



DATA SCIENCE

PROCESS MODEL

The background image shows a typical office environment. In the foreground, there are several desks with various items on them, including papers, a printer, and a small plant. A white chair is visible on the left side of the frame. In the background, there are large windows with multiple panes, some of which are covered by blinds. A radiator is visible under the windows. The overall lighting is bright, suggesting a well-lit office space.

Florian Dohmann

Data Scientist @ *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



Gartner 2013
Cool|Vendor

The logo is contained within a white rectangular box. It features the word "Gartner" in a bold, dark blue sans-serif font, followed by "2013" in a lighter blue sans-serif font. Below this, the words "Cool|Vendor" are displayed in a dark blue sans-serif font, with a vertical bar separating "Cool" and "Vendor". The text is flanked by two horizontal bars: a solid dark blue bar on the left and a light blue bar with a diagonal cut on the right.

2015



~100

Employees in Berlin & Vienna

Data Science

Custom Big Data Analytics, Machine Learning & Co.

Workshops & Trainings

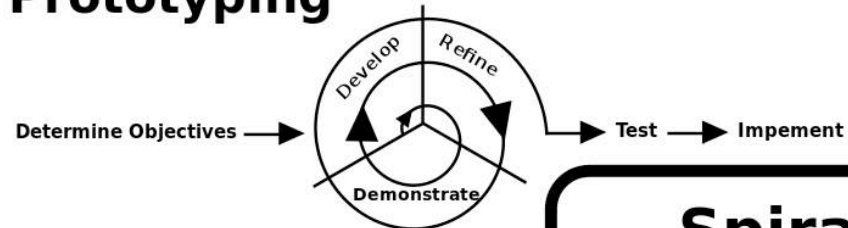
Support in creating own Data Science Teams

= Full-
Service

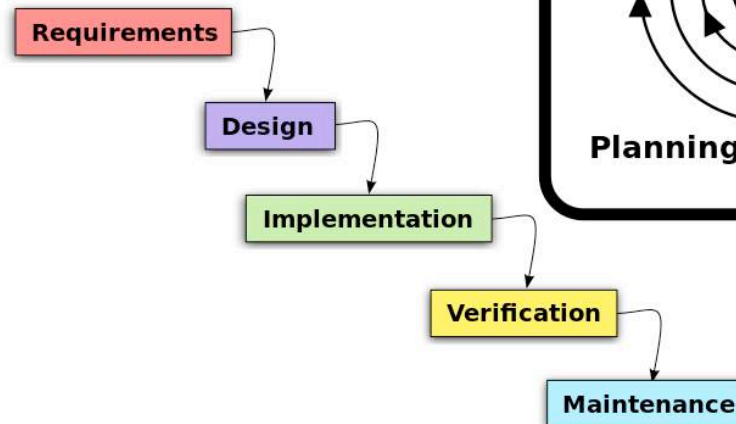
**„We need a perfect
interplay of human &
machine intelligence“**

Process?

Prototyping



Waterfall



Spiral

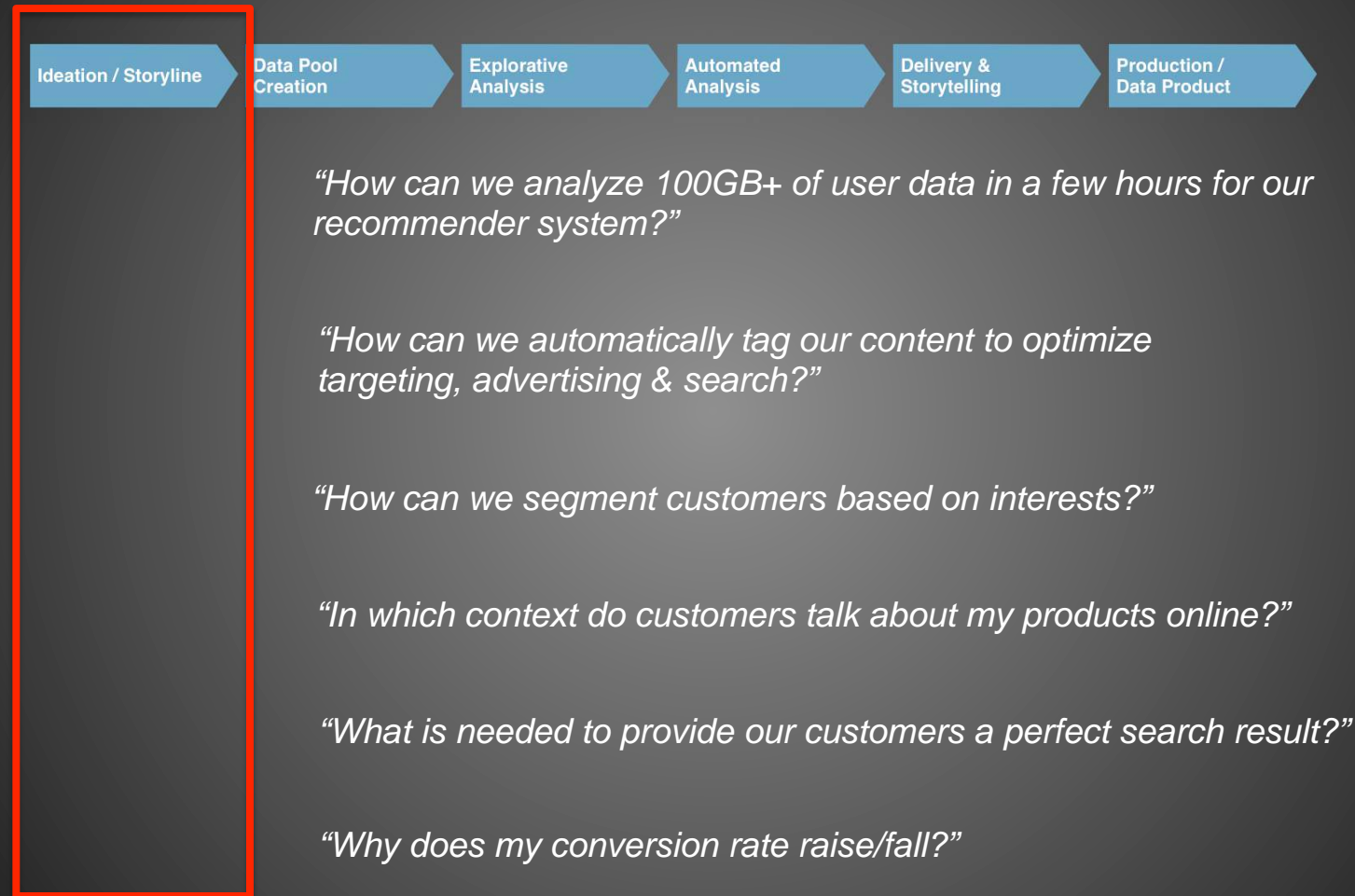


http://en.wikipedia.org/wiki/Software_development_process#mediaviewer/File:Three_software_development_patterns_mashed_together.svg

