

Using a Data Lake Stack in Animal Sciences

12/12/2018

Dirkjan Schokker, Ioannis N. Athanasiadis, Bram Visser, Roel F. Veerkamp, Claudia Kamphuis



Acknowledgements



Big Data - data explosion

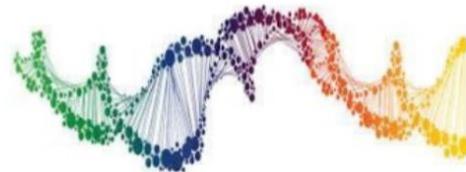


© Fotolia / Maksim Pasko

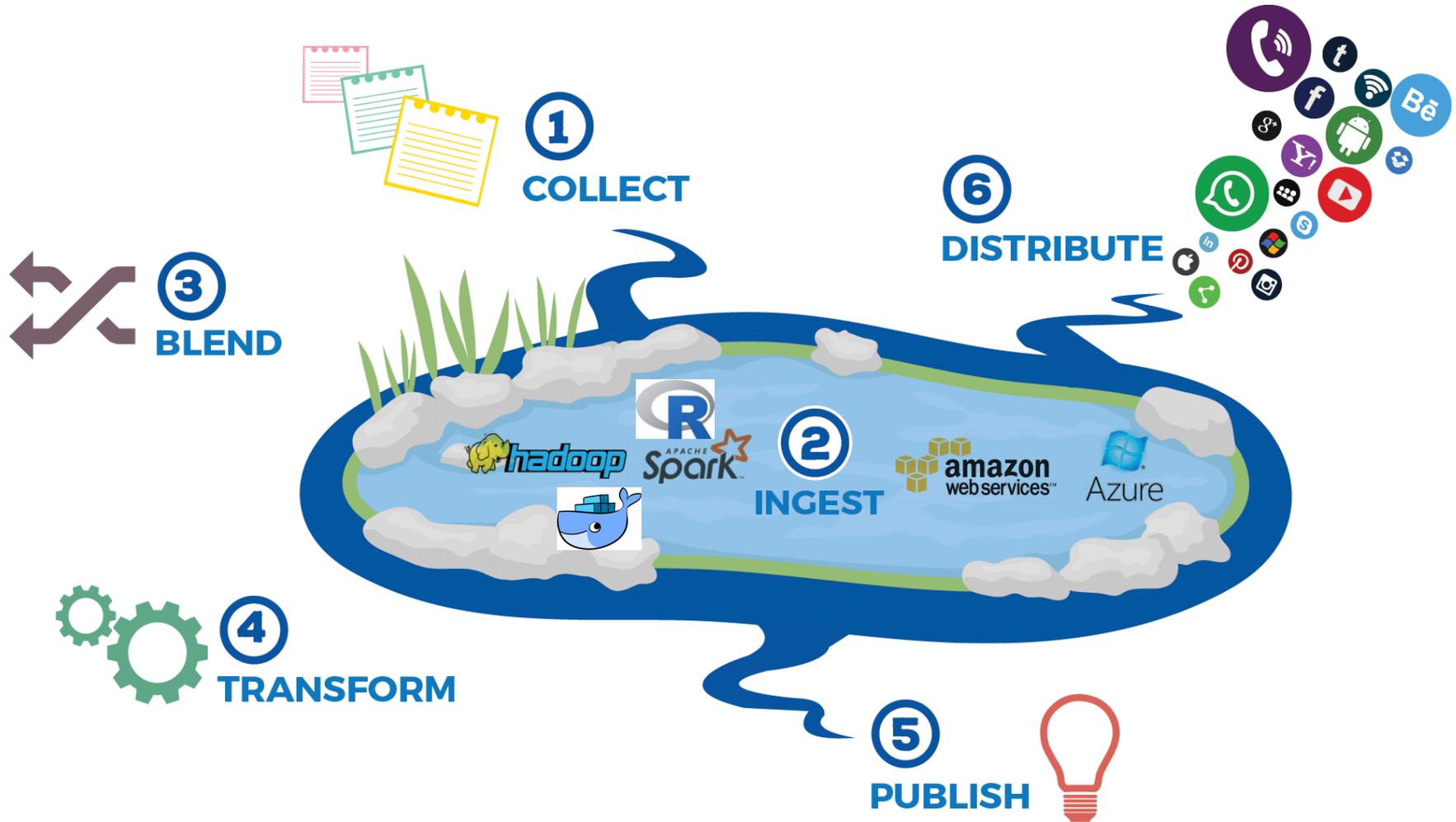
Example in Animal Sciences

Precision Livestock Farming: Dairy

Estimating individual feed intake of cows



