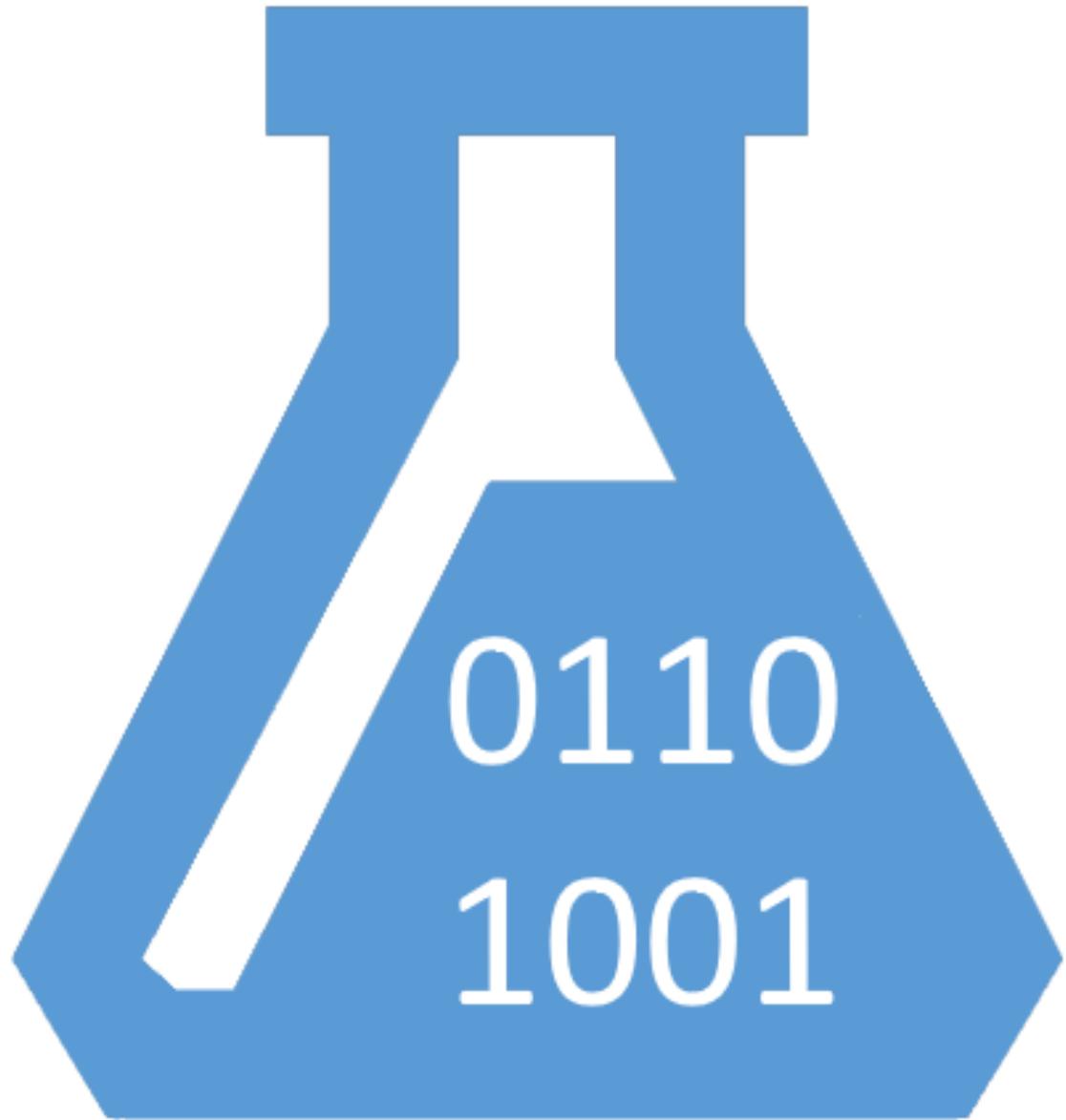
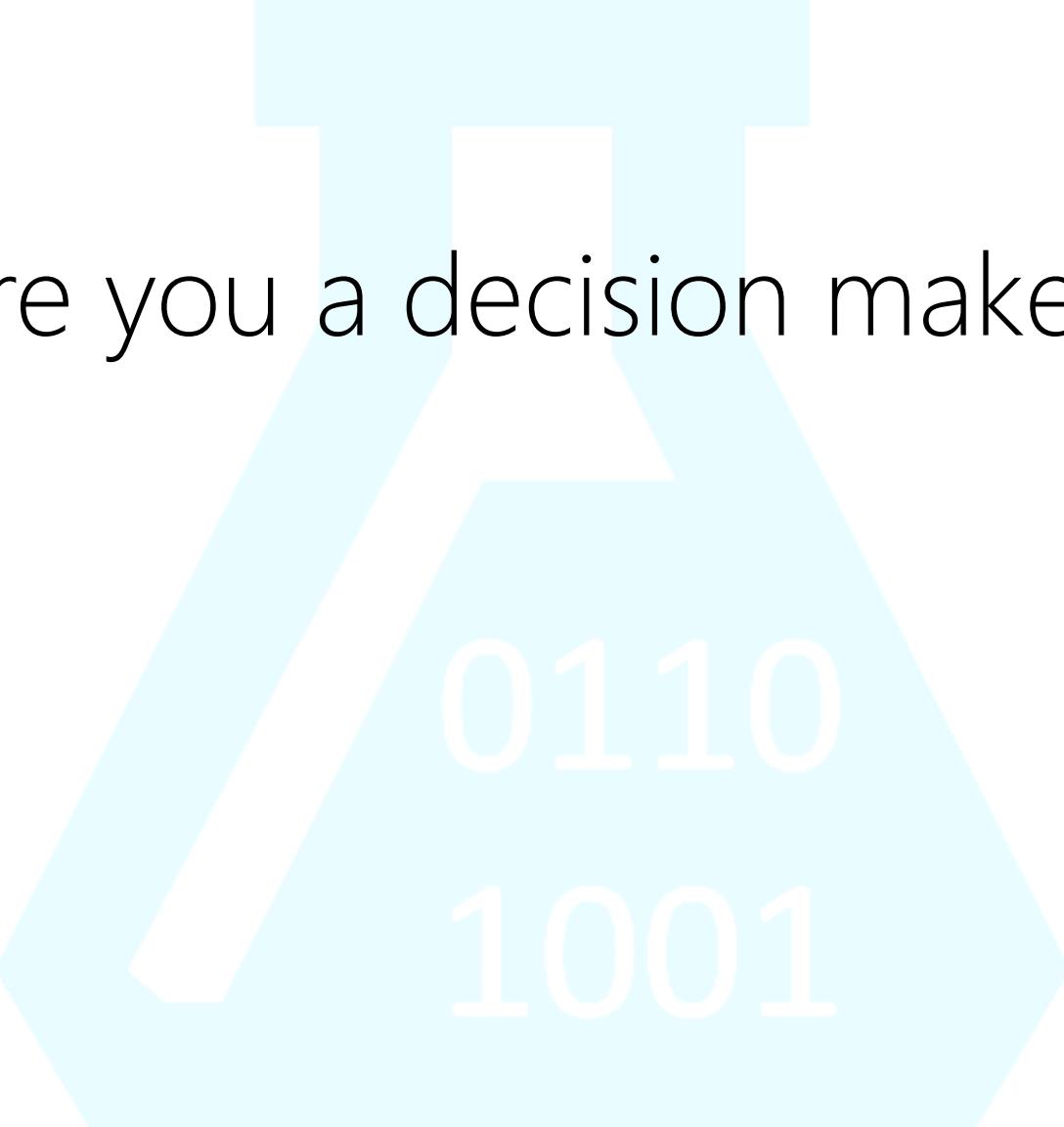


