leonardo

ATR
50% Leonardo
50% Airbus Group

Vitrociset
100% Leonardo

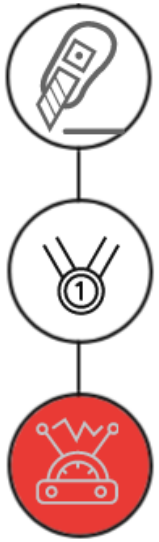
AGENDA

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?

1 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

Next Generation Products

1

New technologies, new hardware and increased complexity means more data than ever

New Development Strategy

2

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

New Customer Environment

3

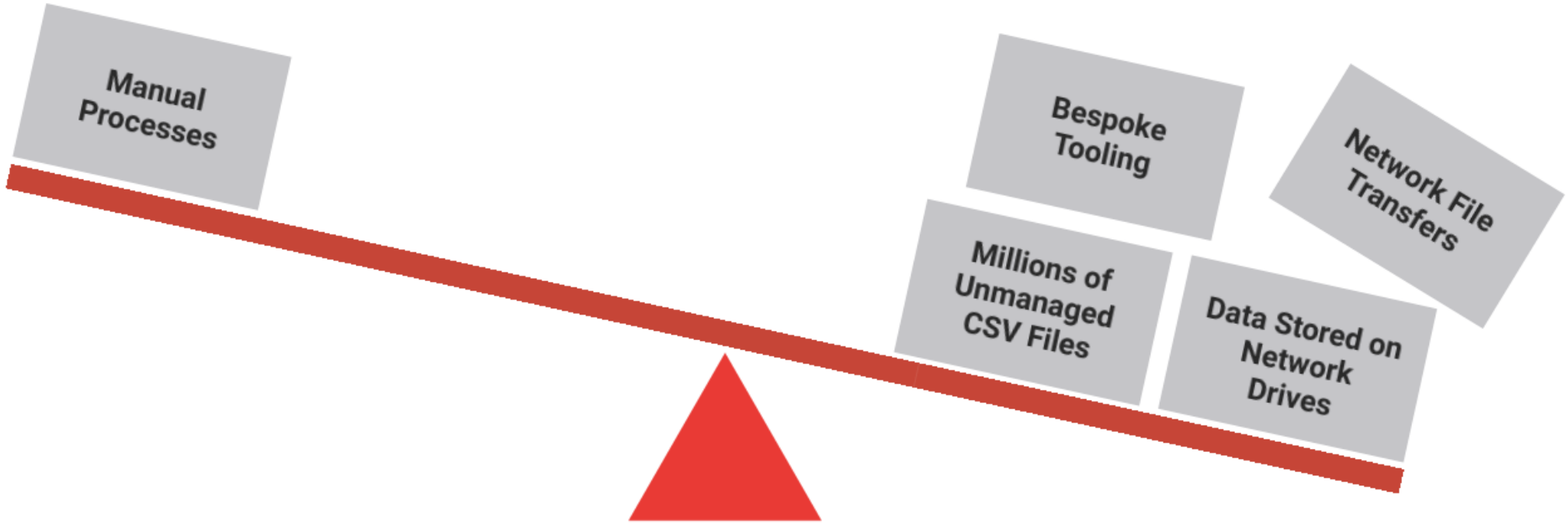
Modern technologies and products increase the demand for product flexibility and so extensibility.

Multi-decade Programmes

4

Long term effective management and utilisation of data is key to unlocking the business investment in data

Project Analysis: what was the Status Quo?



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.

2 Our Solution

Use Cases, Data Architecture, Hardware, Software.