Data lake stack



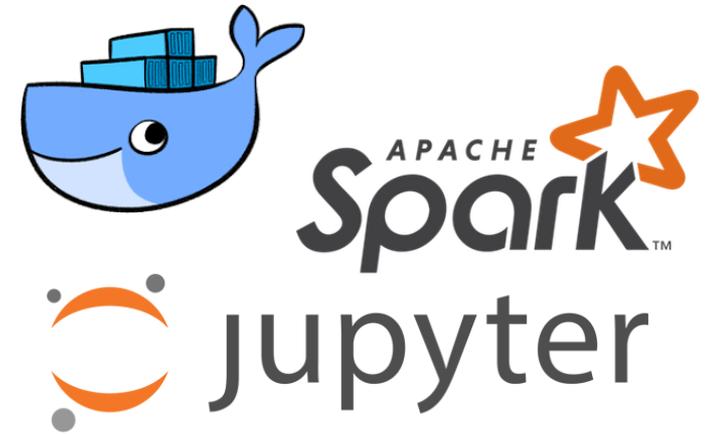
Key drivers of data lake

■ Need to handle:

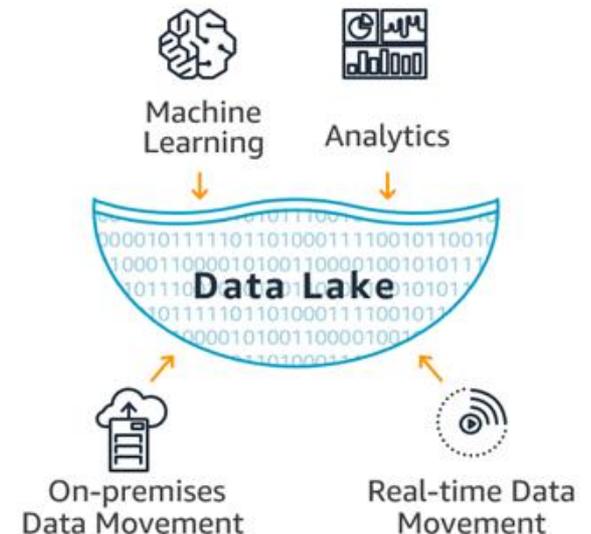
- Ever increasing datasets
- Varying data structures
- Heterogeneous
- Multimodal data

■ Improved:

- Scalability
- Modularity
- Interoperability



Extract Transform Load

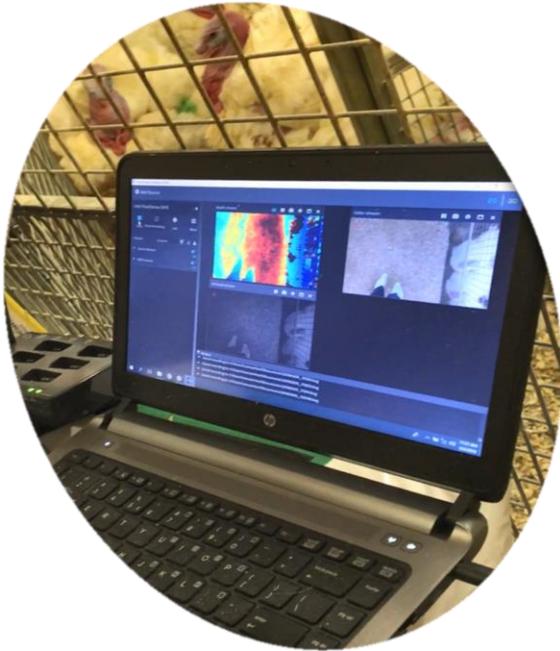


Current study - Locomotion



Gait score in action – ‘catwalk’