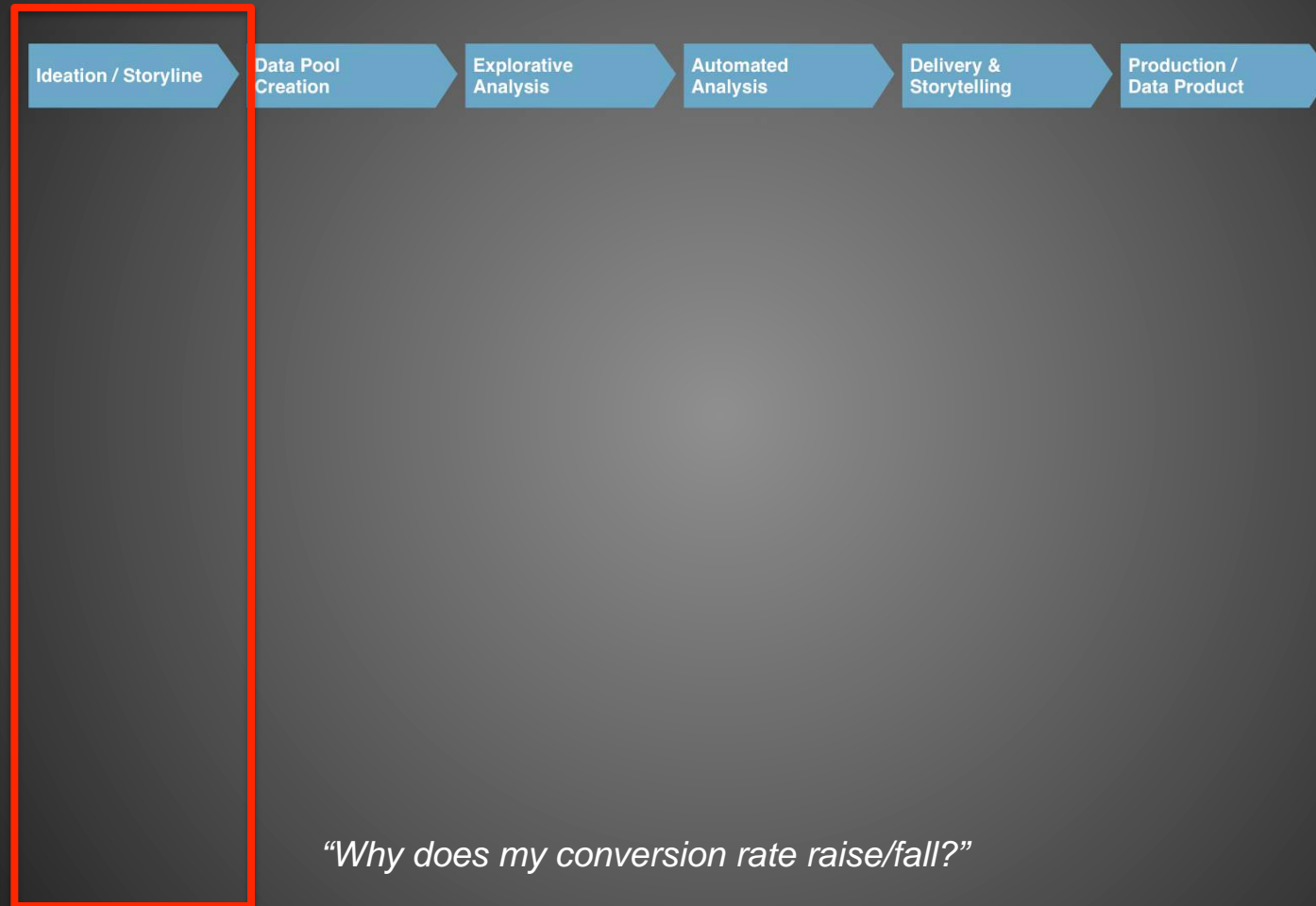
Data Science Process Model: Stages



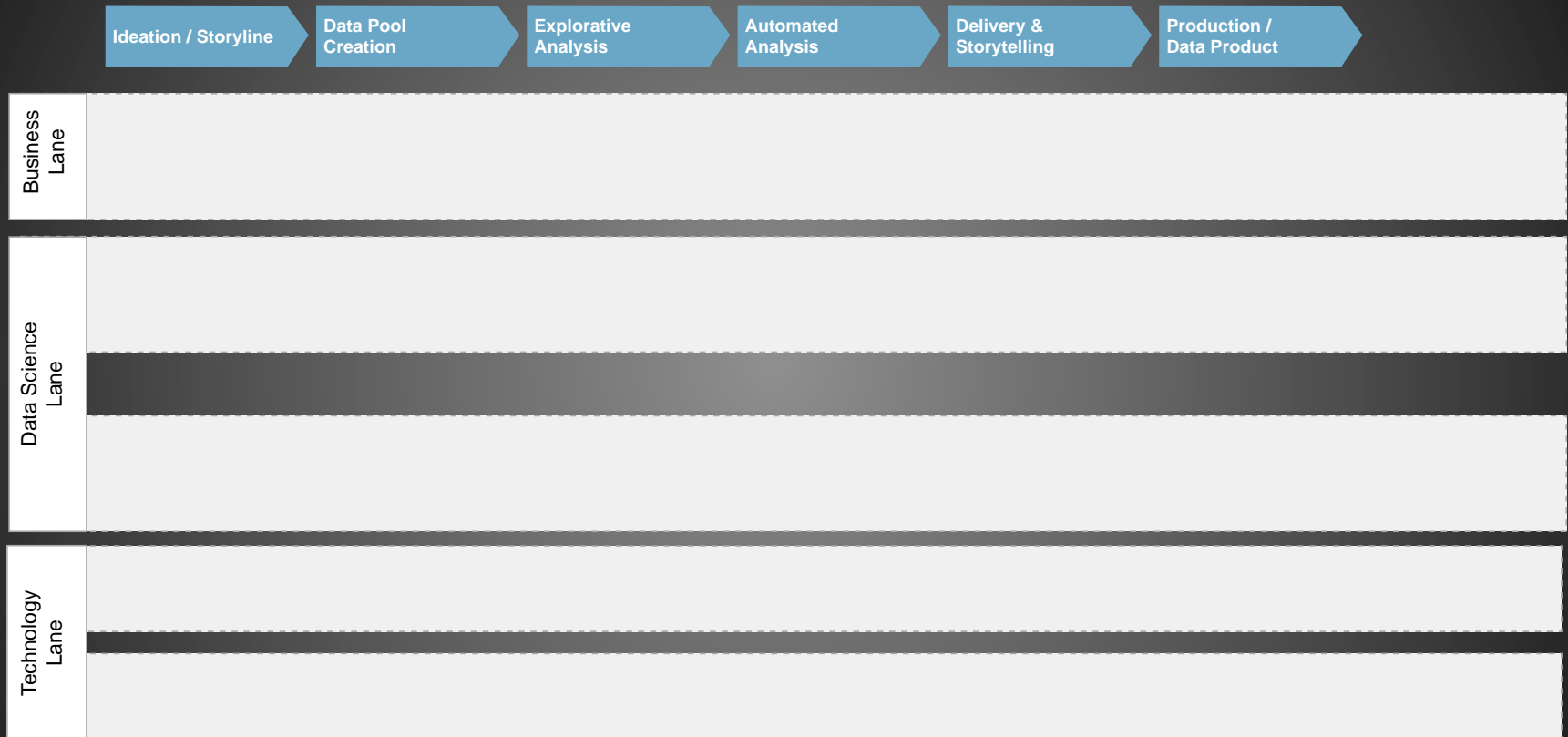
Data Science Process Model



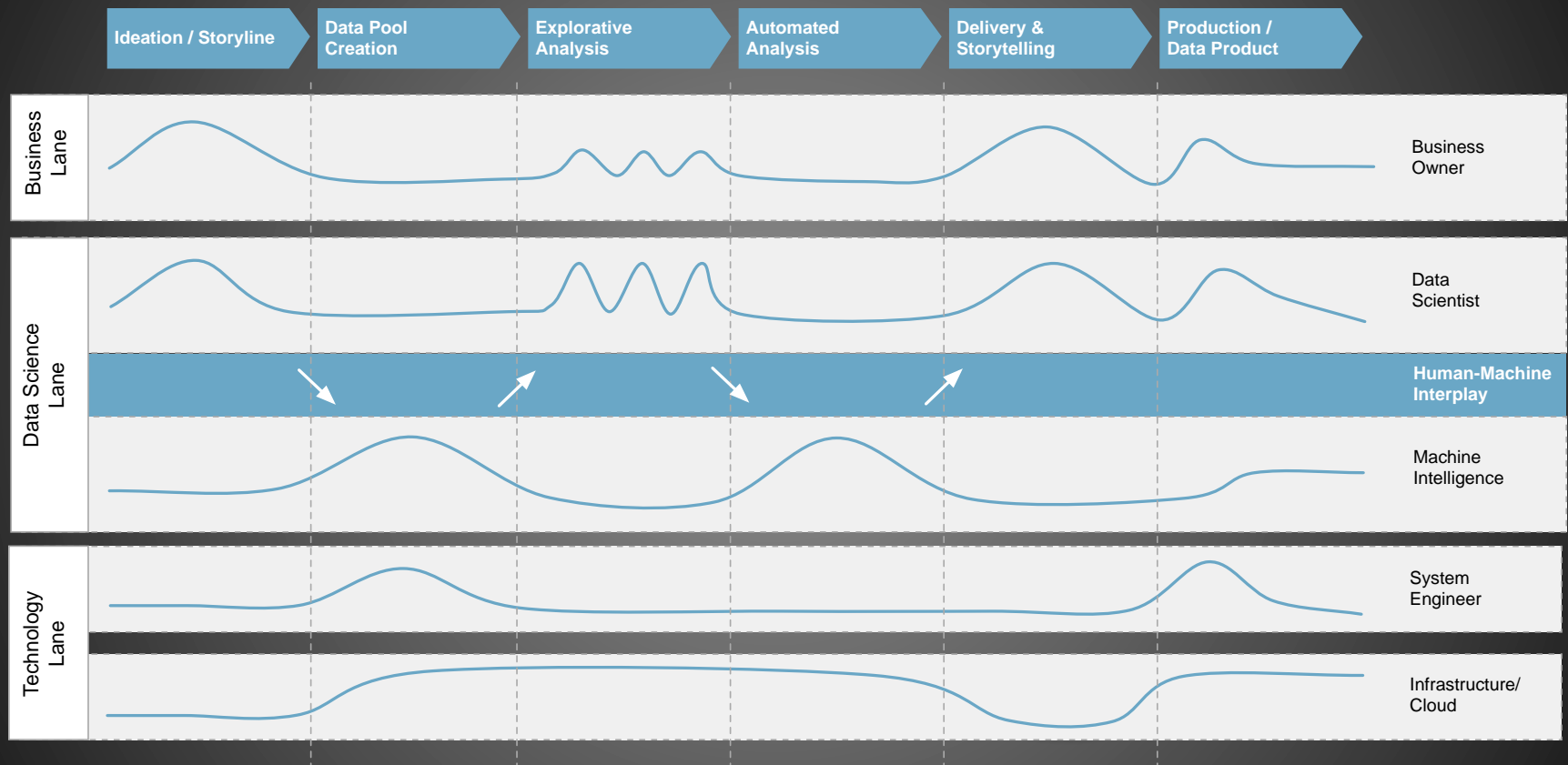
Data Science Process Model



