Università degli Studi
di Milano-Bicocca

DATA SCIENCE
Master's Degree Programme

CLOUD BUSINESS INTELLIGENCE

DATA TRENDS THAT SUPPORT THE NEW DIGITAL IMPERATIVES

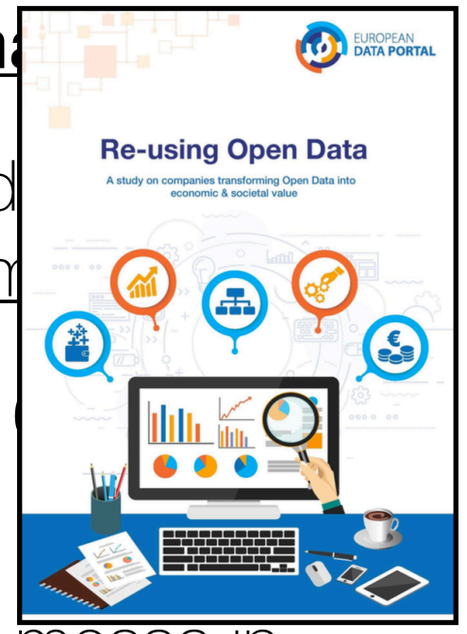
Mauro Pelucchi

mauro.pelucchi@gmail.com

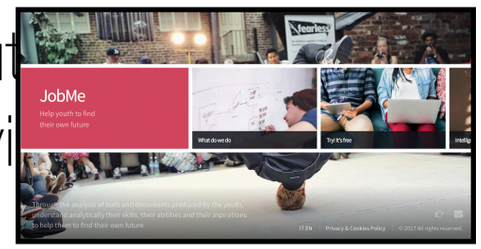
mauro.pelucchi@unimib.it

About me

- Mauro Pelucchi **mauro.pelucchi@unimib.it** **mauro.pelucchi@gma**



- Laureato in **Ingegneria Informatica** presso Università degli Studi d con una tesi sull'utilizzo degli OpenData (www.hammer-project.com
- Master in **Business Intelligence & Big Data Analytics** (2015/201 lavoro finale all'indirizzo web www.job-me.it)
- Mi occupo di **Data Science** e **Data Engineering**: progettazione e messa in qualità di grandi basi di dati, estrazione di patterns da dati non strut realizzazione di sistemi direzionali statistici e strumenti per la data vi



- In passato ho progettato e realizzato sistemi di Process Integration e Business Intelligence per grandi realtà industriali e per la pubblica amministrazione
- Main interests: **OpenData e Open Government, Big Data Analysis, Data Presentation techniques, Artificial Intelligence, Machine Learning Modelling & Deep Learning, Applied Artificial Intelligence**

Agenda

- BI Tools & Features
- The future of Business Intelligence Tools
- Cloud fundamentals
- Cloud & Data Science
- Databricks Workshop
- BI Team
- Data trends

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

What features do you want from Modern Business Intelligence tool?

Features

Data Discovery Capabilities

Governed Data Discovery

Visual-Based Investigation

Mashup and Embedded BI

Metadata Management

Self-service analytics

Cloud BI

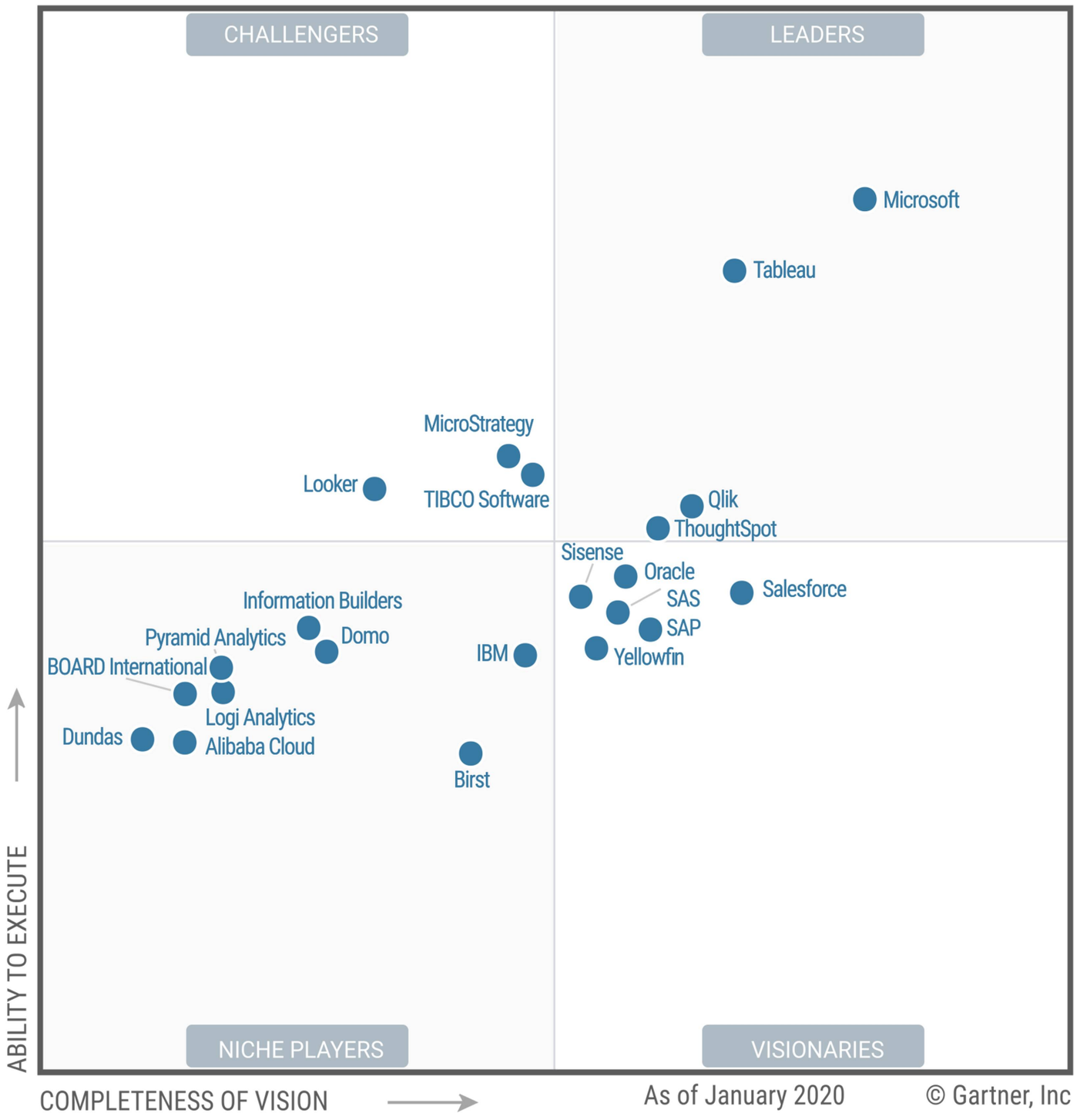
Self-service BI

Features

Agile BI Development	Data Integration for business users	Embedded Advanced Analytics	Mobile Exploration
Ease of Use and Visual Appeal, Data storytelling	Real time BI and real time analytics	Spatial location intelligence	Publish,Share and Collaborate

Magic Quadrant for Business Intelligence and Analytics Platforms (2020)

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms

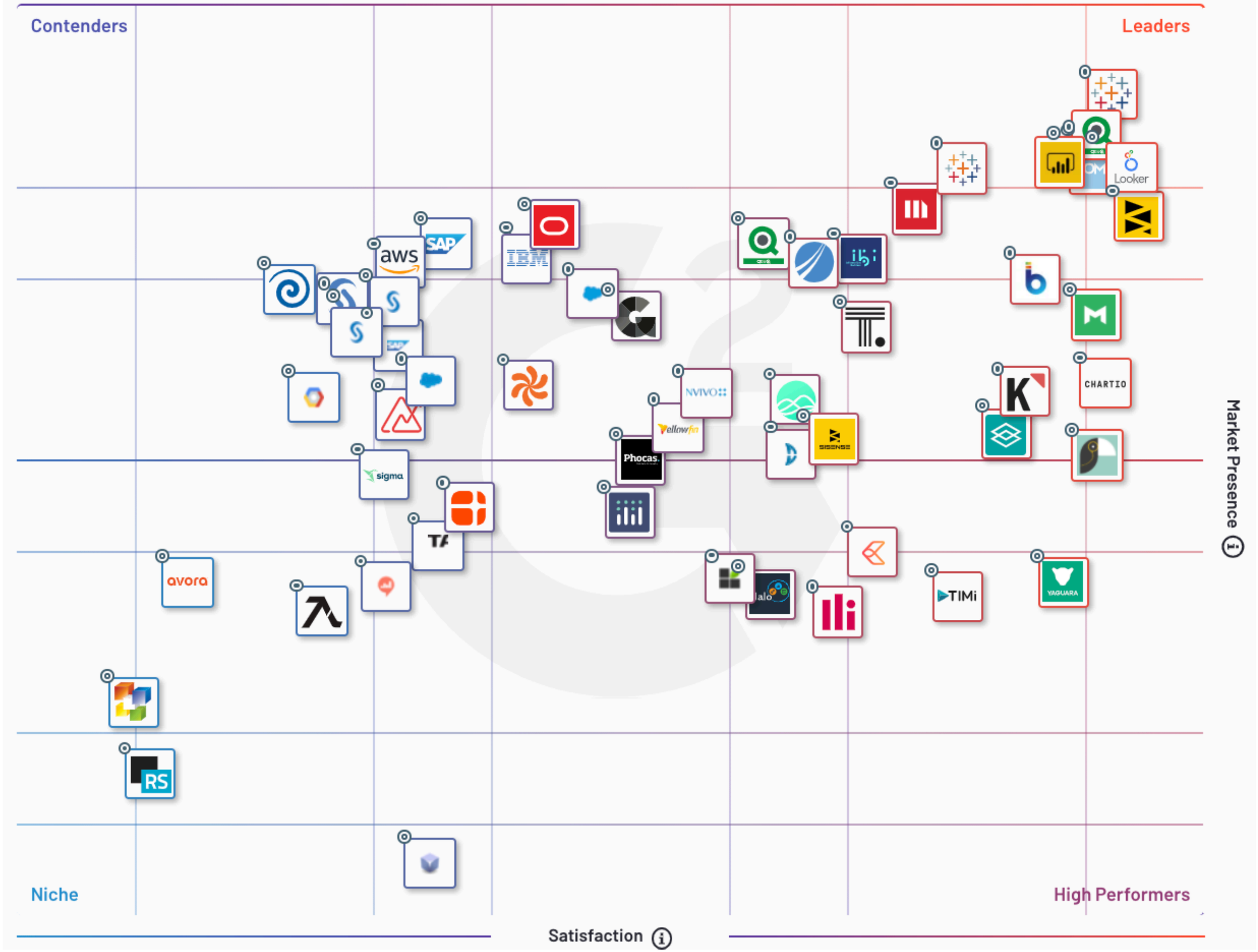


Source: Gartner (February 2020)

Magic Quadrant for Data Science and Machine Learning Platforms (2020)

Source: Gartner (February 2020)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA



https://www.g2.com/categories/analytics-platforms#grid

New Features & Trends

Hadoop & Spark integration

Natural-language generation

Catalog of internal and external data

Analytics integration with process, app or IoT

Machine-Learning Integration

Graph Discovery and investigation

Natural-language processing or voice

Systems of Insight

Evolution of BI Tools

Augmented analytics

Augmented analytics is the next wave of disruption in the data and analytics market. It uses machine learning (ML) and AI techniques to transform how analytics content is developed, consumed and shared.

