Create real business impact, where Data Science and PaaS come together
Merwin de Jongh

‘I have a passion for turning concepts into plans and concrete tasks. The combination with my natural love for added value ensures that we always deliver high quality results in time.’

Founder & CTO

BSc Econometrics & OR
MSc Operations Research

Tilburg
Building Blocks introduction

Data Science

33

Tilburg & Amsterdam

2013
How to create real business impact with Data Science?
What is Data Science?
What is Data Science?

Data Science is gaining popularity!

Source: Google Trends
What is Data Science?

Data Science versus Big Data and Business Intelligence

Data analytics sequence:
- Descriptive: What has happened?
- Diagnostive: Why did it happen?
- Predictive: What will happen?
- Prescriptive: What should I do?

Complexity of data structures:
- Simple
- Complex

- Business Intelligence
- Big Data
- Data Science
What is Data Science?

Data Science in the world around us

Spotify

NETFLIX

Gemeente Amsterdam
What is Data Science?
Because you care about consumers and want to maximize their value

We guide you in finding, converting and keeping your ideal customers by placing them at the center of your business.

Product centric
“What do I want to sell?”

Customer centric
“What does my customer need?”
What is Data Science?

We empower you to perfectly act on your consumers’ needs.

We believe that every consumer is unique and demands a personal approach that addresses its individual needs.

Personal products, promotion...

...pricing and recommendation
What is Data Science?

We make impact on our clients’ businesses

Retail

- SWAROVSKI
- Blocks: Stock, Pricing, Taste, Demand

Insurance

- Nationale Nederlanden
- Blocks: Loyalty, Pricing, Risk, Competitor

Travel

- Corendon
- Blocks: Loyalty, Recommendation, Stock, Demand
What is Data Science?

We make impact on our clients’ businesses

**Retail**

- Turnover: +7.5%

**Insurance**

- Combined ratio: +14%

**Travel**

- Conversion: +54%
Challenges in Data Science
What is a Data Scientist?

MODERN DATA SCIENTIST

Data Scientist, the ancient job of 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand what a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

**MATH & STATISTICS**
- Machine learning
- Statistical modeling
- Experimental design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

**PROGRAMMING & DATABASE**
- Computer science fundamentals
- Scripting languages: e.g., Python
- Statistical computing packages: e.g., R
- Databases: SQL and NoSQL
- Linear algebra
- Parallel processing and parallel query processing
- MapReduce concepts
- Hadoop and Hive Pig
- Data mungers
- Experience with data like AWS

**DOMAIN KNOWLEDGE & SOFT SKILLS**
- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative, and collaborative

**COMMUNICATION & VISUALIZATION**
- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot2 or lattice
- Knowledge of any of visualization tools: e.g., Tableau, Qlik

2 important consequences
Data Scientists are scarce

Challenges in Data Science

**MODERN DATA SCIENTIST**

Data Scientist, the new job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand what a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

**MATH & STATISTICS**
- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

**PROGRAMMING & DATABASE**
- Computer science fundamentals
- Scripting languages: e.g. Python
- Statistical computing packages e.g. R
- R-Shiny, SQL, and NoSQL
- Linear algebra
- Parallel algorithms and parallel computing
- MapReduce concepts
- Hadoop and Hive/Figi
- Cloud runtimes
- Experience with cloud platforms

**DOMAIN KNOWLEDGE & SOFT SKILLS**
- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovation and collaboration

**COMMUNICATION & VISUALIZATION**
- Ability to engage with senior management
- Story telling skills
- Transform data-driven insights into conversations and actions
- Visual art design
- R packages like ggplot2 or lattice
- Knowledge of any visualization tools e.g. Tableau, D3.js, Tableau

**JHERONIMUS ACADEMY OF DATA SCIENCE**

Source: McKinsey
Challenges in Data Science

A typical Data Scientist has limited IT knowledge

MODERN DATA SCIENTIST

(Data Scientist, the unique job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand what a data scientist is, is equally hard. So here is a little checklist about who the modern data scientist really is.