Data Science Process Model: Stages & Lanes

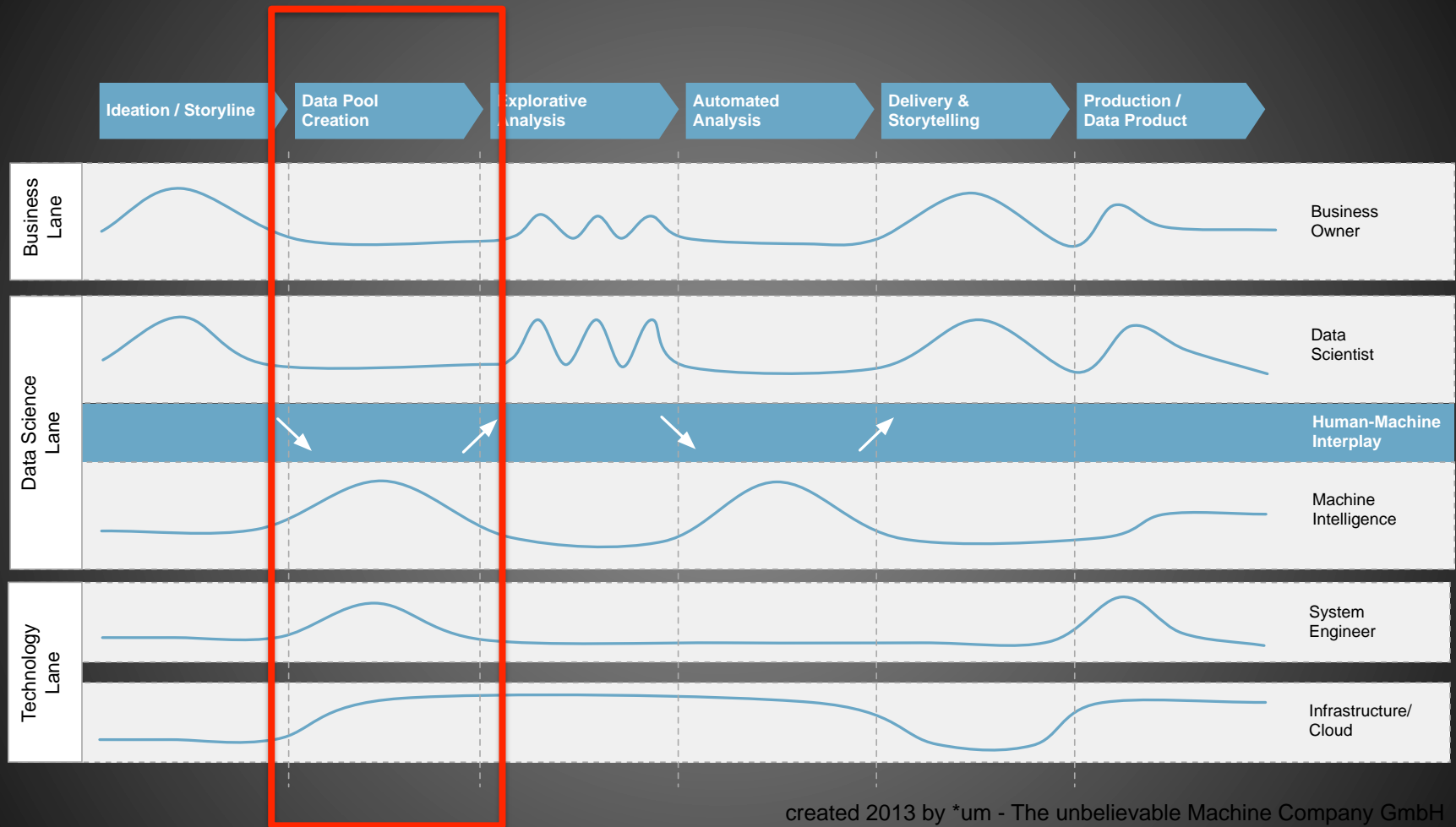


Data Science Process Model



created 2013 by *um - The unbelievable Machine Company GmbH

Data Science Process Model

