
DATA
SCIENCE


PROCESS
MODEL

Florian Dohmann

Data Scientist @ *um

*um

The unbelievable Machine Company GmbH

S P E C I A L I S T F O R

C L O U D S E R V I C E S
I N T E R N E T A P P L I C A T I O N S
B I G D A T A

F R O M B E R L I N

We create data solutions.

From idea to cable.

BORN

2008

DATA SCIENCE

2011

2013

2015

~100

Employees in Berlin & Vienna

Data Science

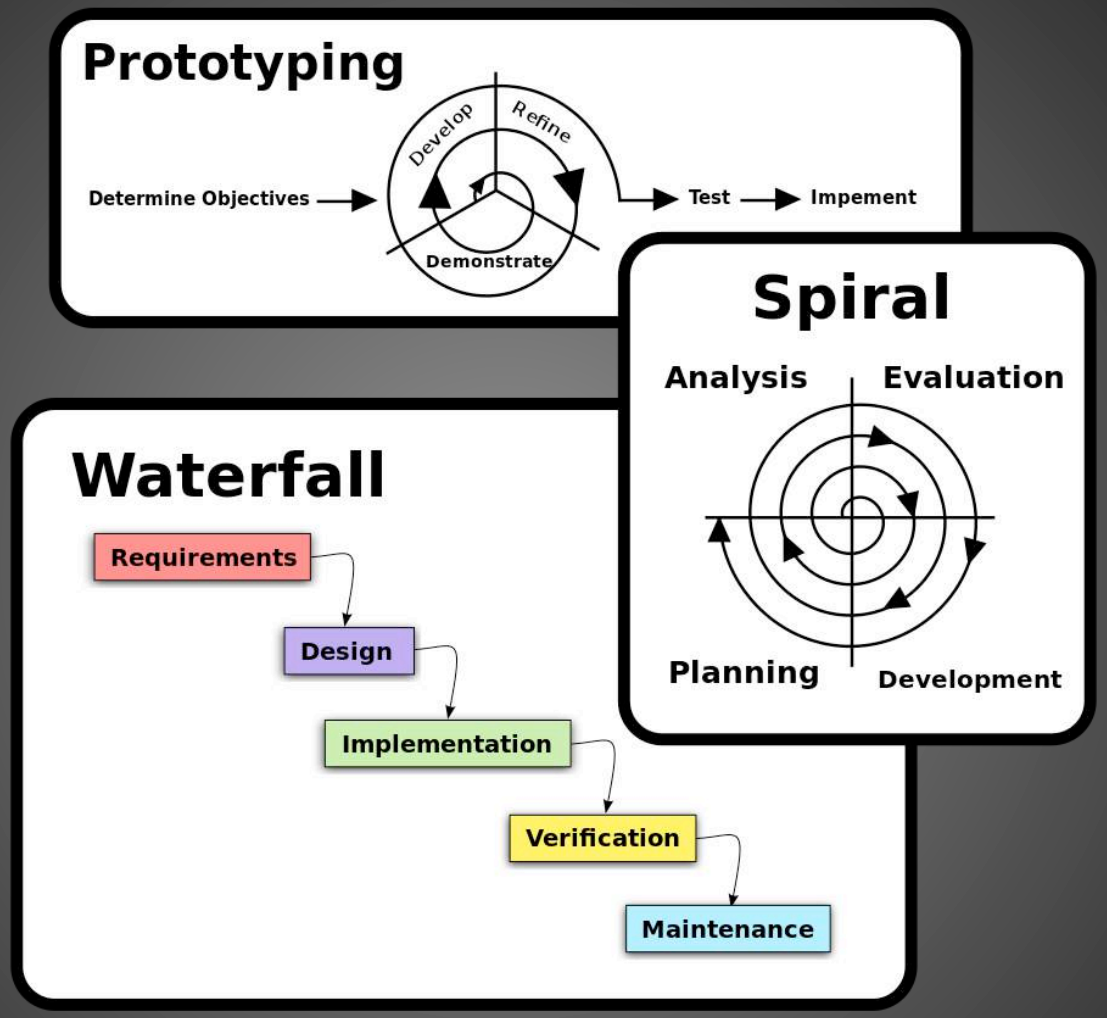
Custom Big Data Analytics, Machine Learning & Co.

Workshops & Trainings
Support in creating own Data Science Teams

= Full-Service

„We need a perfect interplay of human & machine intelligence“

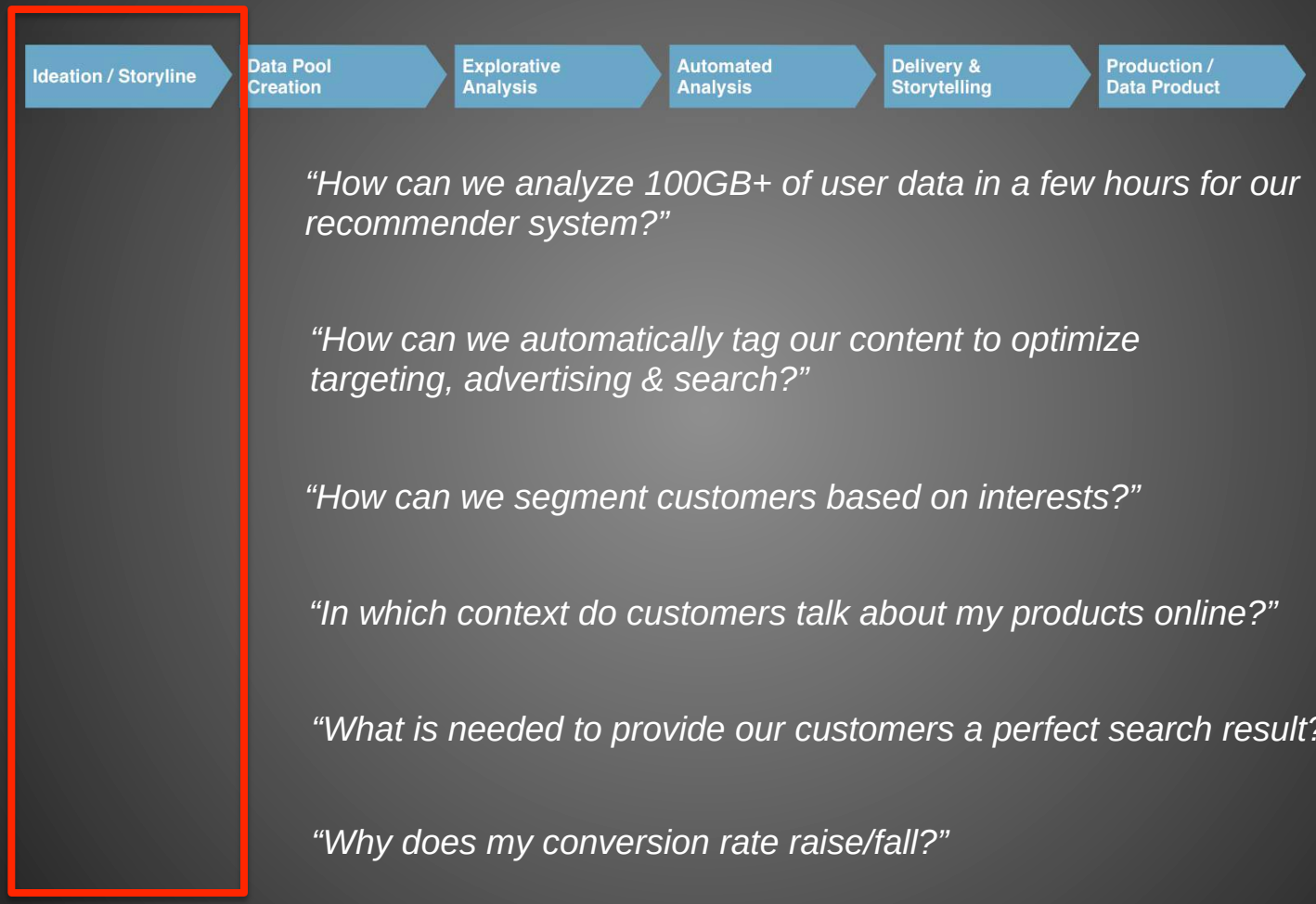
Process?



http://en.wikipedia.org/wiki/Software_development_process#mediaviewer/File:Three_software_development_patterns_mashed_together.svg

Data Science Process Model: Stages

Data Science Process Model



"How can we analyze 100GB+ of user data in a few hours for our recommender system?"

"How can we automatically tag our content to optimize targeting, advertising & search?"