Data Science: Becoming a Data-driven Organization

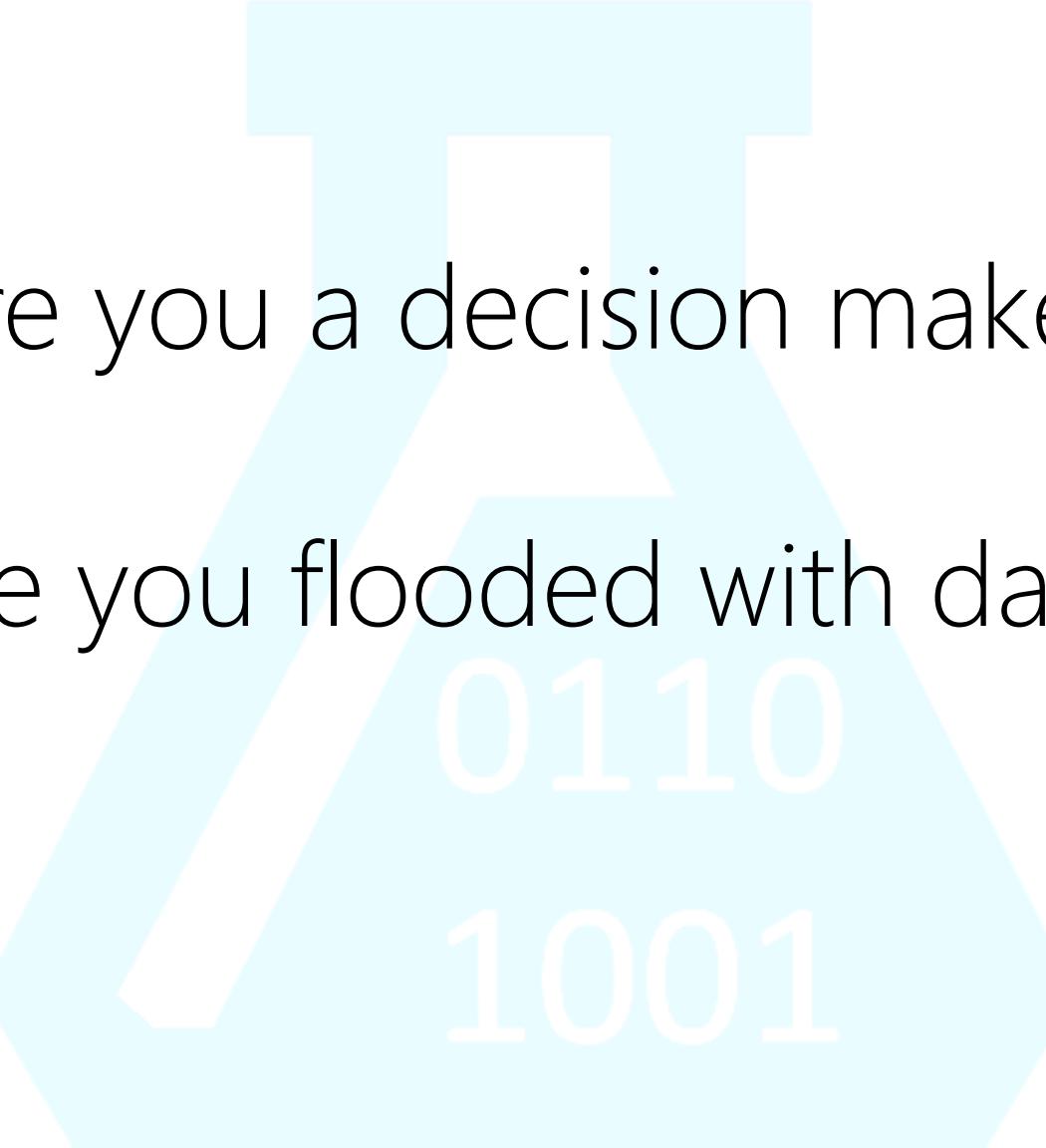
@MatthewRenze

#MicrosoftUAE



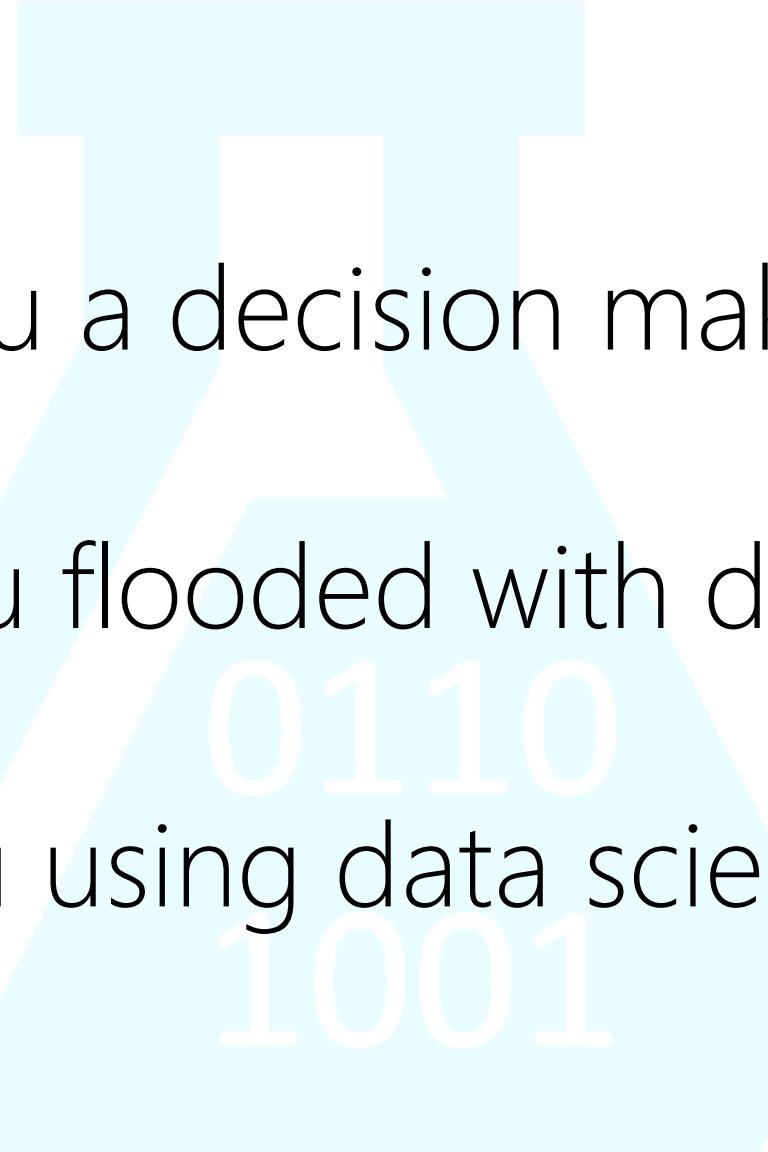


Are you a decision maker?



Are you a decision maker?

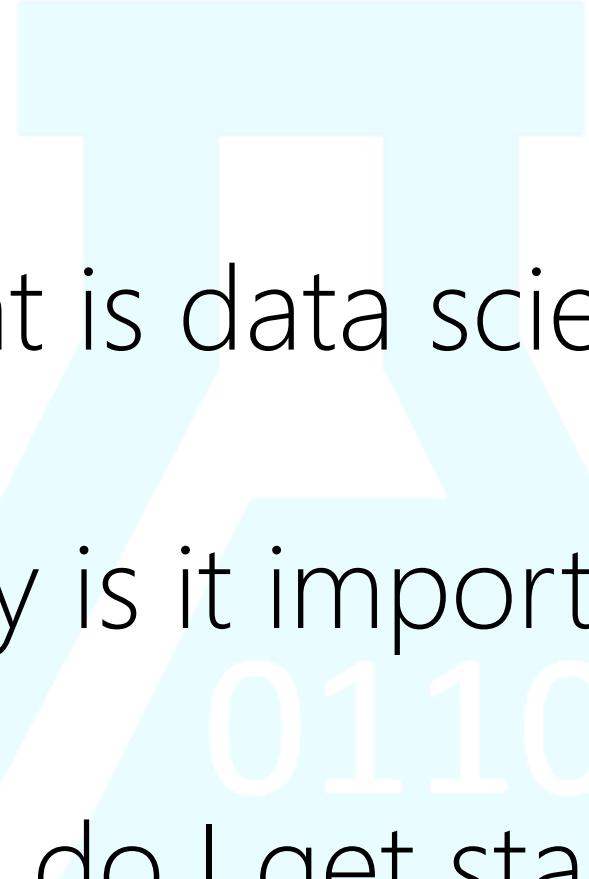
Are you flooded with data?



Are you a decision maker?

Are you flooded with data?

Are you using data science?



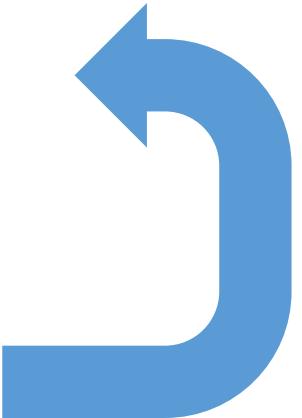
What is data science?



Why is it important?



How do I get started?



Why is it important?

What is data science?

How do I get started?



Why is data science important?

0110

1001







The Economist

FEBRUARY 27TH-MARCH 5TH 2010

Gordon Brown's pitch
What went wrong at RBS
Genetically modified crops blossom
The EU woos Russia
The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LORIN
Published: August 5, 2009

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

- [TWITTER](#)
- [LINKEDIN](#)
- [COMMENTS \(58\)](#)
- [SIGN IN TO E-MAIL](#)

Data Scientist: *The Sexiest Job of the 21st Century*

Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport
and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early.”