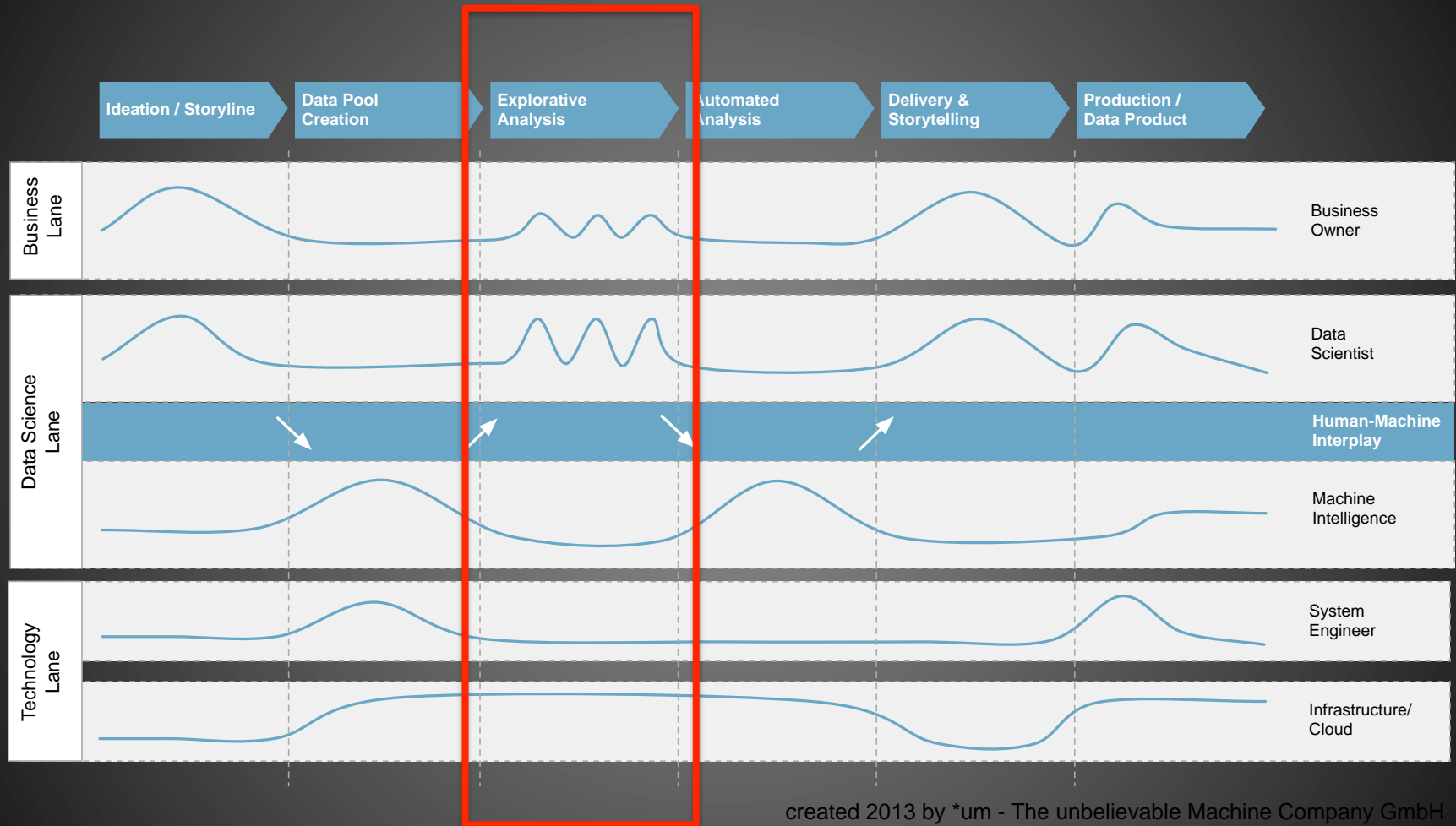




Sample: 200 K

log files
~ 100 GB
~ 2 Monate

Data Science Process Model

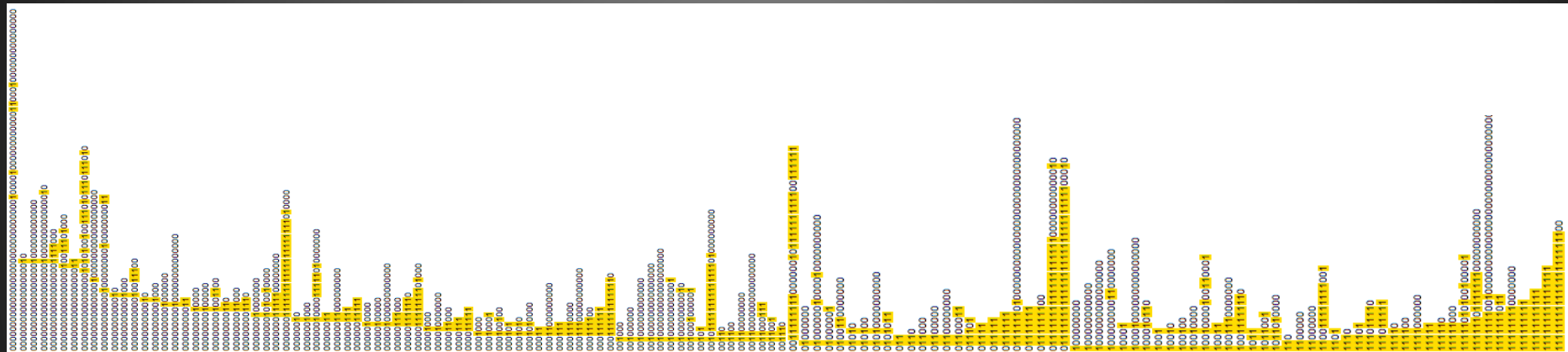




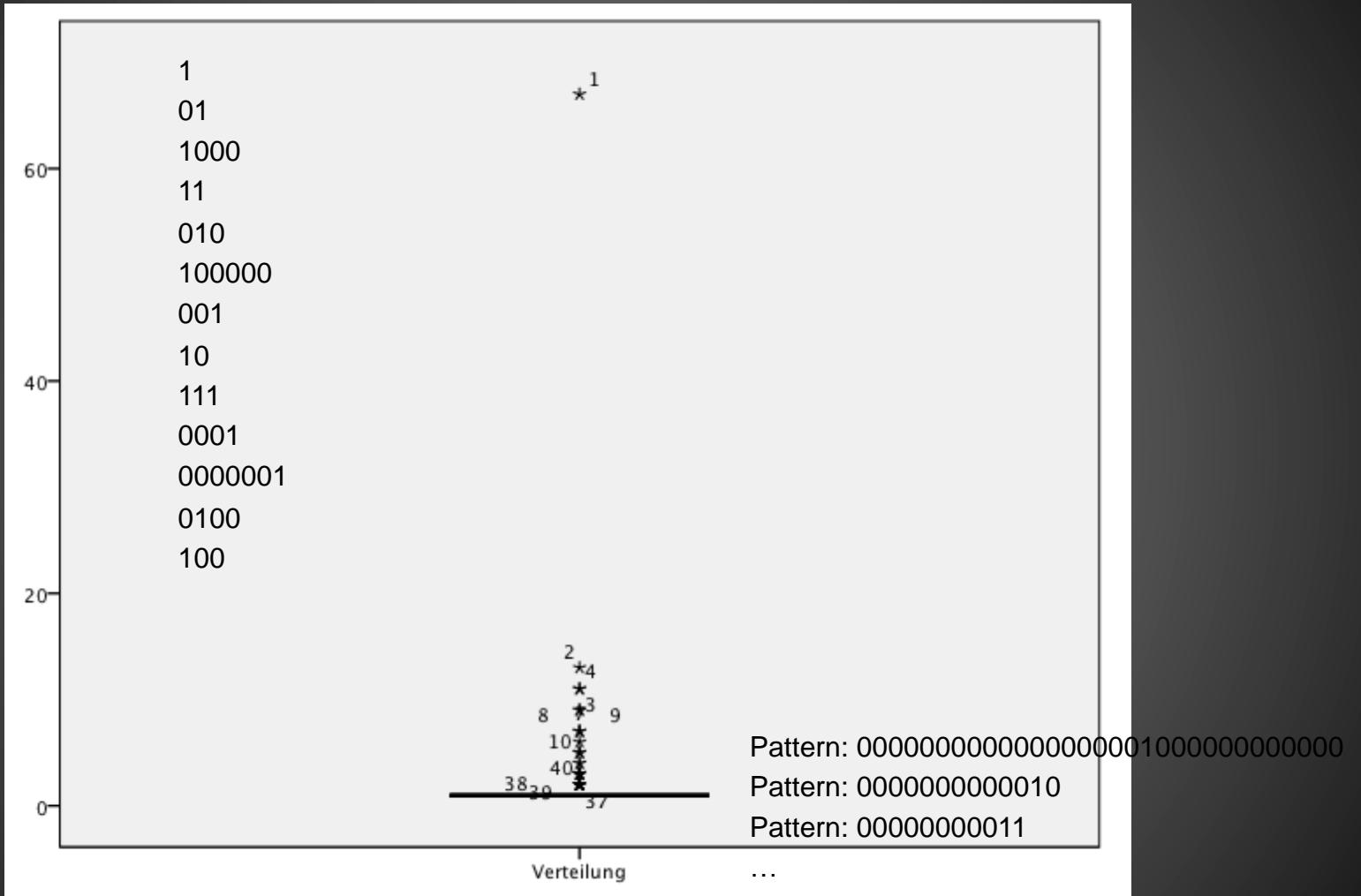
Conversion Rate Based On ..