MATH & STATISTICS
- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE
- Computer science fundamentals
- Scripting languages: e.g. Python
- Statistical computing: package e.g. R
- Databases: SQL and NoSQL
- Linear algebra
- Parallel processing and parallel computing
- MapReduce concepts
- Hadoop and HBase
- Python
- Experience with cloud platform AWS

DOMAIN KNOWLEDGE & SOFT SKILLS
- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative, and collaborative

COMMUNICATION & VISUALIZATION
- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art designs
- R packages like ggplot or lattice
- Knowledge of any visualization tools: e.g. Tableau, D3.js, Tableau
Challenges in Data Science

How to go from experiments to business impact?

How to bring Data Science from *experiments* to *production level*, and create real *business impact*?
How to solve these challenges?
How to solve these challenges?

Challenges in Data Science

Data Scientists are **scarce**

Typical Data Scientist has **limited IT knowledge**
How to solve these challenges?

Eliminate repetitive Data Science work

Data Scientists are scarce
How to solve these challenges?

Eliminate repetitive Data Science work → Introduction of Blocks structure

Data Scientists are scarce
How to solve these challenges?

The Blocks structure

**Combine** the blocks into solution

How our **blocks work**

- Data processing
- Variable selection
- Model selection
- Validation & checks
Challenges in Data Science

Data Scientists are **scarce**

Typical Data Scientist has **limited IT knowledge**
How to solve these challenges?

Introduction of the Data Engineer

Typical Data Scientist has limited IT knowledge
How to solve these challenges?
Also eliminate repetitive Data Engineering work, and make the Data Scientist life easier

Solution

Client Teams
Applied Innovation
Technical Innovation

Developed in:

Existing software

Executable composition

Execution engine

Block executor

Block registry

Tailor-made code

Platform as a Service (PaaS)

Public

Private

Hybrid

Infrastructure

Existing containerized software
How to solve these challenges?

The Data Scientist has an easy toolbox to create business impact!

Developed in:
- Client Teams
- Applied Innovation
- Technical Innovation

Data Scientist Toolbox

- Executable composition
- Block
- Existing containerized software
- Tailor-made code
- Block registry
How to solve these challenges?

This leads to the need for a Container Platform

Developed in:
- Client Teams
- Applied Innovation
- Technical Innovation

Solution

Platform as a Service (PaaS)

- Public
- Private
- Hybrid

- Existing software
- Executable composition
- Execution engine
- Block executor
- Tailor-made code
- Block registry
How to solve these challenges?

This leads to the need for a Container Platform

Developed in:
- Client Teams
- Applied Innovation
- Technical Innovation

Solution
Containerization is a completely new concept in the field of Data Science!

Source: Gartner
Data Science and Container Platform

Some things are different from ‘regular’ software development

**DevOps** in Data Science is different

Data Scientist think of **different** type of **scaling**
How to incorporate the experimentation phase in DevOps?

**DevOps** in Data Science is different

- Develop
- Test
- Operations

Experiment → Develop → Test → Operations
Some things are different from ‘regular’ software development

**DevOps** in Data Science is different

Data Scientist think of *different* type of *scaling*
How to scale up 100x for just 30 minutes?

Data Scientist think of different type of scaling
First tests are really promising!
Luckily, we are not alone in this journey
The journey of Data Science automation is just only the beginning!

Source: Gartner
The journey of Data Science automation is just only the beginning!

Source: Gartner
Data Science and Container Platform

Thank you! Questions?

Ask us!

Yoran van Oirschot
Data Engineer

Merwin de Jongh
CTO & Founder

Come visit us...

Sint Josephstraat 139
5017 GG Tilburg

Barbara Strozzilaan 101
1083 HN Amsterdam

www.building-blocks.nl

...and follow us on LinkedIn!

www.linkedin.com/company/building-blocks
Ask us!

Yoran van Oirschot  
Data Engineer

Merwin de Jongh  
CTO & Founder

This afternoon you can find us at...