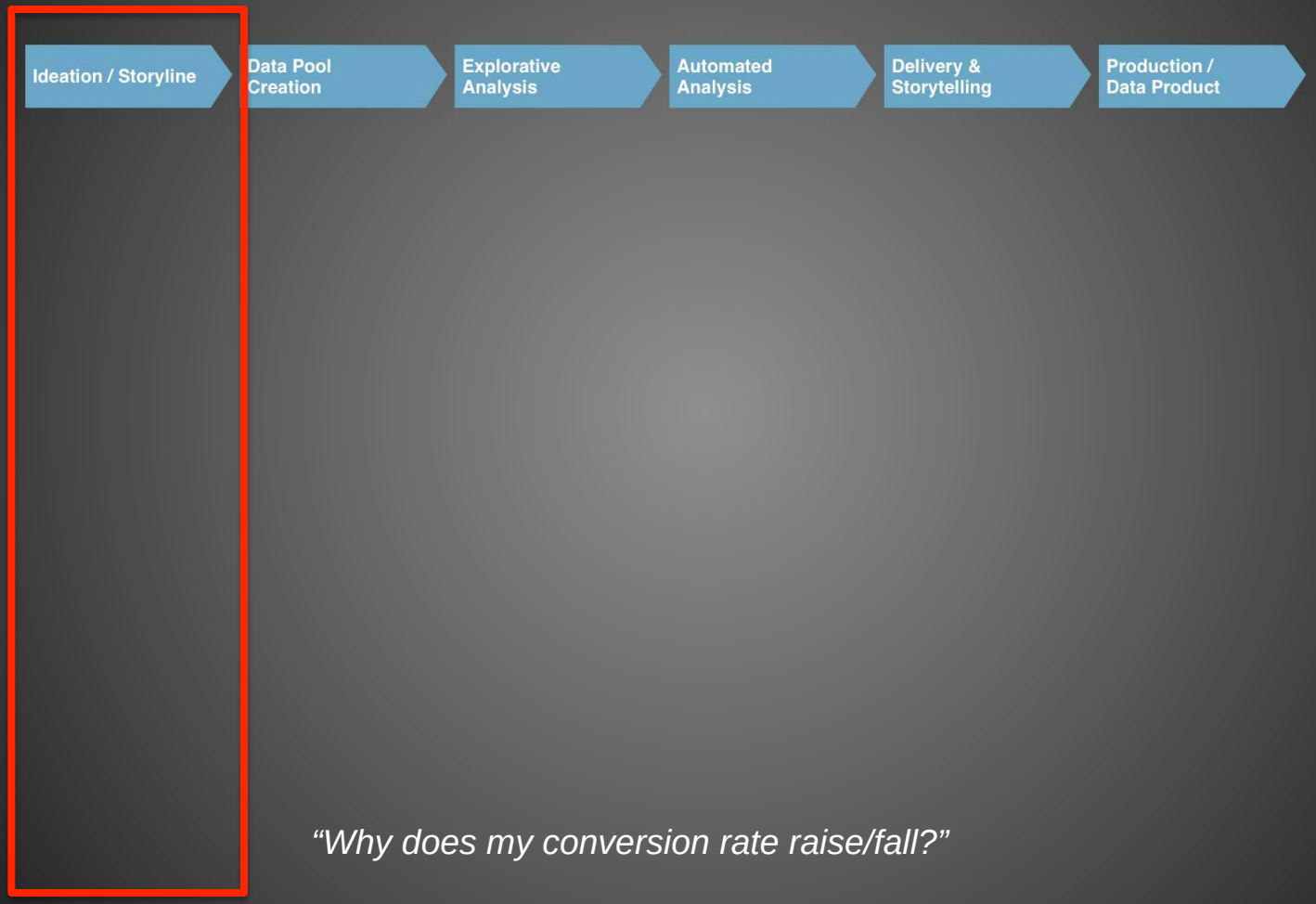
"How can we segment customers based on interests?"

"In which context do customers talk about my products online?"

"What is needed to provide our customers a perfect search result?"

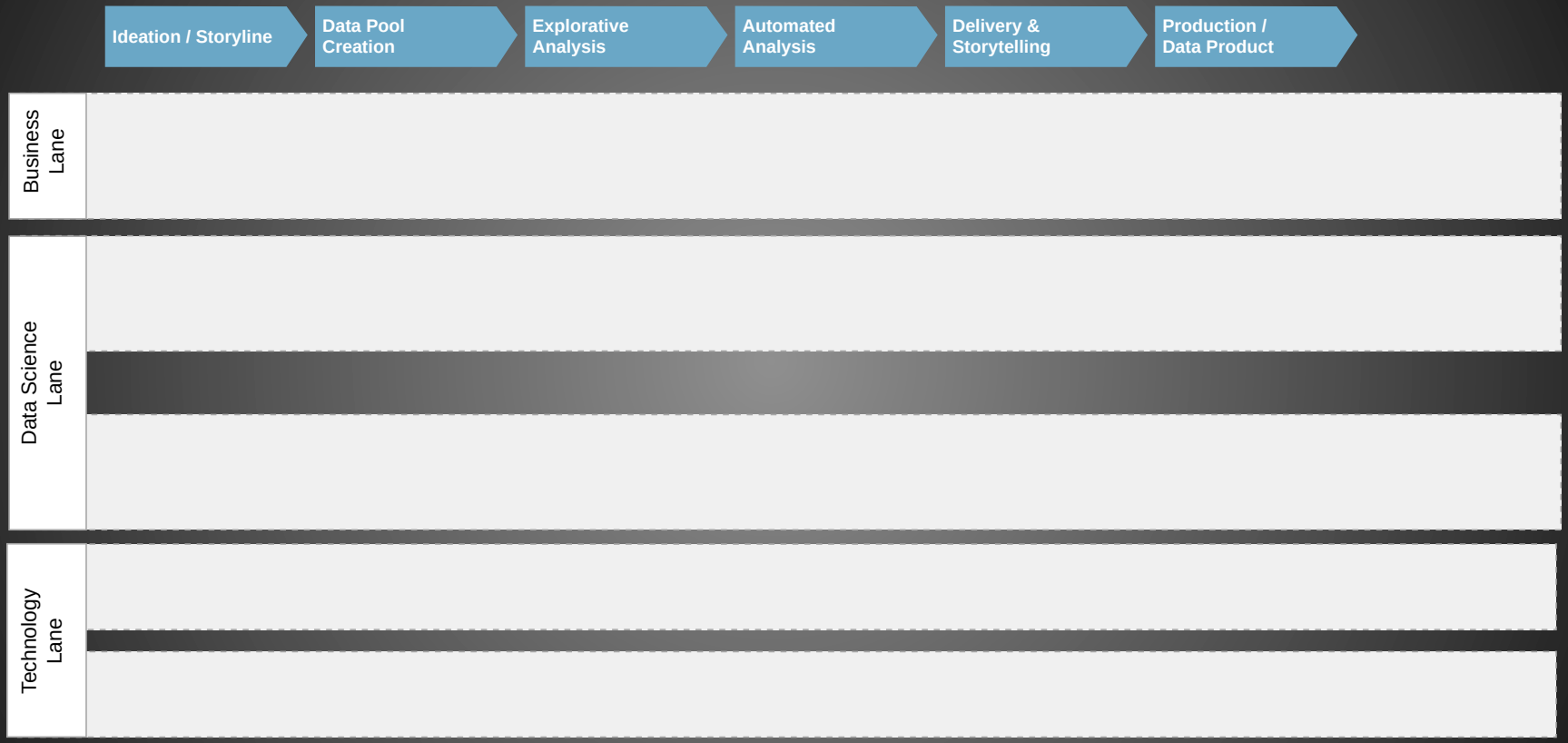
"Why does my conversion rate raise/fall?"