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

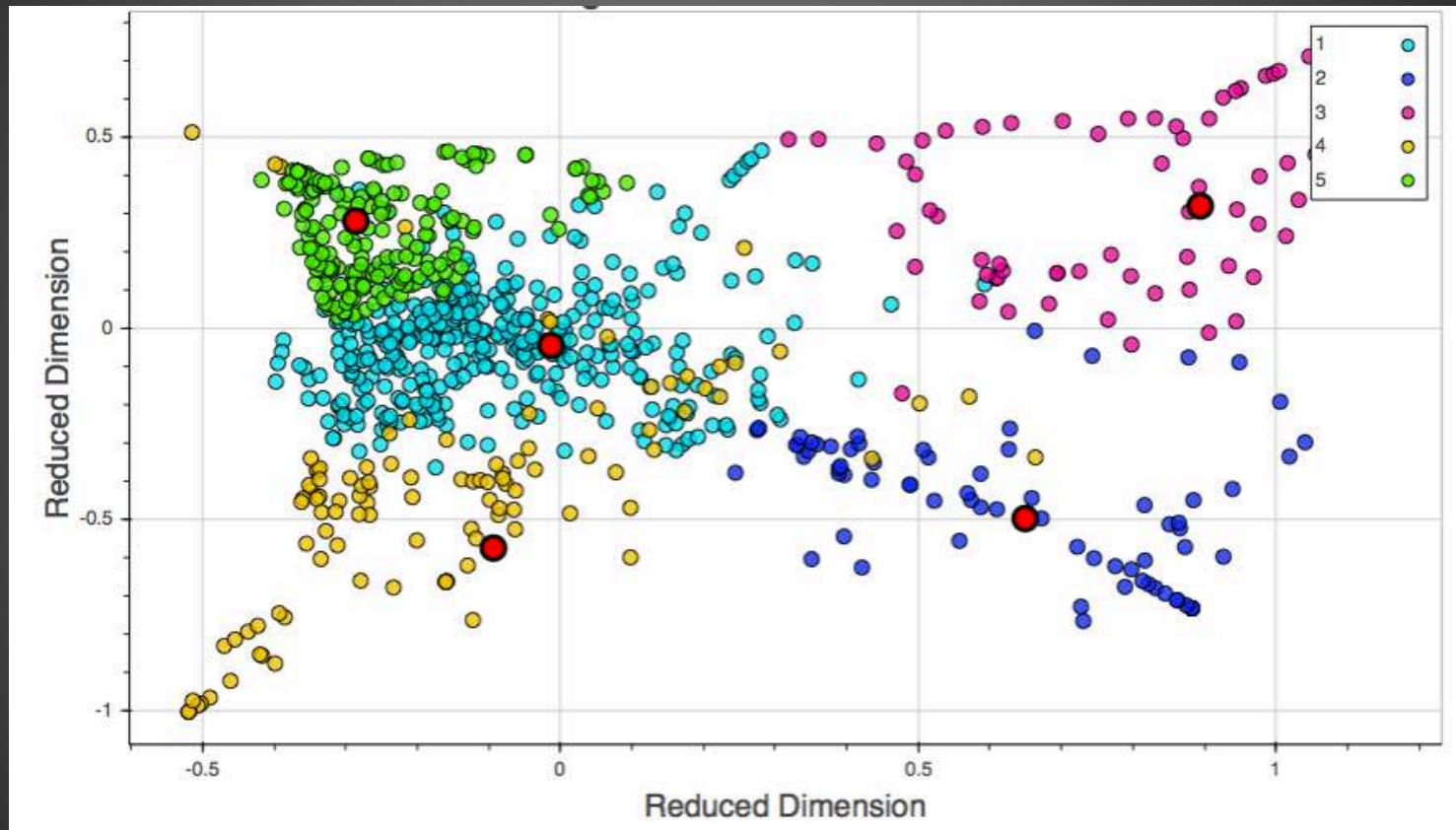
Exploration



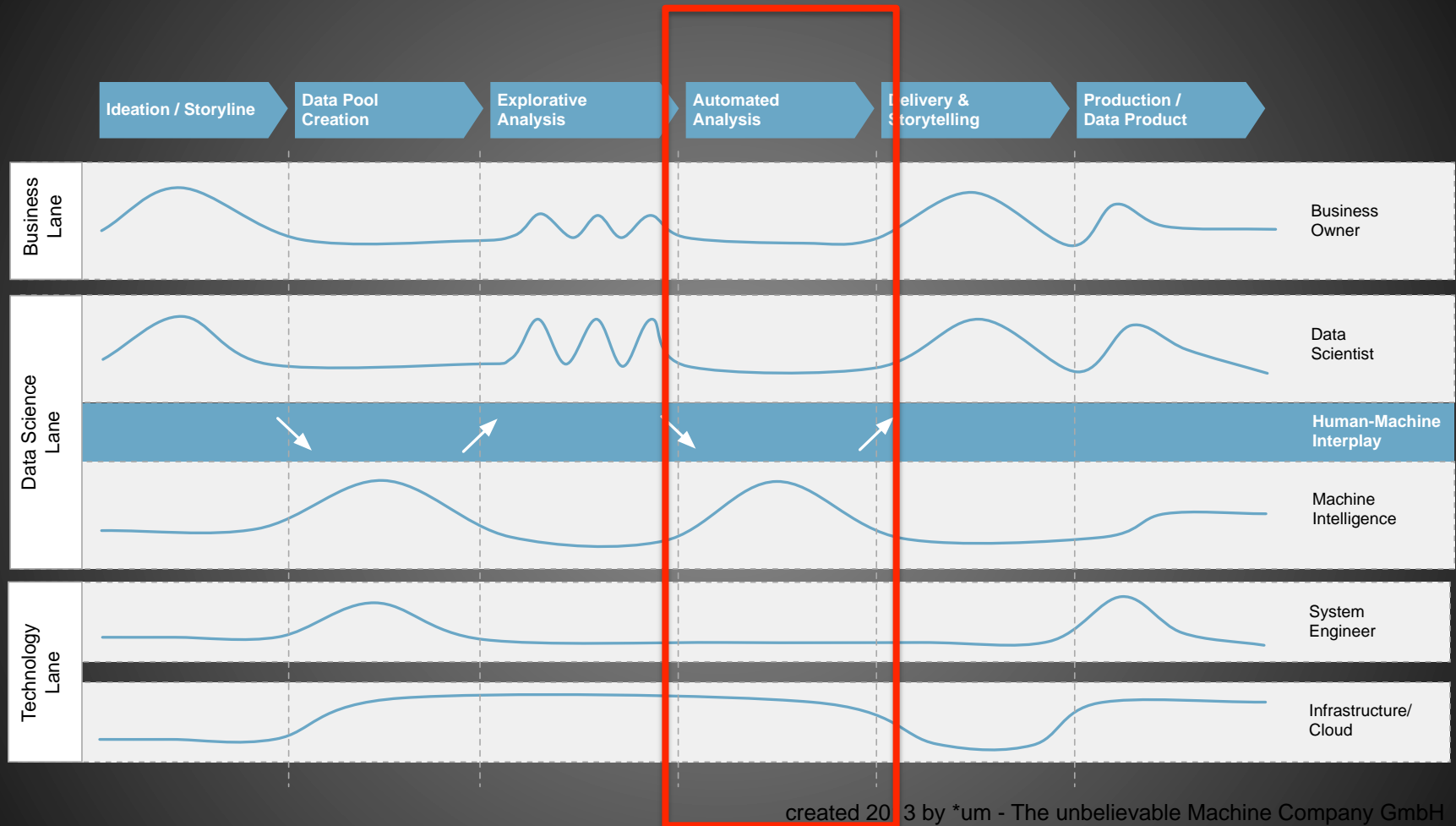
Exploration



Exploration



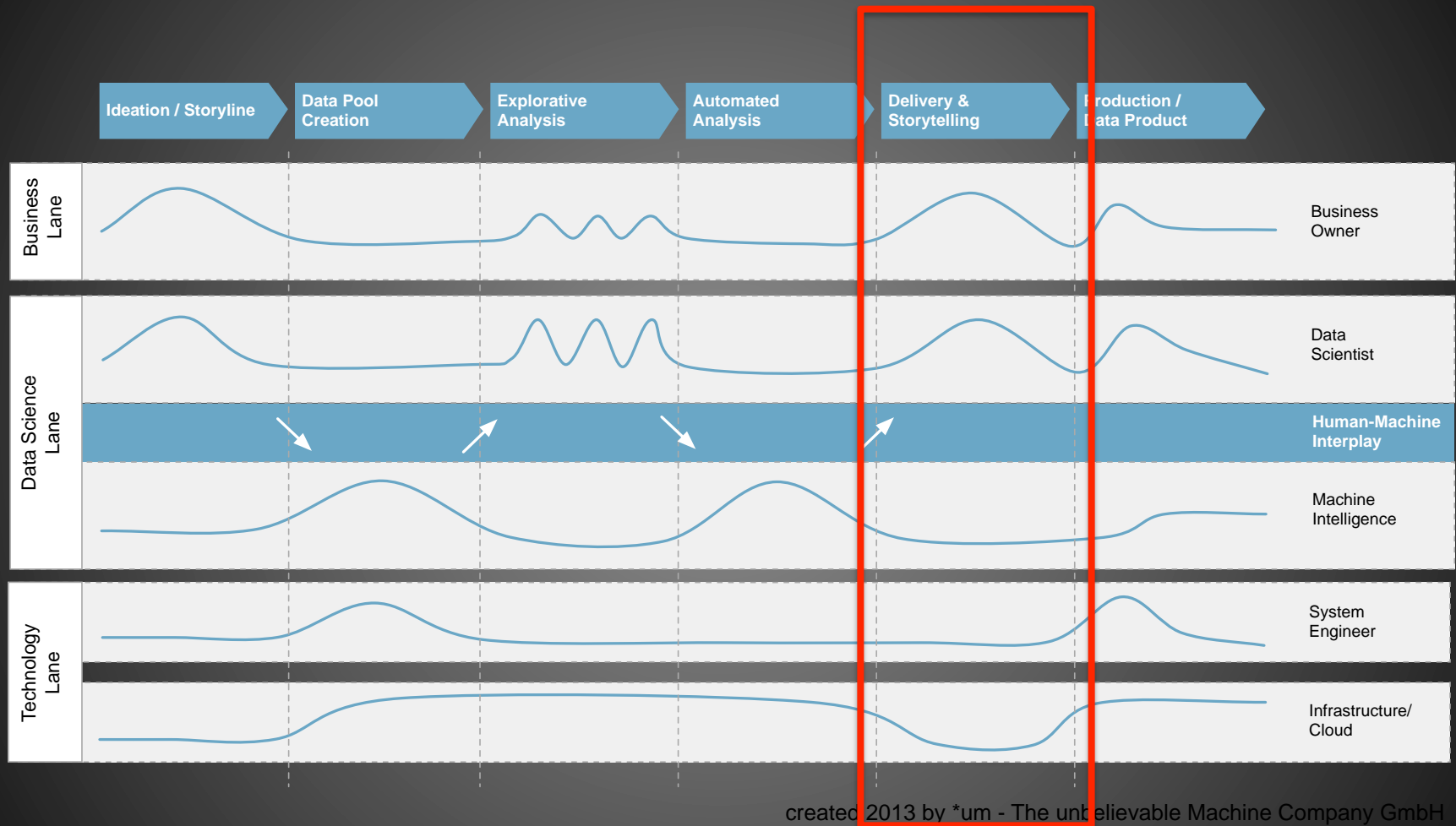
