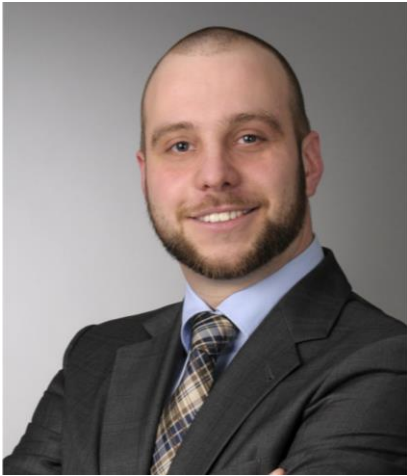


**BioRN Lounge with:**  
**Marc Stanke, [accantec consulting AG](http://accantec.com)**  
Topic: “The Data Scientist”

**March 20, 2018, 7pm**

At the restaurant Urban Kitchen, Poststr. 36/5 in Heidelberg

**Register** by emailing to Sibylle Geilenberg: [sg@biorn.org](mailto:sg@biorn.org)



*Dr. Marc Stanke*

**About:**

Empowered by state of the art technologies and new methodologies the data scientist taps into the treasure of academic research. Today the application of certain algorithms becomes widely available, this used to be formerly unthinkable in production due to the complexity or the sheer computational time. Infrastructure and platform-as-a-service products are available at reasonable prices and open source software provides support for the full spectrum of modern analytics.

Naturally, the data scientist is the combination of a Software and Data Engineer, which enables him to access new methodologies, technologies and a business-related role with strong statistical knowledge, like the well-known Bioinformatician.

Marc Stanke from accantec will give a brief overview of a modern analytical platform and showcase the important junctions and how the data scientist fits in.

**Speaker: Dr. Marc Stanke, accantec consulting AG**

Dr. Marc Stanke studied Life Science at the University of Hanover, before he graduated from the University of Hohenheim Institute of Food Chemistry and Biotechnology with his Ph.D. He started his career at accantec consulting AG in 2014 as Software Engineer for Life Science Analytics. In 2018 he became Head of Data Science. He is a regular speaker and trainer for advanced and predictive analytics / modelling and big data related topics.

# accantec group

The BI company

Hamburg – Düsseldorf – Frankfurt am Main – Heidelberg – München

## Facts

- Founded in 2002 in Hamburg
- Companies accantec consulting AG, accantec finance solutions AG and accantec information solutions AG
- Offices in Dusseldorf, Frankfurt / Main, Heidelberg and Munich
- About € 7 Mio. revenue in 2016 with 57 employees (43 internal, 14 external)
- BI- and SAP-projects (DACH)
- Expert Solution Unit (ACES)
- Managed Services for BI-architectures
- Customers from all branches, i.e., Banks, Insurances, Telecommunications, Life Science, Industries and Public Sector

## The Group

- Joint-Stock company with employees as stock holders
- Associate partners and senior managers as solid fundament for operating business and strategy
- Annual increasing of revenue and employees
- Long-term and successful cooperation with many customers (most are part of TOP500 enterprises in DACH)
- Partnerships with the most important BI-platform suppliers, i.e. SAS, JMP, IBM, SAP and Microsoft
- Accredited trainers and coaches
- Partnerships with other consulting companies for competency enhancements
- Internal Bachelor- and Master programs
- Data Protection responsibility by Mr. Peter Stahlberg, certerius GmbH



# Data Science



## Data Science

State of the Art Tools:

### Predictive Modelling

- Effect / response screening
- Bootstrap Forest
- Boosted Trees
- K-nearest neighbors
- Artificial neuronal networks (Deep Learning)

### Multi linear regression

- Standard least squares
- MANOVA
- Log-Linear Variance
- Logistic regression
- Penalized regression