Continuous intelligence

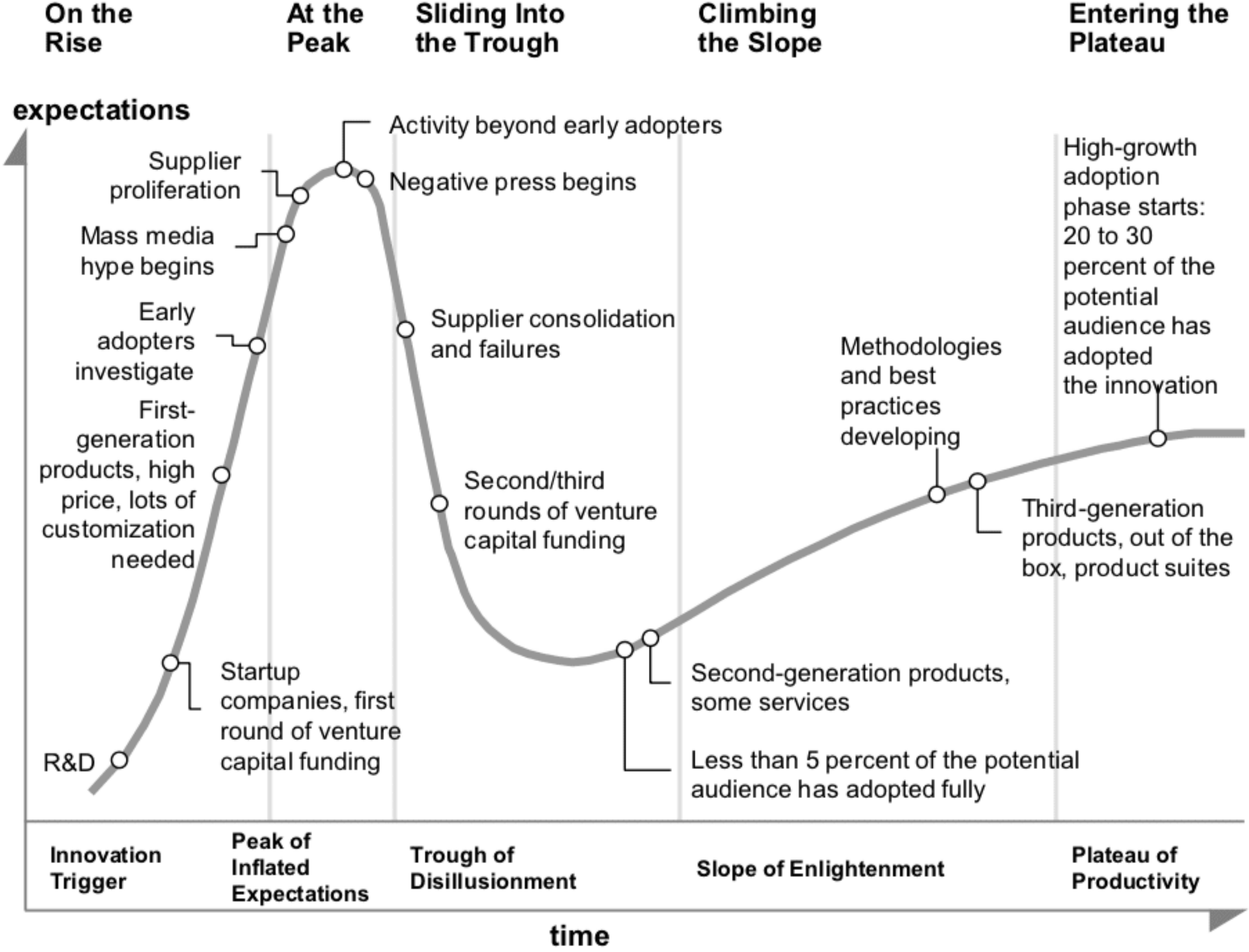
Continuous intelligence is a design pattern in which real-time analytics are integrated within a business operation, processing current and historical data to prescribe actions in response to events.

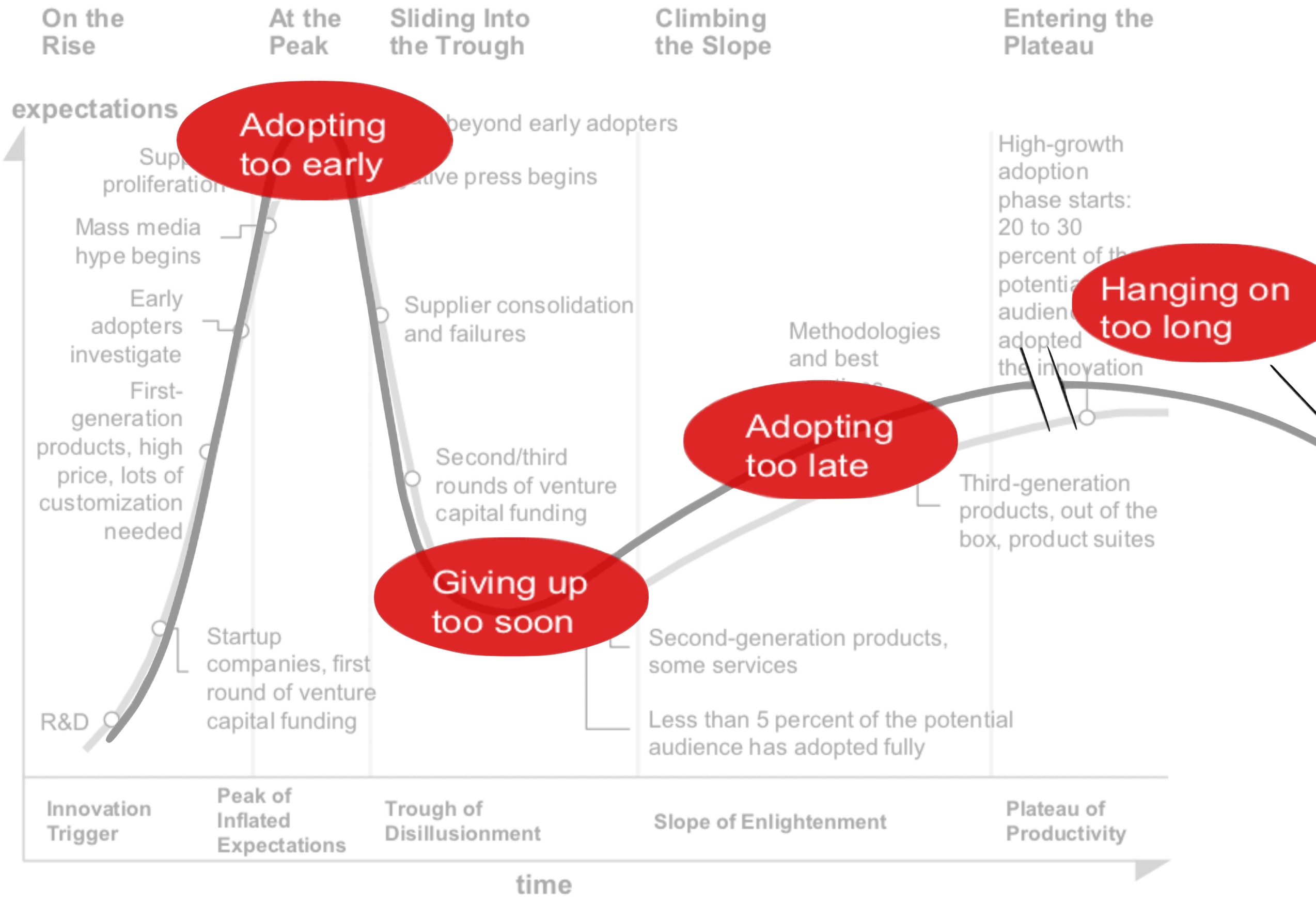
Explainable AI

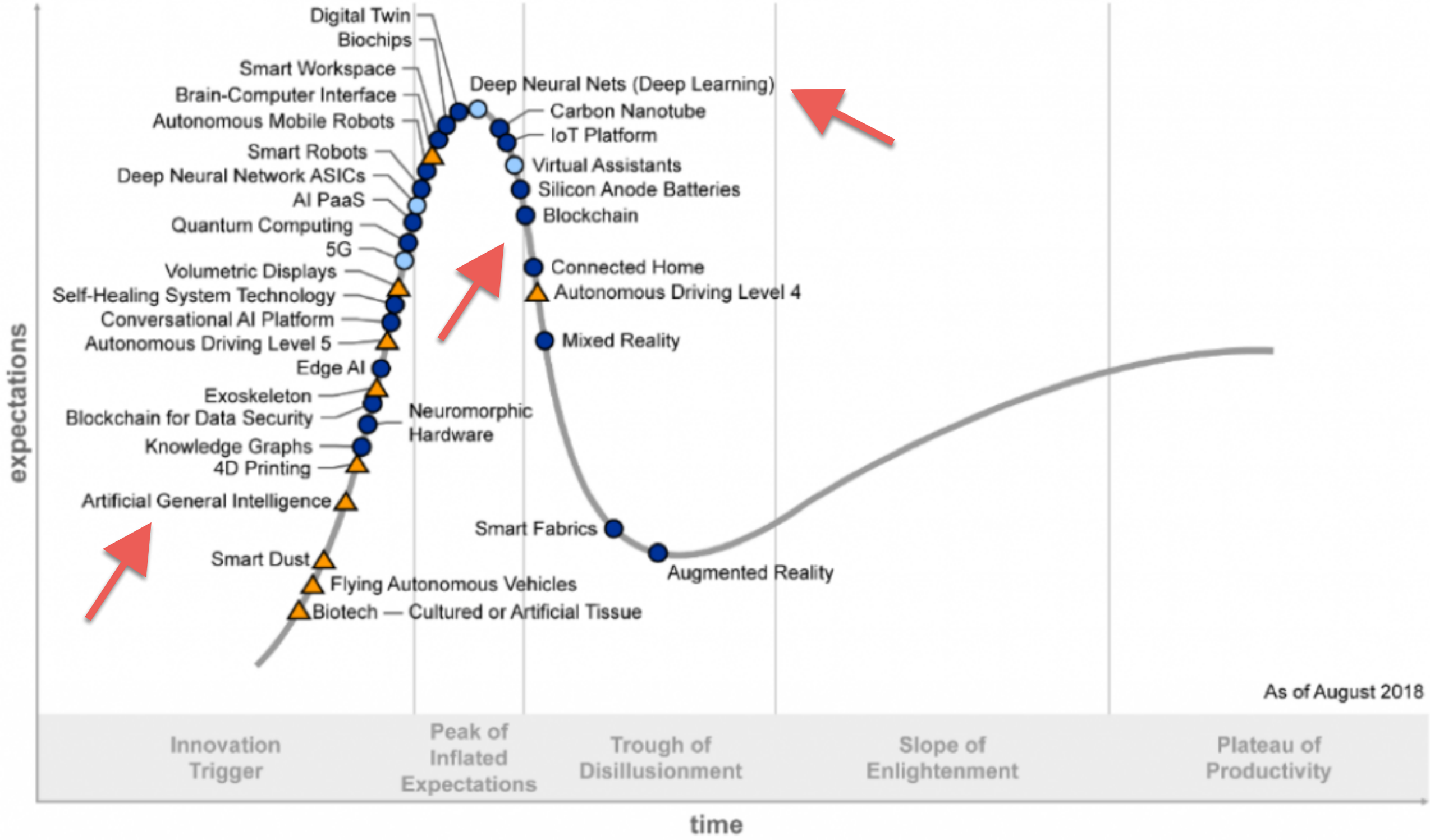
Explainable AI in data science and ML platforms, auto-generates an explanation of models in terms of accuracy, attributes, model statistics and features in natural language.

Blockchain

The core value proposition of blockchain, and distributed ledger technologies, is providing decentralized trust across a network of untrusted participants. The potential ramifications for analytics use cases are significant, especially those leveraging participant relationships and interactions.