Data Science Process Model



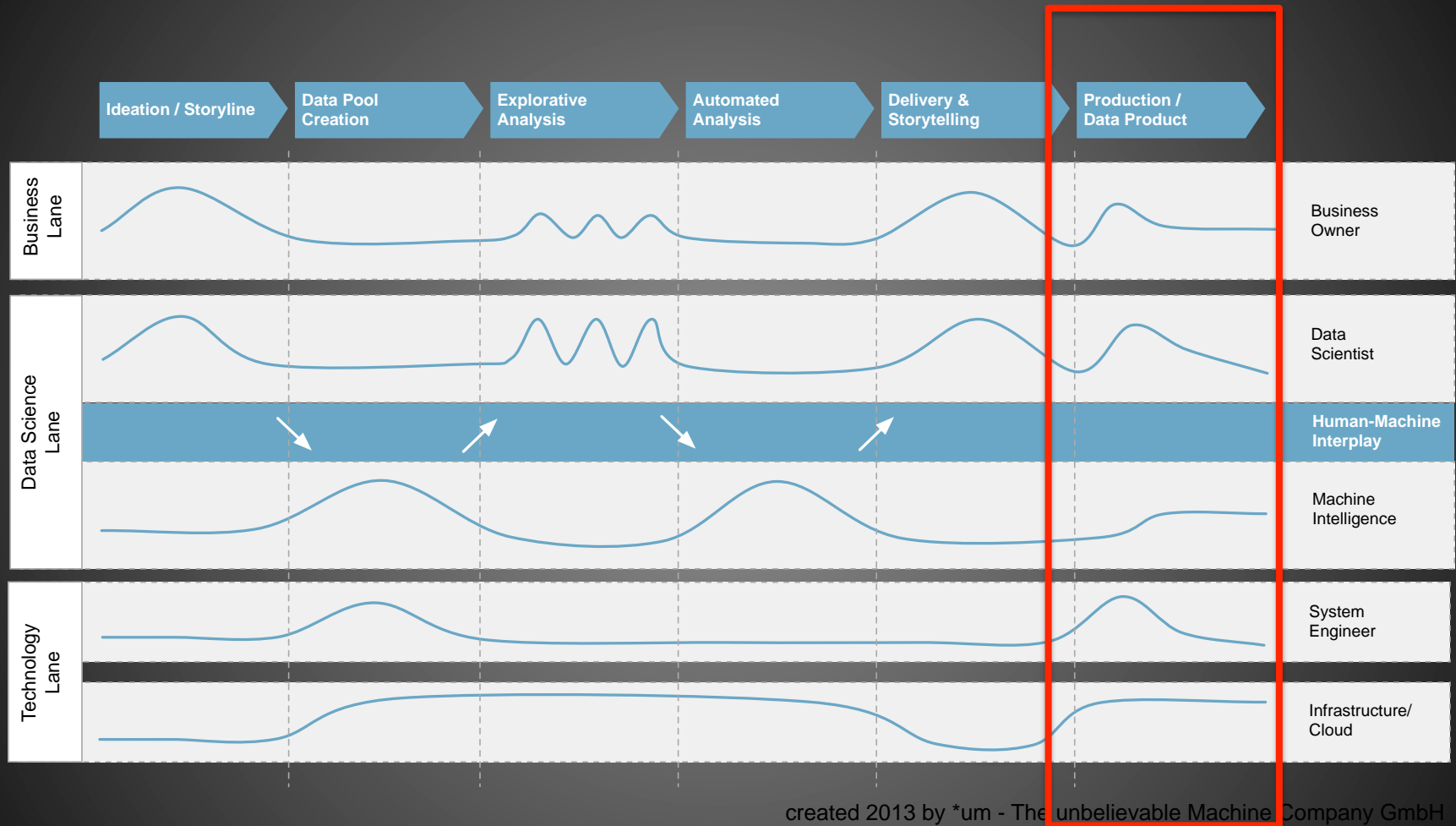
Automation

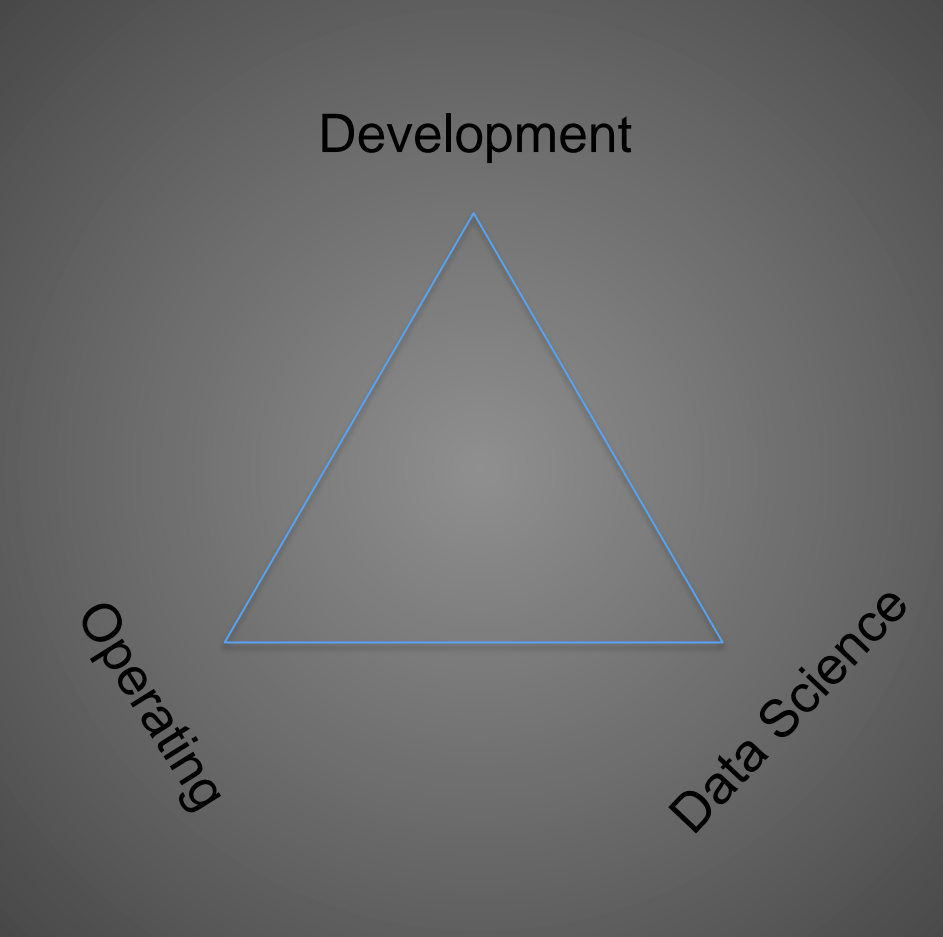


Data Science Process Model



Data Science Process Model





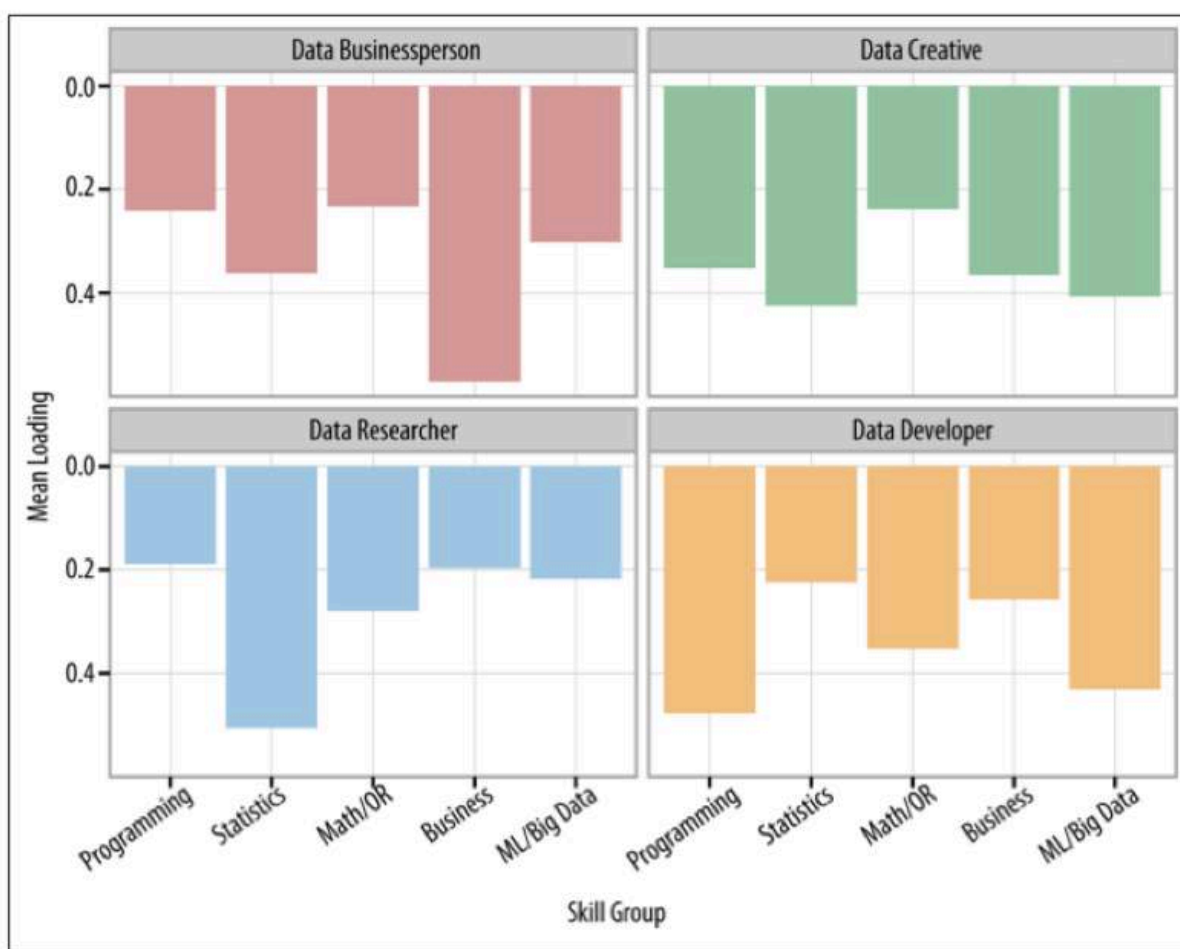
Thanks!