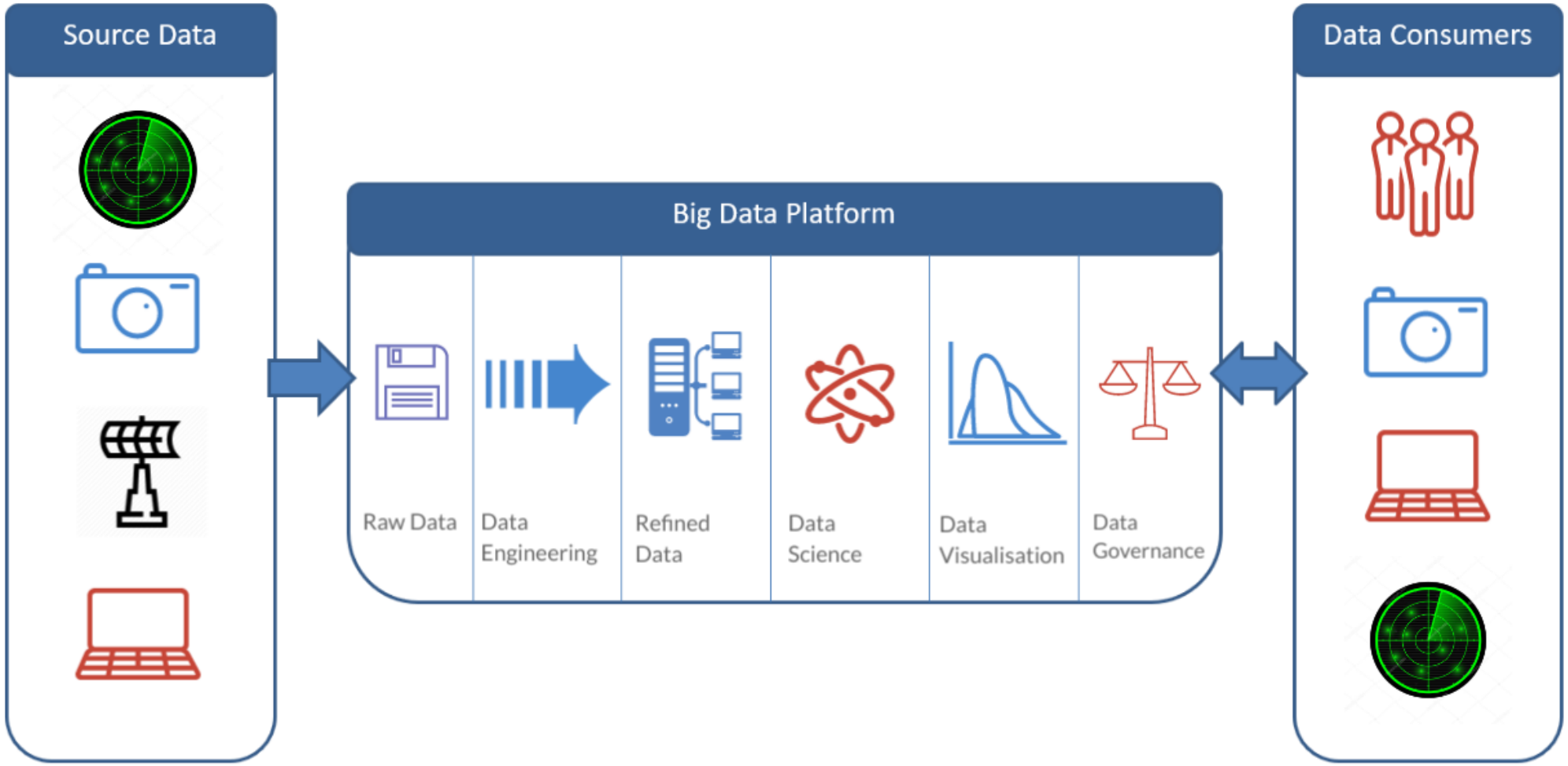
KEY USE CASES

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

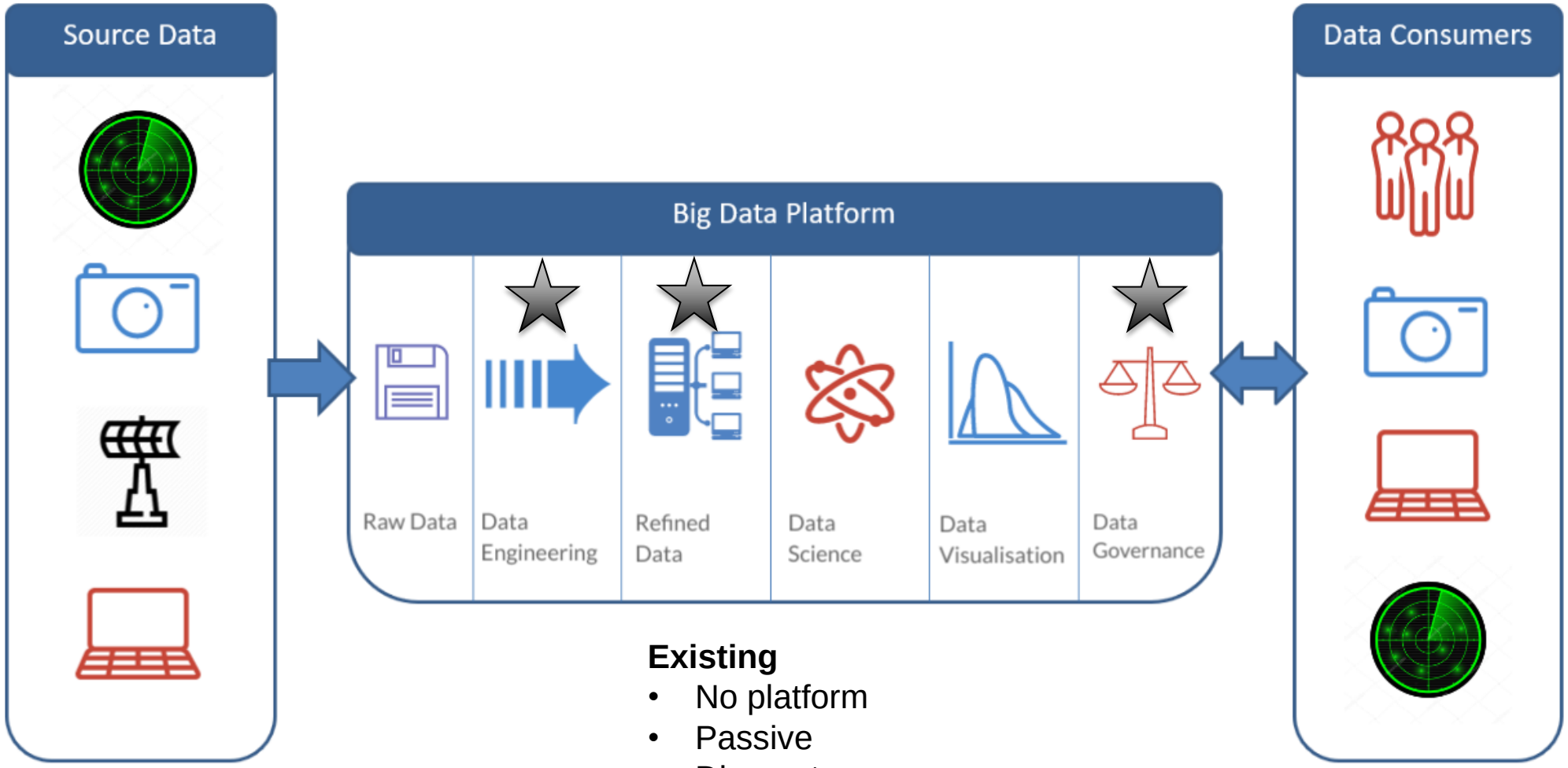
Advanced Search — **Accessibility** - right data for the right problem

Advanced Modelling — Deploy our MATLAB and Simulink models on large volumes of data - improved ability to **experiment** and **validate**

