ABOUT A DATA SCIENTIST

PATRIC HELLGREN
STRATEGIC PARTNERSHIPS & BUSINESS DEVELOPMENT
TEL: + 46 (0)8 5221 70 51
PATRIC.HELLGREN@SAS.COM

Era	1.0: Traditional Analytics	2.0: Big Data	3.0: Data Economy
Timeframe	Mid-1950s to 2000	Early 2000s to Today	Today and in the Future
Culture, Ethos	Very few firms "compete on analytics"..."we know what we know."	Agile, experimental, hacking...new focus on data-based products and services for customers	Agile methods that speed "time to decision"...all decisions driven (or influenced) by data...the data economy
Type of analytics	• 5% predictive, prescriptive • 95% reporting, descriptive	• 5% predictive, prescriptive • 95% reporting, descriptive (visual)	• 90%+ predictive, prescriptive • Reporting automated commodity
Cycle time	Months ("batch" activity)	An insight a week	Millions of insights per second
Data	Internal, structured...very few external sources available or perceived as valuable **"data"**	Very large, unstructured, multi-source...much of what's interesting is external...explosion of sensor data **"Big Data"**	Seamless combination of internal and external...analytics embedded in operational and decision processes... tools available at the point of decision **"Data Economy"**
Technology	Rudimentary BI, reporting tools...dashboards...data stored in enterprise data warehouses or marts	New technologies: Hadoop, commodity servers, in-memory, machine learning, open source..."unlimited" compute power	New data architectures...beyond the warehouse New application architectures...specific apps, mobile
Organization & Talent	Analytical people segregated from business and IT..."Back Room" statisticians, quants without formal roles	Data Scientists are "on the bridge"...talent shortage noted...educational programs on the rise	Centralized teams, specialized functions among team members, dedicated funding... Chief Analytics Officers ... recognized training, education programs

HIGH PERFORMERS ARE ABLE TO REALIZE OUTCOMES BETTER USING ANALYTICS

Low Performers

Focus on Data to Insights

- Only one in five invest at a high level in analytics and only one in 10 expect this to increase significantly in the next three years
- Less than half manage talent from end-to-end; just over half use a multi-faceted approach; few are willing to acquire talent
- Only one in five use seven or more types of data in their analysis; less than half use advanced analytical techniques
- One third or fewer embed analytics into the decision process and struggle with decision making