### Multivariate Methods

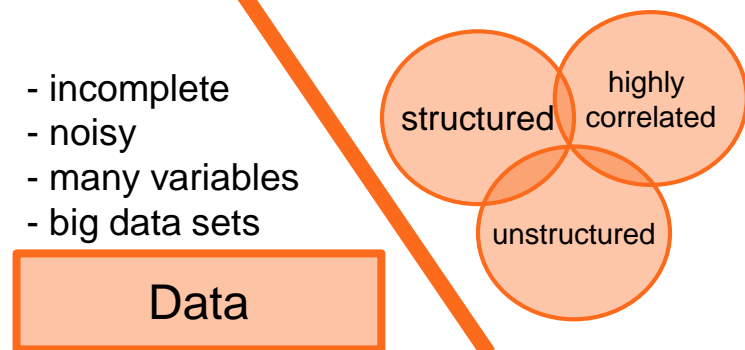
- Clustering
- Partial least squares
- Principal component analysis

Value

Modeling

- Influencing variable identification
- Higher process knowledge
- Predictions regarding the measurement System
- Process optimization

Data



# Parenthesis: Modeling reaction kinetics 1. order

Reaction Time	Reaction Temperature	Yield	
1	0,227	534,9	0,57828613
2	0,152	528,3	0,55181641
3	0,2	540	0,53588003
4	0,209	525,1	0,56584872
5	0,11	525,8	0,43362022
6	0,233	520,6	0,52267612
7	0,164	533,8	0,61119906
8	0,101	530,4	0,49451478
9	0,194	520,1	0,47233341
10	0,25	529,2	0,60262243
11	0,15	538,7	0,61029984
12	0,298	532,1	0,55300836
13	0,113	535	0,58600292
14	0,134	522,6	0,42539388
15	0,263	537,9	0,48491541
16	0,285	524,6	0,59473903

0,6021427542

+ 0,0279042968 ·  $\left( \frac{\text{Reaction Time} - 0,1995}{0,0985} \right)$

+ 0,0351096628 ·  $\left( \frac{\text{Reaction Temperature} - 530,05}{9,95} \right)$

+  $\left( \frac{\text{Reaction Time} - 0,1995}{0,0985} \right) \cdot \left( \frac{\text{Reaction Time} - 0,1995}{0,0985} \right) \cdot -0,065837819$

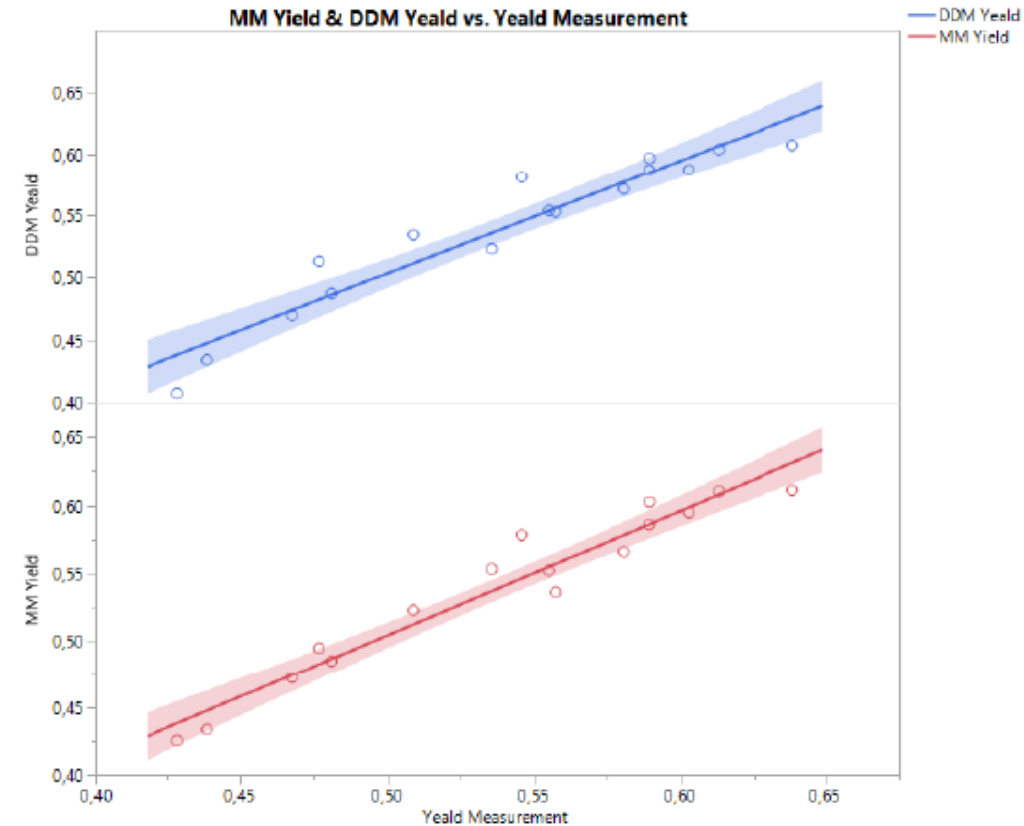
+  $\left( \frac{\text{Reaction Time} - 0,1995}{0,0985} \right) \cdot \left( \frac{\text{Reaction Temperature} - 530,05}{9,95} \right) \cdot -0,138082766$