Data Architecture



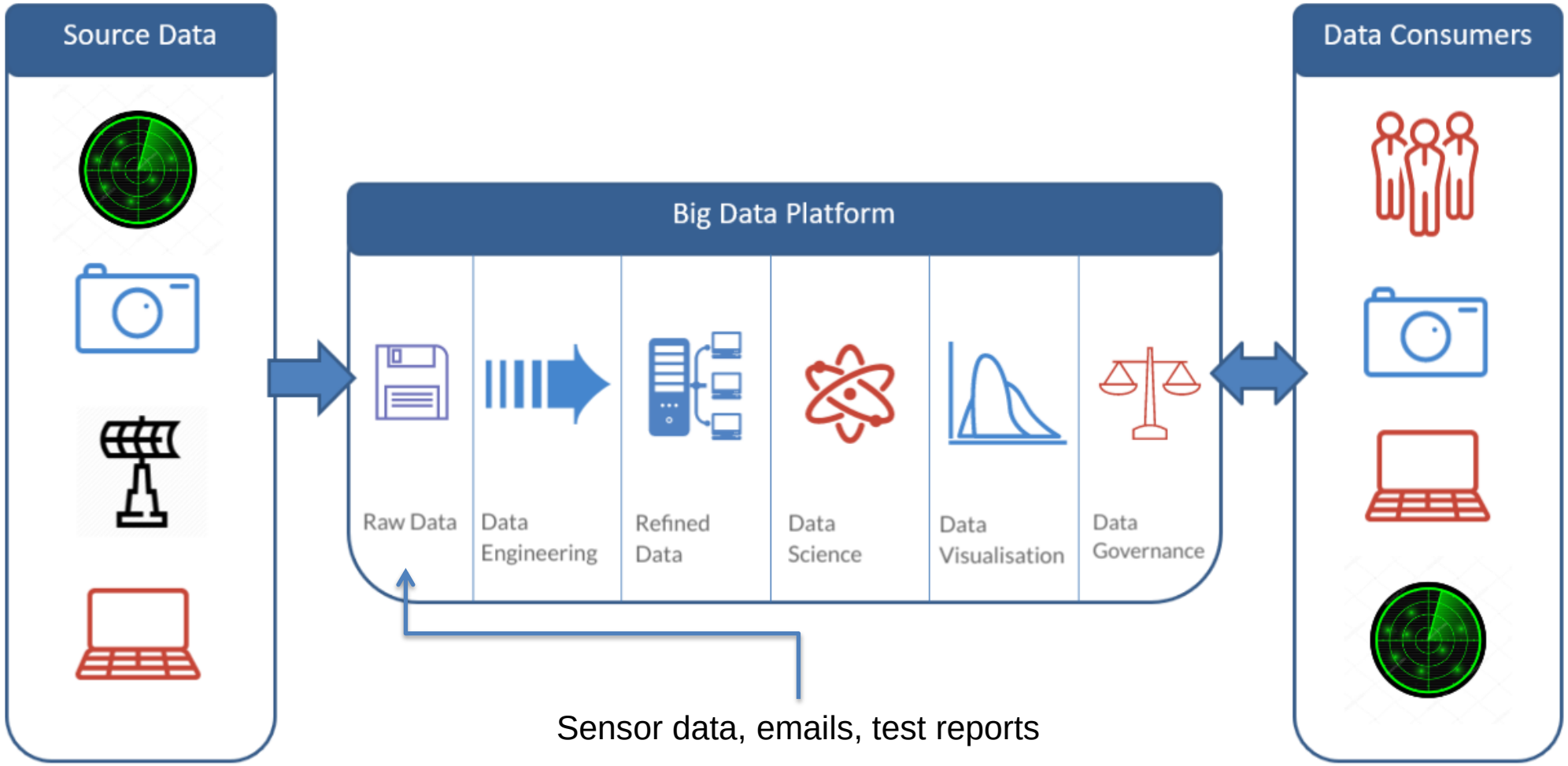
Data Architecture



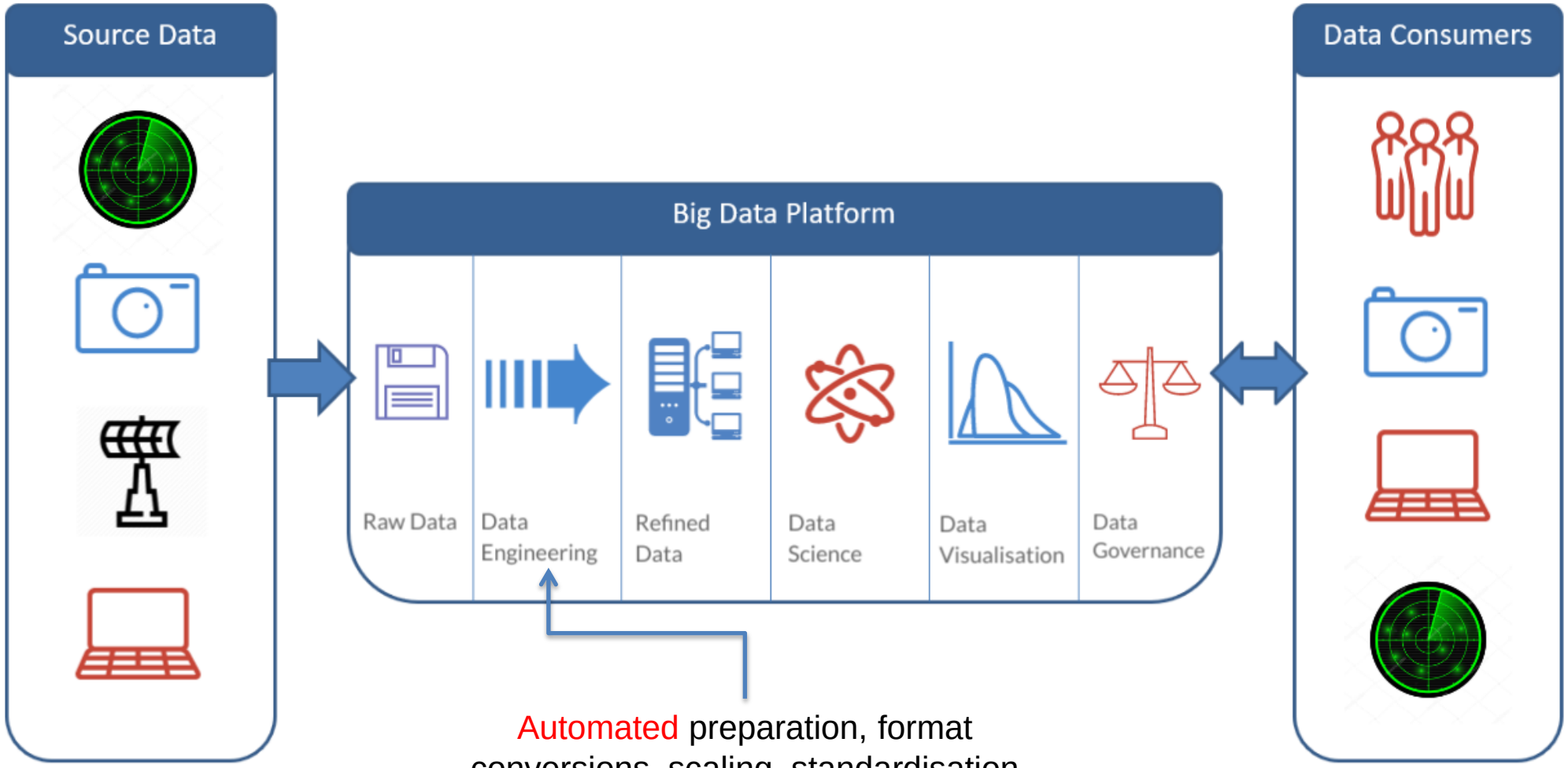
Existing

- No platform
- Passive
- Disparate

Data Architecture