Job Postings for Data Scientists

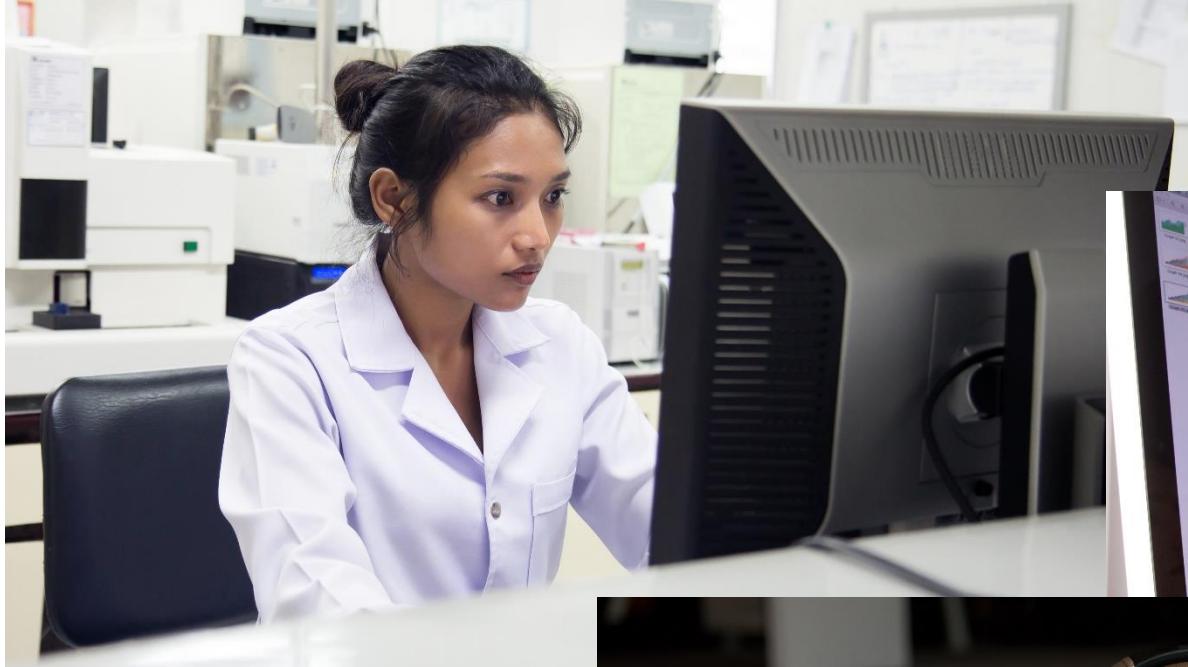


Top-paying Tech Skills

Skill	2016	Change
HANA (High Performance Analytical Application)	\$ 128,958	-3.3%
MapReduce	\$ 125,009	-0.3%
Cloud Foundry	\$ 124,038	n/a
Hbase	\$ 123,934	5.7%
Omnigraffle	\$ 123,782	-1.9%
Cassandra	\$ 123,459	2.2%
Apache Kafka	\$ 122,728	n/a
SOA (Service Oriented Architecture)	\$ 122,094	-1.9%
Ansible	\$ 121,382	n/a
Jetty	\$ 120,978	1.3%
PaaS (Platform as a Service)	\$ 120,403	-4.4%
Elasticsearch	\$ 120,002	n/a
ABAP (Advanced Business Application Programming)	\$ 119,961	0.5%
NoSQL	\$ 119,498	1.3%
CMMI (Capability Maturity Model Integration)	\$ 119,466	-0.6%
Amazon Redshift	\$ 119,197	n/a
Pig	\$ 119,118	-4.2%
Solr	\$ 119,032	0.1%
Cloudera	\$ 118,896	-9.0%
Docker	\$ 118,873	0.2%

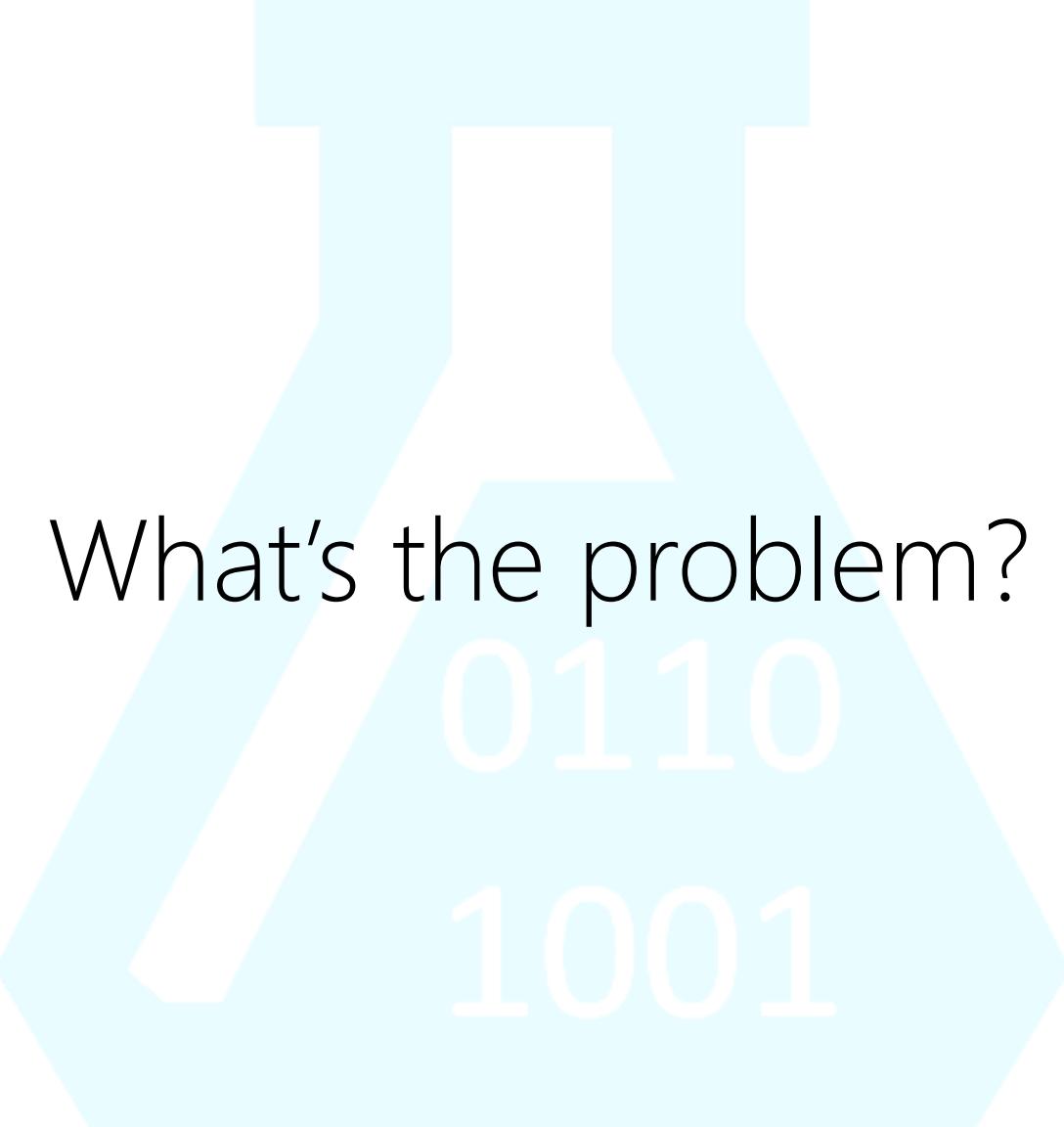
Skill	2016	Change
Amazon Route 53	\$ 118,828	n/a
Hadoop	\$ 118,625	-2.5%
Hive	\$ 118,589	-1.3%
Korn Shell	\$ 118,273	1.4%
PMBok (Project Management Body of Knowledge)	\$ 118,233	0.7%
Dynamo DB	\$ 118,119	n/a
Groovy	\$ 117,897	-0.1%
IaaS (Infrastructure as a Service)	\$ 117,422	n/a
JAX-RS (Java API RestFUL Services)	\$ 116,997	n/a
RabbitMQ	\$ 116,909	n/a
JDBC (Java Database Connectivity)	\$ 116,833	2.0%
SOX (Sarbanes Oxley)	\$ 116,743	0.6%
Objective C	\$ 116,667	2.5%
FCoE (Fibre Channel over Ethernet)	\$ 116,145	7.2%
UML (Unified Modeling Langauge)	\$ 115,285	-3.6%
XSLT (Extensible Stylesheet Language Transformations)	\$ 115,089	3.5%
Redis	\$ 114,922	2.8%
ETL (Extract Transform and Load)	\$ 114,892	2.6%
SDN (Software Defined Network)	\$ 114,739	-2.3%
Informatica	\$ 114,143	1.1%