On the Rise
At the Peak
Sliding Into the Trough
Climbing the Slope
Entering the Plateau
expectations
Activity beyond early adopters
Supplier proliferation
Negative press begins
Mass media hype begins
Early adopters investigate
Supplier consolidation and failures
First-generation products, high price, lots of customization needed
Methodologies and best practices developing
High-growth adoption phase starts: 20 to 30 percent of the potential audience has adopted the innovation
Second/third rounds of venture capital funding
Third-generation products, out of the box, product suites
Startup companies, first round of venture capital funding
Second-generation products, some services
R&D
Less than 5 percent of the potential audience has adopted fully
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time

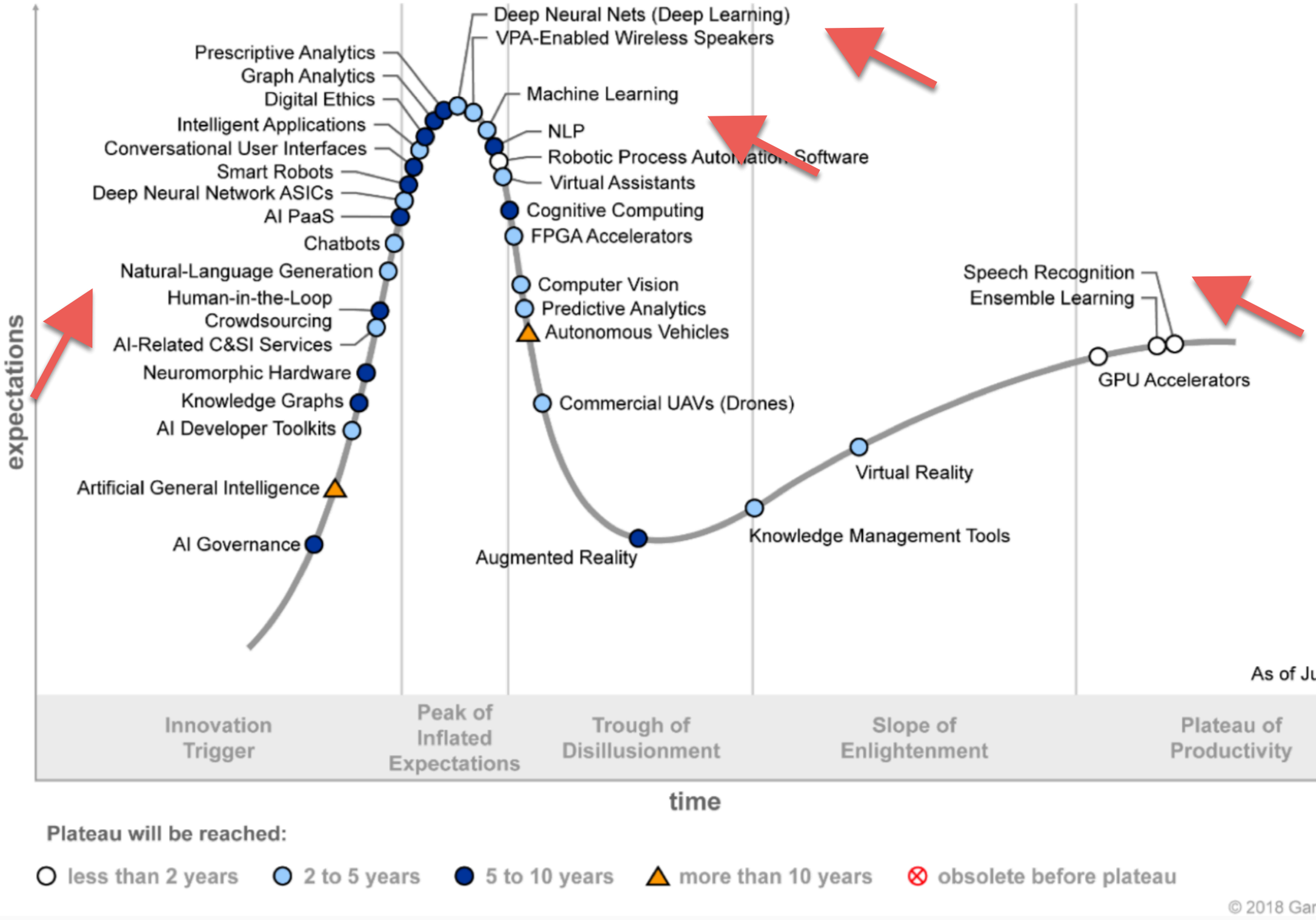

On the Rise
At the Peak
Sliding Into the Trough
Climbing the Slope
Entering the Plateau
expectations
Adopting too early
beyond early adopters
Supp
proliferation
ative press begins
Mass media hype begins
Early adopters investigate
Supplier consolidation and failures
First-generation products, high price, lots of customization needed
Second/third rounds of venture capital funding
High-growth adoption phase starts: 20 to 30 percent of th
potentia
audien
adopted
the innovation
Hanging on too long
Methodologies and best
Adopting too late
Third-generation products, out of the box, product suites
Giving up too soon
Startup companies, first round of venture capital funding
Second-generation products, some services
R&D
Less than 5 percent of the potential audience has adopted fully
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time



Plateau will be reached:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ⊗ obsolete before plateau

Figure 1. Hype Cycle for Artificial Intelligence, 2018



Source: Gartner (July 2018)

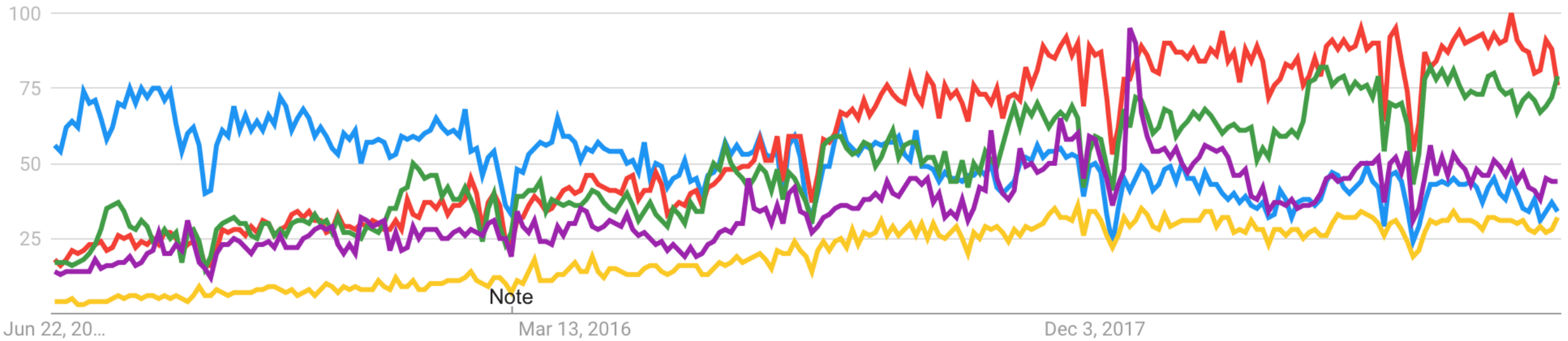
Priority Matrix for Artificial Intelligence, 2018

benefit	years to mainstream adoption: less than 2 years	2 to 5 years	5 to 10 years	more than 10 years
transformational	**Speech Recognition**	**AI-Related C&SI Services** **Chatbots** **Deep Neural Nets (Deep Learning)** **Intelligent Applications** **Machine Learning** **Virtual Assistants** **VPA-Enabled Wireless Speakers**	Cognitive Computing Conversational User Interfaces Neuromorphic Hardware NLP	Artificial General Intelligence Autonomous Vehicles
high	**Ensemble Learning** **GPU Accelerators** **Robotic Process Automation Software**	AI Developer Toolkits Commercial UAVs (Drones) Computer Vision Deep Neural Network ASICs Natural-Language Generation Predictive Analytics	AI Governance AI PaaS Augmented Reality Digital Ethics Graph Analytics Human-in-the-Loop Crowdsourcing Knowledge Graphs Prescriptive Analytics Smart Robots	
moderate		FPGA Accelerators Knowledge Management Tools Virtual Reality		
low				

As of July 2018

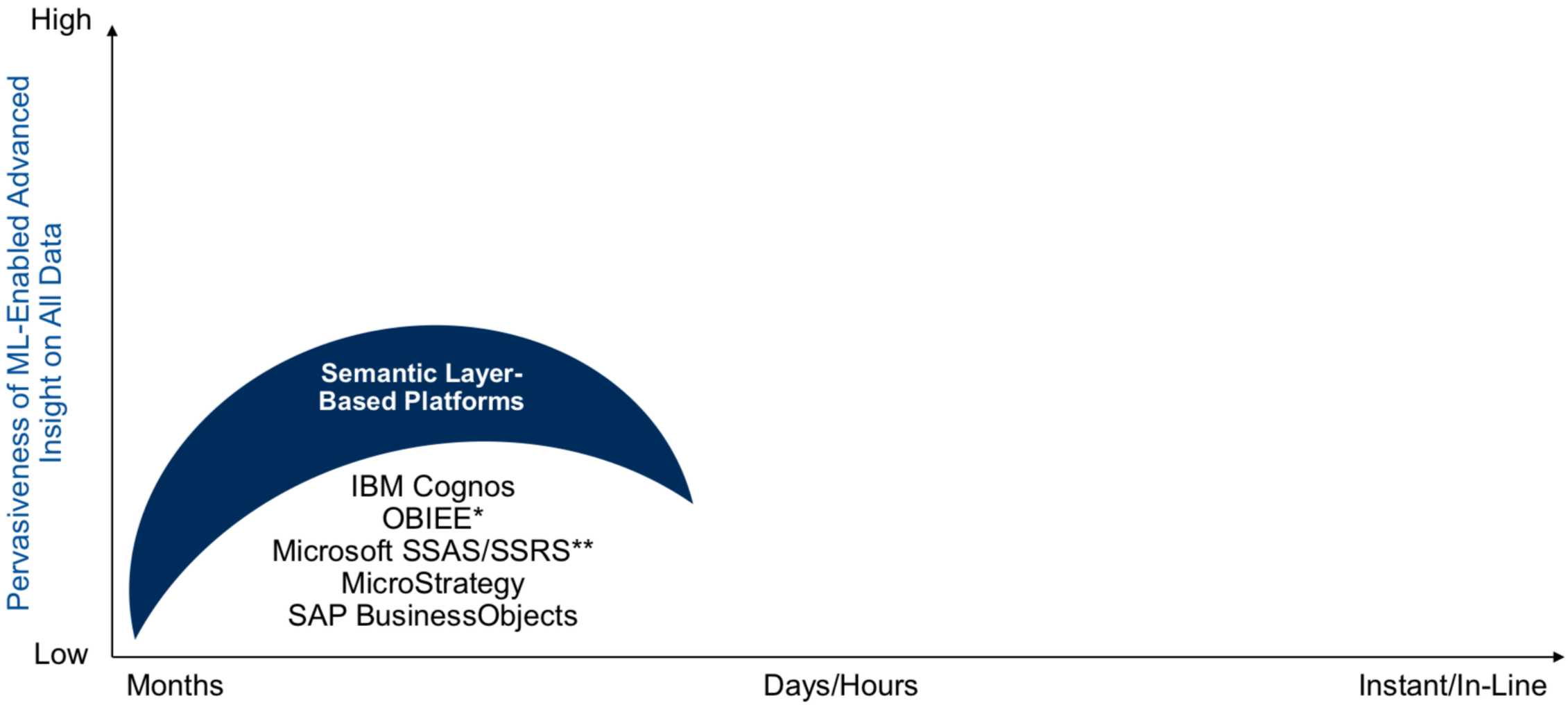
Trend

- big data
  Search term
- machine learning
  Search term
- deep learning
  Search term
- data science
  Search term
- artificial intelligence
  Search term



UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

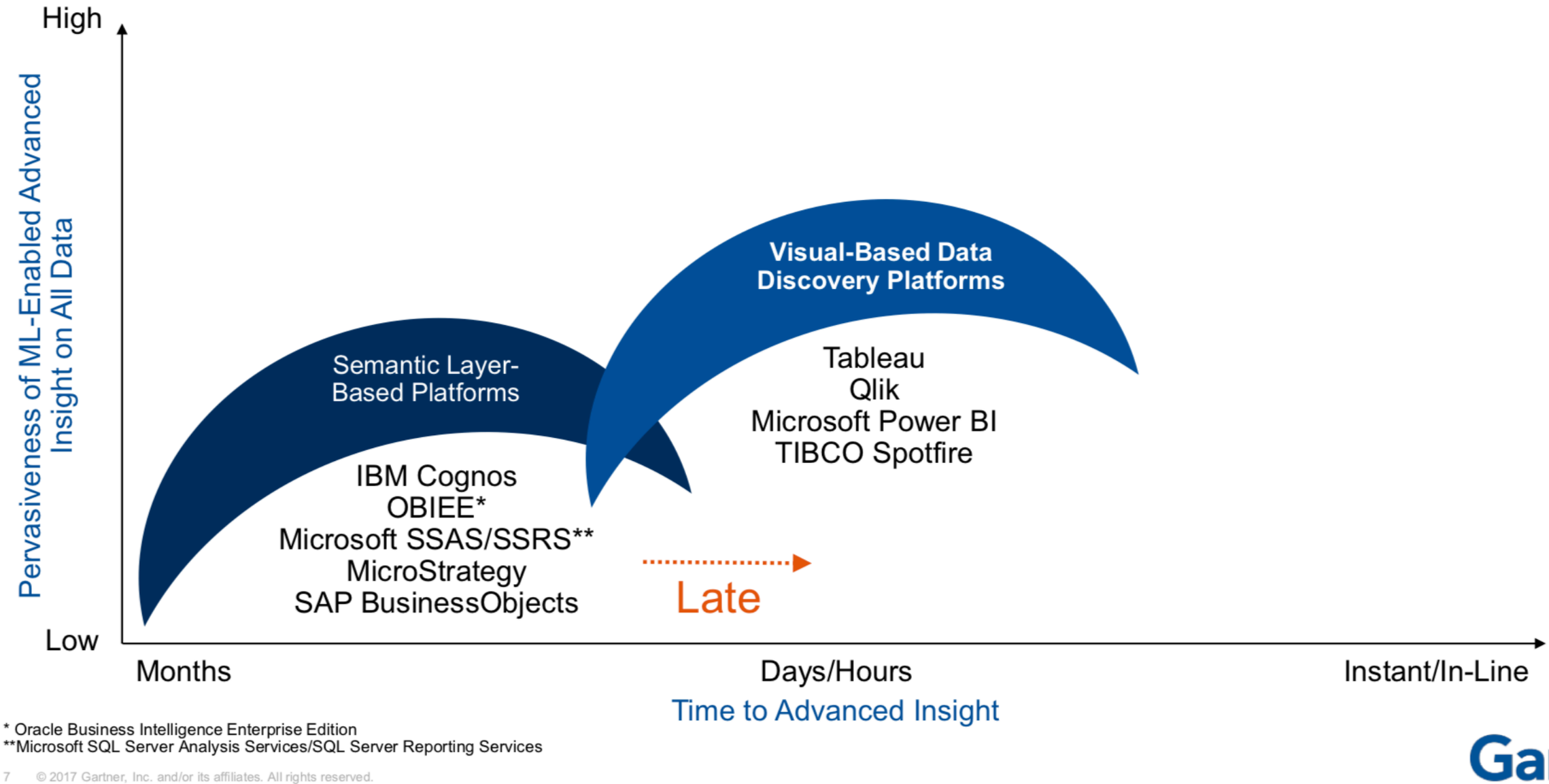
First Wave of Disruption



* Oracle Business Intelligence Enterprise Edition
**Microsoft SQL Server Analysis Services/SQL Server Reporting Services