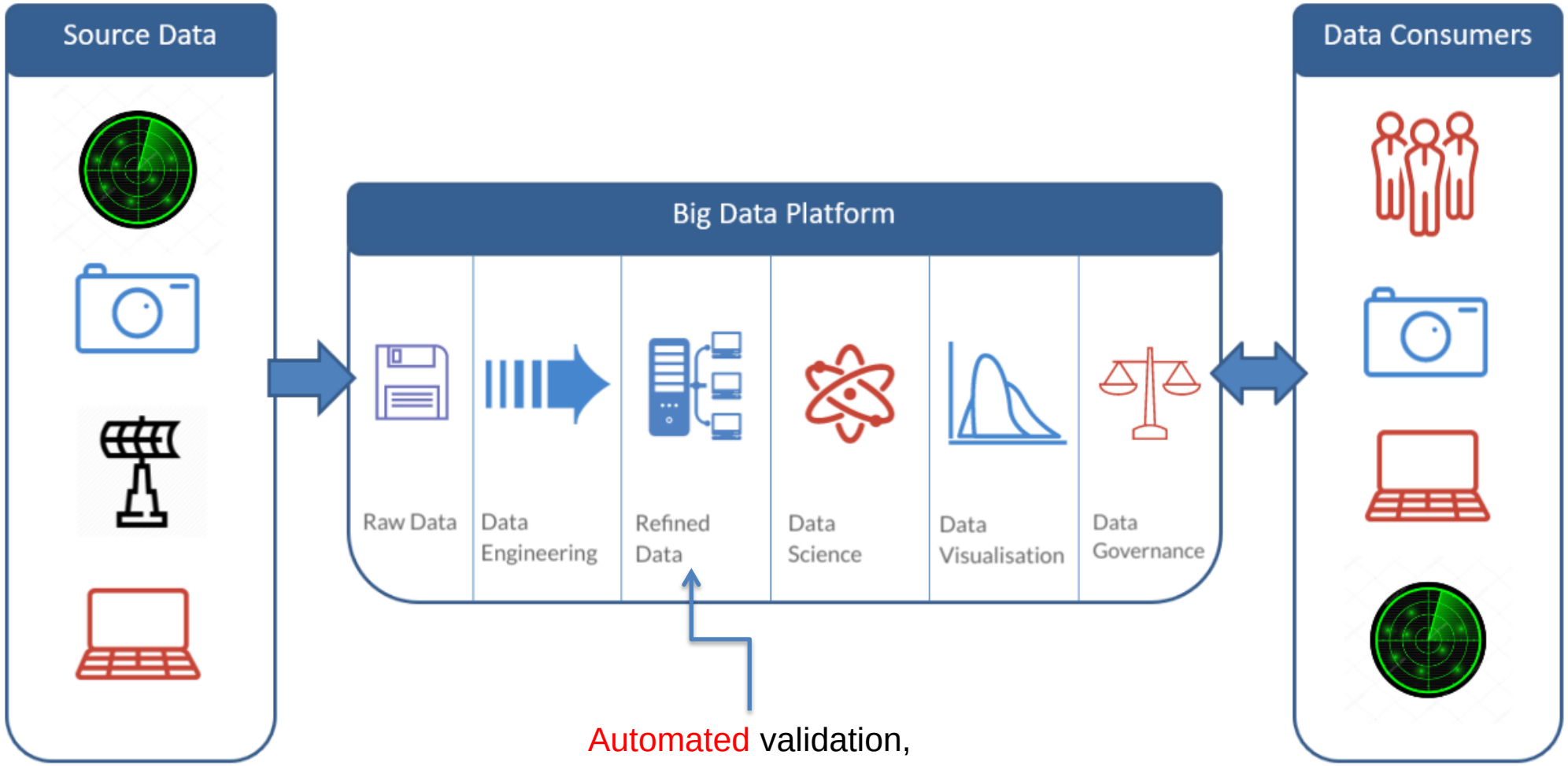


Data Architecture

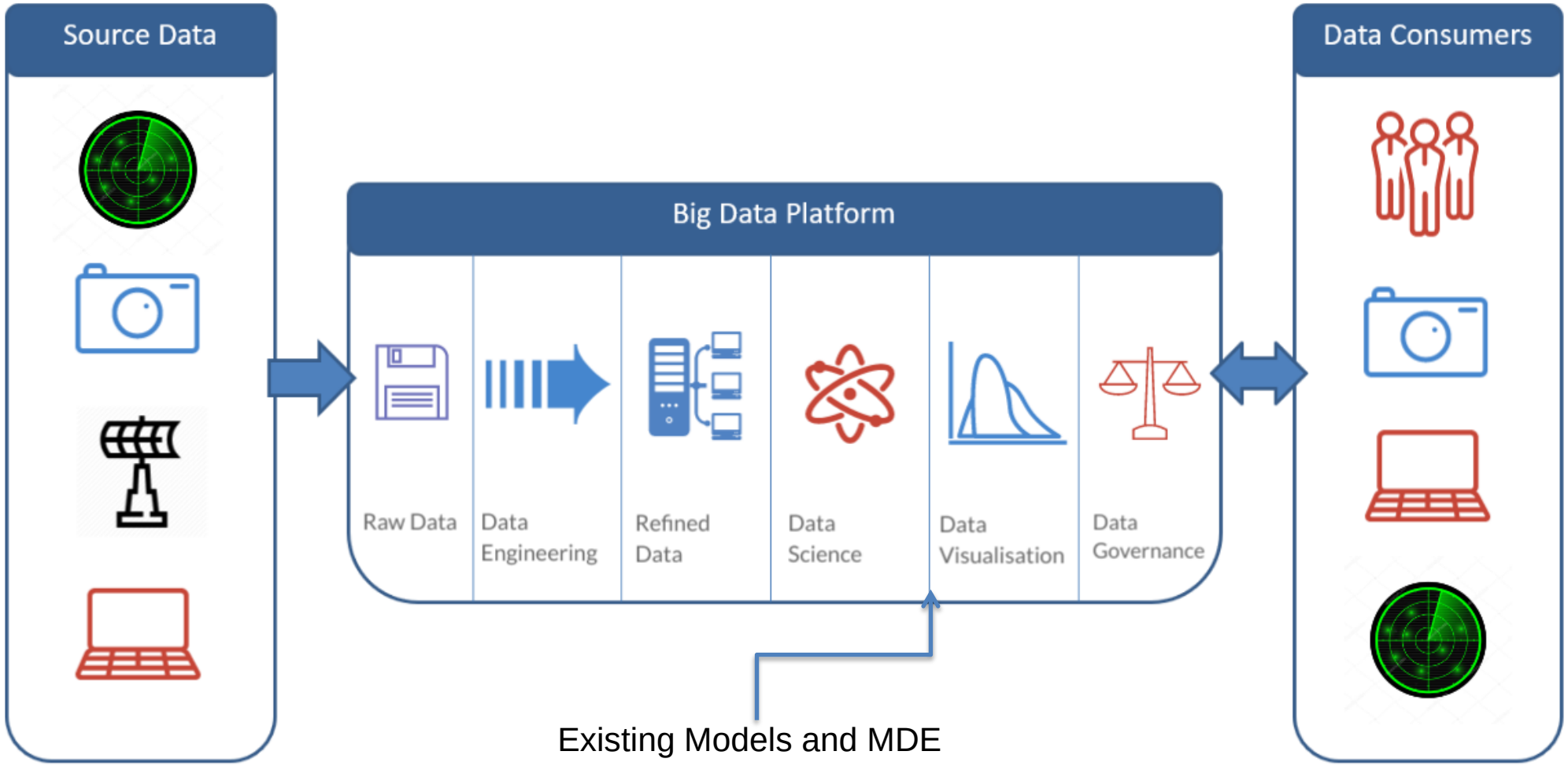


Automated preparation, format conversions, scaling, standardisation

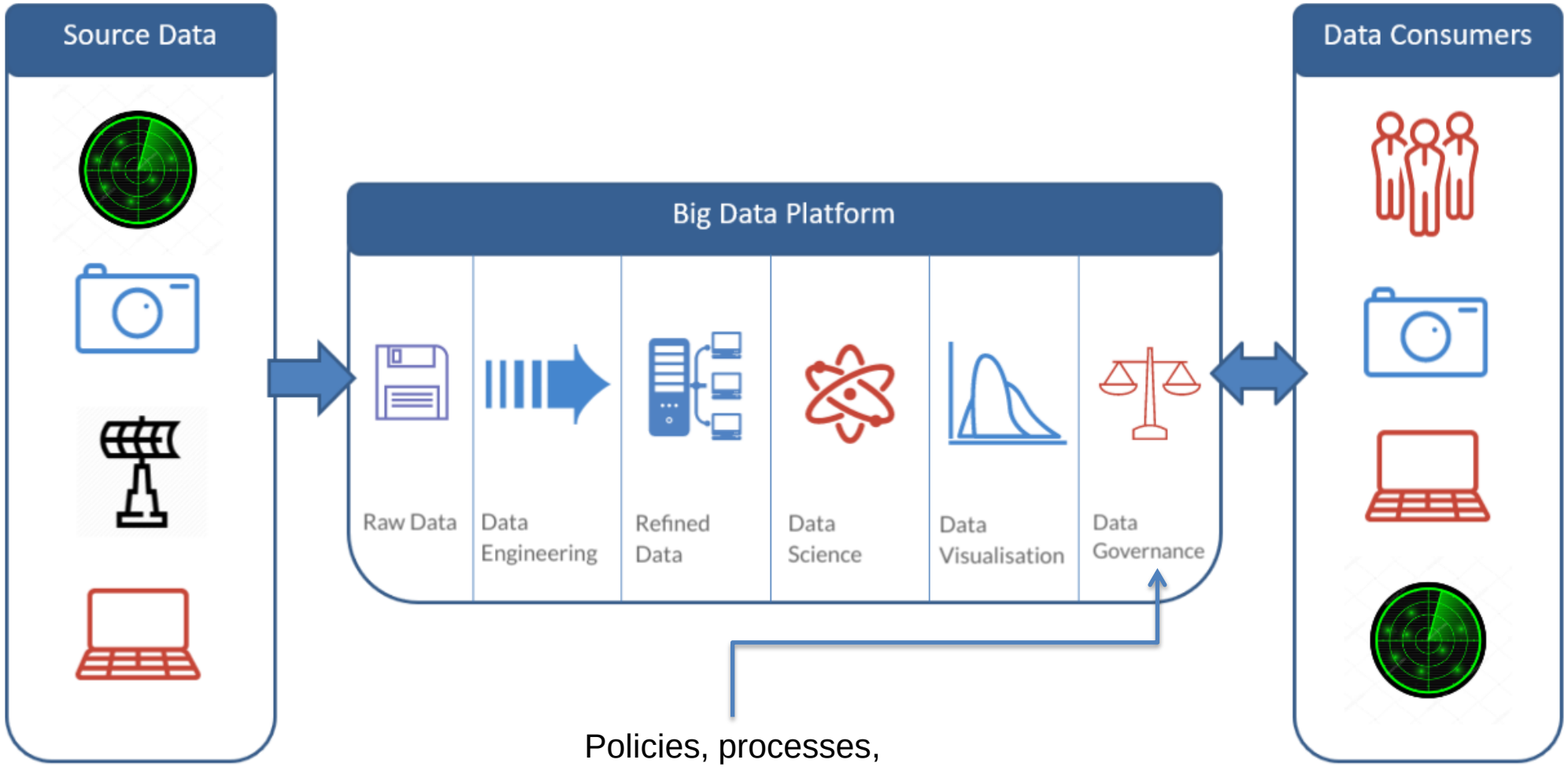
Data Architecture



Data Architecture



Data Architecture



Solution: BigData Platform as a Service

Leonardo Big Data Platform

Number of Server Racks

Space for expansion with COTS hardware.