Data Science Process Model

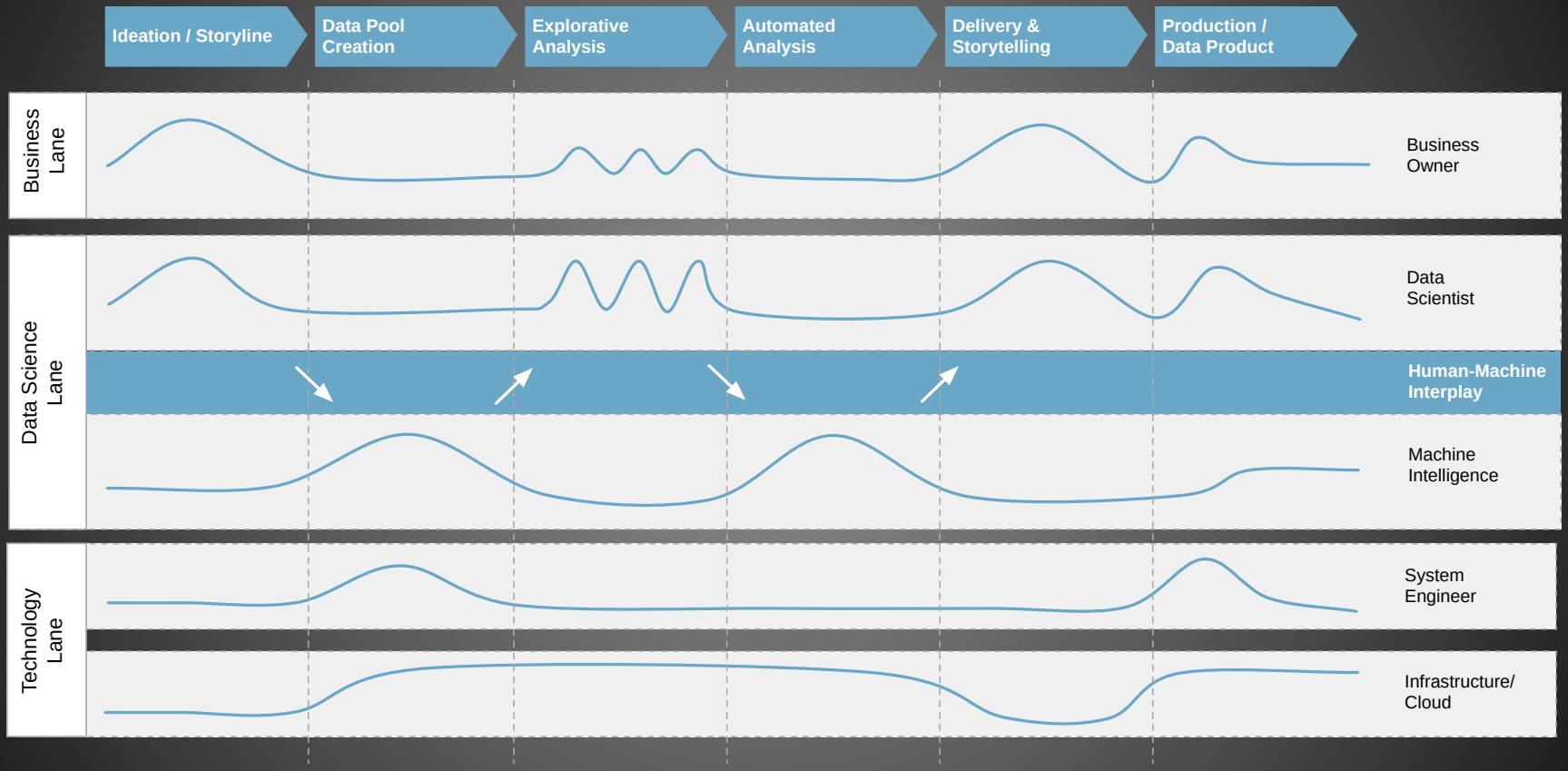


“Why does my conversion rate raise/fall?”