Gartner

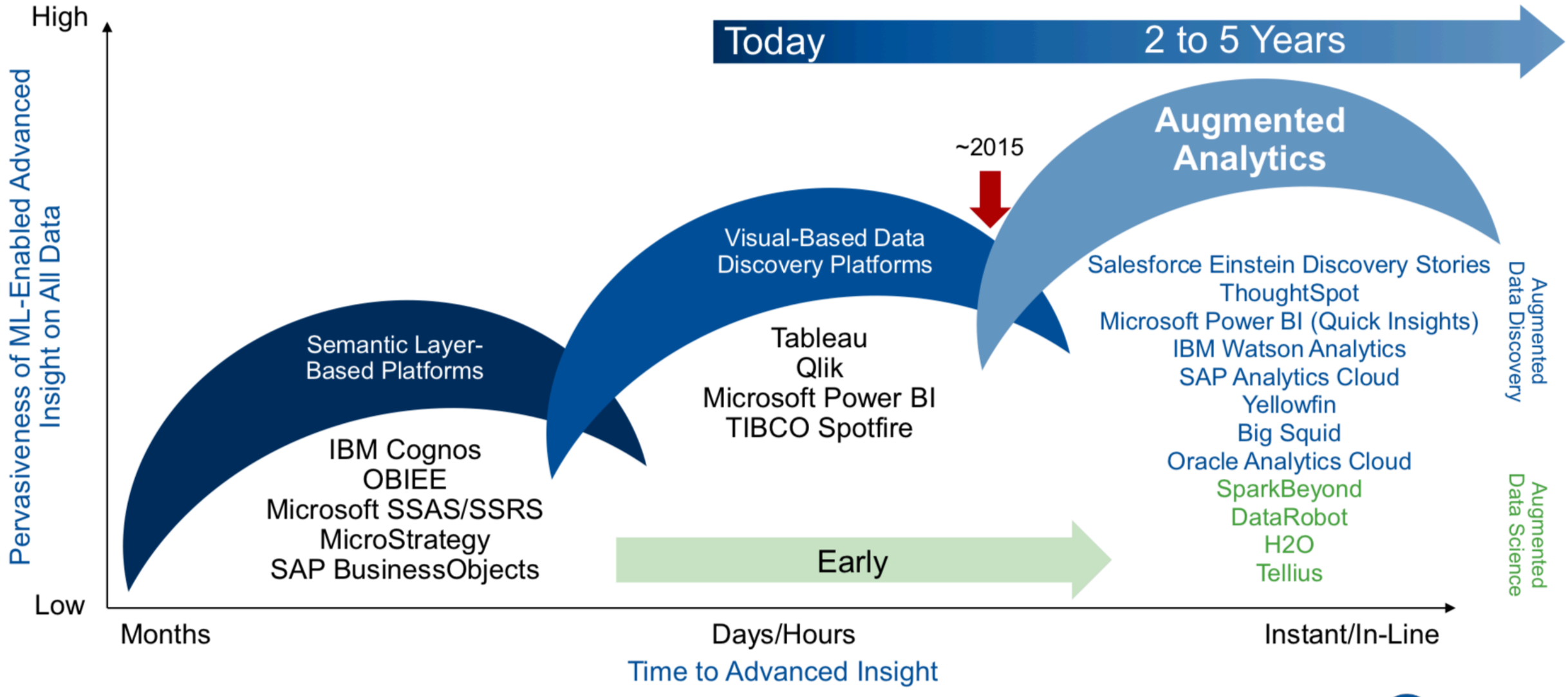
The Visual-Based Data Discovery Market Disruption



* Oracle Business Intelligence Enterprise Edition
**Microsoft SQL Server Analysis Services/SQL Server Reporting Services

Gartner

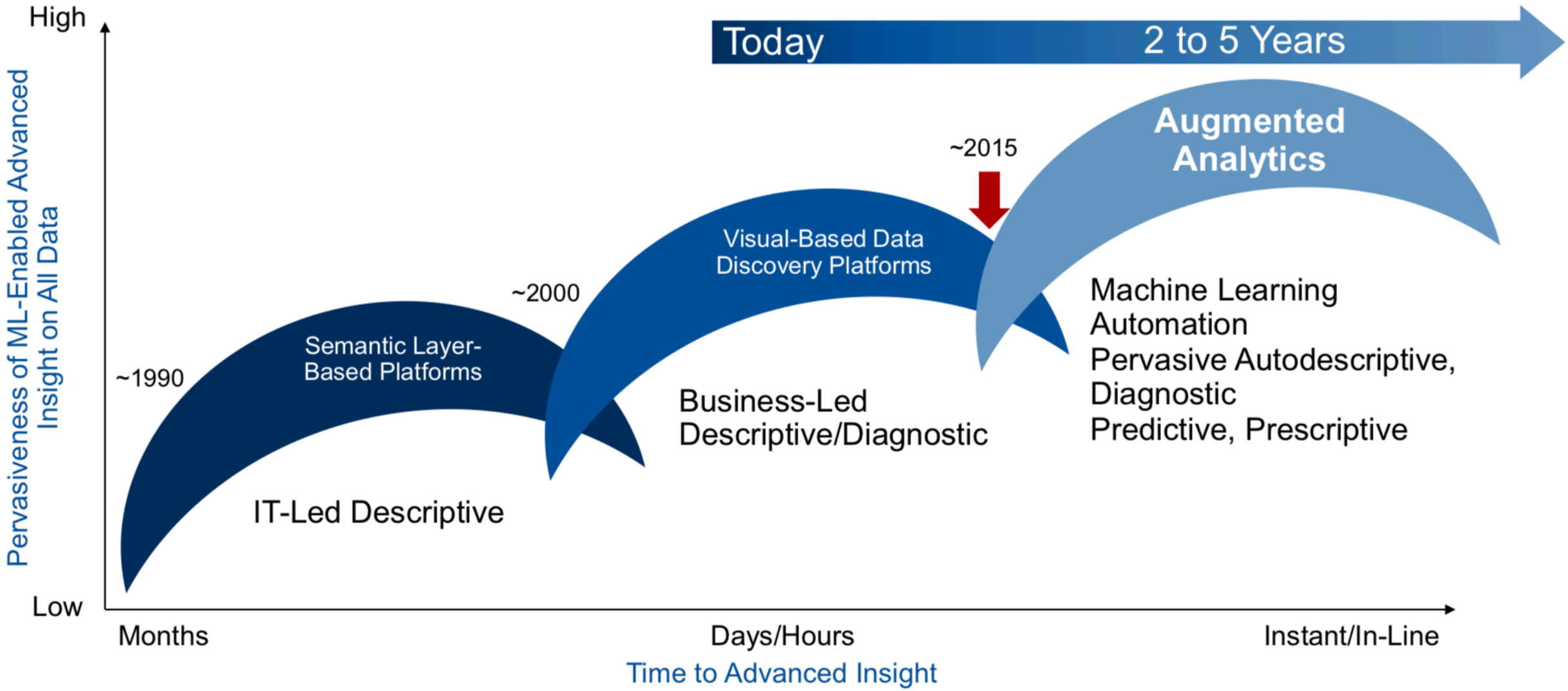
We Are at a Market Disruption Point



*Sample of vendors not exhaustive

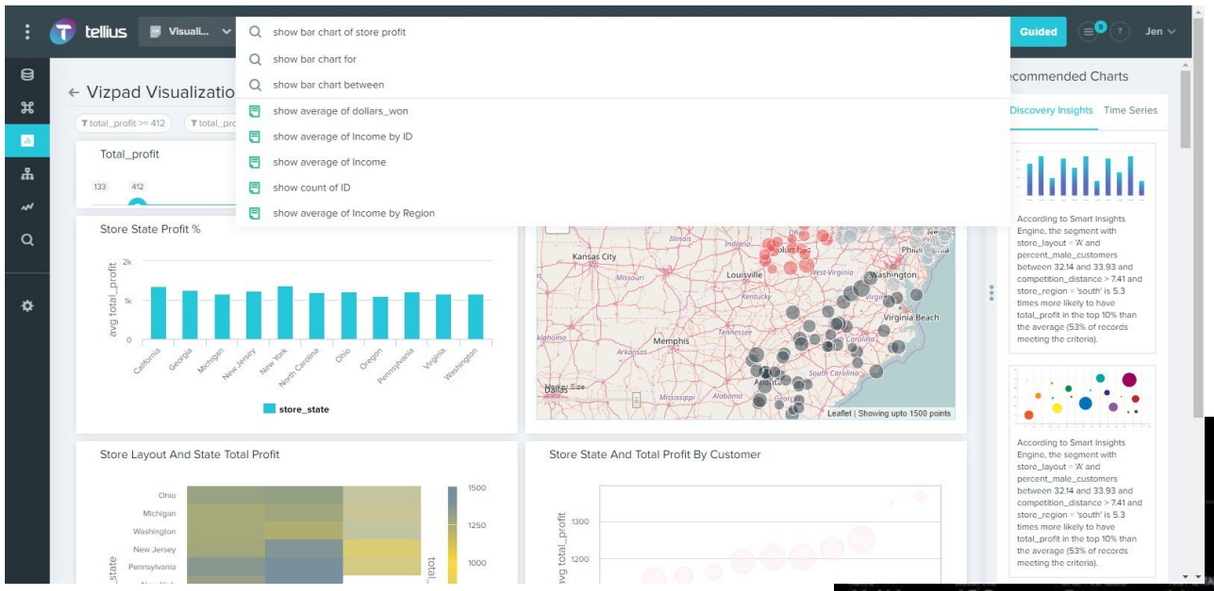
Gartner

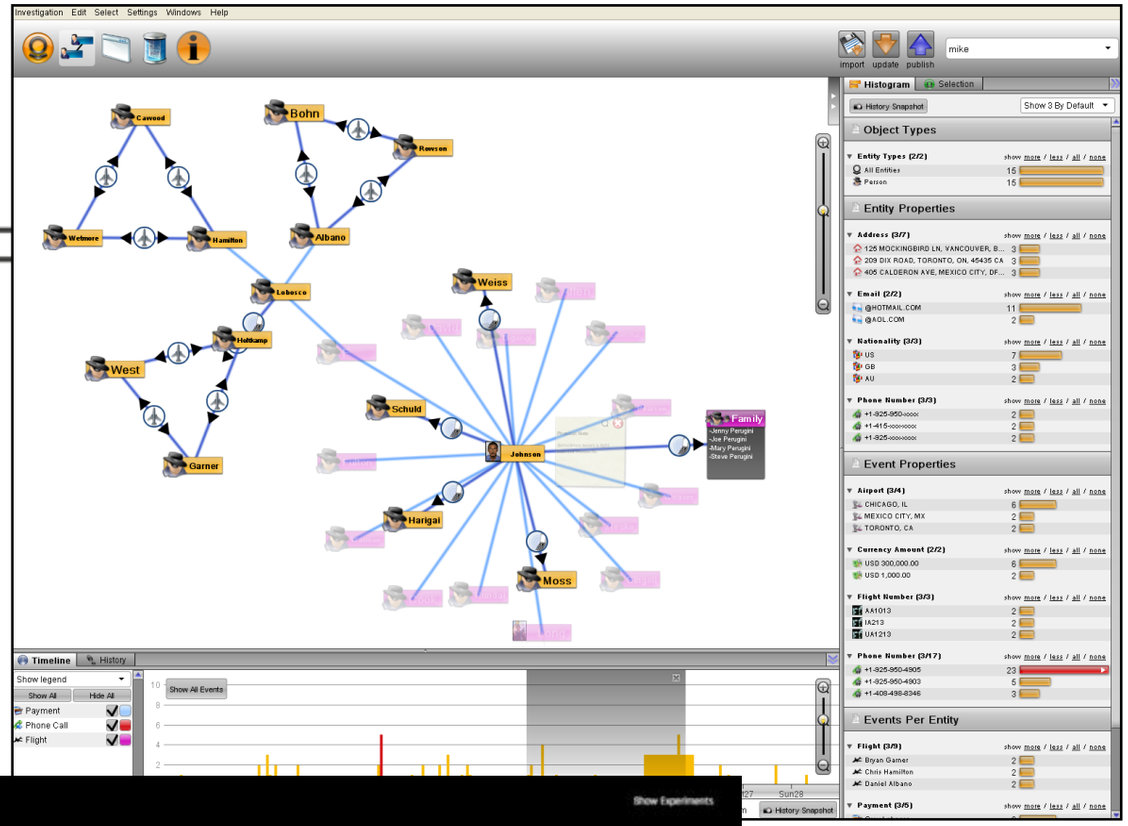
The Future of Data & Analytics Is Augmented Analytics