2

20

Number of Processing Nodes

3 management nodes, 2 edge, 15 workers

TB of Total Installed Storage

300TB of usable storage after accounting for distributed file system

900

6

TB Memory

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**.

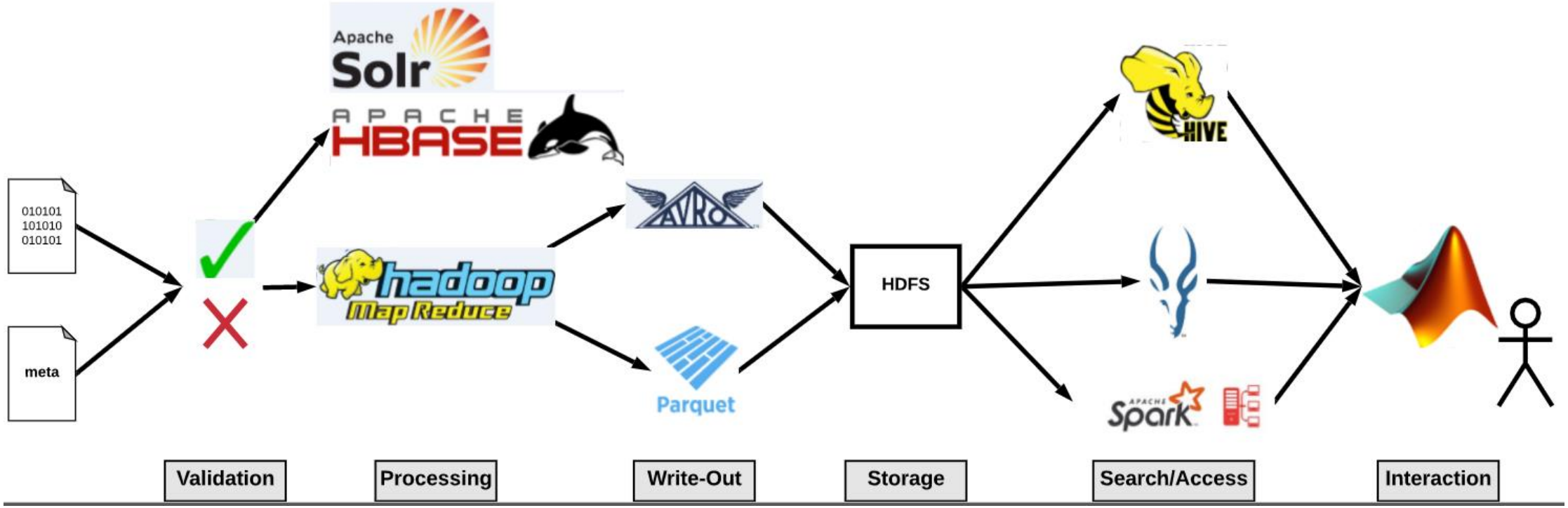
MATLAB® & SIMULINK®

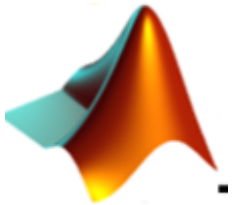
Core products in our **Model Driven Engineering** strategy and analytics

APACHE Spark™

Processing engine optimised for **distributed processing** in MATLAB, python, scala.

Processing Architecture...





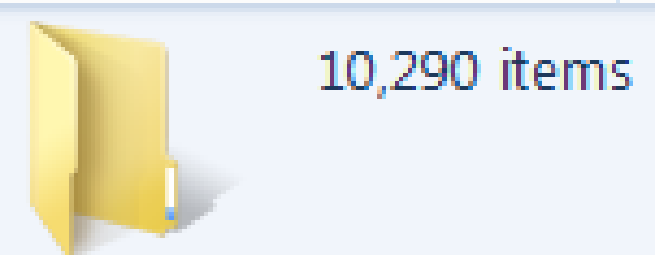
From this...

Name	Type	Size
metadataCaptureRecording_1_5.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_1_7.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_2_1.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_2_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_3_8.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_3_10.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_4_1.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_4_3.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_5_5.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_5_7.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_6_6.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_6_7.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_7_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_7_7.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_8_2.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_8_5.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_9_1.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_10_3.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_10_5.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_11_1.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_11_2.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_12_8.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_13_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_14_7.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_14_9.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_15_5.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_15_8.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_16_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_16_6.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_17_2.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_17_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_18_2.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_18_10.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_19_3.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_19_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_20_4.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_20_10.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_21_3.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_21_5.csv	Microsoft Excel Comma S...	1 KB
metadataCaptureRecording_22_4.csv	Microsoft Excel Comma S...	1 KB

How do I load lots of files?

what about different data formats?

how do I process huge volumes?



To this...