+  $\left( \frac{\text{Reaction Temperature} - 530,05}{9,95} \right) \cdot \left( \frac{\text{Reaction Temperature} - 530,05}{9,95} \right) \cdot -0,090329968$

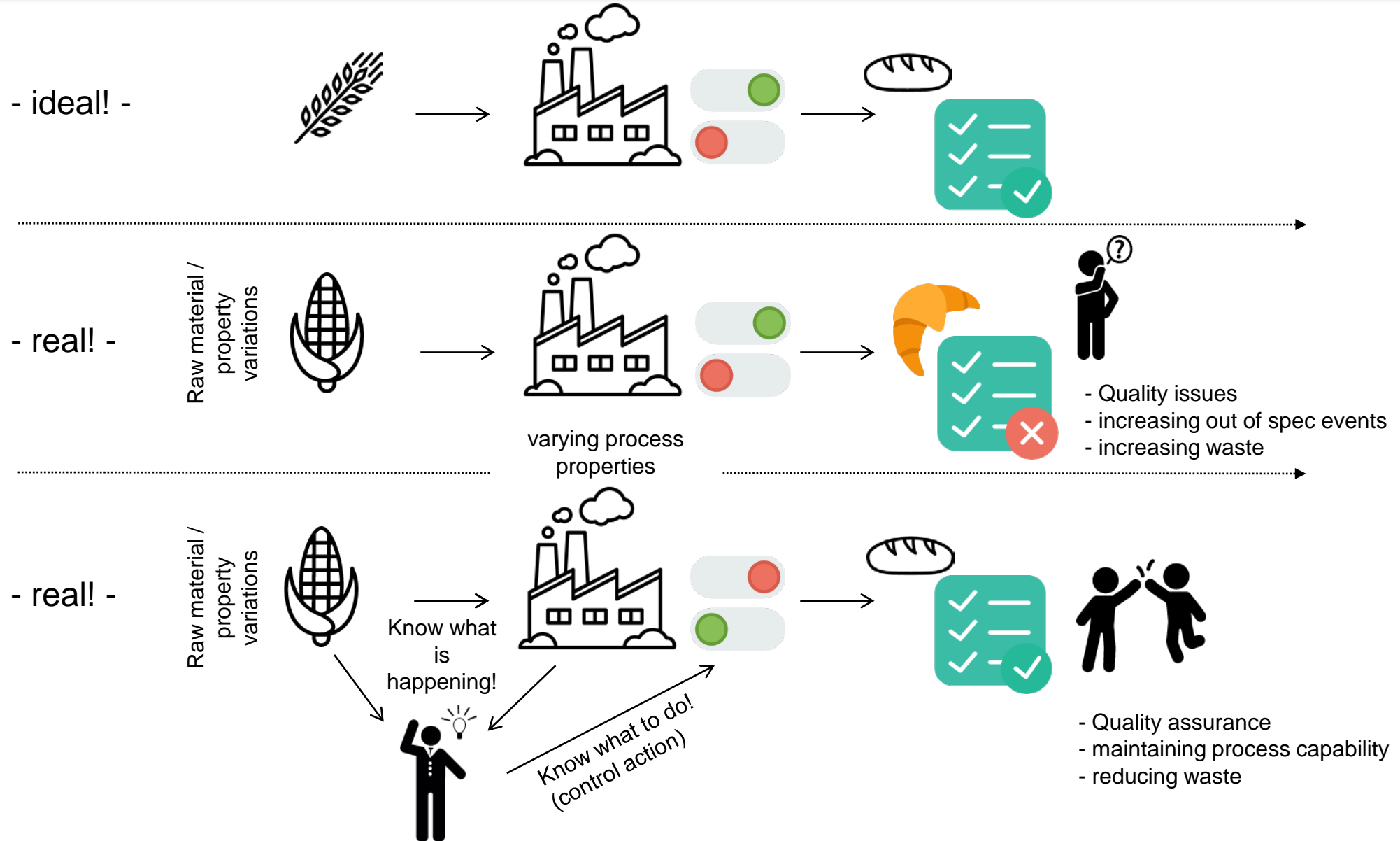
**DDM**  
(Data Driven Model)