High Performers

Focus on Insights to Actions

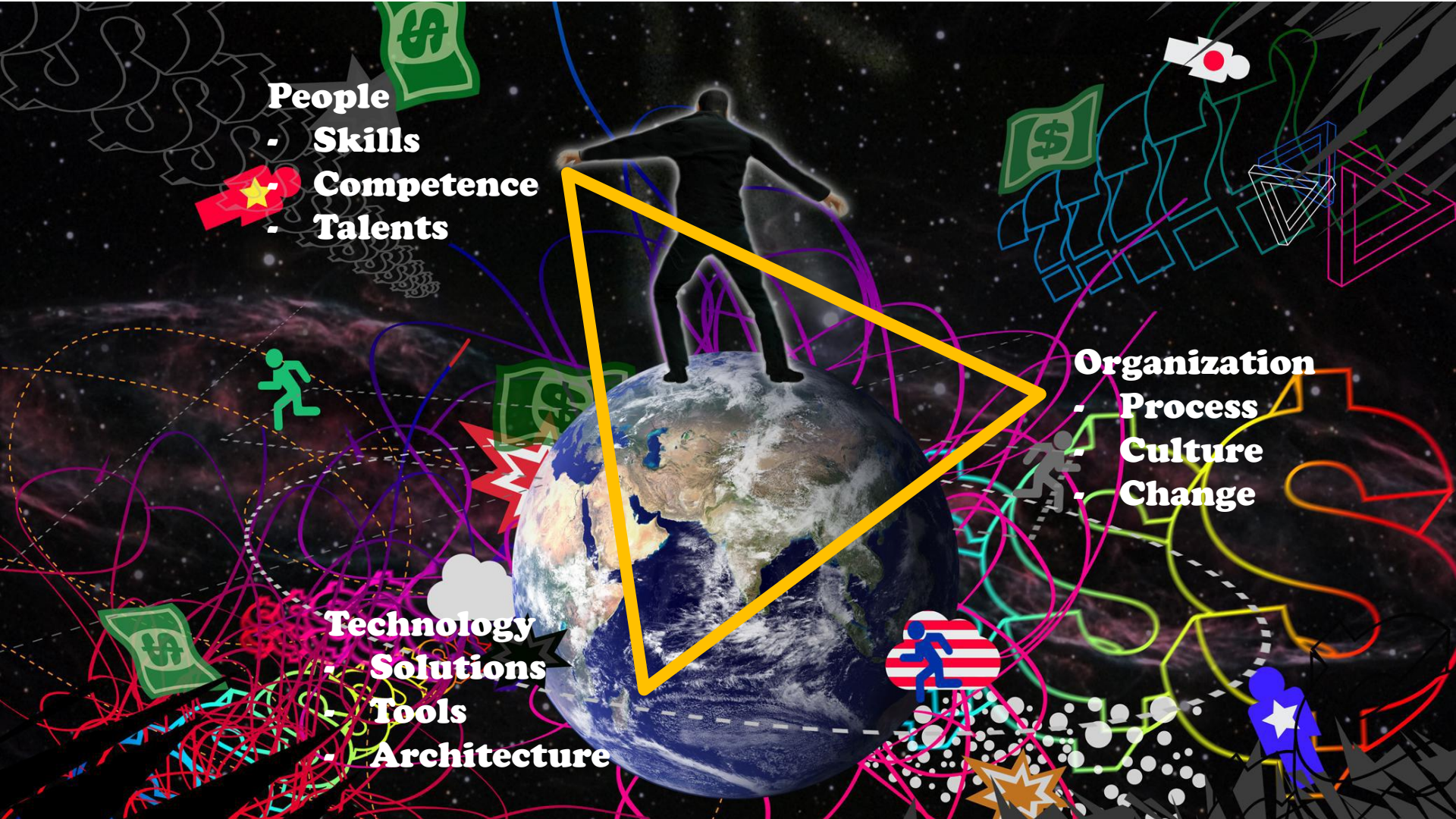
- The majority invest considerably in their analytical capability and this is expected to significantly increase over the next three years
- Nearly all manage talent from end-to-end and four out of five source talent using a multi-faceted approach; more than twice as many high performers are willing to acquire talent compared to low performers
- Majority use seven or more types of data in analyses; four out of five use advanced analytical techniques
- Four out of five embed analytics into the decision process

Source: Accenture/MIT High Performance Analytics Study, May 2014

90% or more High Performing companies are satisfied with the contribution analytics has made to

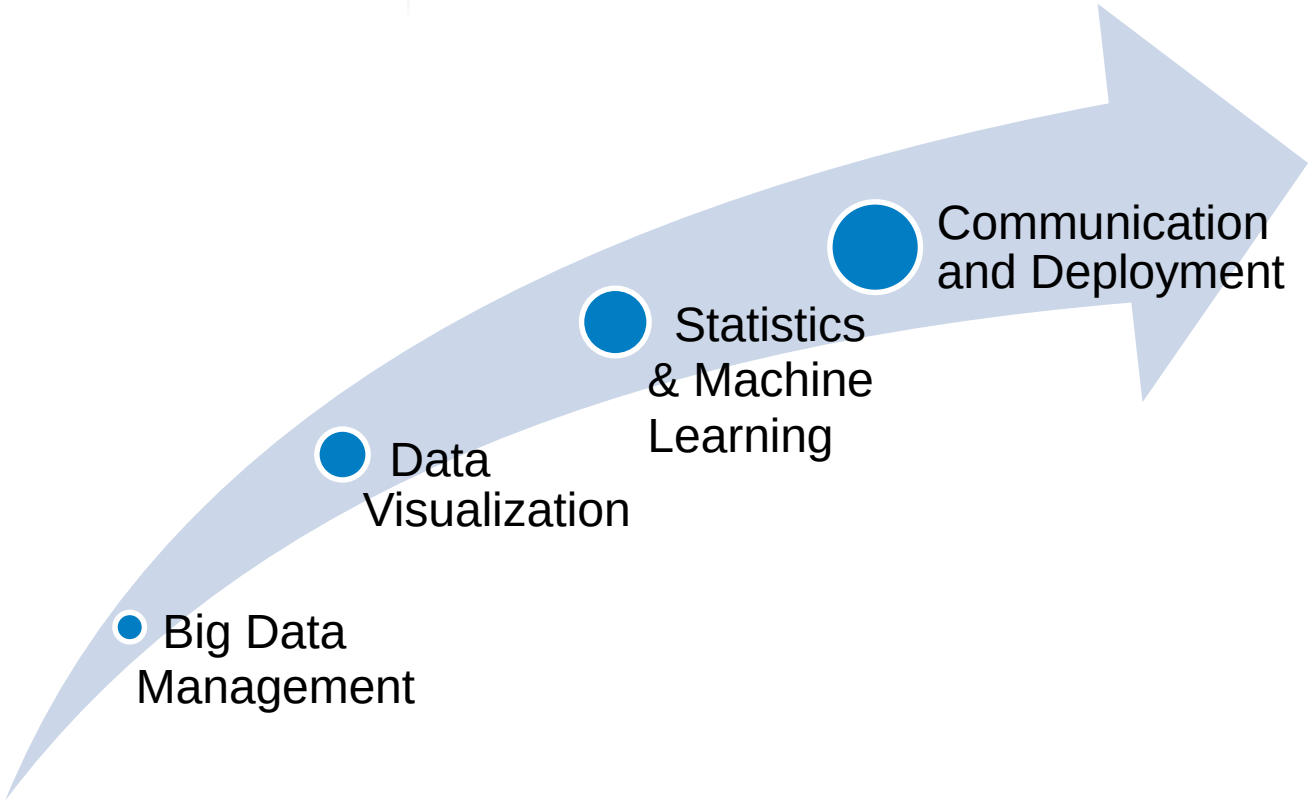
- **financial performance**
- **strategic direction**
- **addressing growth opportunities**
- **informing critical decisions and**
- **managing risk**