Source: Dice Salary Survey 2017





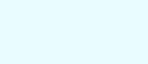




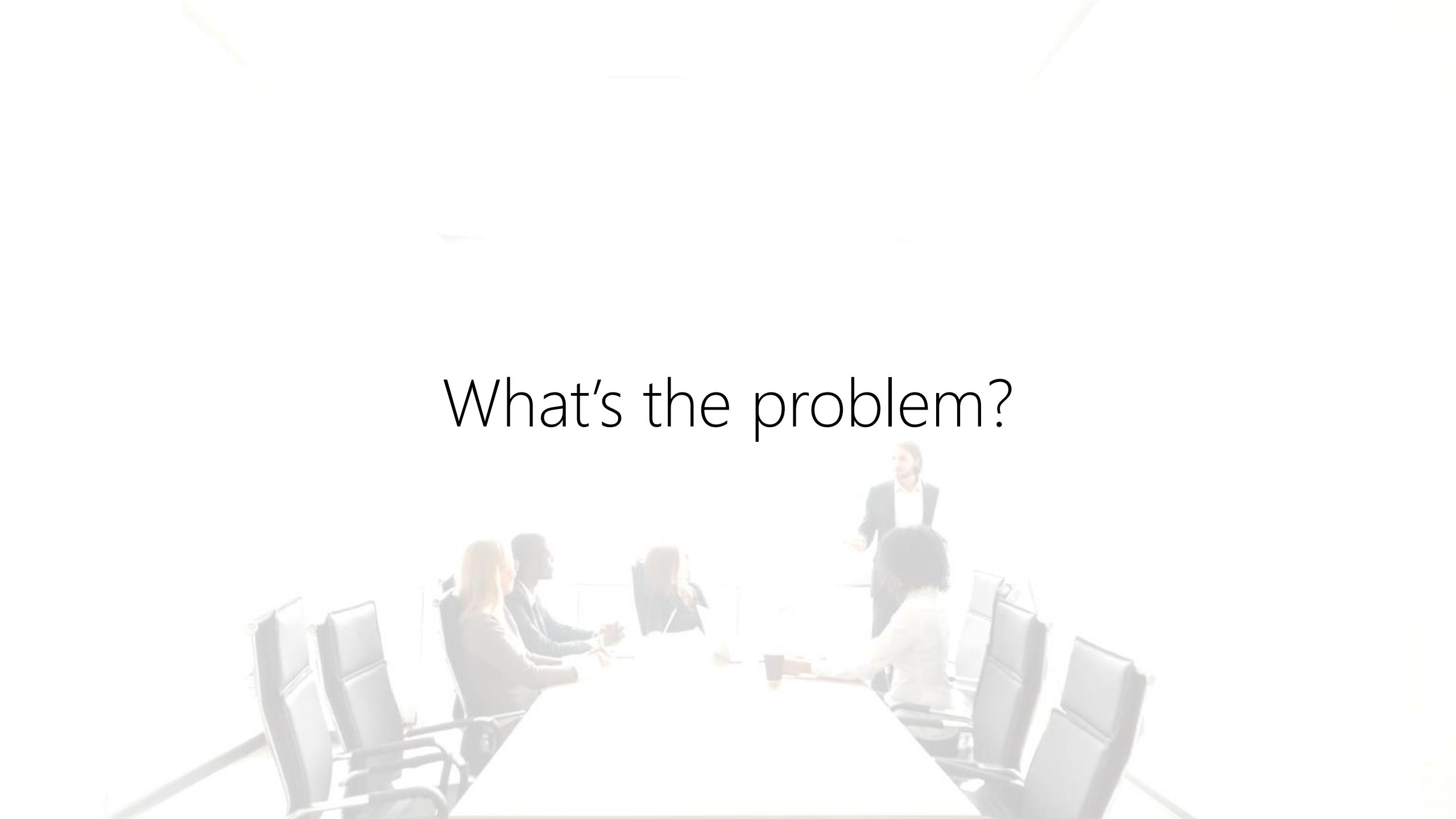
What's the problem?

0110

1001



What's the problem?

A faint, grayscale background image of a business meeting. A man in a suit stands at the head of a long conference table, gesturing with his hands as if speaking. Several people are seated around the table, facing him. The scene is set in a modern office environment with large windows.