tellius

VISALLO



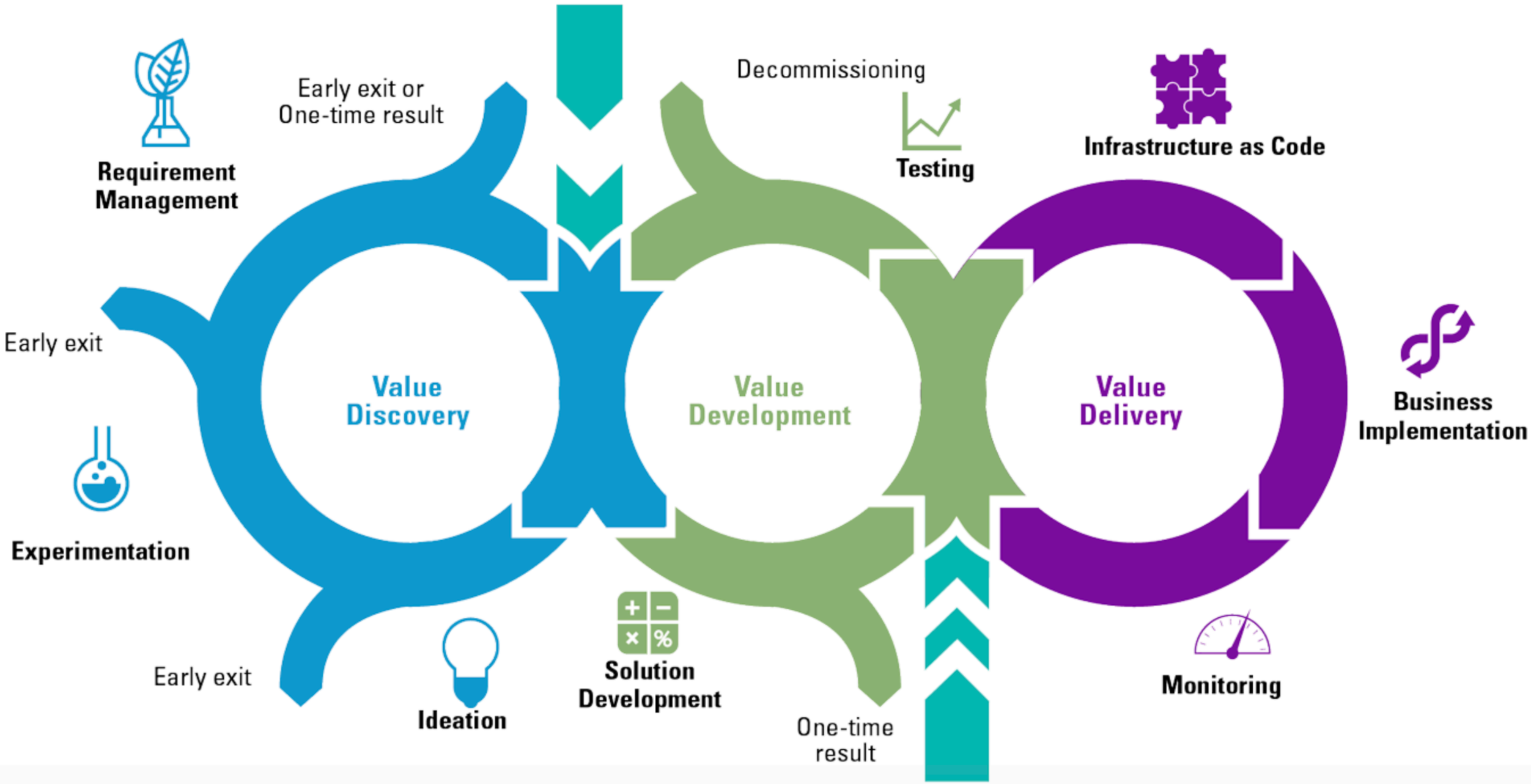


H2O.ai

Palantir

Cloud & Data Science

Modern software development

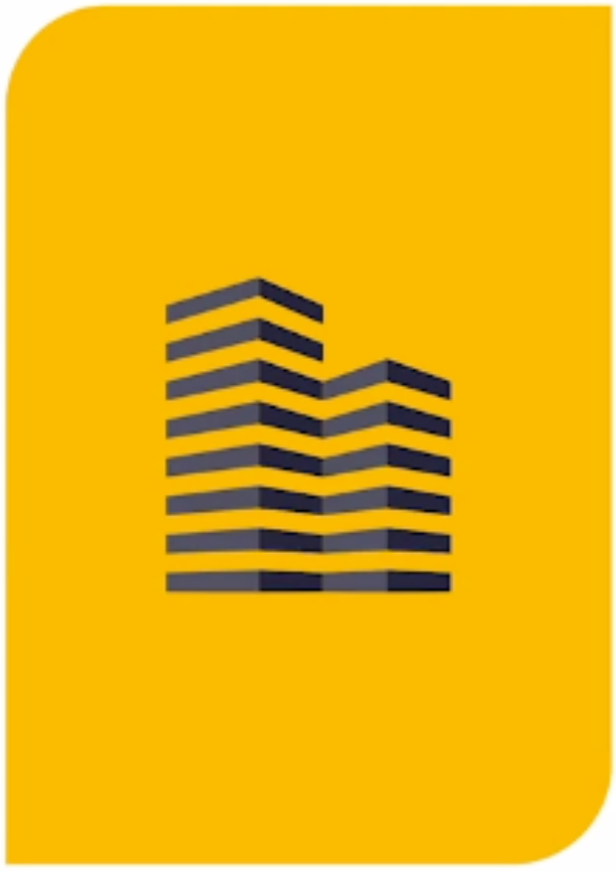


https://www.compact.nl/en/articles/modern-software-development/

It is all about quality and speed

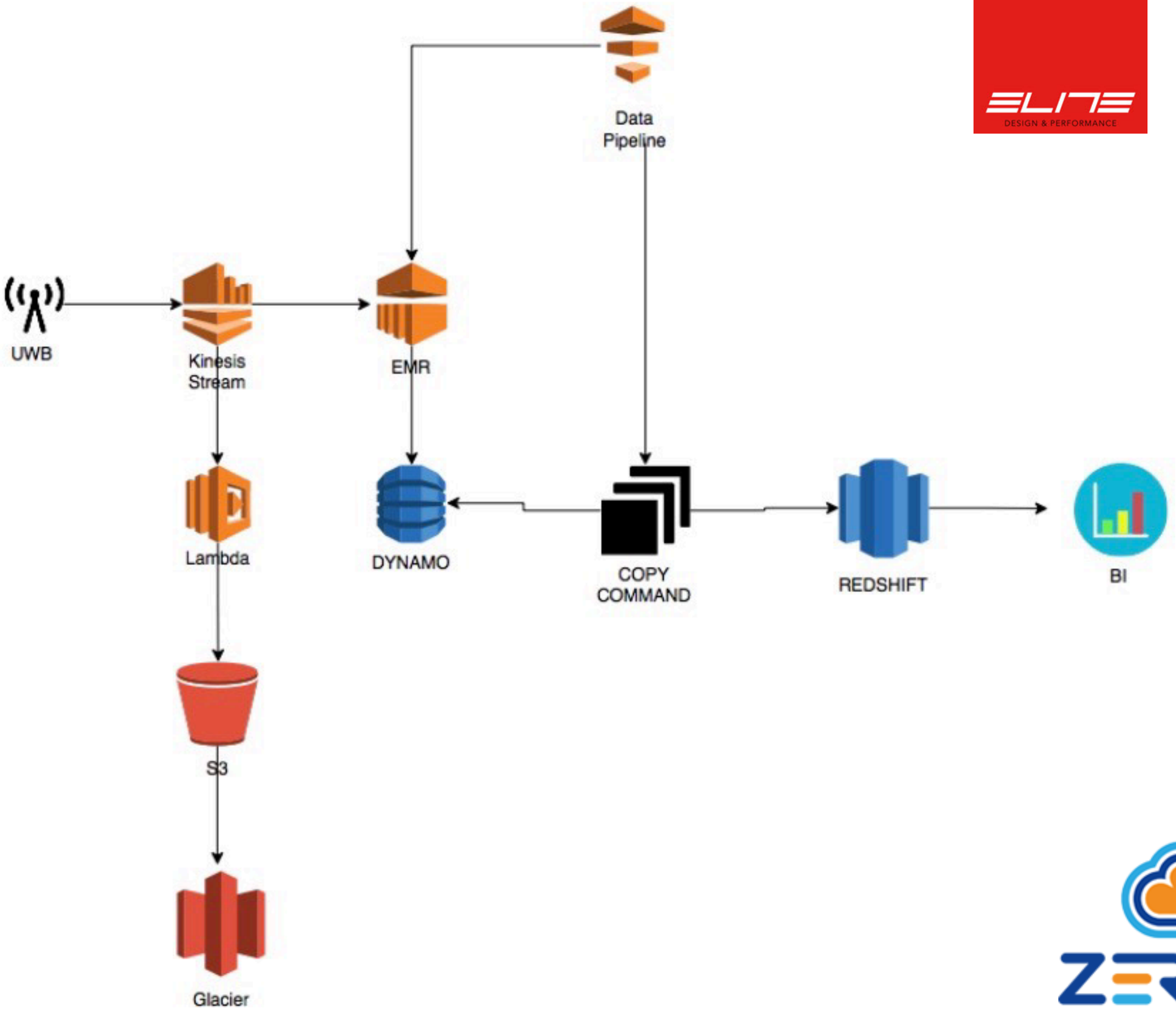
Old World

One size fits all



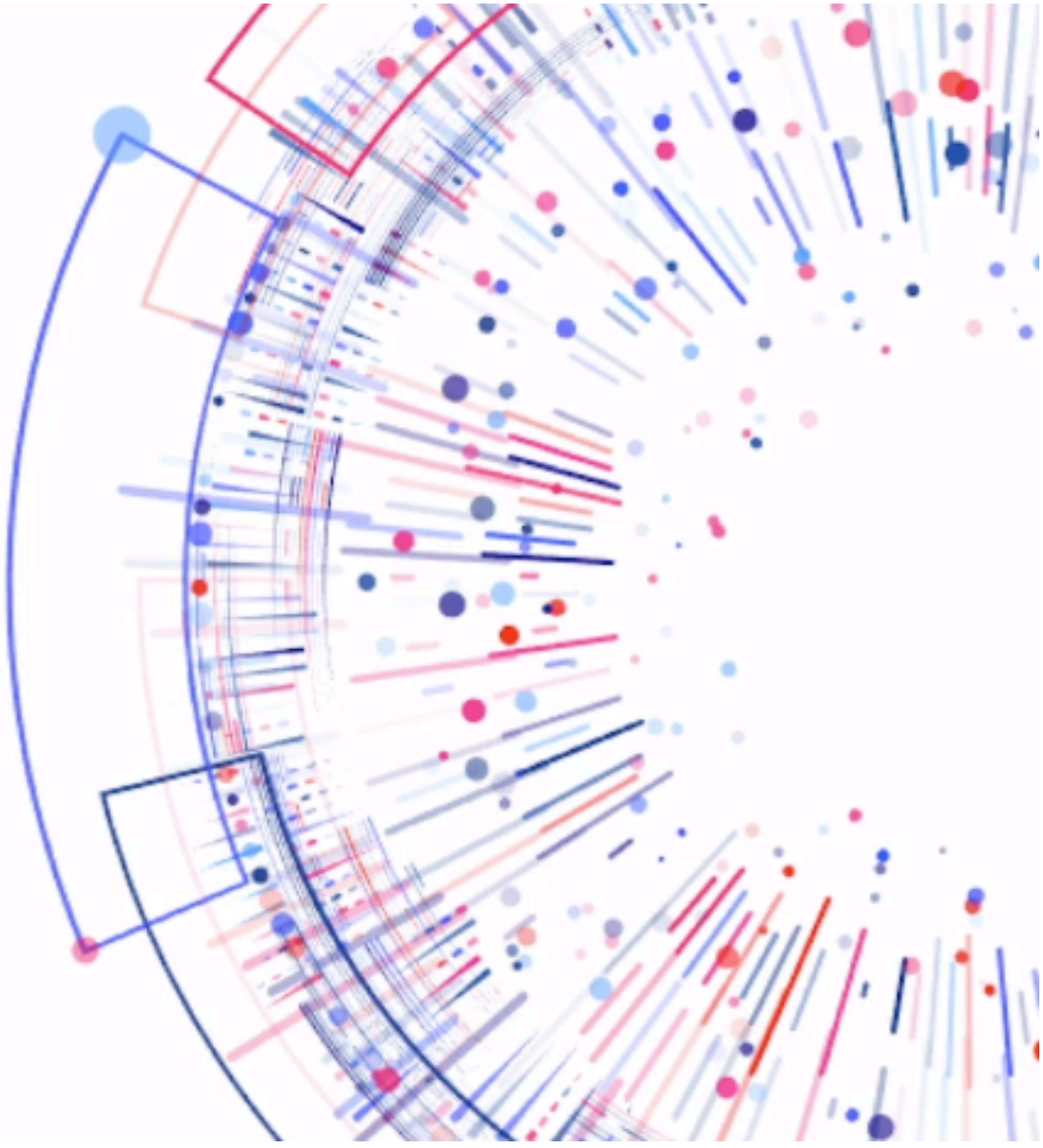
New World

Best of breed


Data
Pipeline
UWB
Kinesis
Stream
EMR
Lambda
DYNAMO
COPY
COMMAND
REDSHIFT
BI
S3
Glacier

Critical drivers for modern architecture

- Data volume
- Data variety
- Use case complexity
- Data access
- Breaking down silos



Cloud & Data Science

https://github.com/mauropelucchi/mauropelucchi-unimib_business_intelligence_cloud_workshop_2020

https://community.cloud.databricks.com/login.html

Cloud computing

Computing paradigms

Mainframe computing
1 computer / many users

"The network is the computer"
(John Gage, Sun Microsystems, 1984)

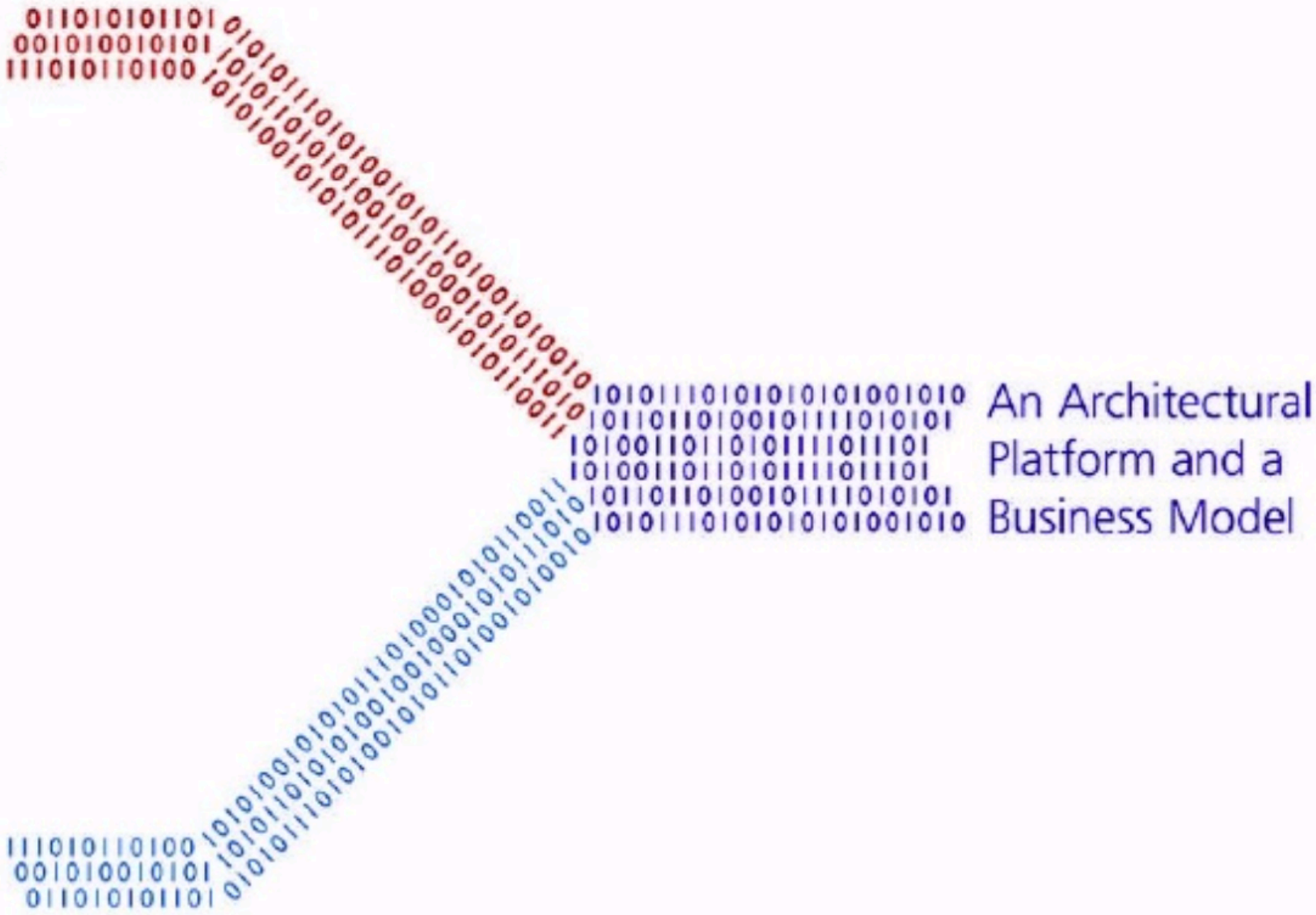