compared with 39% of low performers (on average)


People
- Skills
- Competence
- Talents
Organization
- Process
- Culture
- Change
Technology
- Solutions
- Tools
- Architecture

ABOUT A DATA SCIENTIST

FOR THE 21ST CENTURY WORKFORCE



ANALYTICS

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

The world needs more Data Scientists

Harvard Business Review

The shortage of data scientists is becoming a serious constraint in some sectors.

Who Are These People?

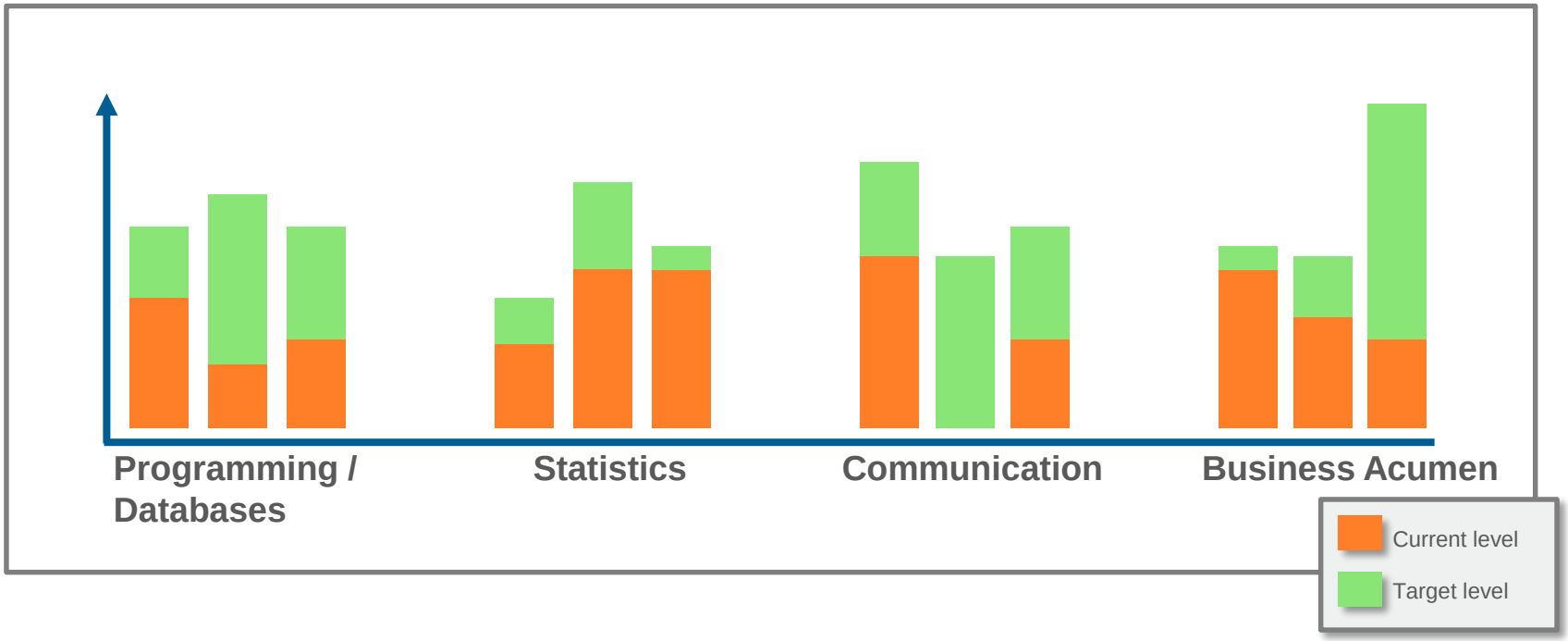
If capitalizing on big data depends on hiring scarce data scientists, then the challenge for managers is to learn how to identify that talent, attract it to an enterprise, and make it productive. None of those tasks is as straightforward as it is with other, established organizational roles. Start with the fact that there are no university programs offering degrees in data science. There is also little consensus on where the role fits in an organization, how data scientists can add the most value, and how their performance should be measured.

