Agile Data Science
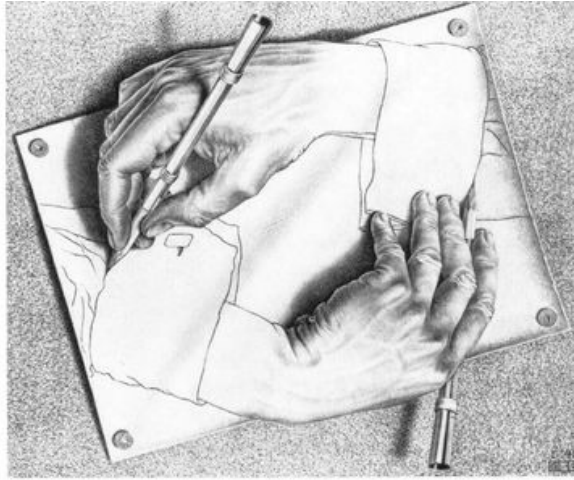
Waclaw Kusnierczyk | waclaw.kusnierczyk@gmail.com

Outline

1. What is Agile Data Science?

2. What is Agile?

3. What is Data Science?

4. What is Agile Data Science?

5. What Agile Data Science Is Not?
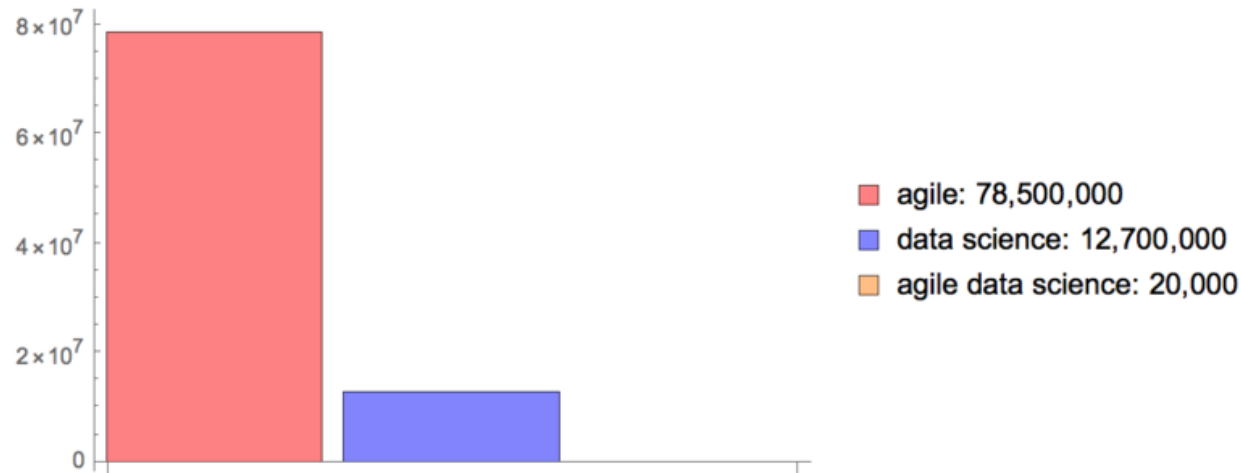
What is Agile Data Science?

Everybody knows what **agile** and **data science** are.

- 'Data science' is no longer a buzzword (or is it?), so we **need another one**.
- What's better for insight into agile data science than agile data science?

What is Agile Data Science?

What does Google search tell us about agile data science?

- agile: 78,500,000
- data science: 12,700,000
- agile data science: 20,000

Question | Have **you** ever heard about agile data science?

What is Agile?

Agile is...

- alternative to traditional sequential development;
- meant to help in **responding to unpredictability**;
- **incremental**, **iterative** work with empirical feedback;
- focused on potentially **shippable product** increments in short iterations;
- **continually revisiting** every aspect of development throughout the lifecycle;
- **continuously re-evaluating the direction** with the possibility of **change**.

What is Agile?

Agile is...

- a **well-established** software development methodology;
- somewhat **unpopular** before Christmas.

Interest over time ⊙

100

75

50

25

Dec 21 - Dec 27 2014	
agile	51

Sep 11, 2...   Mar 17, 2013   Sep 21, 2014   Mar 27, 2016

What is Data Science?

Data science is...

- the "Sexiest Job of the 21st Century";
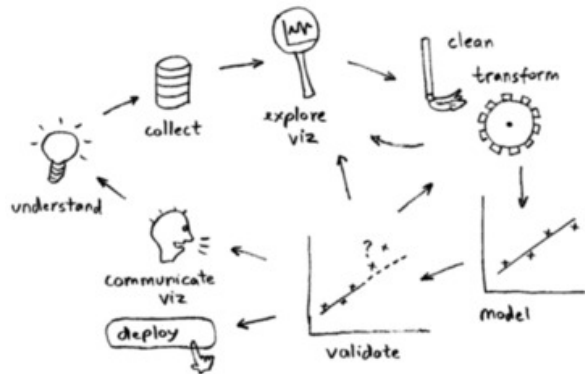- particularly exciting after vacation.

Interest over time ❓

What is Data Science?

Data science involves:

- extraction, transformation, loading of data (ETL), cleaning, preprocessing;
- **exploratory** data analysis (EDA);
- statistical **inference and modeling**, machine learning;
- model testing, validation, tuning, optimisation;
- visualisation, **delivery of insight**;
- **design and prototyping** of data-based products.

What is Data Science?

A data scientist is...

- **better statistician** than most programmers;
- **better programmer** than most statisticians.