Client-server computing
Many computers / many users

"The cloud is the computer"

Cloud computing

A Business Model

- Pay as you go
- Virtualized resources
- On-demand
- Multi-tenancy
- Scalability, etc.



An Architectural Platform and a Business Model

A Delivery Platform

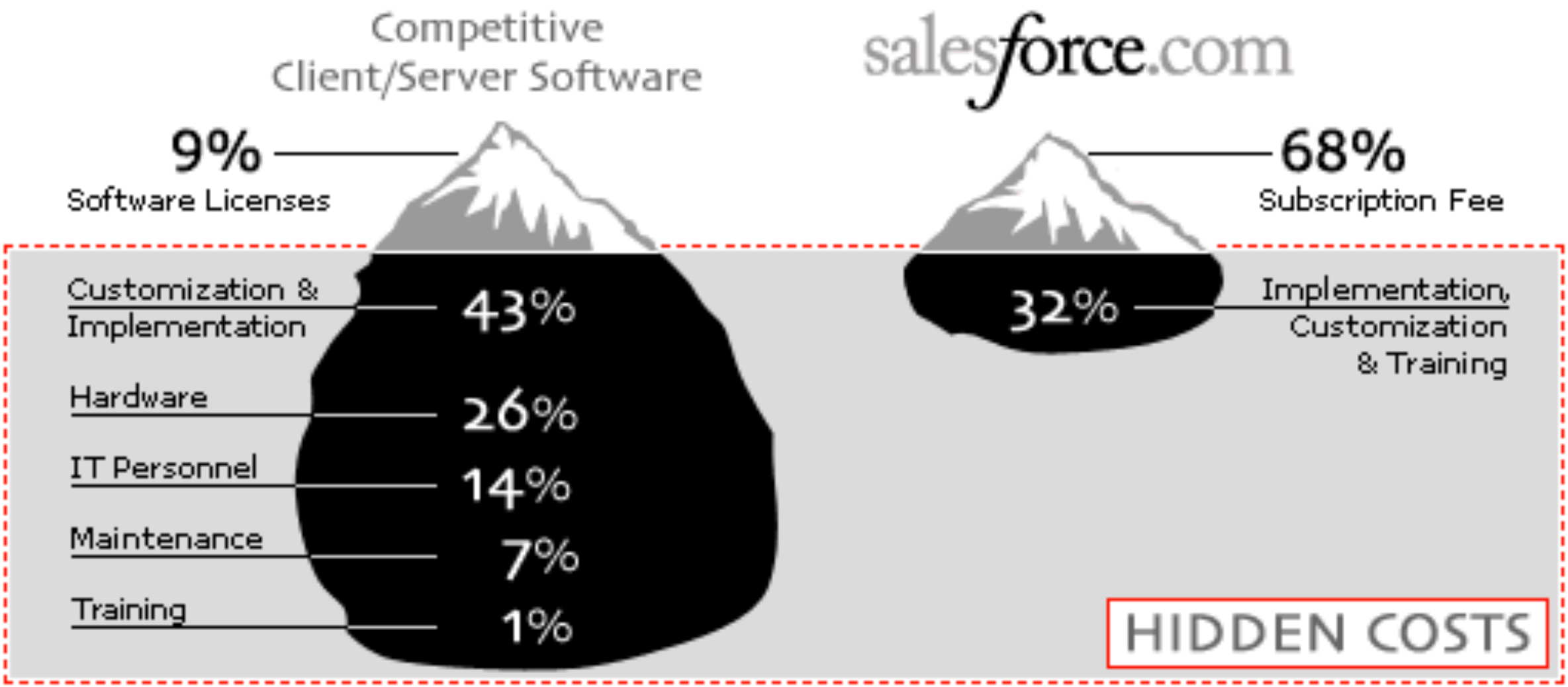
Enable users to access IP services and on-demand applications

Yankee Group, 2010

A style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet.

[Wikipedia]

Avoid the hidden costs of traditional CRM software



Benefits and Challenges

Pay per Use	Infinity capacity	Security (Compliance)	Service Assurance
Scalability	Services	Trasparency	Integration With IT
Complexity	Continue Innovation	Portability	Licenses
Minimize time to market	Updated	Costs	

Service Models

SaaS

Software as a Service

The client uses applications on an infrastructure accessible from various client devices through an interface (API, web interface, dedicated client)