More than anything, what data scientists do is make discoveries while swimming in data. It's their preferred method of navigating the world around them. At ease in the digital realm, they are able to bring structure to large quantities of formless data and make analysis possible. They identify rich data sources, join them with other, potentially incomplete data sources, and clean the resulting set. In a competitive landscape where challenges keep changing and data never stop flowing, data scientists help decision makers shift from ad hoc analysis to an ongoing conversation with data.

What kind of person does all this? What abilities make a data scientist successful? Think of him or her as a hybrid of data hacker, analyst, communicator, and trusted adviser. The combination is extremely powerful—and rare.

ABOUT A DATA SCIENTIST

PROFILE



ABOUT A DATA SCIENTIST

THE RISE OF CITIZEN DATA SCIENTISTS

Microsoft Azure Machine Learning

IBM Watson Analytics

SAP Lumira

SAS Visual Statistics

Angoss

Predixion

Alpine

Ayasdi

Emcien

BeyondCore

- Citizen data scientists are well positioned to "drive" data-driven culture
- Domain expertise
- They can be recruited in-house as well as from many "adjacent" disciplines

- Tools Are Getting Easier to Use
- Expert Data Scientists Will Remain Rare

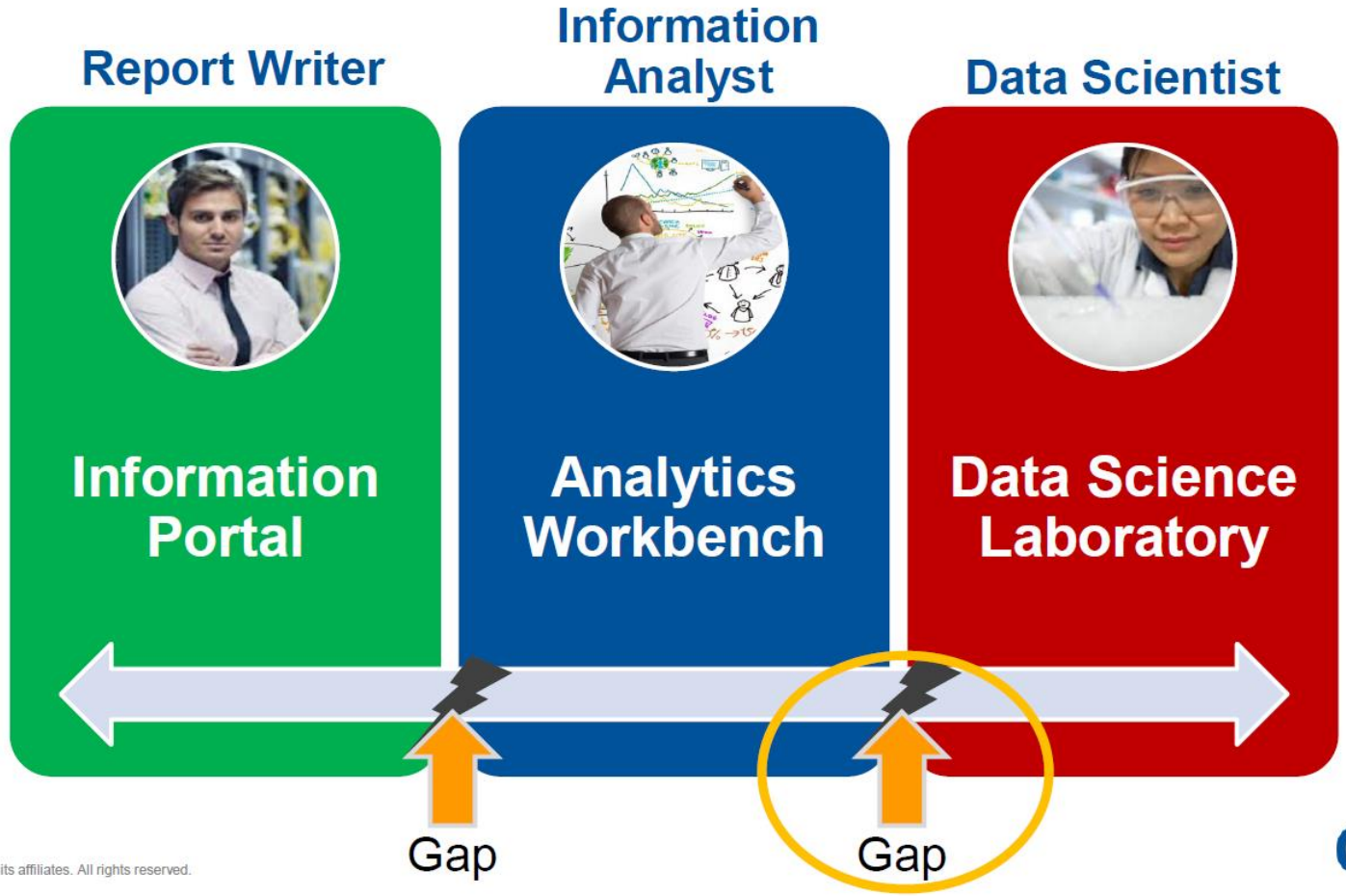
Gartner

sas THE POWER TO KNOW.

Gartner Predicts

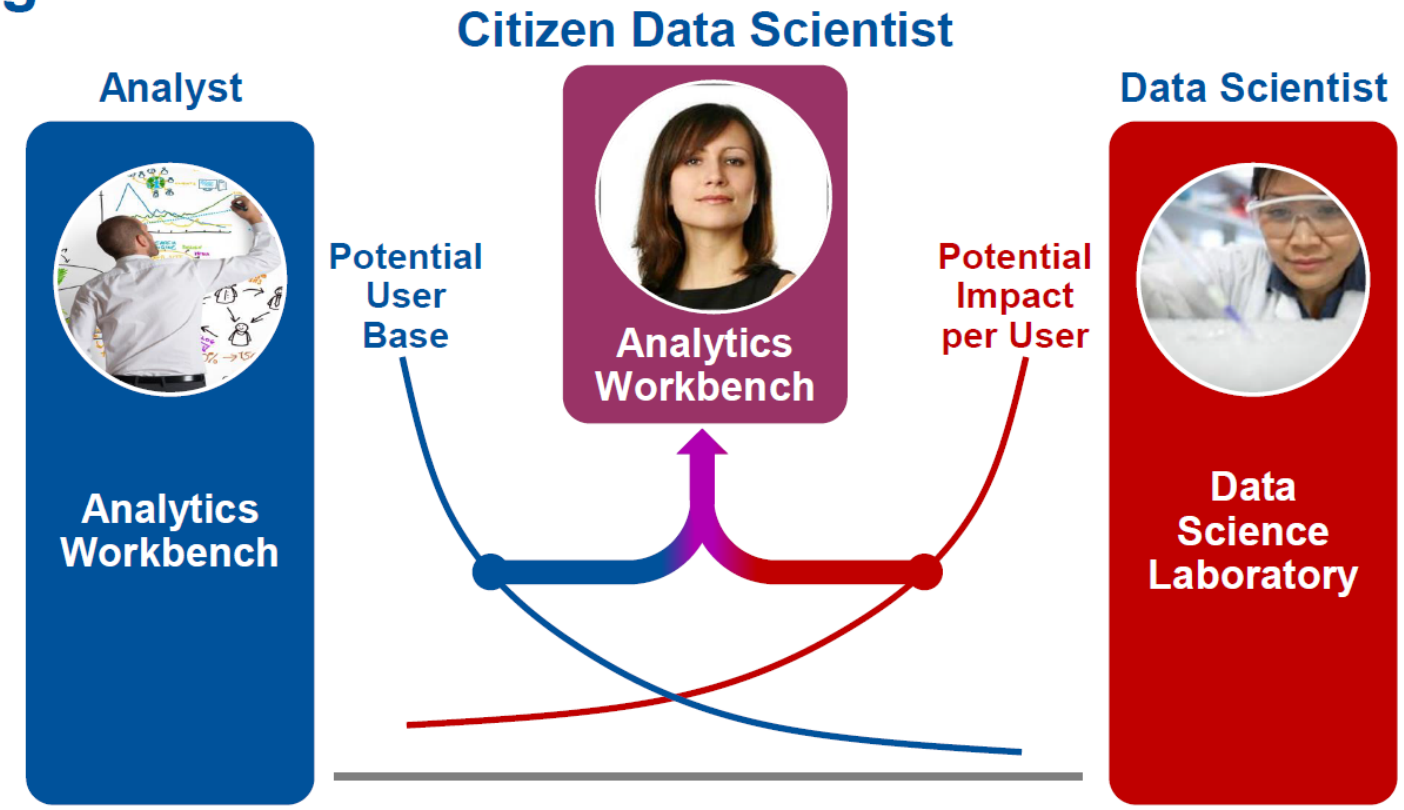
2017: The number of citizen data scientists will grow 5x the rate of specialized data scientists