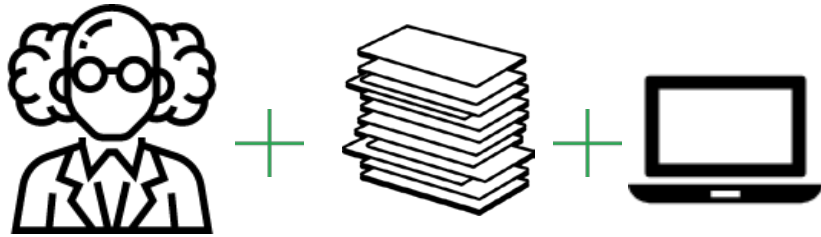
$$\frac{d[A]}{dt} = k \cdot [A] = [A_0] \cdot \frac{k_1 \cdot e^{-k_1 \cdot t} \cdot e^{-k_2 \cdot T}}{(k_1 - k_2)}$$

**MM**  
(Mechanistic Model)



# Application of Data Science (PAT)





## Data Science

- ... terminology → not new!
- ... statistical methods in principle → not new!
- ... way of looking at problems → not new!

Noting new, and yet...



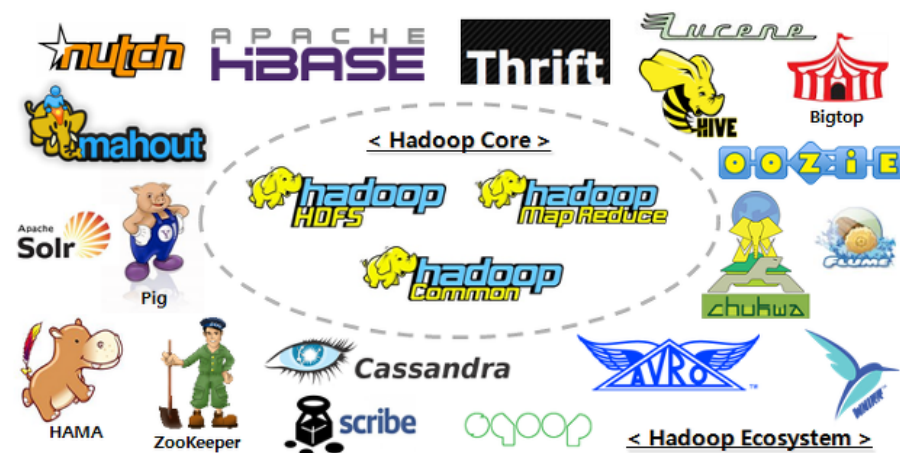
new! ← methodology ...  
 new! ← performance ...  
 new! ← Scalability ...