Data Science Process Model: Stages & Lanes

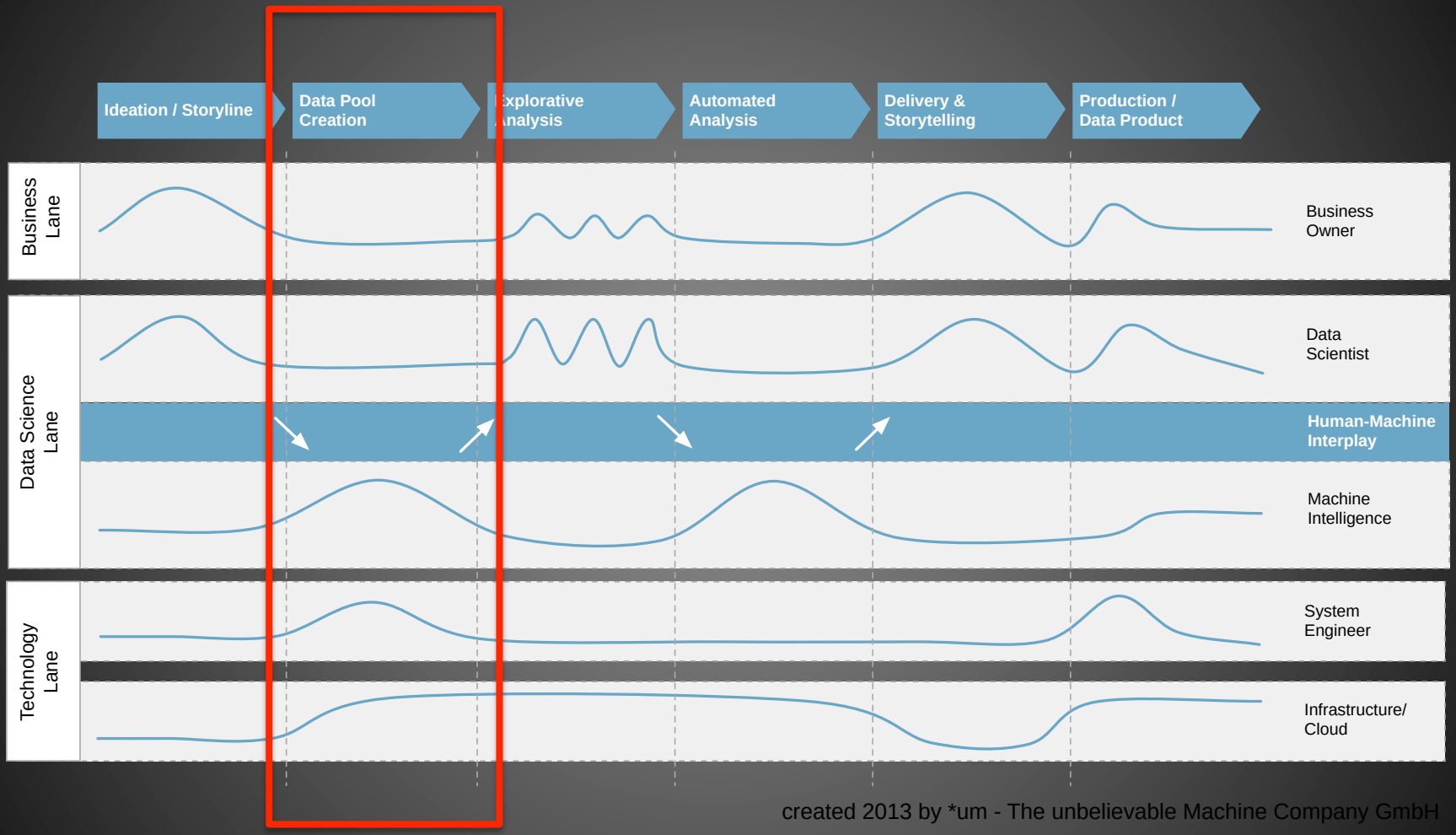


Data Science Process Model



created 2013 by *um - The unbelievable Machine Company GmbH

Data Science Process Model



created 2013 by *um - The unbelievable Machine Company GmbH

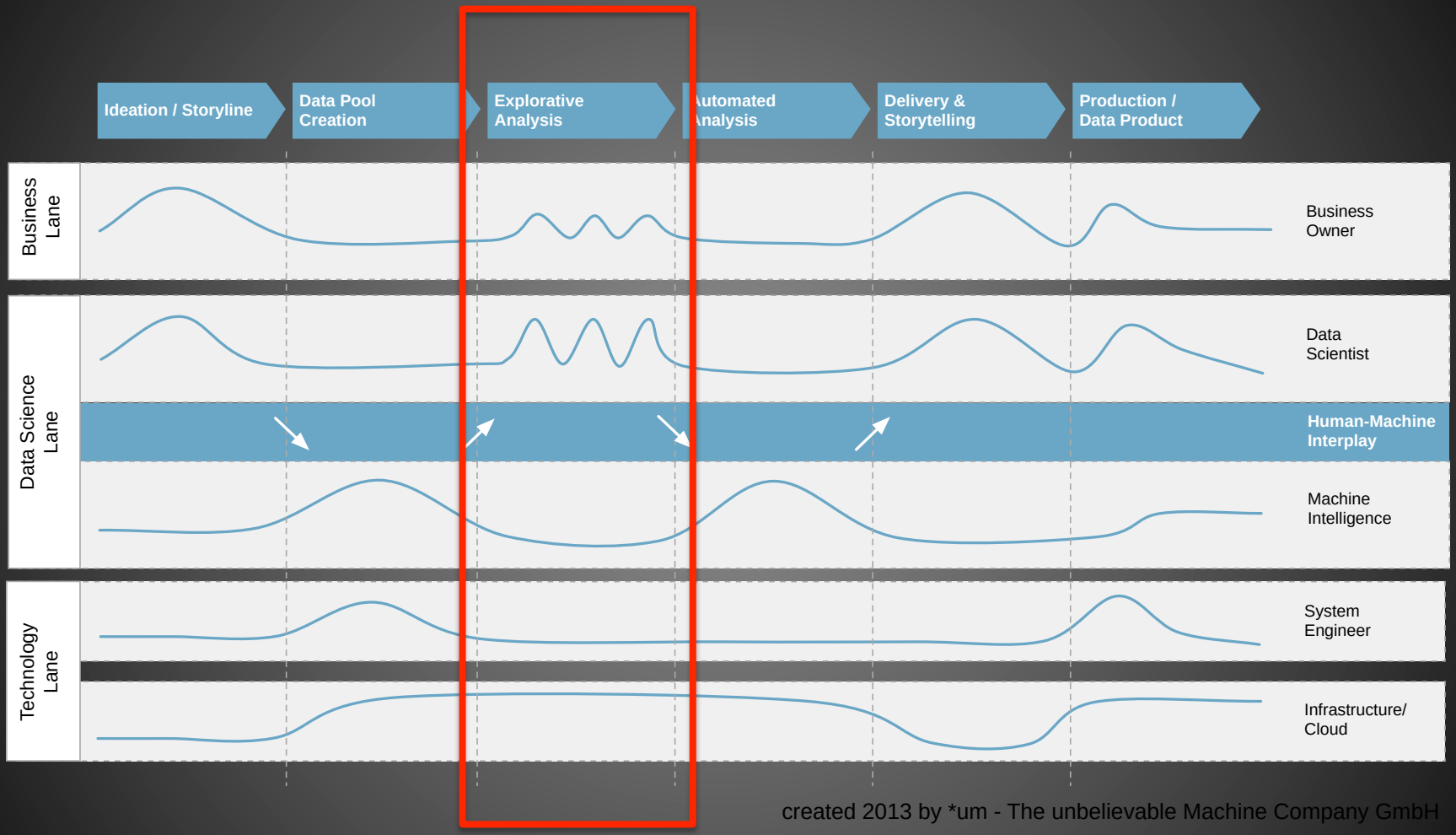
log files

~ 100 GB

~ 2 Monate

Sample: 200 K

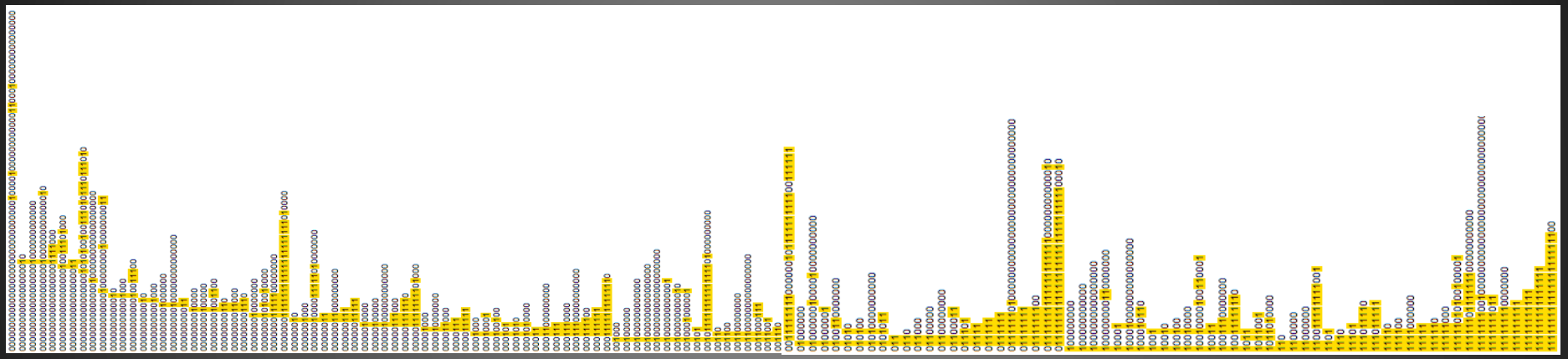
Data Science Process Model



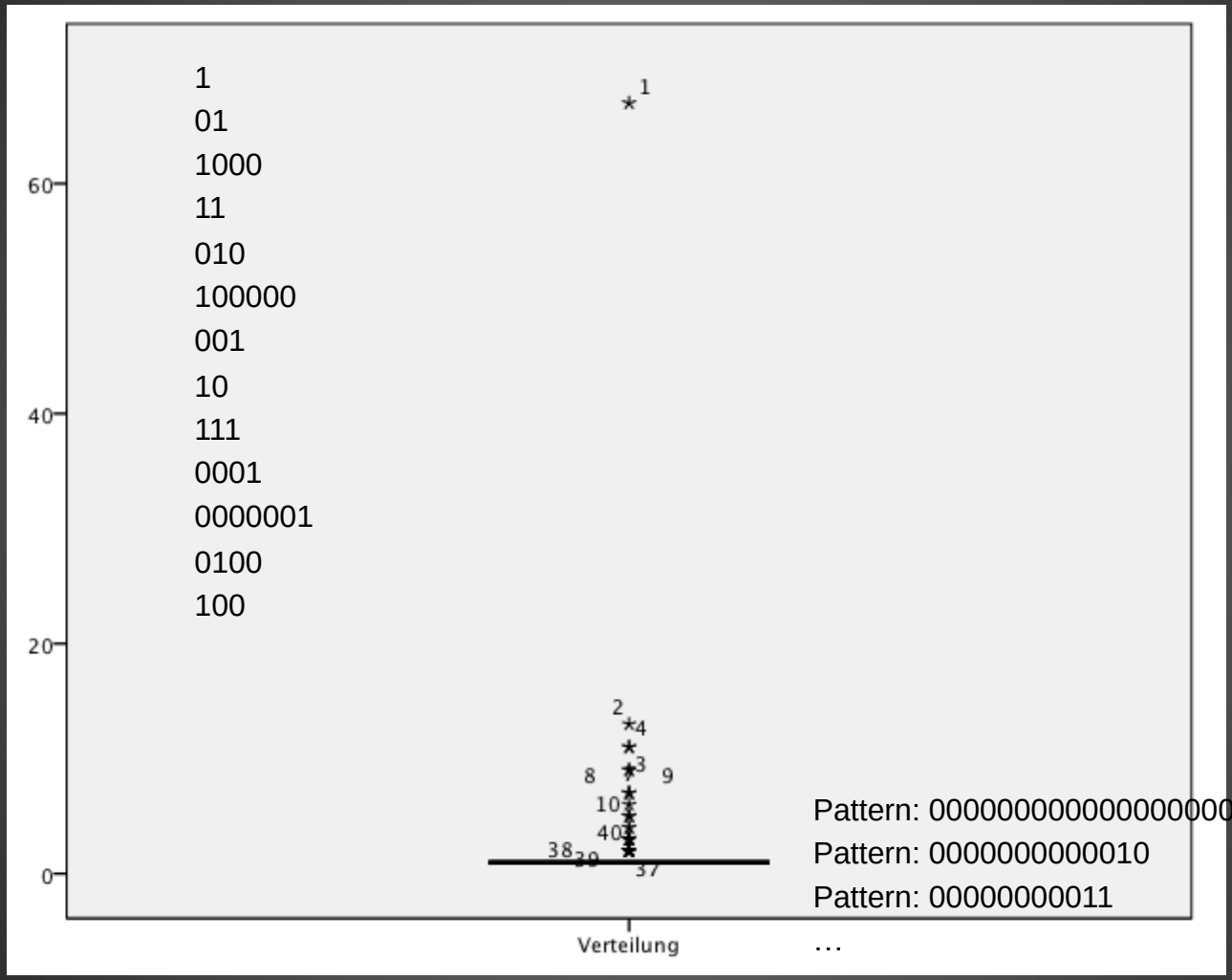
Conversion Rate Based On ..

80.18xxxx - - [19/Nov/2014:15:55:21 +0100] "GET /hd_de/checkout/onepage/success/ HTTP/1.1" 200 10341 "https://adomain/hd_de/checkout/onepage/" "Mozilla/5.0 (iPad; CPU OS 7_0_4 like Mac OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11B554a Safari/9537.53" 1635552 https