Roles in Information Exploration



Gartner

A New Role to Increase the Impact of Analytics on the Organization



Gartner

ABOUT A DATA SCIENTIST

PROGRAMS

- SAS Data Discovery Scientist (18 days during 12 month)
- SAS Academy for Data Science
  - SAS Certified Big Data Professional (30 days classroom training)
  - SAS Certified Data Scientist (30 days classroom training)

SAS DATA DISCOVERY SCIENTIST

18 DAYS CLASSROOM DURING 12 MONTH I STHLM

Month (open scheduled start February and September in Stockholm)

1	2	3	4	5	6
Tier 1 – Tools and Mindset for Business Analytics **VA, VS, EG, (+ essential tools) 9 days, references**				**Tier 2 – Case Value** **3 models**	
7	8	9	10	11	12
Tier 2 – Value of Analytics **Analytics, Communication, Business perspective 9 days, final presentation and test**					

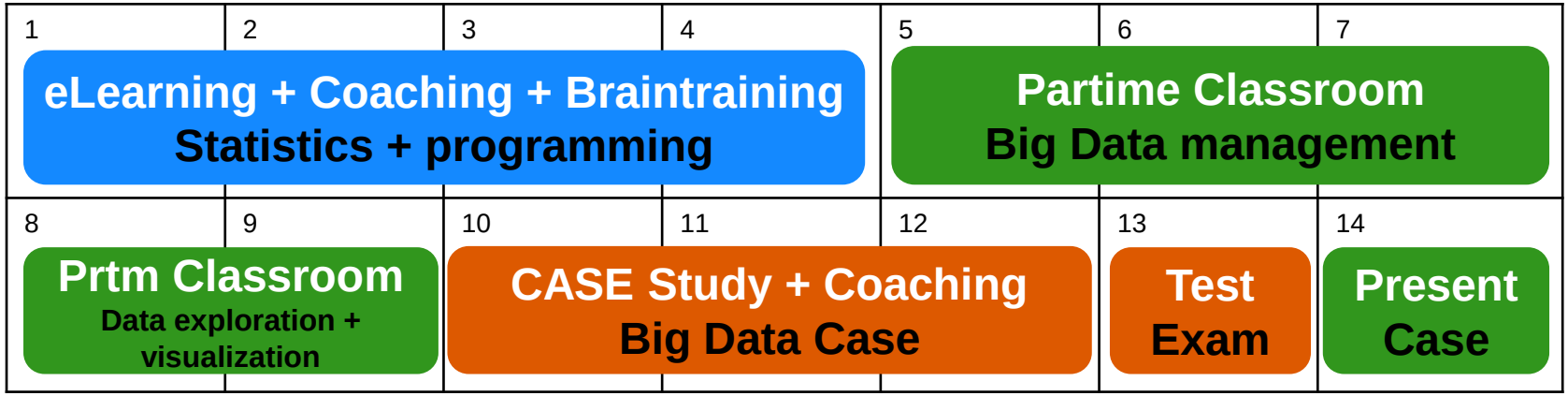
Individual customized case program with focus on mindset of analytics, business value and communication combining classroom, eLearning and selfstudies. Access to local SAS experts, theory and practice in real life case.