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



+ Algorithms



Magic?

A word cloud background with various terms related to machine learning and statistics. The words are in different sizes and colors (shades of brown and gold). The central text 'Machine Learning & Co.' is in a large, white, sans-serif font. Other visible words include 'algorithm', 'complex', 'mach', 'algorithms', 'objects', 'information', 'view', 'science', 'tries', 'massive', 'young', 'network', 'exploratory', 'ones', 'patterns', 'kernels', 'online', 'manifold', 'Exact', 'use', 'amounts', 'relation', 'social', 'extract', 'topics', 'semi-su', 'large', 'arise', 'best', 'Across', 'properties', 'similarity', 'analyze', 'becomes', 'principled', 'discipline', 'methods', 'classification', 'ensure', 'increasingly', 'structures', 'able', 'get', 'phs', 'clustering', 'unimportant', 'important', 'me', 'appe', 'relationships', 'certain', 'Clustering', 'chemical', 'designing', 'computer', 'recognize', 'strengths', 'misleading', 'reproduced', 'text', 'statements', 'example', 'language-translation', 'sensors', 'day', 'web', 'structure', 'give', 'people', 'eractions', 'ional', 'exactly', 'particular', 'ways', 'in', 'one', 'hand', 'label', 'wide', 'either', 'relation', '●', 'social', 'extract', 'topics', 'semi-su', 'large', 'arise', 'best', 'Across', 'properties', 'similarity', 'analyze', 'becomes', 'principled', 'discipline', 'methods', 'classification', 'ensure', 'increasingly', 'structures', 'able', 'get', 'phs', 'clustering', 'unimportant', 'important', 'me', 'appe', 'relationships', 'certain', 'Clustering', 'chemical', 'designing', 'computer', 'recognize', 'strengths', 'misleading', 'reproduced', 'text', 'statements', 'example', 'language-translation', 'sensors', 'day', 'web', 'structure', 'give', 'people', 'eractions', 'ional', 'exactly', 'particular', 'ways', 'in', 'one', 'hand', 'label', 'wide', 'either', 'relation', '●', 'social', 'extract', 'topics', 'semi-su', 'large', 'arise', 'best', 'Across', 'properties', 'similarity', 'analyze', 'becomes', 'principled', 'discipline', 'methods', 'classification', 'ensure', 'increasingly', 'structures', 'able', 'get', 'phs', 'clustering', 'unimportant', 'important', 'me', 'appe'.

+ Programming



Python & Co.

IP[y]: Notebook

spectrogram

Last saved: Mar 07 11:14 PM

File Edit View Insert Cell Kernel Help

 Markdown ▾

Simple spectral analysis

An illustration of the [Discrete Fourier Transform](#)

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

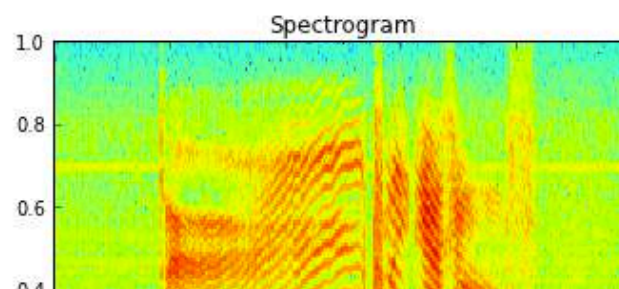
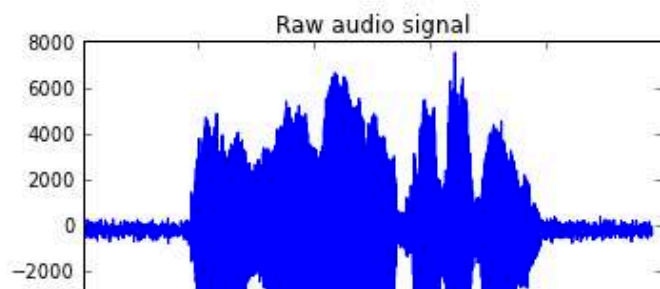
using windowing, to reveal the frequency content of a sound signal.

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 spectrogram routine:

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



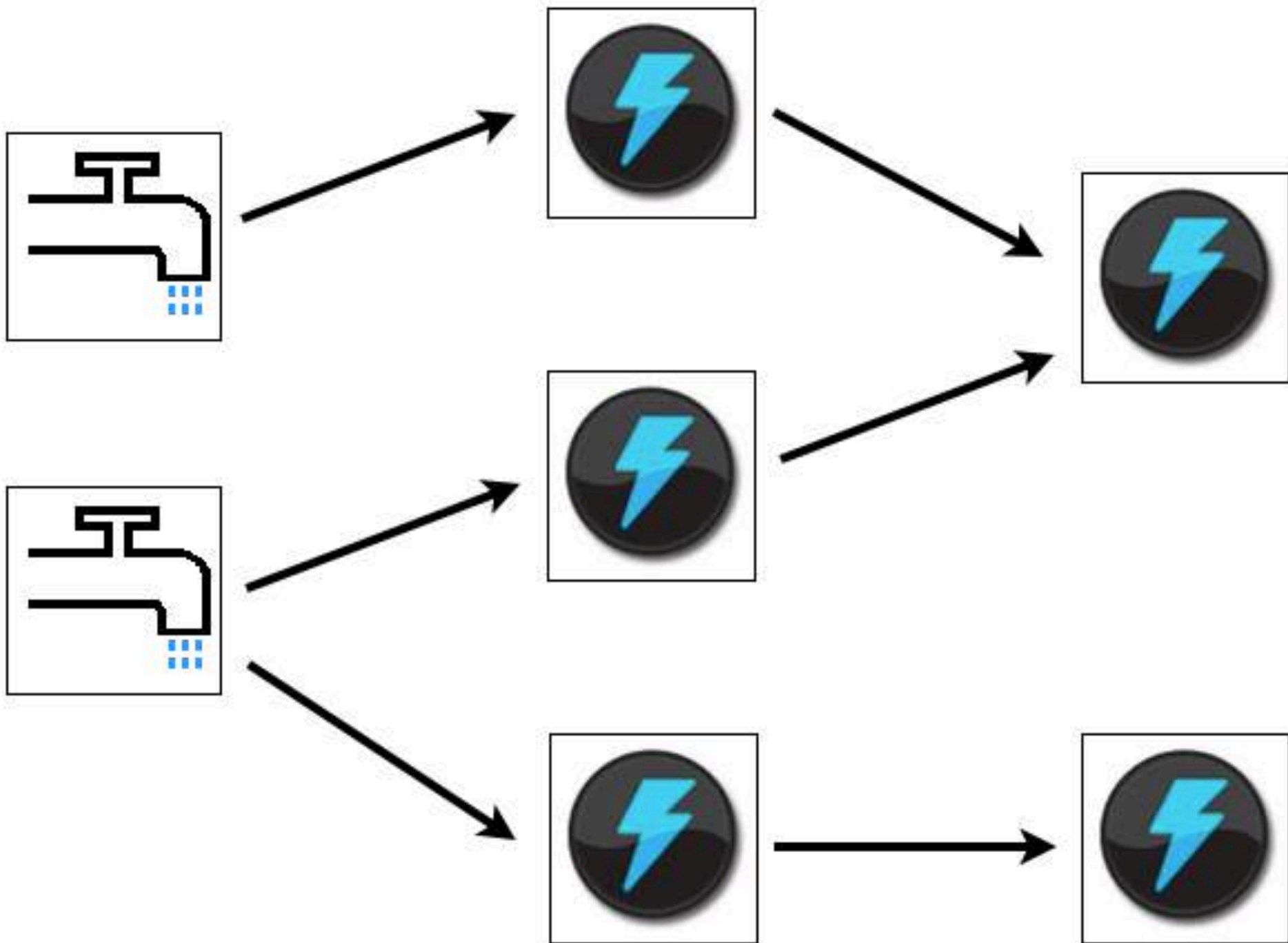
A close-up photograph of an elephant's head, facing left. The elephant has a large, wrinkled ear on the right side of the frame. Its trunk is raised and curled at the tip. Two large, curved tusks are visible at the bottom. The background is a blurred savanna landscape with tall grass and a blue sky. The text "Hadoop & Co." is overlaid in white, bold, sans-serif font across the center of the elephant's face.