PaaS

Platform as a Service

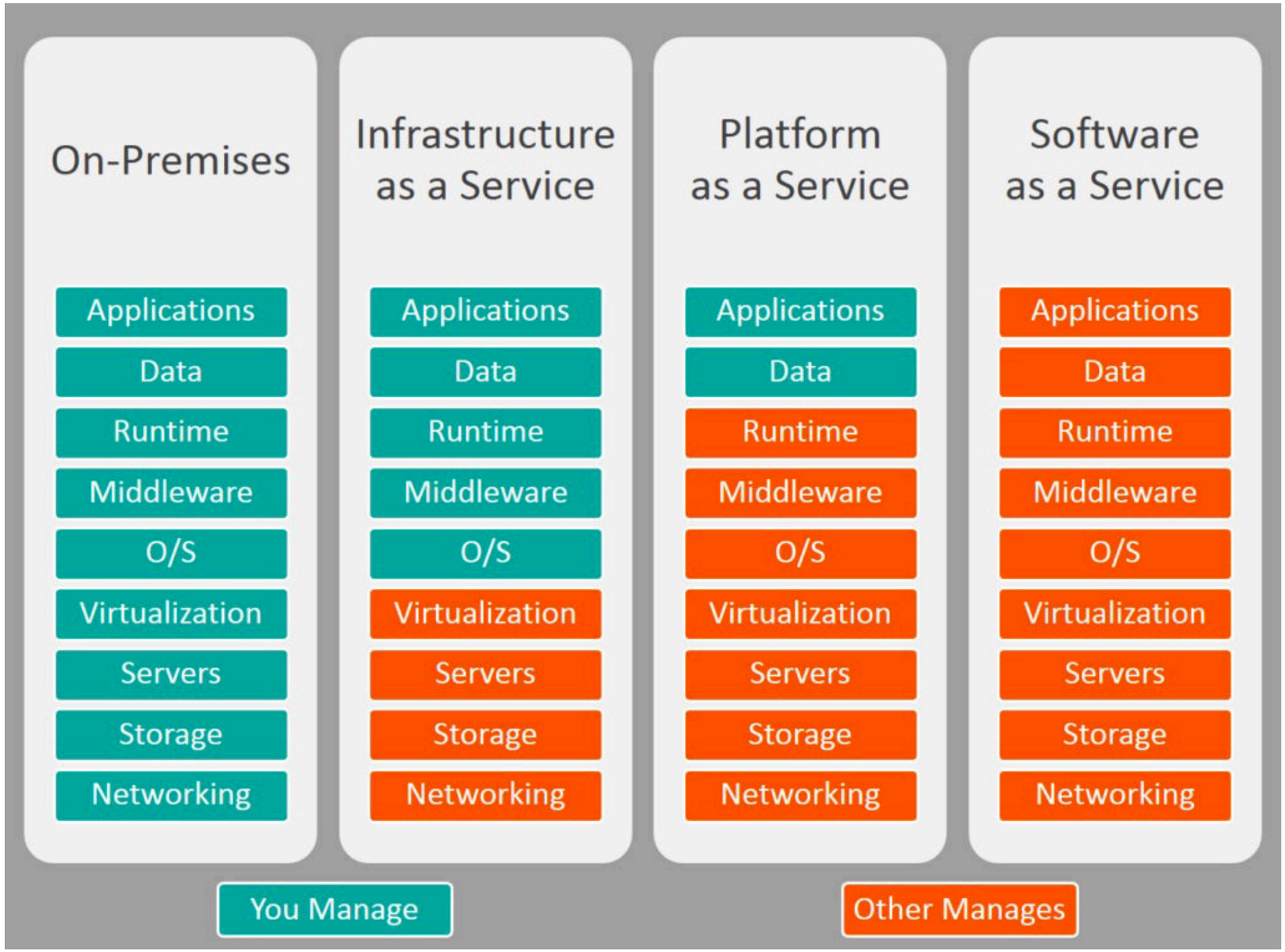
Allows you to develop and distribute applications created using vendor supported programming languages

IaaS

Infrastructure as a Service

Rental capacity of CPU, storage, network and other resources such as operating systems and applications

Service Models



Serverless computing
(also known as *functions as a service*)
is a new cloud computing abstraction that makes it easier to write robust, large-scale web services.

In serverless computing, programmers write what are called serverless functions, which are programs that respond to external events.

Serverless means:

Greater agility

Less overhead

Better focus

Increased scale

More flexibility

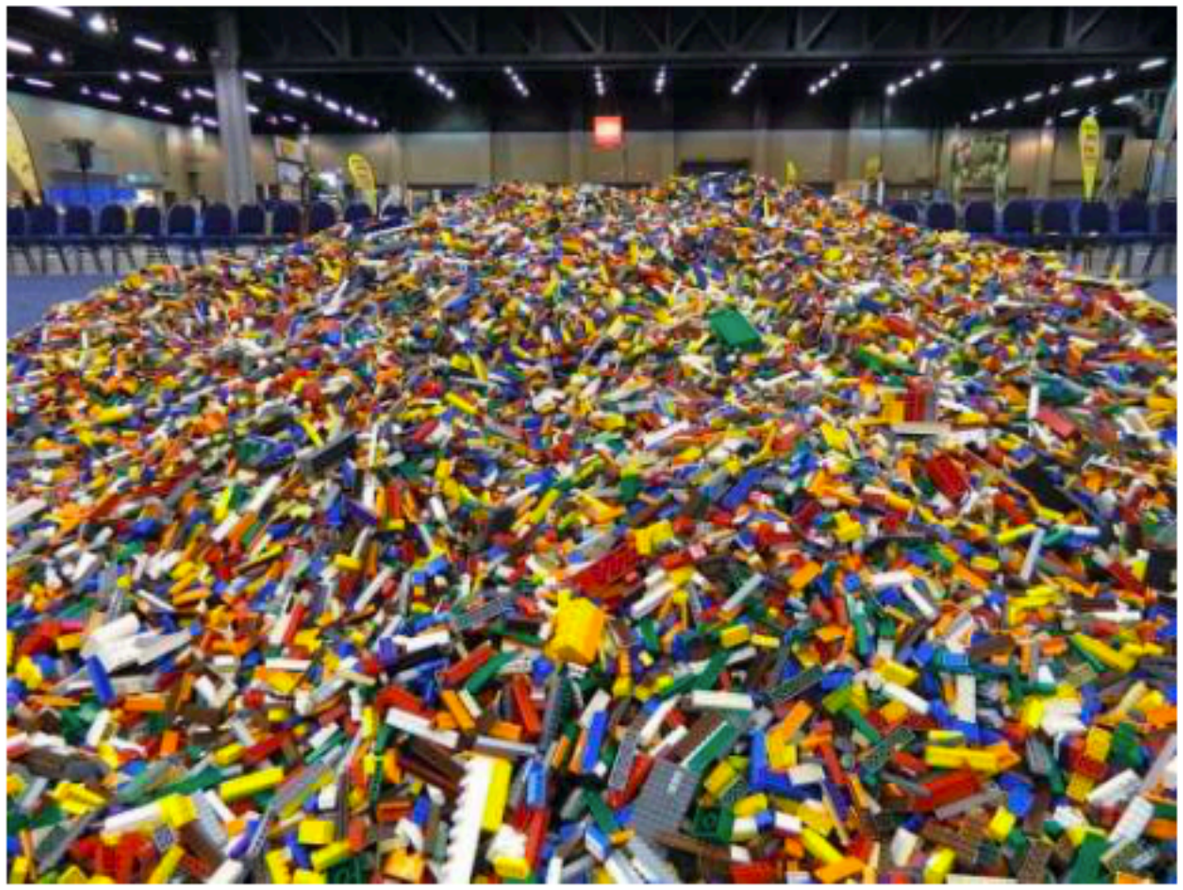
Faster time to market

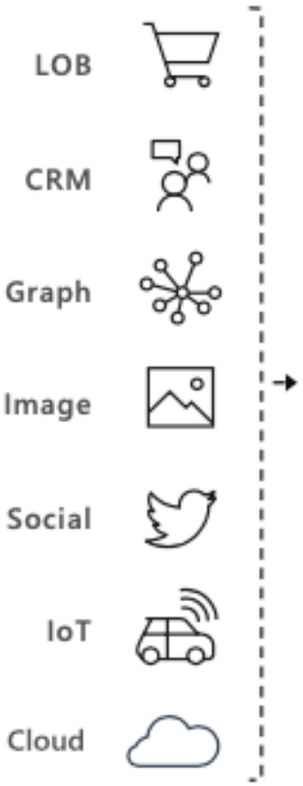
Periodic Table of Amazon Web Services

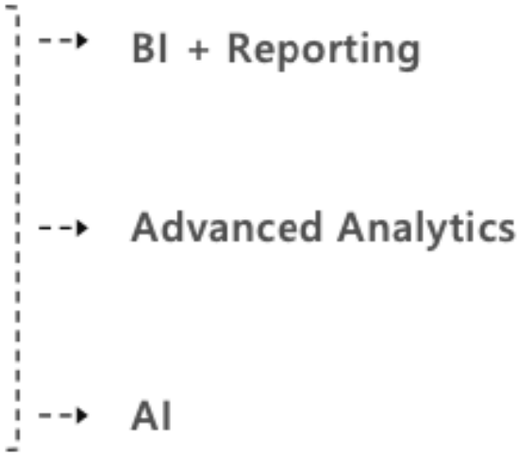
by @awsgeek

- Analytics
- Application Integration
- AR & VR
- AWS Cost Management
- Blockchain
- Business Applications
- Compute
- Customer Engagement
- Database
- Developer Tools
- End User Computing
- Game Tech
- Internet of Things
- Machine Learning
- Media Services
- Migration & Transfer
- Mobile
- Managment & Governanace
- Network & Content Delivery
- Robotics
- Satellite
- Storage
- Security, Identity & Compliance

amazon A athena																	amazon S3 simple storage service	aws Sg storage gateway
amazon Cs cloudsearch	aws Gs ground station													aws Wa well-architected tool	aws Lm license manager	amazon Et elastic transcoder	aws Mh migration hub	aws Am amplify	amazon V virtual private cloud	amazon Bs elastic block store
amazon Mr elastic map reduce	aws G glue													amazon Cw cloudwatch	aws Mc management console	amazon Kv kinesis media streams	aws Ad application discovery service	amazon Ag api gateway	amazon Cr cloudfront	amazon Fs elastic file system
amazon Es elasticsearch	aws Lf lake formation	aws Ce cost explorer	amazon Ax alexa for business	amazon Cs elastic container service	aws L lambda	amazon Co connect	aws Cb codebuild	amazon Au aurora	amazon Ds relational database service	aws I iot core	aws Dd iot device defender	aws Ac auto scaling	aws Ms managed services	aws Mc elemental mediaconnect	aws Dm database migration service	amazon Po pinpoint	amazon 53 route 53	amazon Fx fsx for lustre
amazon K kinesis	aws Sf step functions	aws Bu budgets	amazon Ch chime	amazon Ks elastic container service for kubernetes	aws O outposts	amazon Pi pinpoint	aws Cc codecommit	amazon D dynamodb	amazon Rs redshift	aws Os iot freertos	aws Dm iot device management	aws Cf cloudformation	aws Ow opsworks	aws Mv elemental mediaconvert	aws Ds datasync	aws As appsync	amazon Pl vpc privatelink	amazon Fw fsx for windows file server
amazon Ka managed streaming for kafka	amazon Mq activemq	aws Cu cost & usage report	amazon Wm workmail	amazon Ls lightsail	aws Sa serverless application repository	amazon E simple email service	aws Cd codedeploy	amazon Do documentdb	amazon Ti timestream	aws Gg iot greengrass	aws Ev iot events	aws Ct cloudtrail	aws Ph personal health dashboard	aws Ml elemental medialive	aws Sm server migration service	aws Df device farm	aws Am App Mesh	amazon Gl s3 glacier
amazon R redshift	amazon N simple notification service	aws Ri reserved instance reporting	amazon C elastic cloud compute	aws Ba batch	aws Lb elastic load balancing	aws Cs codestar	aws Cp codepipeline	amazon Ec elasticache	aws Ms database migration service	aws 1c iot 1-click	aws Sw iot sitewise	aws Co config	aws Sc service catalog	aws Mp elemental mediapackage	aws Sn snow family	aws Ga global accelerator	aws Cm cloud map	aws Ba backup
amazon Qs quicksight	amazon Q simple queue service	amazon Bc managed blockchain	amazon Au ec2 auto scaling	aws Eb elastic beanstalk	aws Vm vmware cloud on aws	amazon Co correto	aws Cli command line interface	amazon Ne neptune	amazon Gl gamelift	aws An iot analytics	aws Tg iot things graph	aws Ct control tower	aws Sm systems manager	aws Ms elemental mediastore	aws Sf transfer for sftp	aws Tg transit gateway	aws Dc direct connect	aws Sn snow family
aws Dp data pipeline	aws Ap appsync	amazon Su sumerian	amazon Cr elastic container registry	aws F fargate	aws Rm robomaker	aws C9 cloud9	aws Sd tools & sdks	amazon Ql quantum ledger db	amazon Ly lumberyard	aws Bu iot button	aws Dc iot partner device catalog	aws Cm console mobile app	aws Ta trusted advisor	aws Mt elemental mediatailor	amazon Wd workdocs	amazon Wo worklink	amazon Ws workspaces	amazon As appstream 2.0
	amazon Sm sagemaker	amazon Co comprehend	amazon Ei elastic inference	amazon Fc forecast	amazon Lx lex	amazon Pe personalize	amazon Po polly	amazon Rk rekognition	amazon Gt sagemaker ground truth	amazon Tx textract	amazon Tr translate	amazon Ta transcribe	aws Dl deep learning amis	aws Dl deeplens	aws Dr deepracer	aws If inferentia	aws Mx apache mxnet on aws	aws Tf tensorflow on aws
	aws Im identity & access management	amazon Cd cloud directory	amazon Co cognito	amazon Gd guardduty	amazon I inspector	amazon M macie	amazon Ar artifact	aws Cm certificate manager	aws Hs cloudhsm	aws Ds directory service	aws Fm firewall manager	aws Km key management service	aws O organizations	aws Sm secrets manager	aws Sh security hub	aws S shield	aws Ss single sign-on	aws Wf web application firewall







INGEST

Data orchestration and monitoring

- Azure Data Factory
- SSIS

STORE

Big data store

- Azure Data Lake Storage Gen2
- Blob Storage
- Azure Data Lake Storage Gen1
- SQL Server 2019 Big Data Cluster