DATA SCIENCE ACADEMY

LEVEL 1: CERTIFIED BIG DATA PROFESSIONAL

Program

Weeks (company specific or scheduled courses in NL, UK, US)



eLearning
Classroom or LWC
Certification

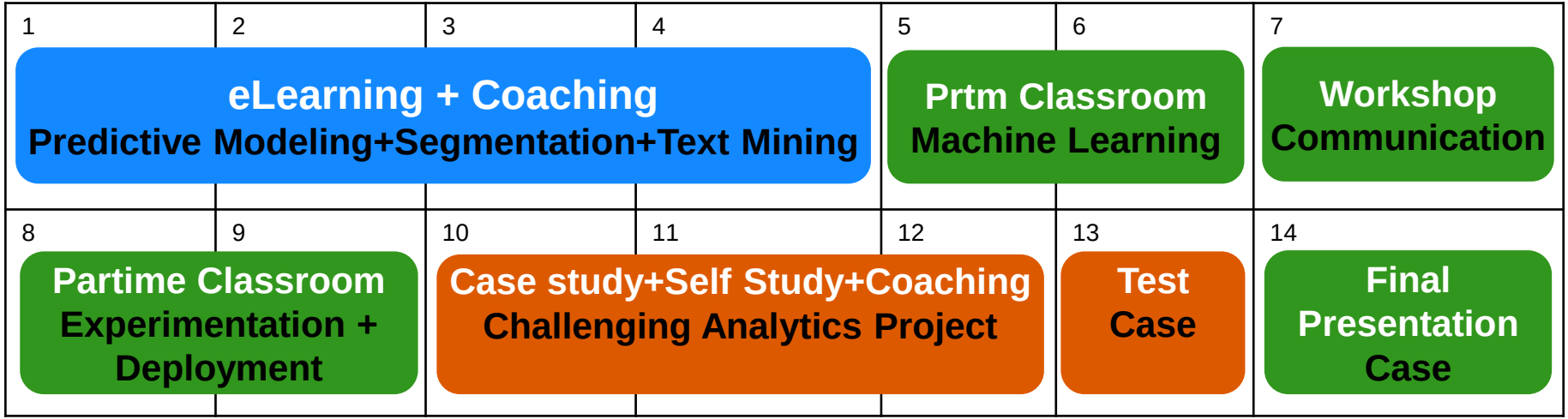
All classroom training via hybrid connected classroom

DATA SCIENCE ACADEMY

LEVEL 2: CERTIFIED DATA SCIENTIST

Program

Weeks (company specific or scheduled courses in NL, UK, US)



eLearning

Classroom or LWC

Certification

Classroom training via hybrid connected classroom

SAS ACADEMY FOR DATA SCIENCE

COURSES IN THE ACADEMY PROGRAM

1. Big Data Professional

- *Statistics 1*
- *Programming 1+ 2*
- Accessing Hadoop
- Data step 2
- Descriptive and Inferential Statistics
- ANOVA, Regression, and Logistic Regression
- Working with In-Memory Data with PROC IMSTAT
- Hadoop Big Data Analysis with Pig, Hive, and SAS
- Visual Analytics
- Communicating your findings

Open Source Tooling; R, Phyton

2. Data Scientist

- *Predictive Modeling (PMLR)*
- *Enterprise Miner and Text Miner*
- Logistic Regression, Decision Trees, and Neural Networks (AAEM)
- Cluster Segmentation and Association Analysis (AAEM)
- SAS In-Memory Statistics (IMPM)
- Visual Statistics (SVSO72)
- Big Data Experimentation
- Communicating Technical Findings to a Nontechnical Audience