CLIENT / OPTIONS







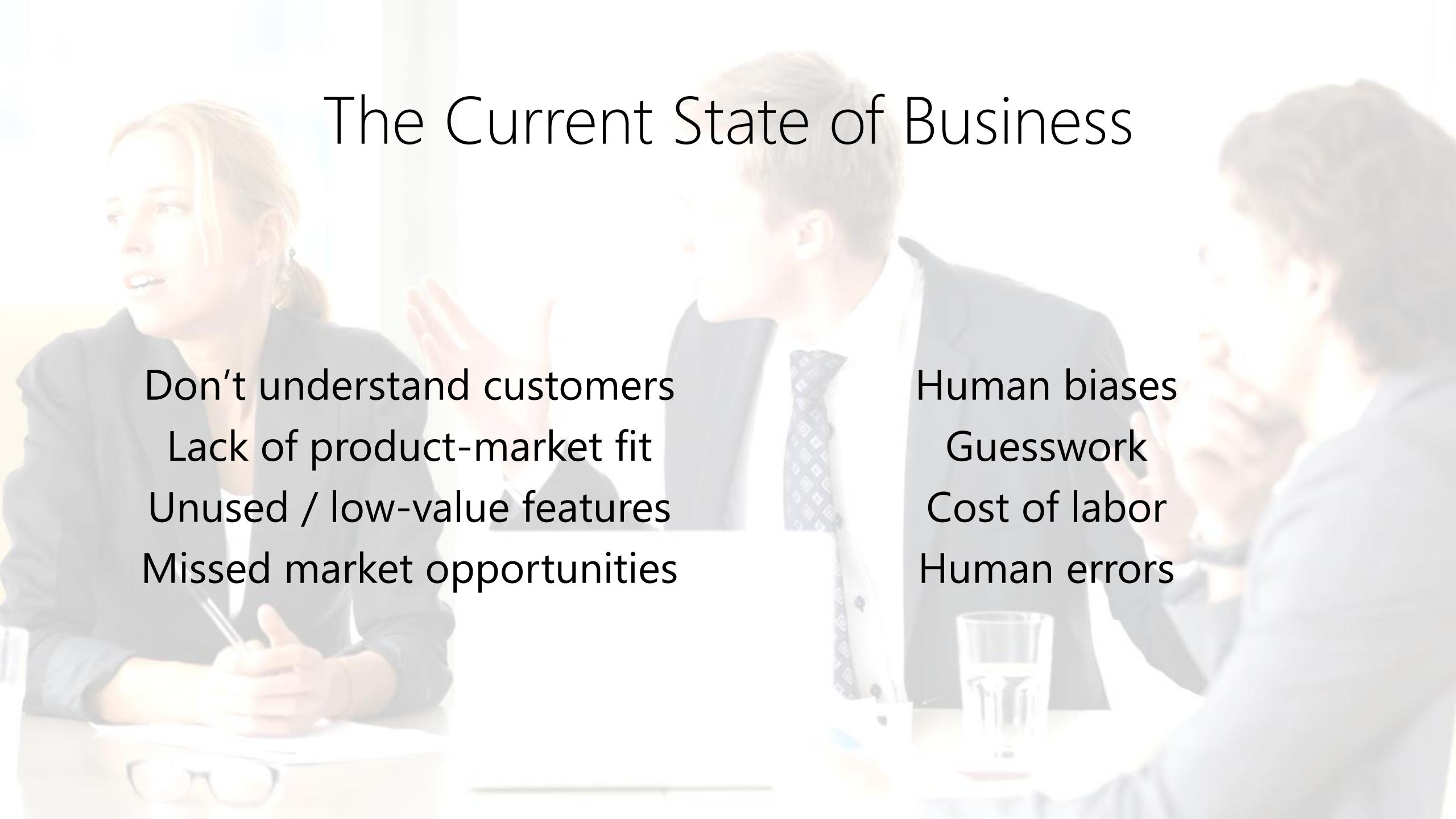
CONTRACT







The Current State of Business

A photograph showing four business professionals in a meeting. A woman in a dark blazer and white shirt is gesturing with her hands while speaking. Behind her, a man in a suit and tie looks on. To the right, another man in a suit is partially visible, and a woman's face is seen in profile. The scene is set in an office environment with a table and glasses in the foreground.

- Don't understand customers
- Lack of product-market fit
- Unused / low-value features
- Missed market opportunities