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

Easy to change processing environment
Local,
Local-Parallel,
Distributed...

```
function errorCode = mapReduceFramework(binaryFile, configFolder, outputFolder, varargin )
% Input options
%
% 'RunAsLocal' - true/ false specify running the framework as a compiled
% local job (runs as LocalMapReduce by default)
%
% 'DefineJavaHome' - Override the path set for JAVA_HOME with a user specified value


%% Handle any optional arguments passed in
iParse = inputParser();
addParameter(iParse,'RunAsLocal', 'false', @ischar)
addParameter(iParse, 'DefineJavaHome', '/usr/java/jdk1.8.0_191-amd64', @ischar)
addParameter(iParse, 'IntermediateFolder','/data_staging_2/IntermediateFolder',@isfolder)
parse(iParse,varargin{:})

ingestionFolder = fileparts(binaryFile);

%% Setup Data Access Layer & Logging

dal = setup_dal(binaryFile, ingestionFolder, configFolder, iParse.Results.IntermediateFolder, outputFolder);

dal.LogLocation = setup_logging(ingestionFolder,binaryFile);

print_dal_debug_info(dal);
%% Set Environment variable settings up
Logger.TRACE('mapReduceFramework:EnvironmentSetup','Setting Up Hadoop Environment');
setupHadoopEnv(iParse.Results.DefineJavaHome);

if isdeployed && strcmpi(iParse.Results.RunAsLocal, 'false')

    isDistributed = true;

    fileSizeInMb = computeYarnContainerSize(dal.IngestionFolder,'.bin');
    Logger.DEBUG('mapReduceFramework:config setup','Hadoop Map Container Size set to %d',fileSizeInMb)

    config = setup_hadoop_config(fileSizeInMb);

    % Temporary files can be written to HDFS if this a deployed Hadoop application
    [~,name] = fileparts(binaryFile);
    tmpFolder = fullfile('hdfs://nsprod1/tmp',name);
else
    isDistributed = false;
    config = 0;
```

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

Datastore 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

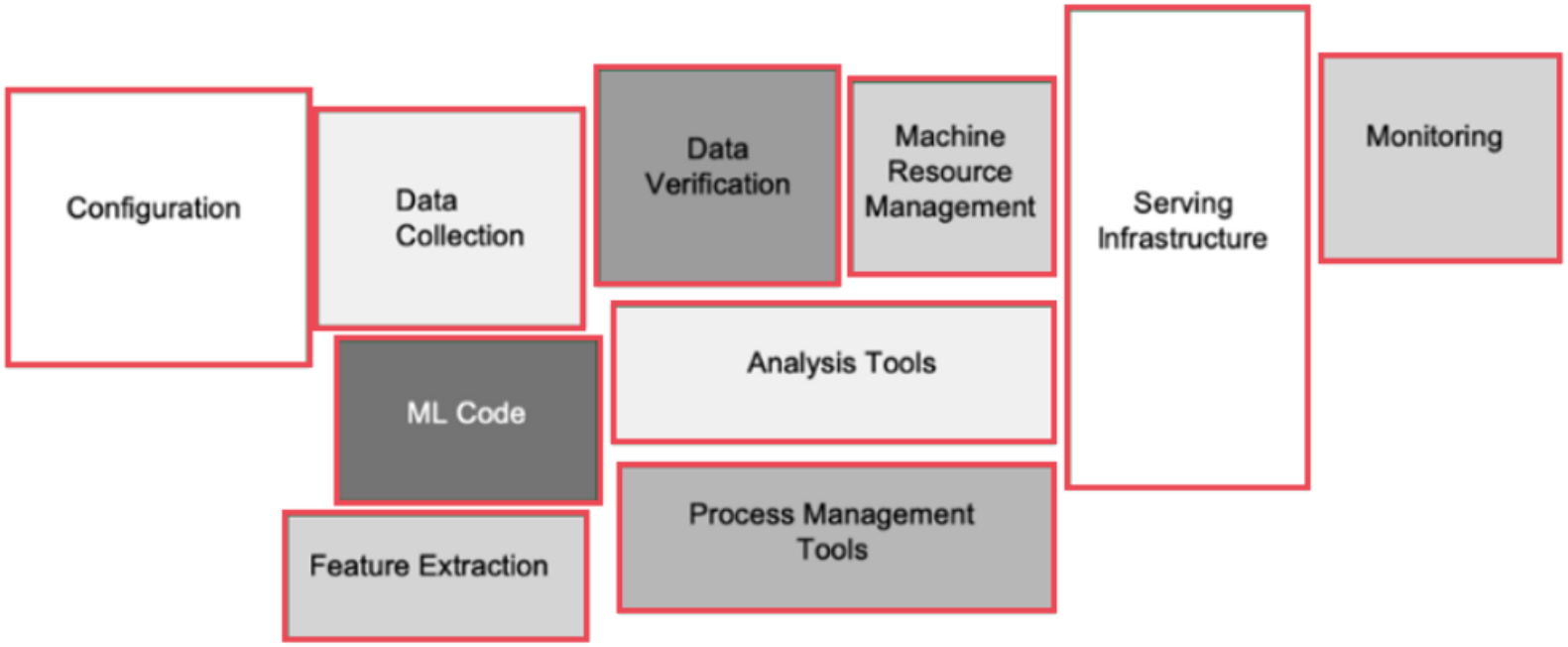
Improved Model Driven Engineering capability with deployment of MATLAB models

3 Future

What's next.

A great model isn't enough...