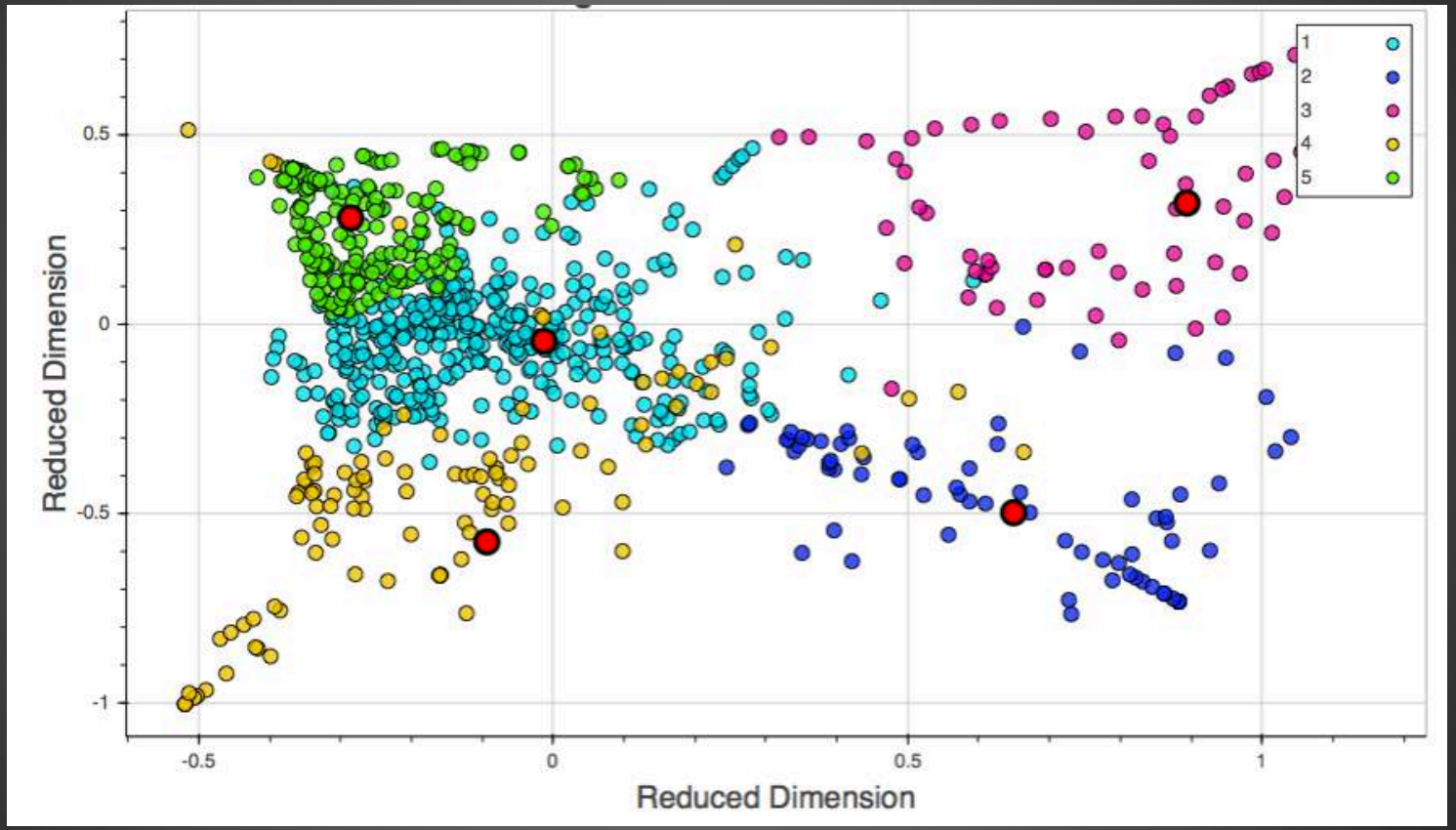
Exploration



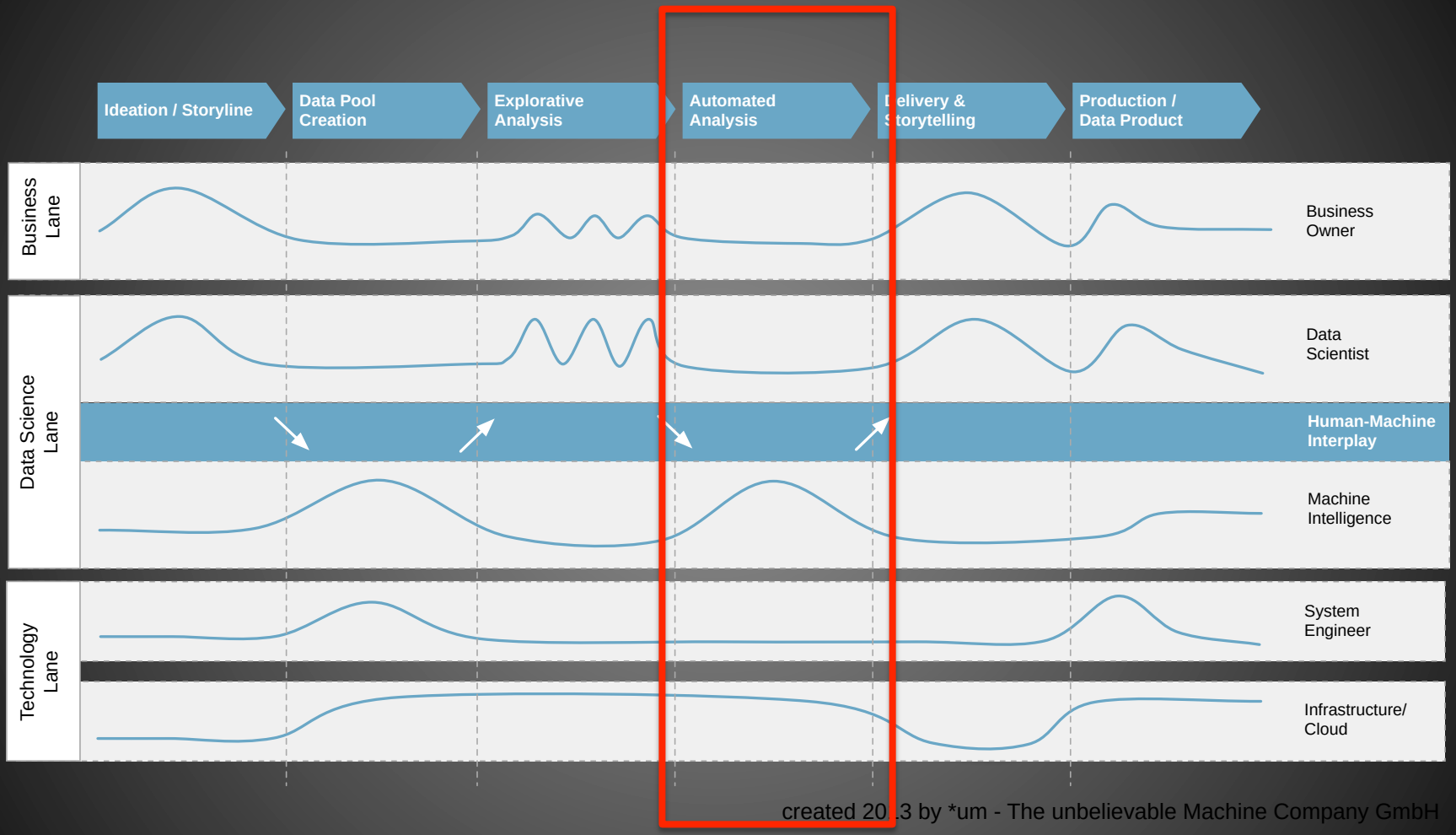
Exploration



Exploration

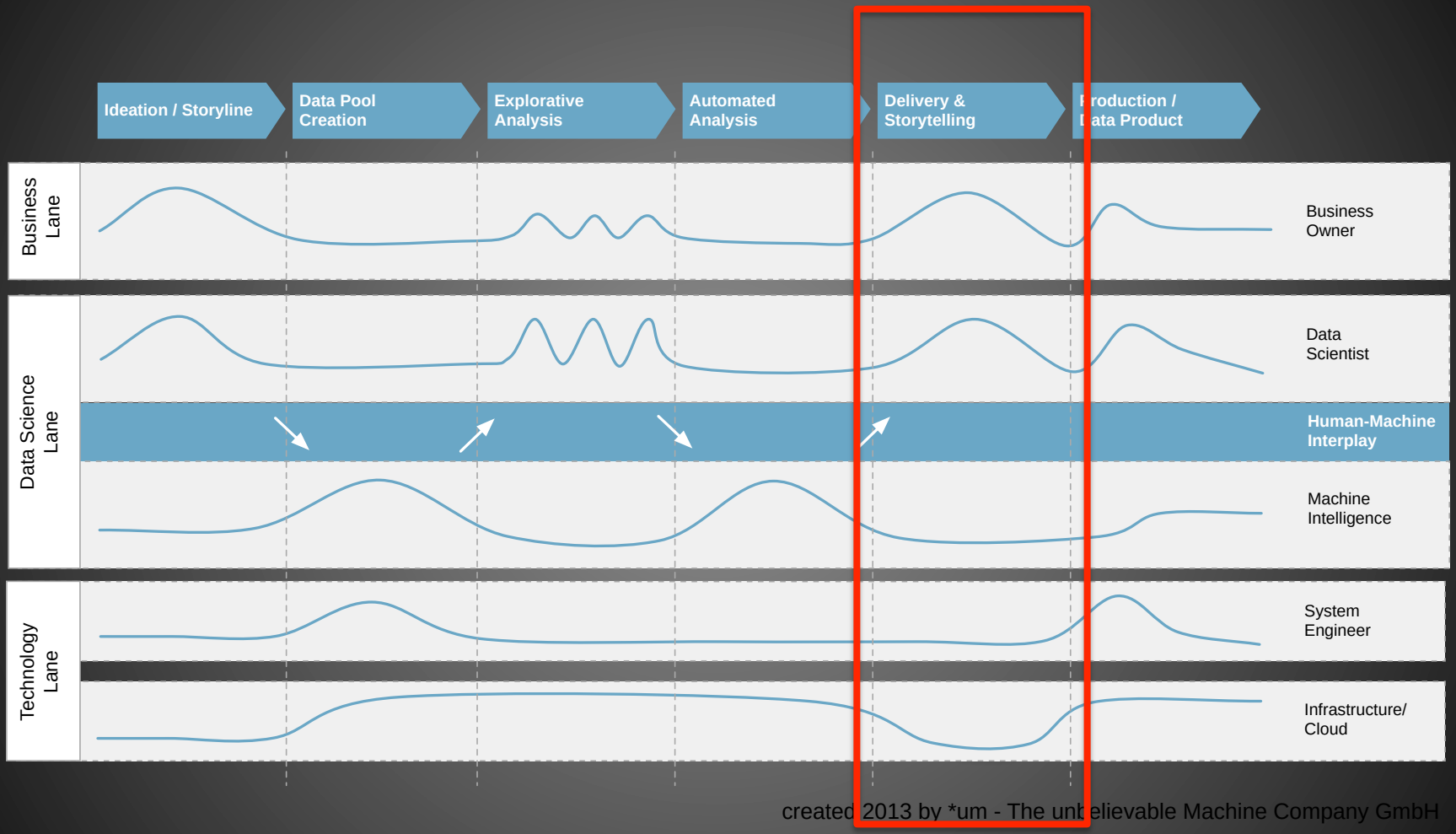


Data Science Process Model

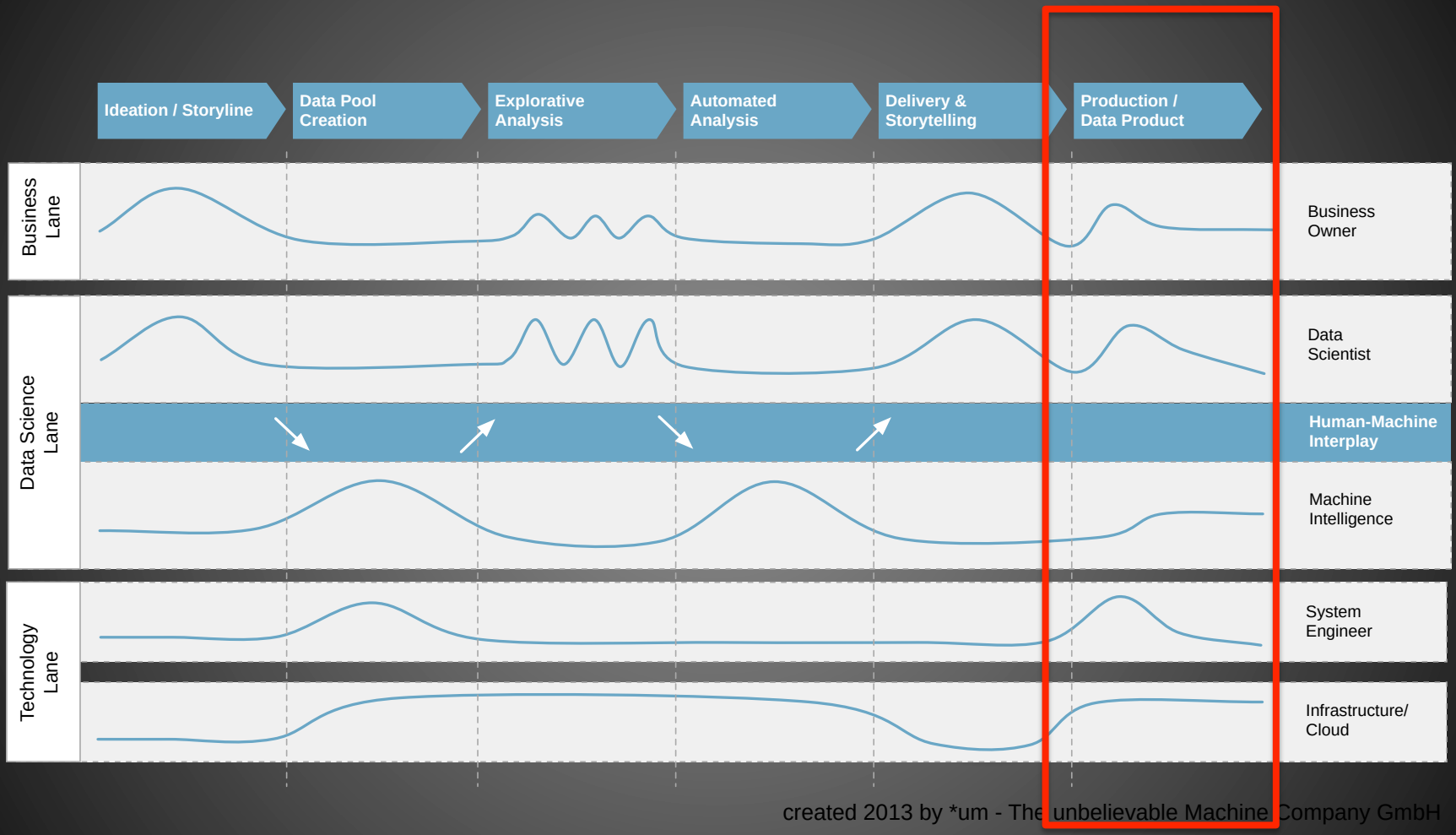