© morganimation - Fotolia.com   #83642646
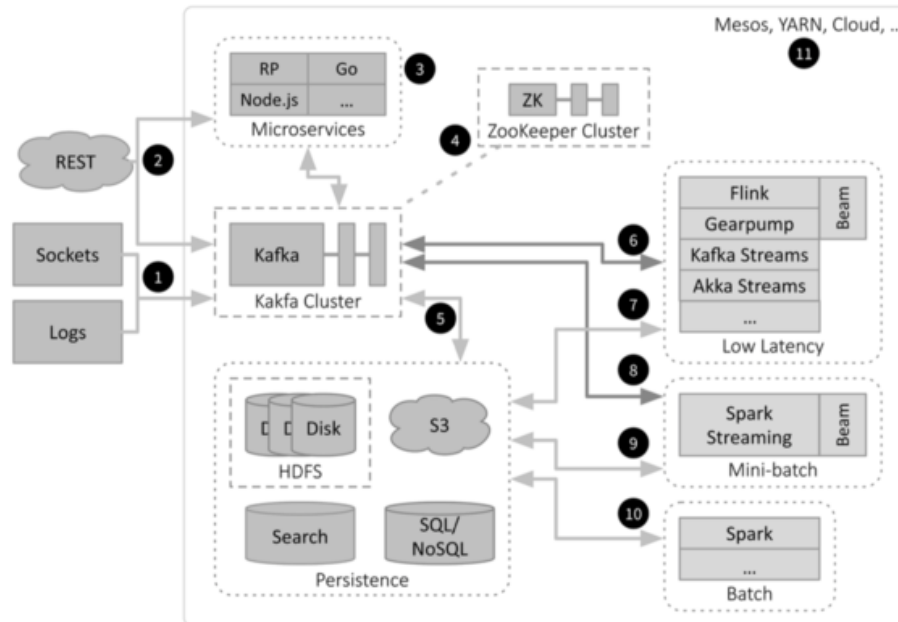
What is Agile Data Science?

Agile data science...

- focuses on **delivery** of insight and predictions **efficiently**;
- creates **research plans to build MVPs**;
- prefers **simplistic but efficient models** to elaborate but slow ones;
- uses **off-the-shelf tools** to the extent possible;
- evaluates the results against **business objectives**.

What is Agile Data Science?

Agile data science requires **modular**, **fast architecture** for storing and processing event and other data.
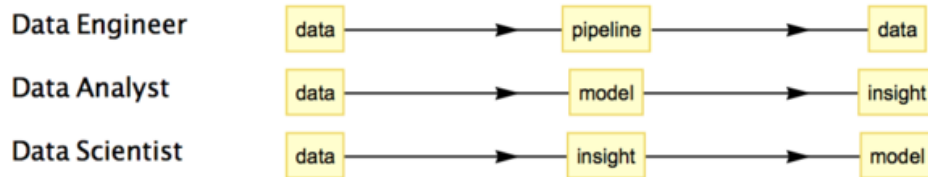
What is Agile Data Science?

Agile data science requires **mixed competence** teams.

Very **roughly**,

- a **data engineer** focuses on data;
- a **data scientist** focuses on models;
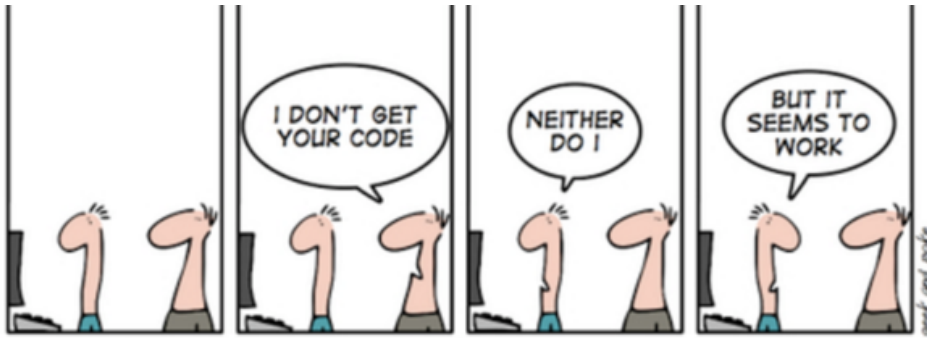- a **data analyst** focuses on insights.

Question | Have you heard about a **full-stack data scientist**?

What Agile Data Science Is Not?

Some actual responses from data scientists:

- On my laptop, the results were better.
- The code ran successfully in PyCharm.
- Change alpha to 0.2 in line 324 of the script.
- I'm not sure what alpha was.
- I'm not sure what alpha is.
- The code seemed to run fine before the last update I committed.
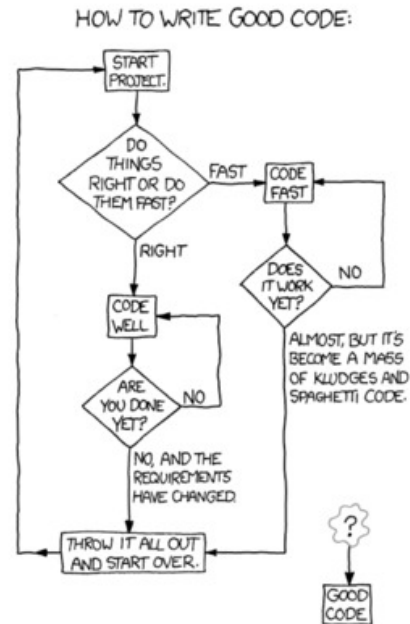- You need to install RServe on your production cluster.

What Agile Data Science Is Not?

Some actual responses from data scientists:

- The library I import is used by many, why would I test the code.
- It worked but that library changed.
- It worked with a small batch of data.
- It makes no sense to try linear regression.
- The new model may perform better than the old one.
- I can tell you how to feed the model.
- Let me recompile those LaTeX report sources for you.

Thank you!