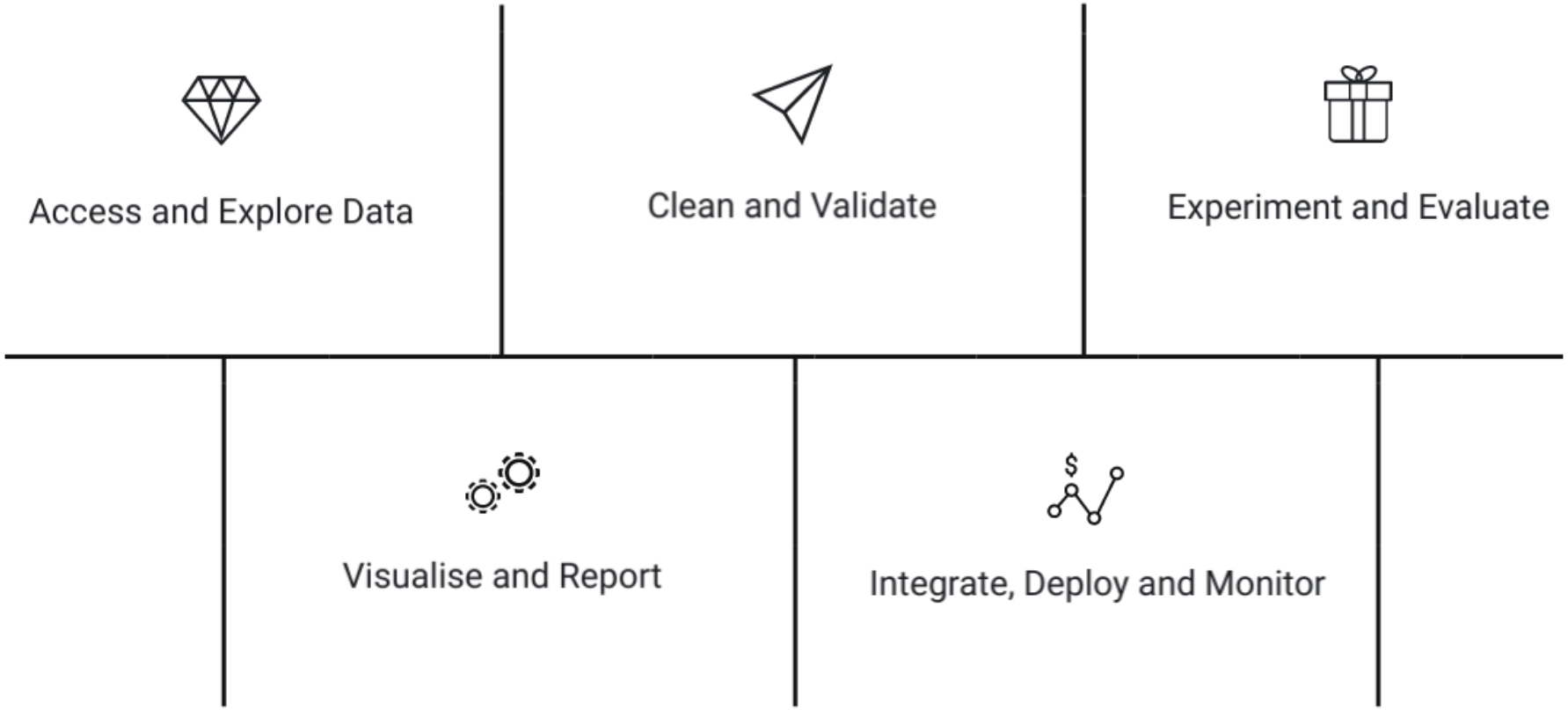
End to end solution to support a data-driven workflow



Source: *Hidden Technical Debt in Machine Learning Systems*, **Advances in Neural Information Processing Systems 28 (NIPS 2015)**

Towards DataOps

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

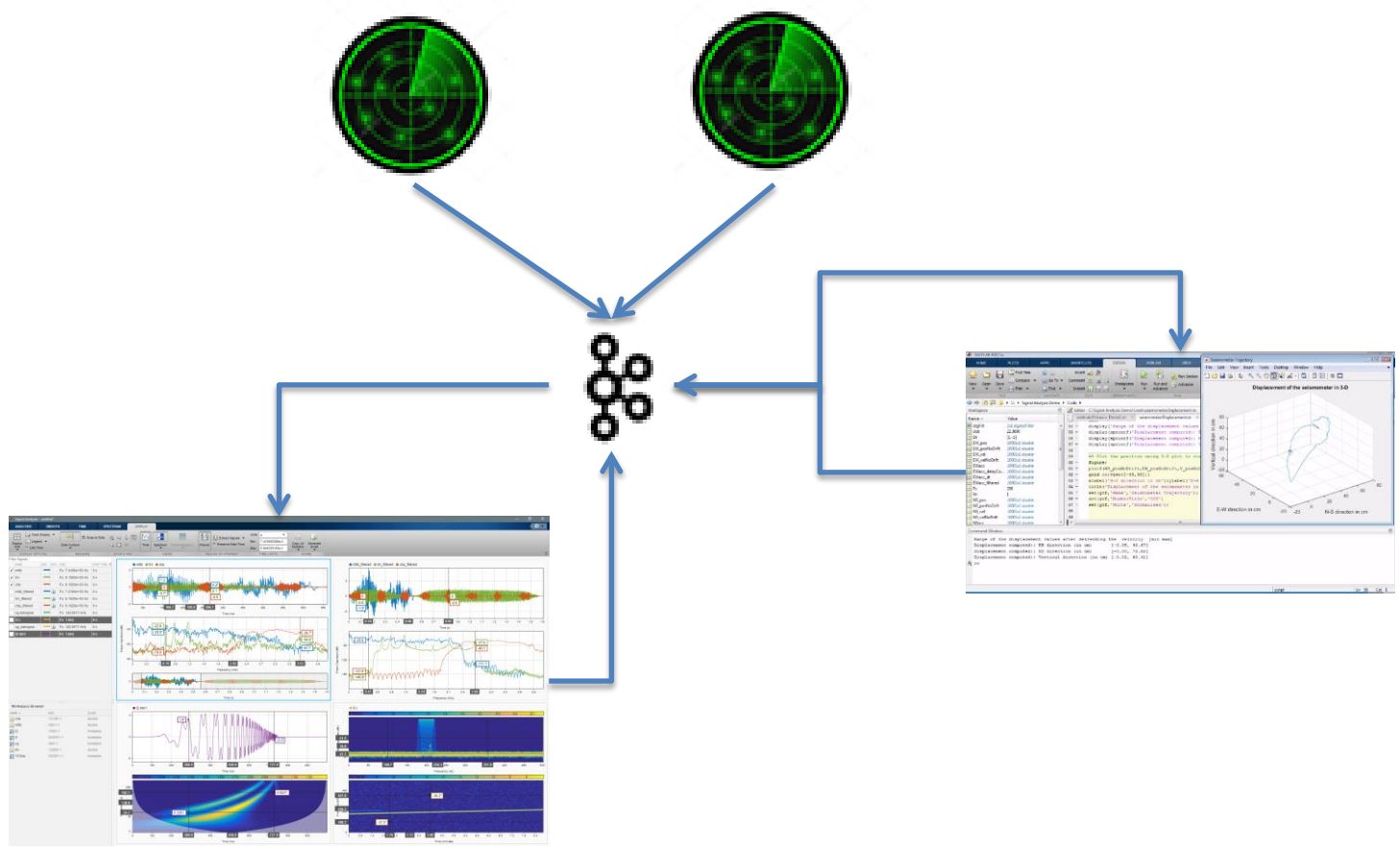


Future Technologies...

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