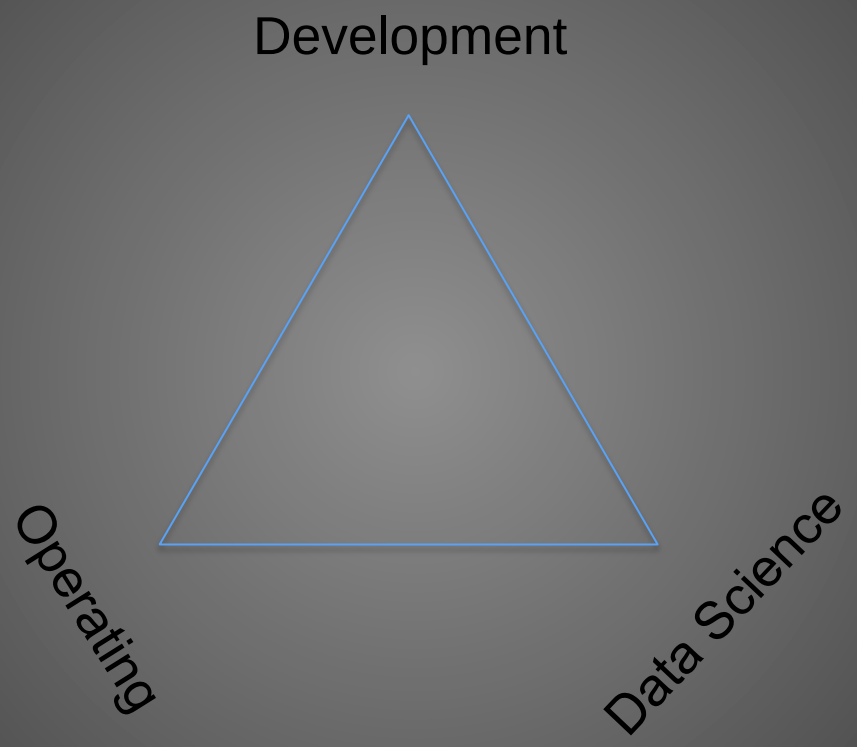
Automation

Data Science Process Model



Data Science Process Model




Development
Operating
Data Science

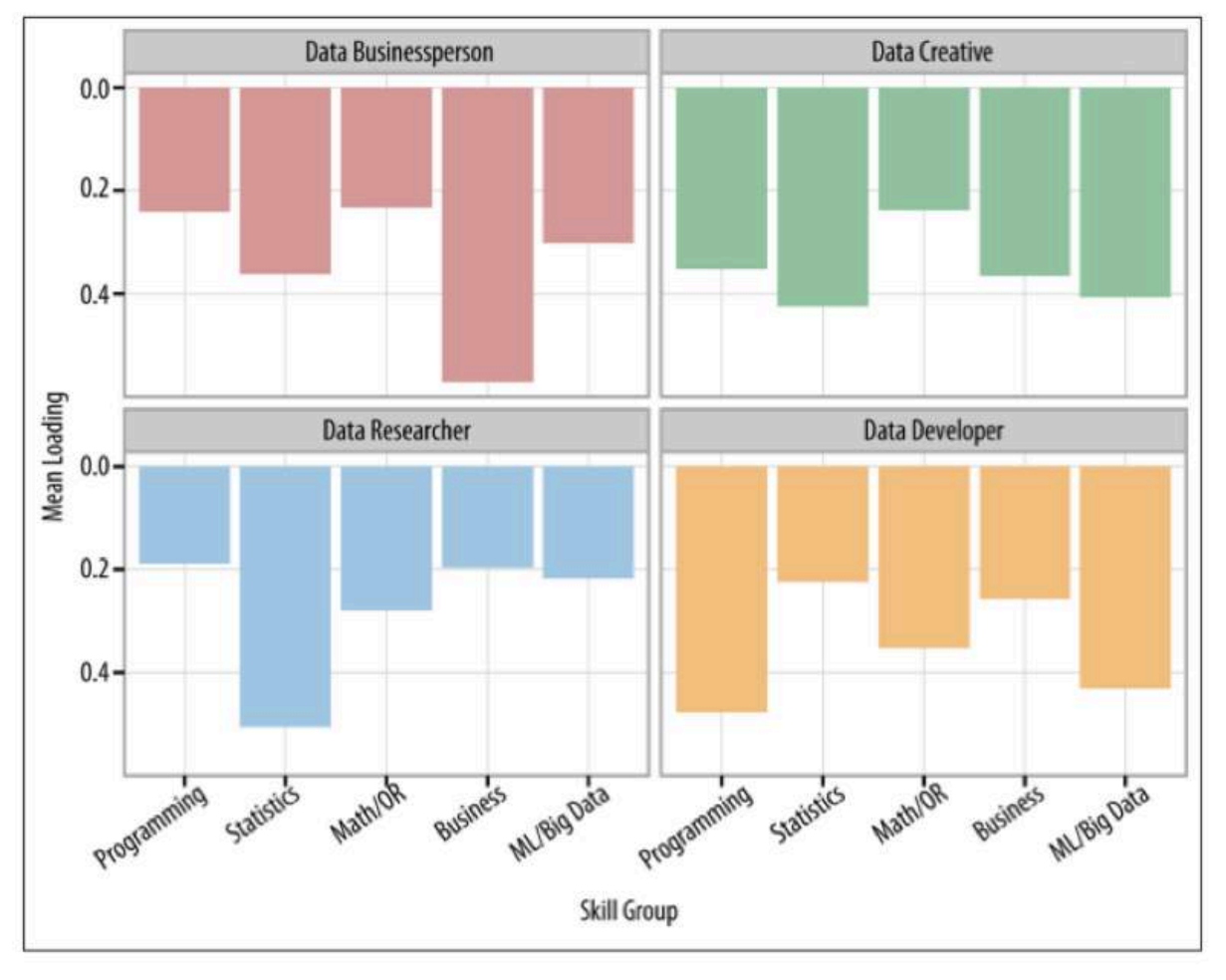
Thanks!