Hadoop & Co.

+ Stream processing



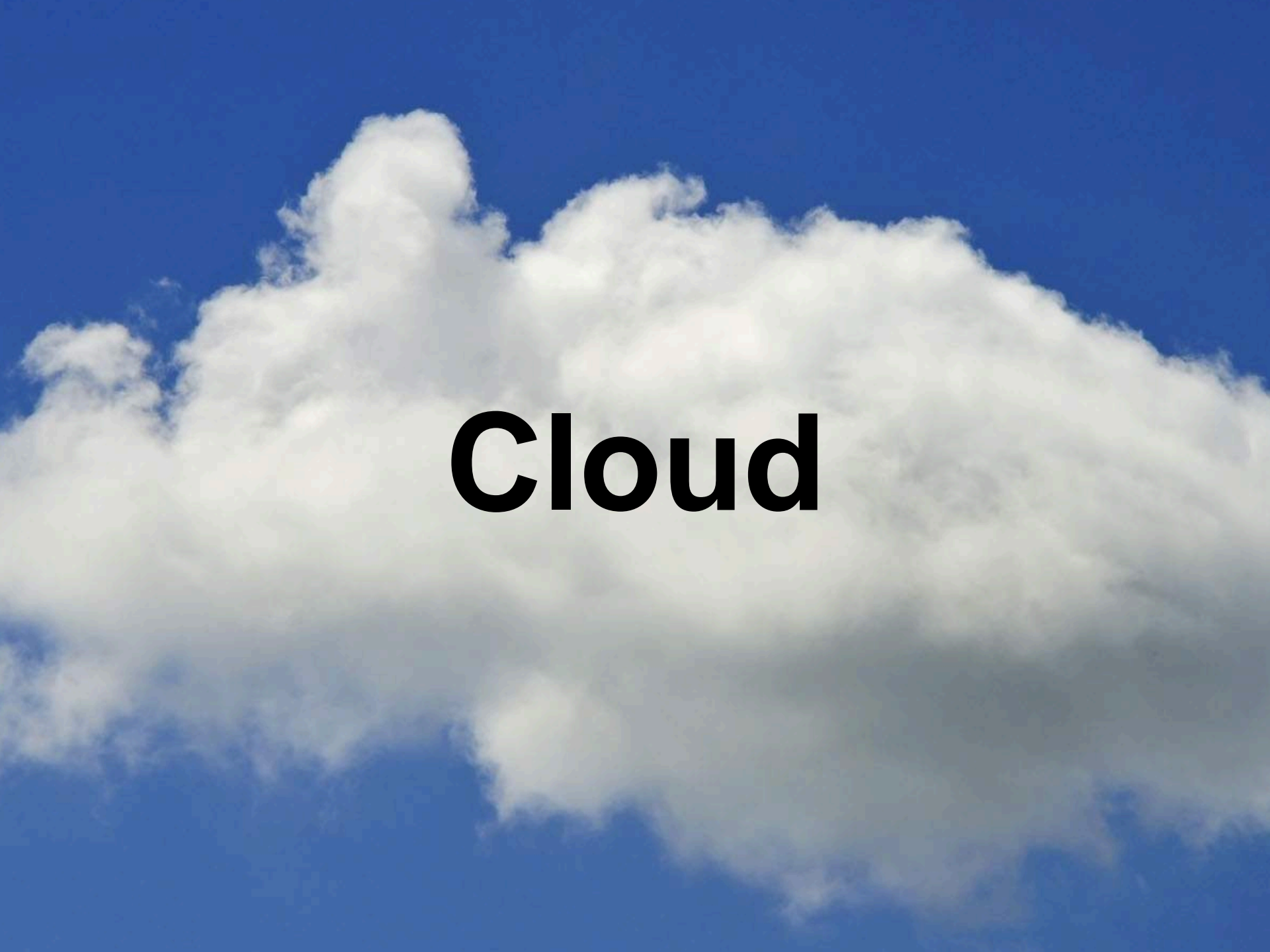
Storm & Co.



+ Infrastructure

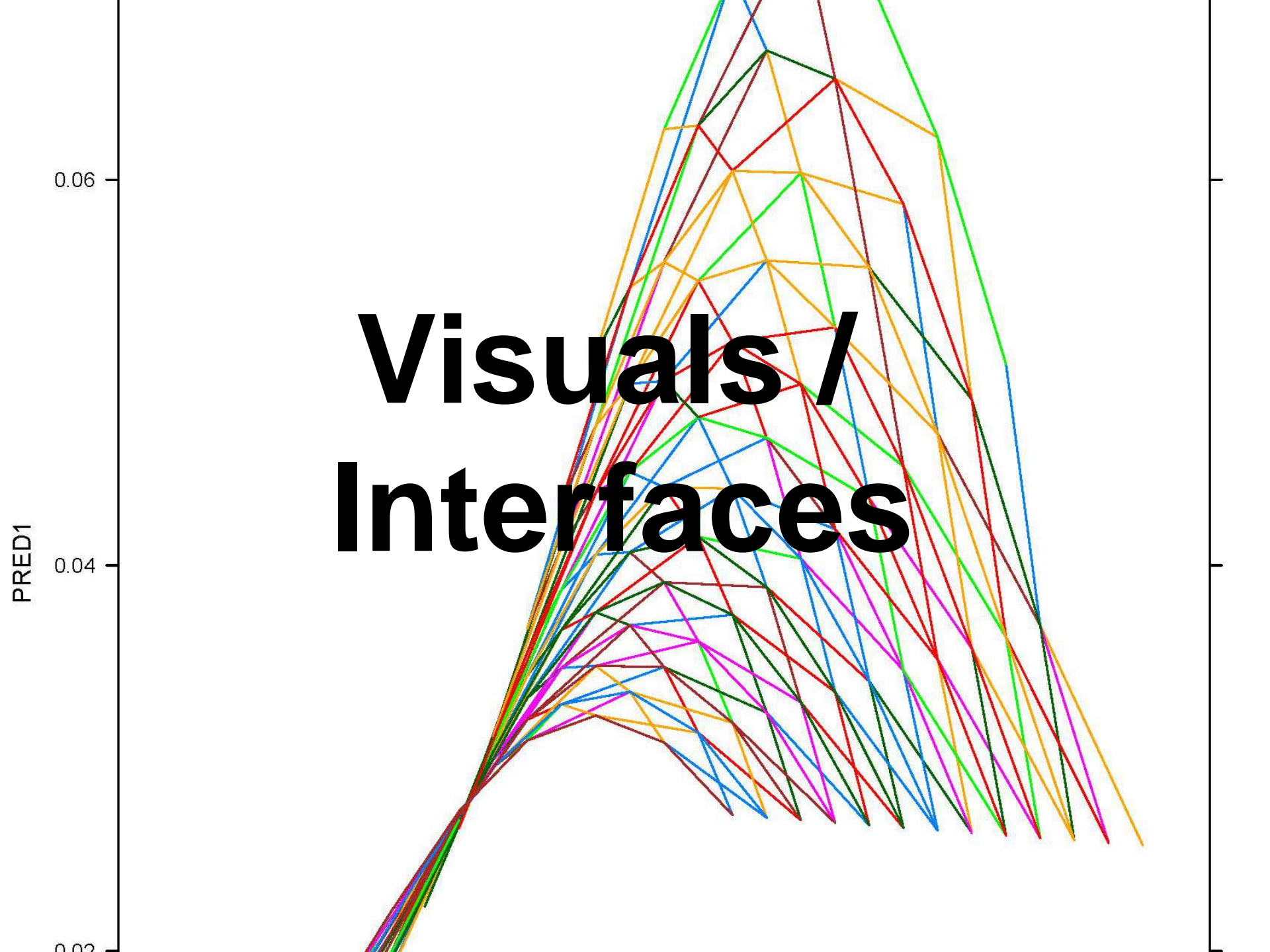
A photograph of a server room. In the foreground, a blue metal server rack is filled with numerous black server units. The rack is open, showing the internal components. To the left, another rack is visible, but it is partially obscured. The background shows more racks and a person's arm reaching into one of them. The floor is light-colored and reflective. The word "Dedicated" is overlaid in large white text.

Dedicated

A large, fluffy white cloud with soft, irregular edges, filling much of the frame. The cloud is set against a clear, vibrant blue sky. The lighting is bright, giving the cloud a soft, airy appearance.

Cloud

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