- Human biases
- Guesswork
- Cost of labor
- Human errors



Three Main Approaches

Create
better
products

Make
smarter
decisions

Reduce
labor
costs

Three Main Approaches

Create
better
products

Make
smarter
decisions

Reduce
labor
costs

Three Main Approaches

Create
better
products

Make
smarter
decisions

Reduce
labor
costs

Three Main Approaches

Create
better
products

Make
smarter
decisions

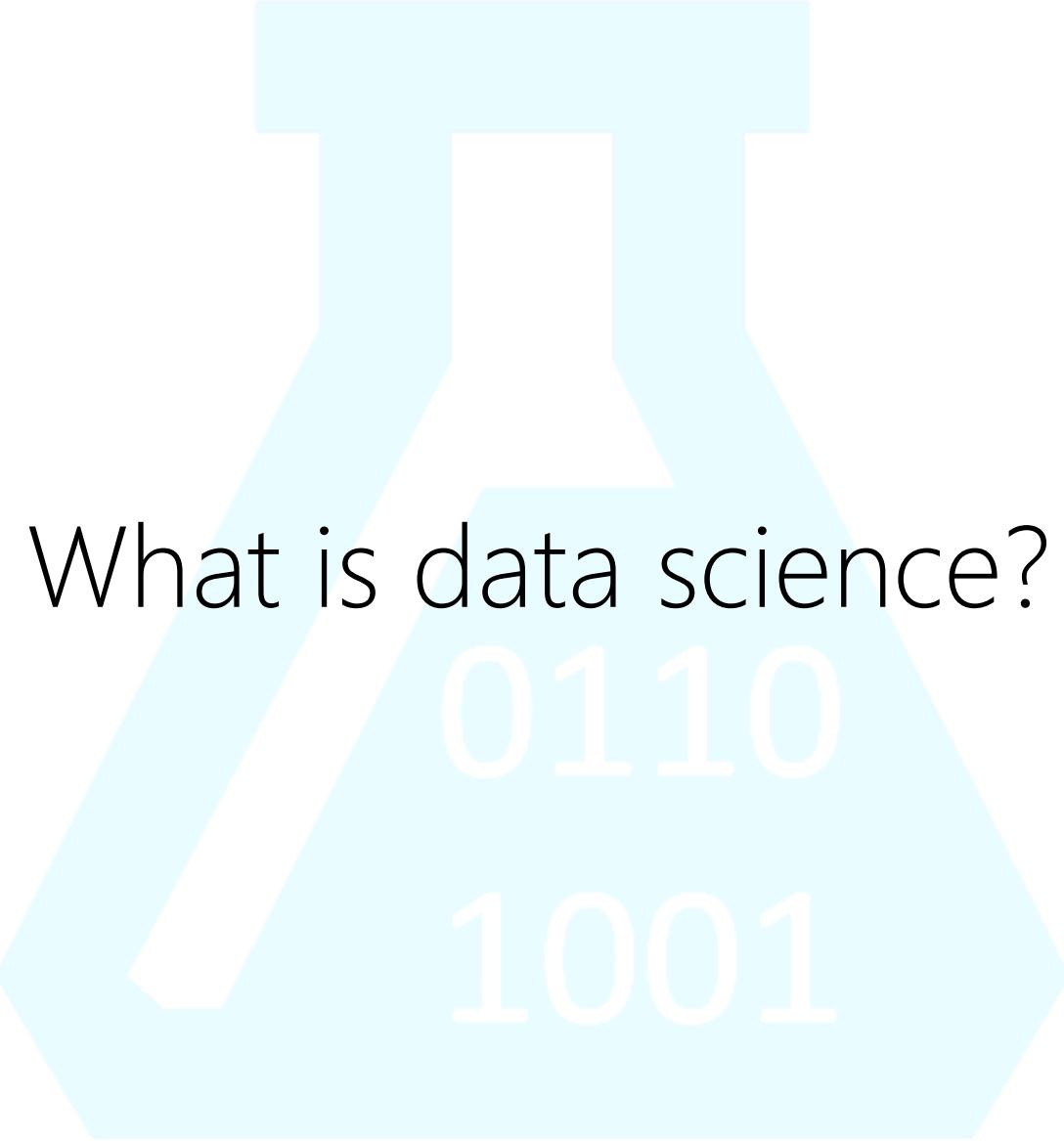
Reduce
labor
costs

Three Main Approaches

Create
better
products

Make
smarter