The unbelievable Machine Company GmbH
Grolmanstr. 40
10623 Berlin

Contact:
Florian Dohmann
florian.dohmann@unbelievable-machine.com
Tel. +49-30-889 26 56 – 36
Mobile +49-173 75 22 140

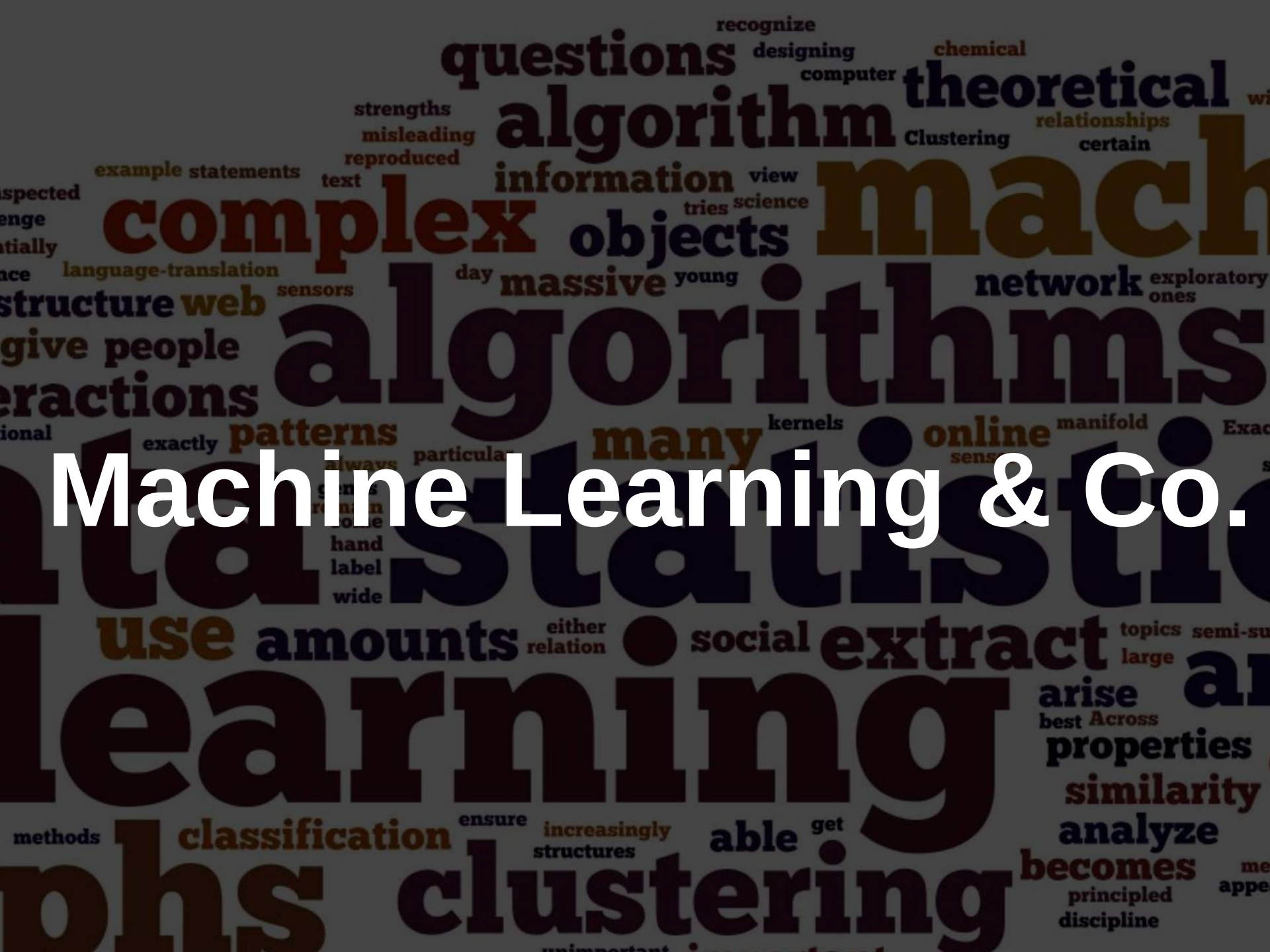
What we also need ..

A Great Team


Data Businessperson
Data Creative
Data Researcher
Data Developer
Mean Loading
0.0
0.2
0.4
Programming
Statistics
Math/OR
Business
ML/Big Data
Skill Group

+ Algorithms


Magic?



Machine Learning & Co.

+ Programming


Python & Co.

Simple spectral analysis

An illustration of the Discrete Fourier Transform

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \qquad k = 0, \ldots, N-1$$

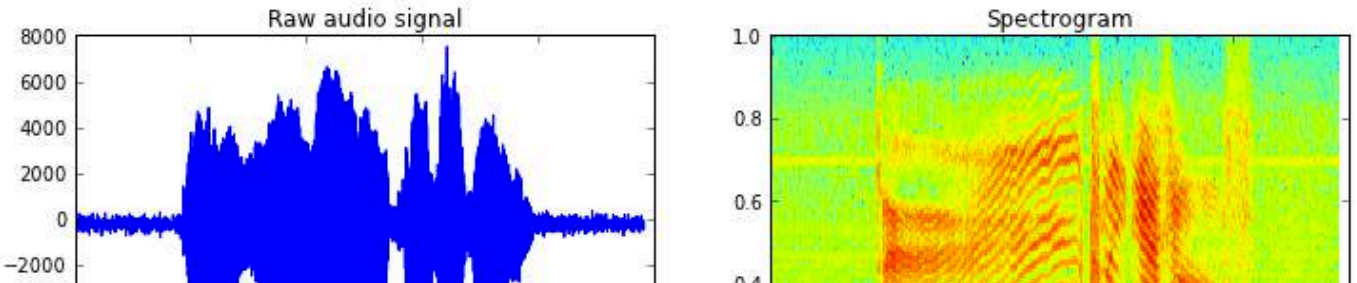
using windowing, to reveal the frequency content of a sound signal.

We begin by loading a datafile using SciPy's audio file support:

In [1]:
```
from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view its spectral structure using matplotlib's builtin `specgram` routine:

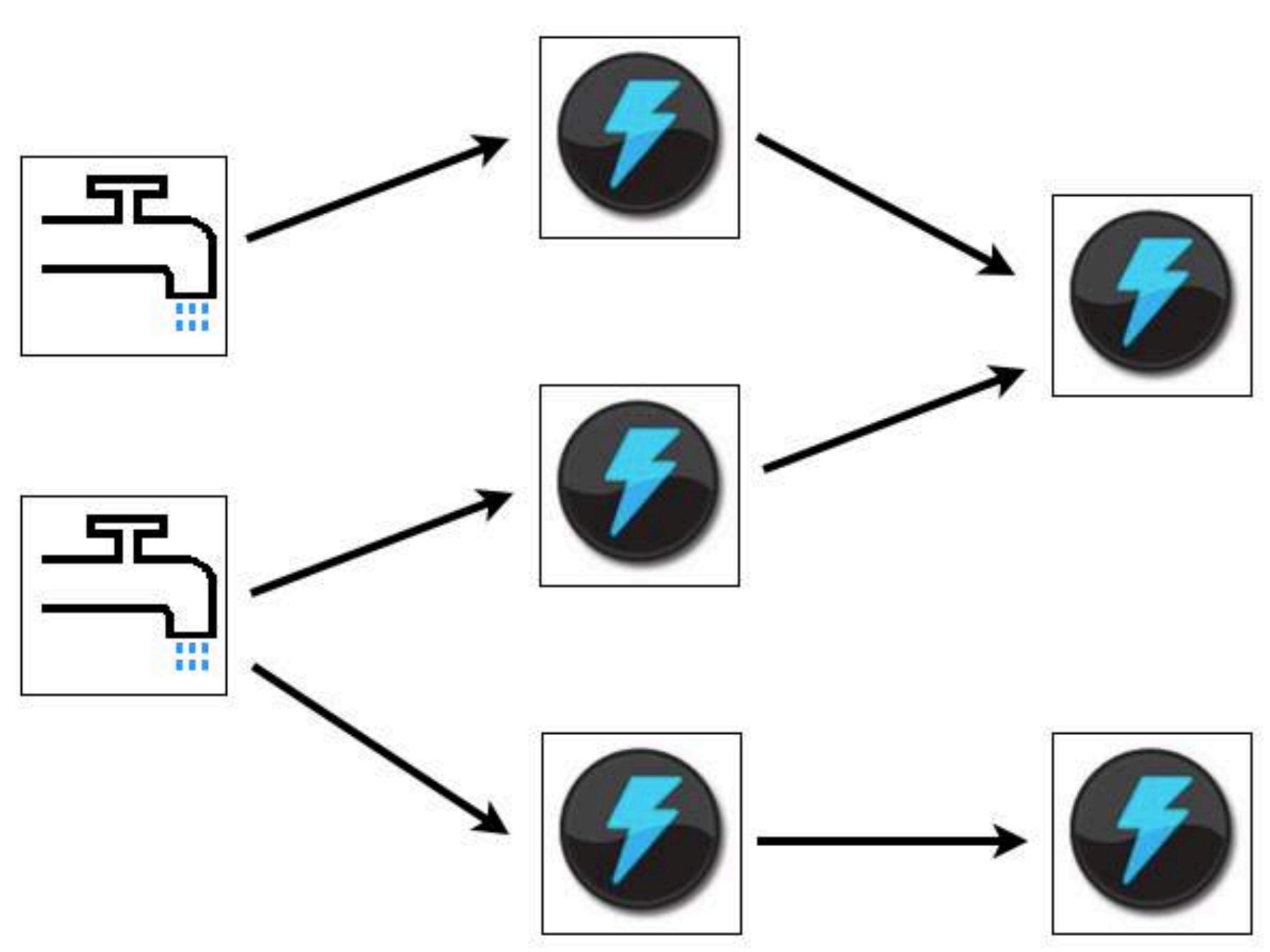
In [2]:
```
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.specgram(x); ax2.set_title('Spectrogram');
```




Hadoop & Co.

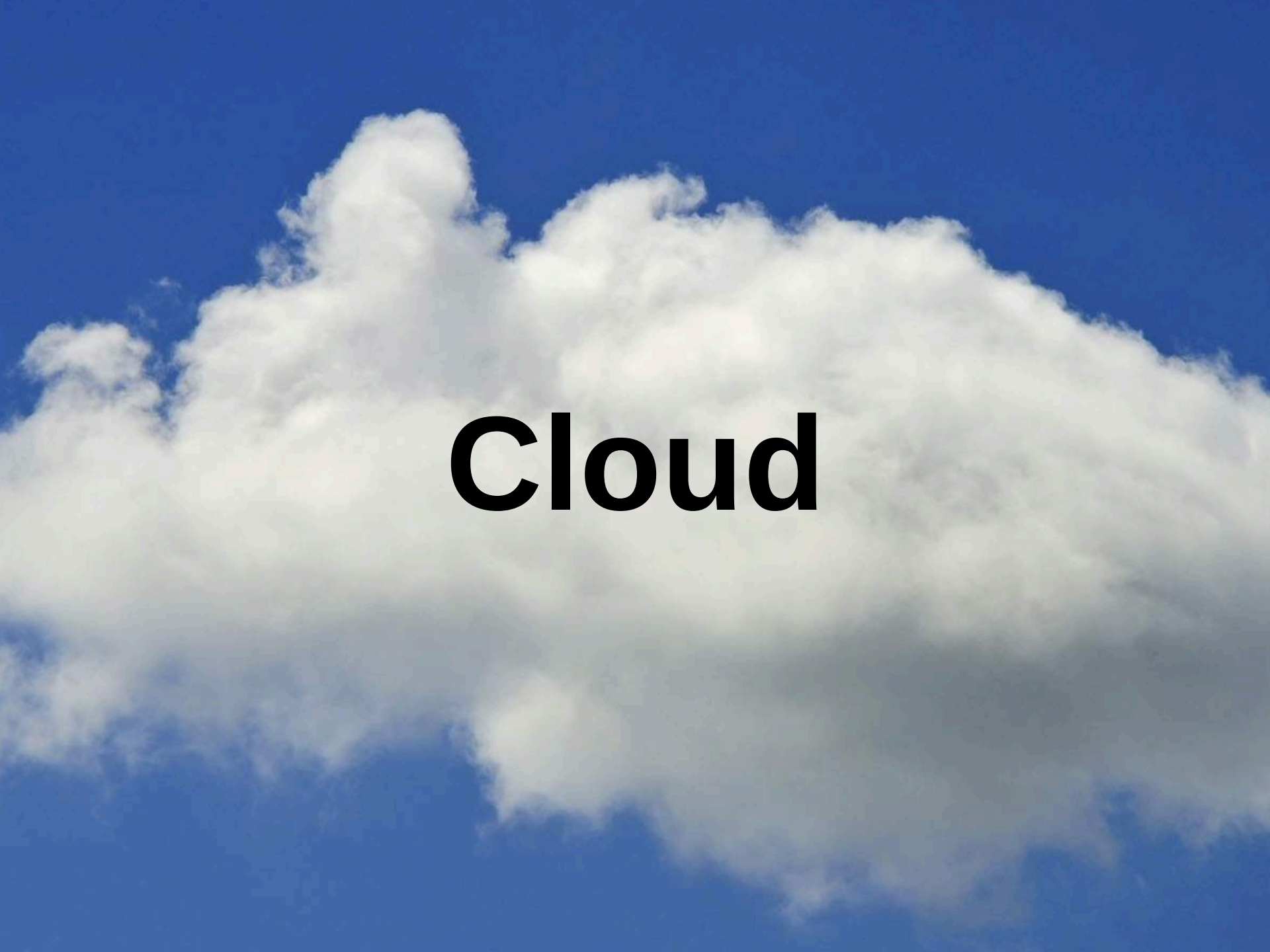
+ Stream processing

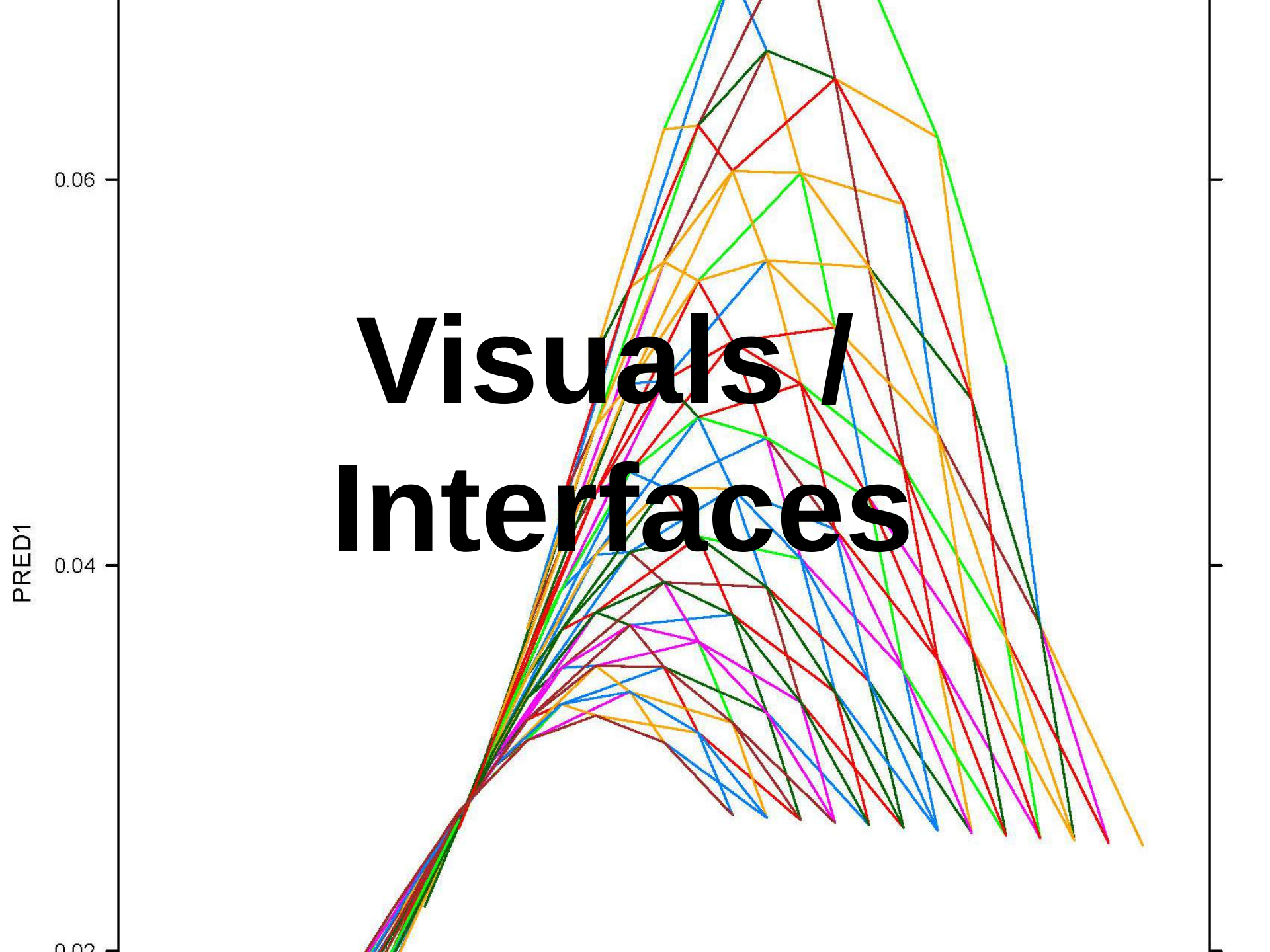

Storm & Co.



+ Infrastructure


Dedicated


Cloud


0.06
0.04
0.02
PRED1
Visuals /
Interfaces

Thanks 2!

The unbelievable Machine Company GmbH
Grolmanstr. 40
10623 Berlin

Contact:
Florian Dohmann
florian.dohmann@unbelievable-machine.com
Tel. +49-30-889 26 56 – 36
Mobile +49-173 75 22 140