PREP

Transform & Clean

- Azure Databricks
- Azure HDInsight
- PolyBase & Stored Procedures
- Power BI Dataflow
- Azure Data Lake Analytics

MODEL & SERVE (& store)

Data warehouse

- Azure SQL Data Warehouse
- Azure Analysis Services
- SQL Database (Single, MI, HyperScale, Serverless)
- SQL Server in a VM
- Cosmos DB
- Power BI Aggregations

Benefits to BI in the Cloud

- Flexibility to scale computing resources with few barriers
- Ability to shorten BI implementation windows
- Reduced cost for BI programs
- Ability to add environments for testing, proof-of-concepts and upgrades
- Geographic scalability

BI Team


Data Engineers
Business Analysts /
Domain Experts
Data Scientists

Business Analyst
Business Intelligence Analyst

Uses BI tools and applications to understand business conditions and drive business processes

Top Necessary skills by demand

Skill	Salary Premium	Job Postings Requesting (Last 12 Months)	Projected Growth (2 Years)
Project Management	✓	11,894	-5.8%
Business Process	✓	11,099	-7.9%
Microsoft Power BI		7,009	
Business Analysis		6,701	-4.8%
Python		5,784	+25.7%
Relational Databases		5,262	-0.5%
Data Visualization		5,218	+30.3%
Budgeting		4,735	-5.4%
Data Management		4,710	+8.1%
SAS	✓	4,638	+7.0%

Top Defining skills by demand

Skill	Salary Premium	Job Postings Requesting (Last 12 Months)	Projected Growth (2 Years)
SQL		21,297	-5.8%
SAP		20,010	+8.0%
Business Intelligence		16,418	-5.4%
Data Analysis		15,496	+7.4%
Tableau		12,334	+32.7%
Data Warehousing		6,602	-14.3%
Enterprise Resource Planning (ERP)		6,084	+2.0%
Oracle	✓	5,338	-9.7%

Collects and analyzes data that provides an accurate picture of business operations and performance for a company. Completes statistical analysis of current and historic business data, identifies trends and develops projections. Presents data analysis that informs planning and strategic decision making for a company.

Data Scientist

Uses advanced algorithms and interactive exploration tools to uncover non-obvious patterns in data

Top Necessary skills by demand

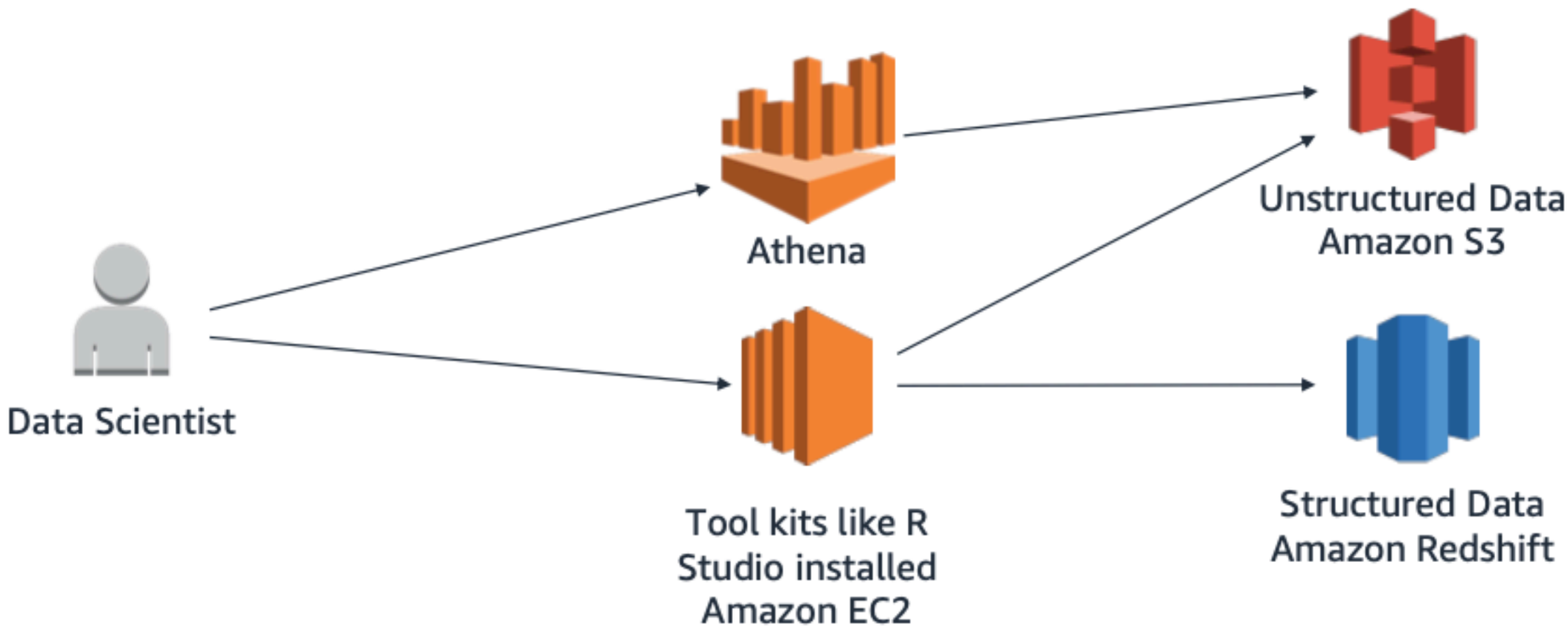
Skill	Salary Premium	Job Postings Requesting (Last 12 Months)	Projected Growth (2 Years)
Tableau		7,159	+28.2%
SAS	✓	6,798	+4.3%
Economics		6,063	+14.4%
Experiments		5,450	-12.7%
Statistics		5,378	-2.6%
Statistical Analysis		5,135	-5.3%
Clustering		4,863	+4.7%
Physics	✓	4,586	-14.3%
C++		3,699	-12.5%
Cluster Analysis		3,647	+0.2%

Top Defining skills by demand

Skill	Salary Premium	Job Postings Requesting (Last 12 Months)	Projected Growth (2 Years)
Data Science		36,102	+16.0%
Python		27,655	+21.6%
Machine Learning		24,954	+34.7%
SQL		18,880	+2.8%
Data Analysis		10,497	+1.3%
Big Data		9,732	+0.8%
Predictive Models	✓	9,356	+12.5%
Apache Hadoop		8,701	-1.6%
Artificial Intelligence		8,405	+24.8%
Data Mining		8,226	-6.7%

Utilizes skills and experience to systematically answer questions using data to provide actionable recommendations. Commonly utilizes advanced statistical analysis and machine learning techniques. Common responsibilities also include data cleaning and data management.

Data Scientist



Data Engineer

Designs, builds and manages the information or big data infrastructure. Develops the architecture that helps analyze and process data in the way the organization needs it. Makes sure those systems are performing smoothly.

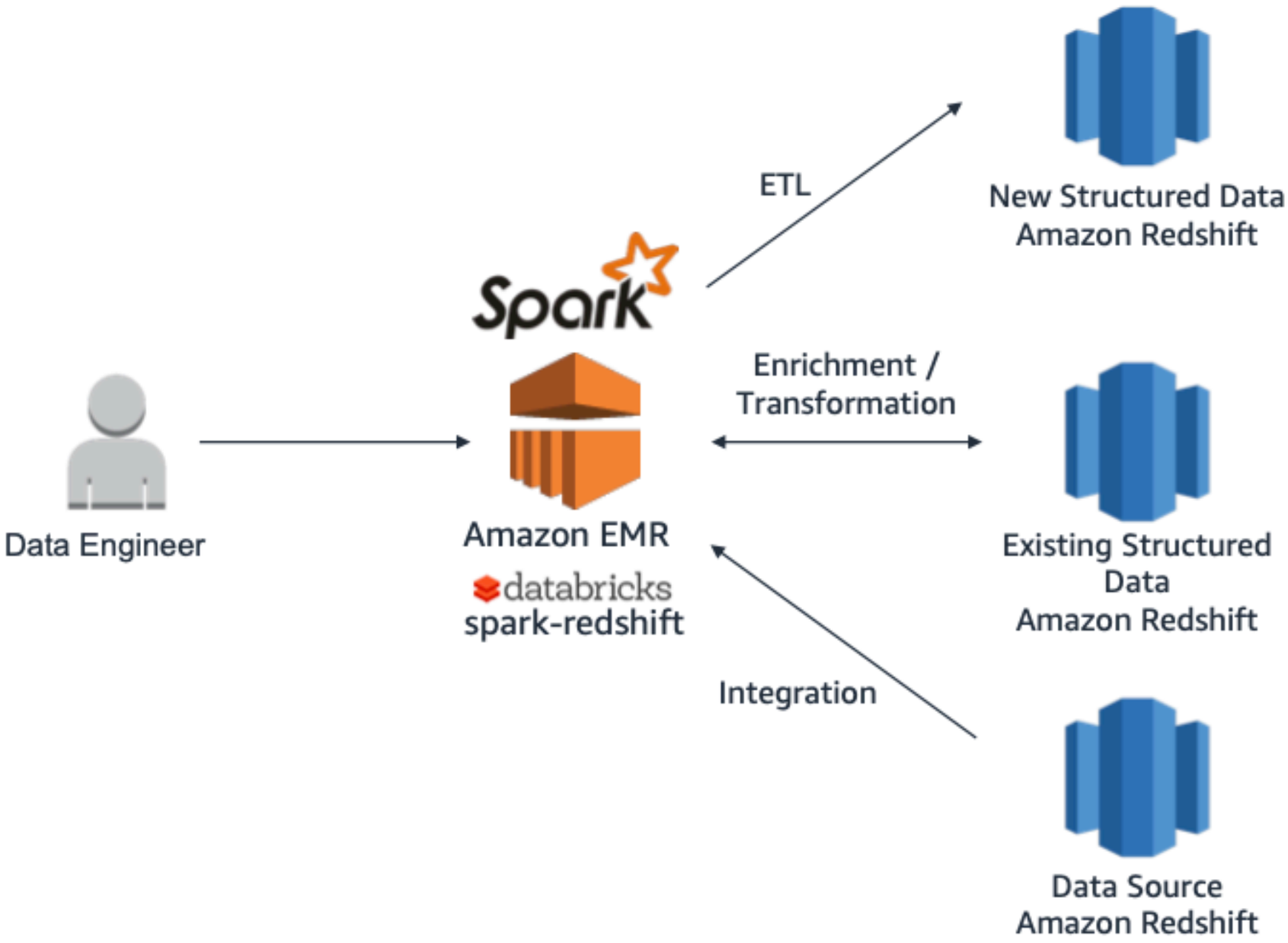
Top Necessary skills by demand

Skill	Salary Premium	Job Postings Requesting (Last 12 Months)	Projected Growth (2 Years)
Apache Webserver		6,642	+23.0%
Software Development		6,559	+3.9%
Relational Databases		5,822	-0.2%
Data Quality		5,545	+13.7%
Tableau		5,064	+33.3%
Oracle		5,060	-8.9%
Data Management		5,023	+6.5%
Linux		4,847	-5.9%
Software Engineering		4,715	+12.7%
Business Intelligence		4,655	-4.9%

Top Defining skills by demand

Skill	Salary Premium	Job Postings Requesting (Last 12 Months)	Projected Growth (2 Years)
Python		26,347	+34.6%
SQL		25,958	-6.1%
Big Data		19,191	+31.5%
Extraction Transformation and Loading (ETL)		17,703	+14.4%
Apache Hadoop		15,516	+23.4%
Java		15,010	+6.2%
Pipeline (Computing)		14,954	+52.4%
Data Warehousing		12,418	-2.9%
Scala		11,864	+51.1%
Data Science	✓	10,885	+40.1%

Data Engineer



BI/Analytics Careers

- Typical BI positions
  - BI solution architects and integration specialists
  - Business and BI analysts
  - BI application developers and testers
  - BI system support specialists
  - Data warehouse specialists
  - Database analysts, developers and testers
  - Data Scientists
  - Data Engineers / Big Data Engineers

Where to put the analytics team?

- Spread throughout the organization
- In IT Department
- In a standalone unit
- In some form of an Analytics Competency Center

Data trends

Data trends

- SaaS is everyone's new best friend
- Self-service has evolved to self-sufficiency
- Shared data, visualizations and storytelling are consumed by the masses
- Up-to-date and business-ready data are more important than ever
- Advanced analytics need to look different
- It's essential to capture and synthesize "alternative" data
- Business process reengineering takes center stage
- The compass for competition, surveillance and security has been recalibrated
- Collaboration has to coalesce earlier in the chain
- The Great Digital Switch may force a generational shift in analytics