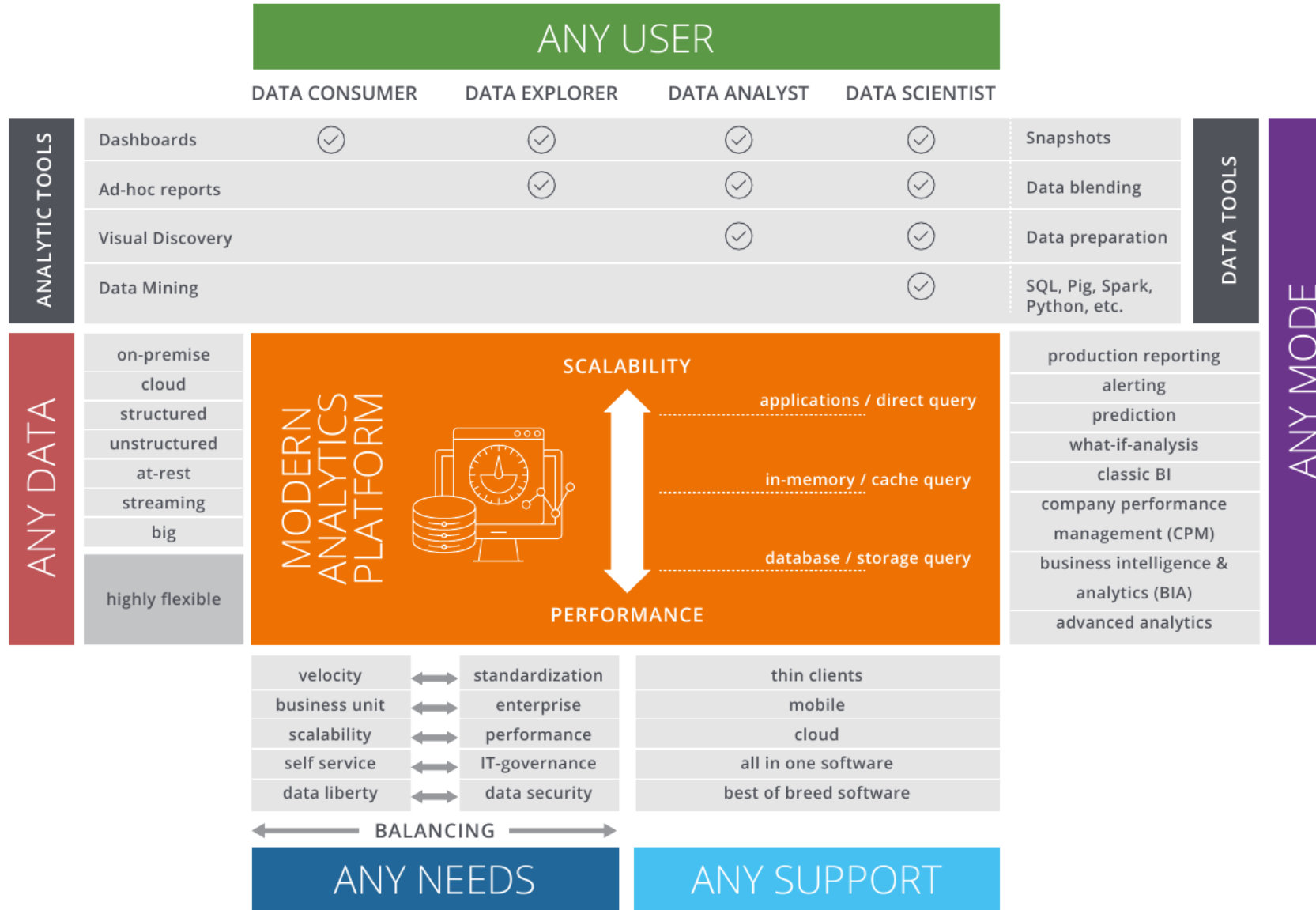
# Data Science

... terminology → not new!  
 ... statistical methods in principle → not new!  
 ... way of looking at problems → not new!

## Noting new, and yet...

Significantly more attention







Something like a statistician

Application of ML & Advanced Analytics

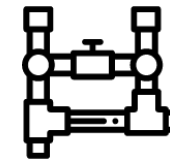


**Banana!**



Software Engineer

Data pipeline implementation on Modern Analytics Platforms



Data Engineer

Understanding and shaping the underlying architecture