- Gait score of 200 turkeys
- Traditionally performed by a trained person

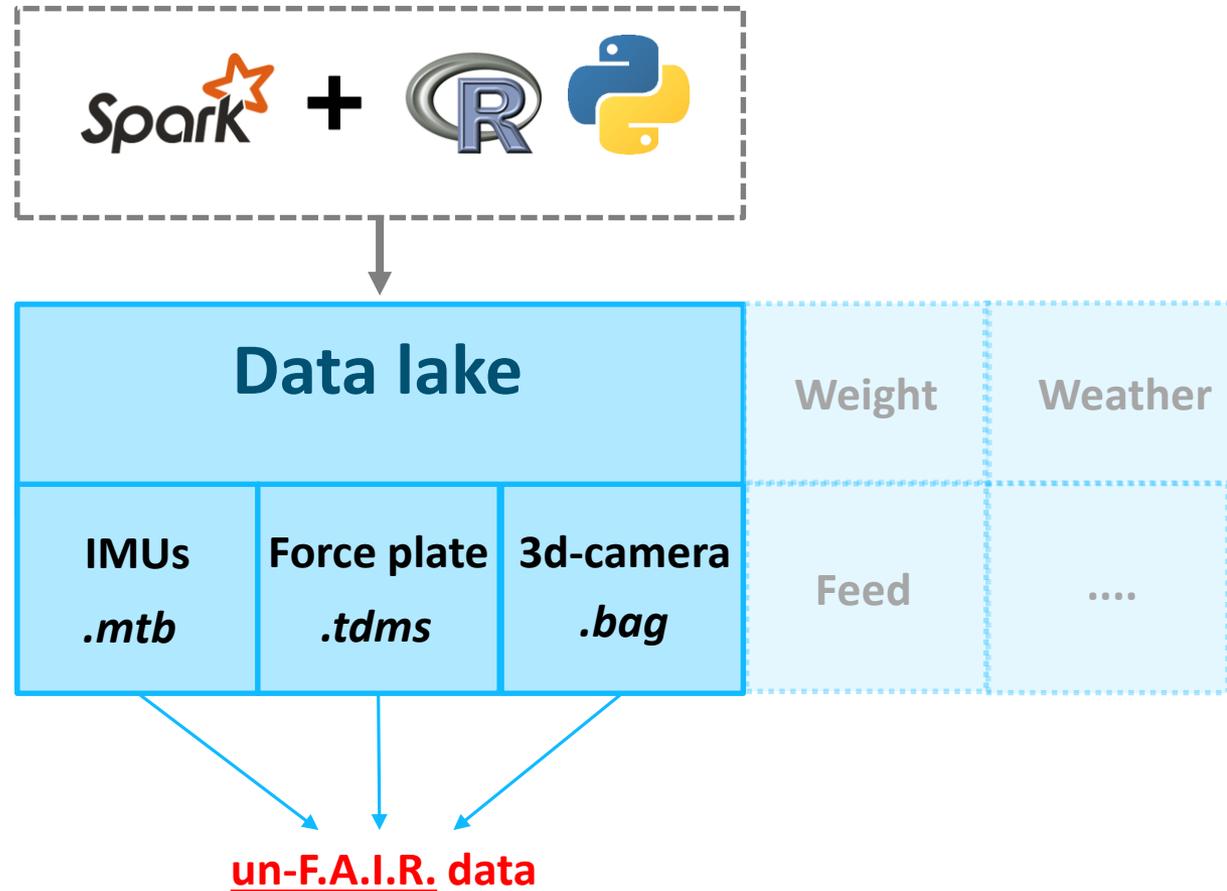


Experimental design and sensor data

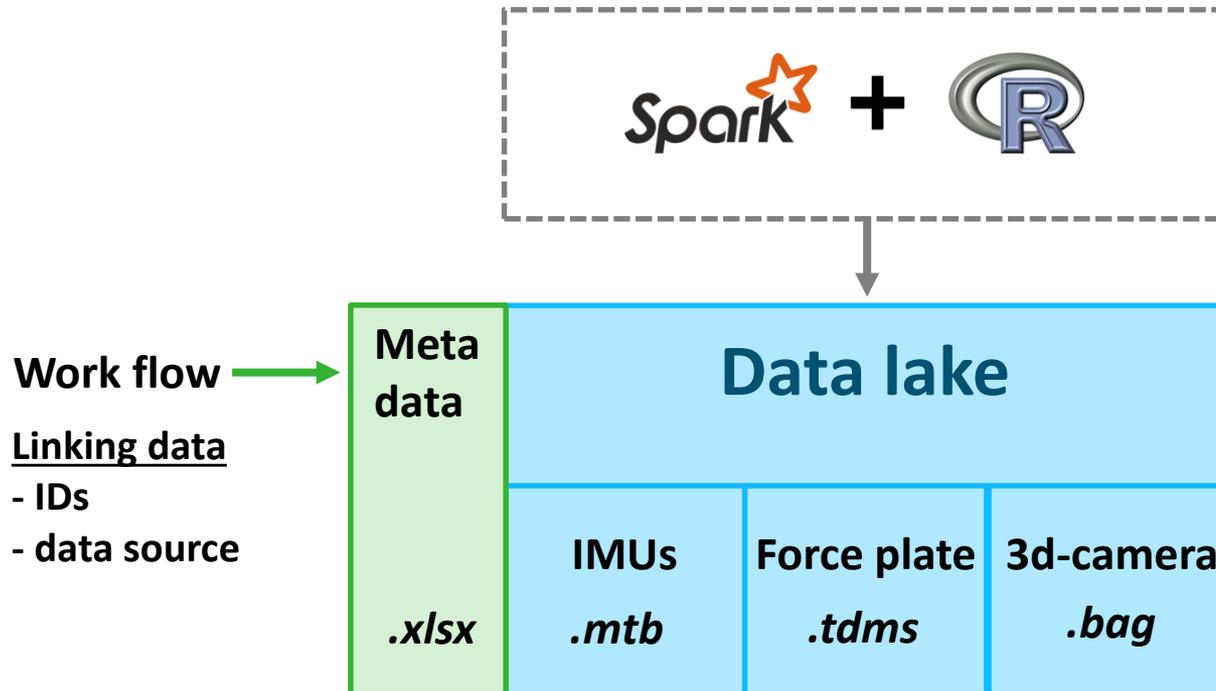
- Different Sensor data types were recorded
 - Inertial measurement units (IMUs)
 - 3D-video camera
 - Force plate



Schematic view of the data lake stack



Metadata not automatized yet



Discussion on FAIR-ness (I)

■ Findable

- Metadata is present, generating a “closed”, IP-protected, FAIR data point

■ Accessible

- Put a lot of effort in generating new open source scripts to make the data accessible

■ Interoperable

- Now it is, readme files in English and scripts in Python/C++/R

■ Reusable

- Scripts are ready to reuse (*gitlab*), data not yet (IP)

Our focus is more on reproducible, scalable ETL pipelines

Discussion on data lake (II)

■ Is it worthwhile to use in animal science?

- Entire 'universe' of data captured and maintained



- No data loss and scalable
 - Stored near native format
 - Can be pushed to cloud services

■ Lessons learned for breeding

- More aware of repeated measures (large volume) and heterogeneity of data (variability)
- Necessity to have open source scripts/pipelines for handling data (automation)

Thank you for your attention



dirkjan.schokker@wur.nl



+31 (0)317 480537



WAGENINGEN
UNIVERSITY & RESEARCH



100years
1918 — 2018