decisions

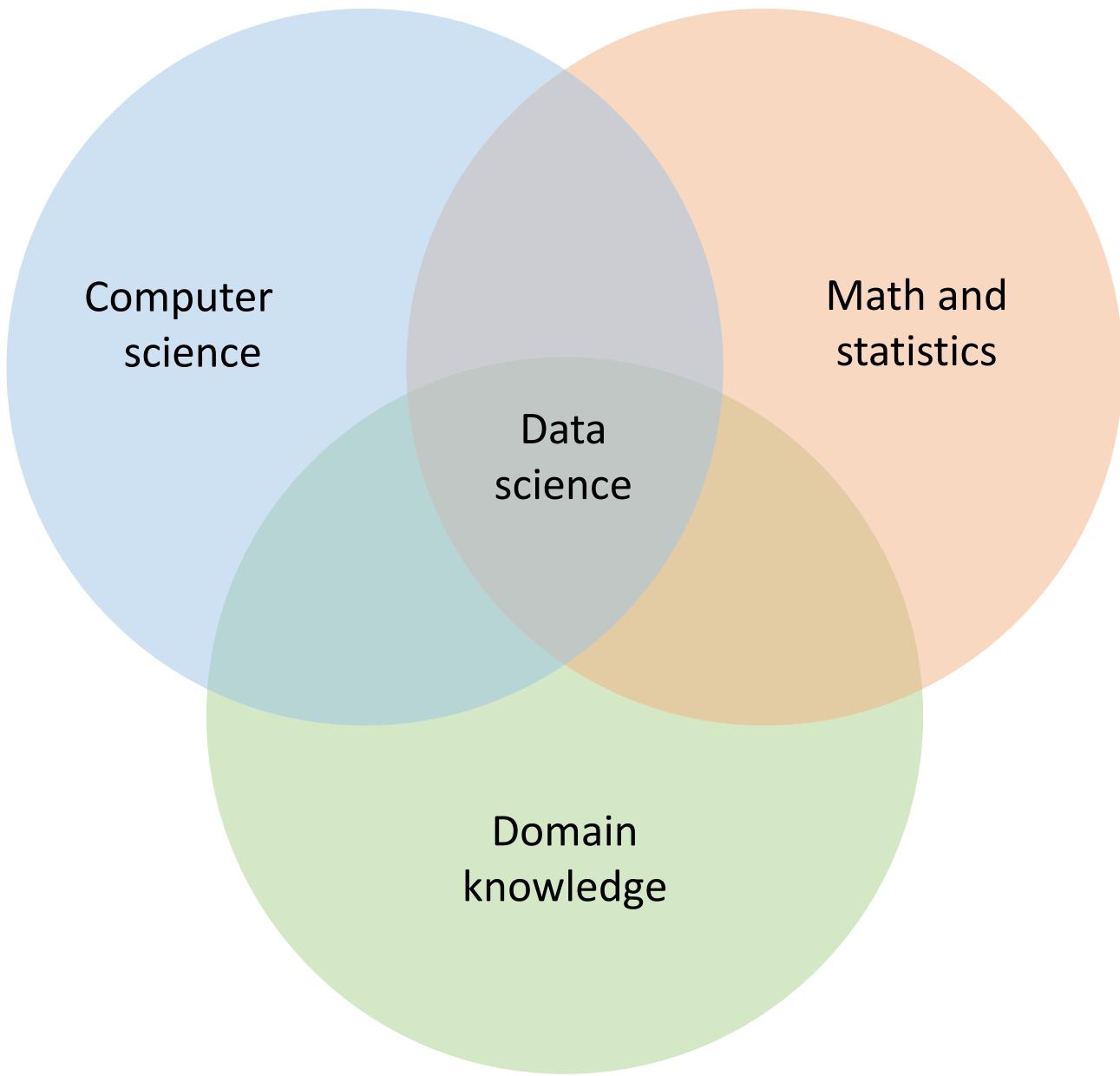
Reduce
labor
costs

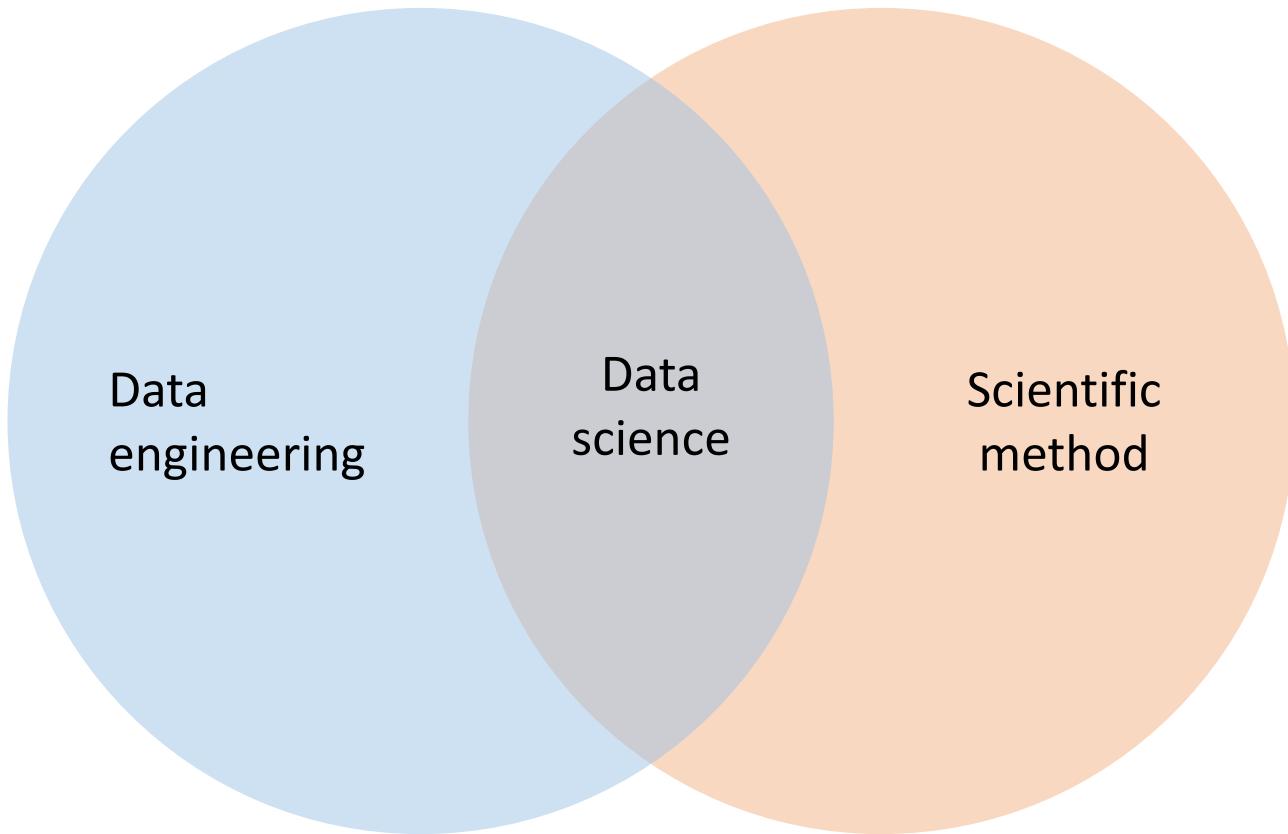


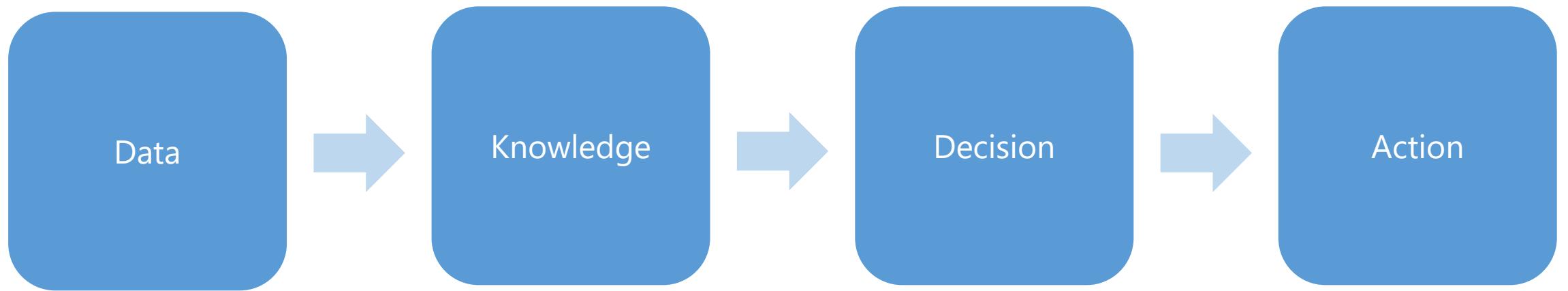
What is data science?

0110

1001







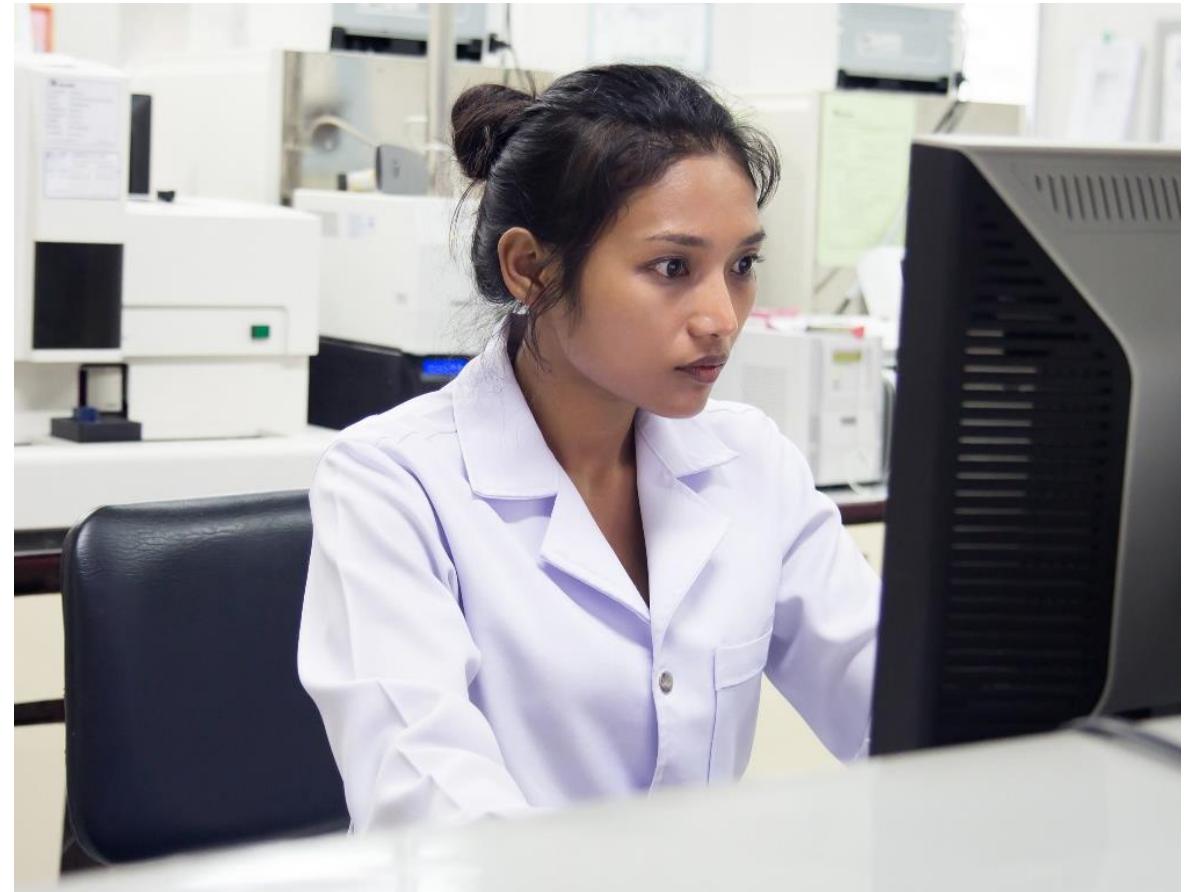
What Is a Data Scientist?

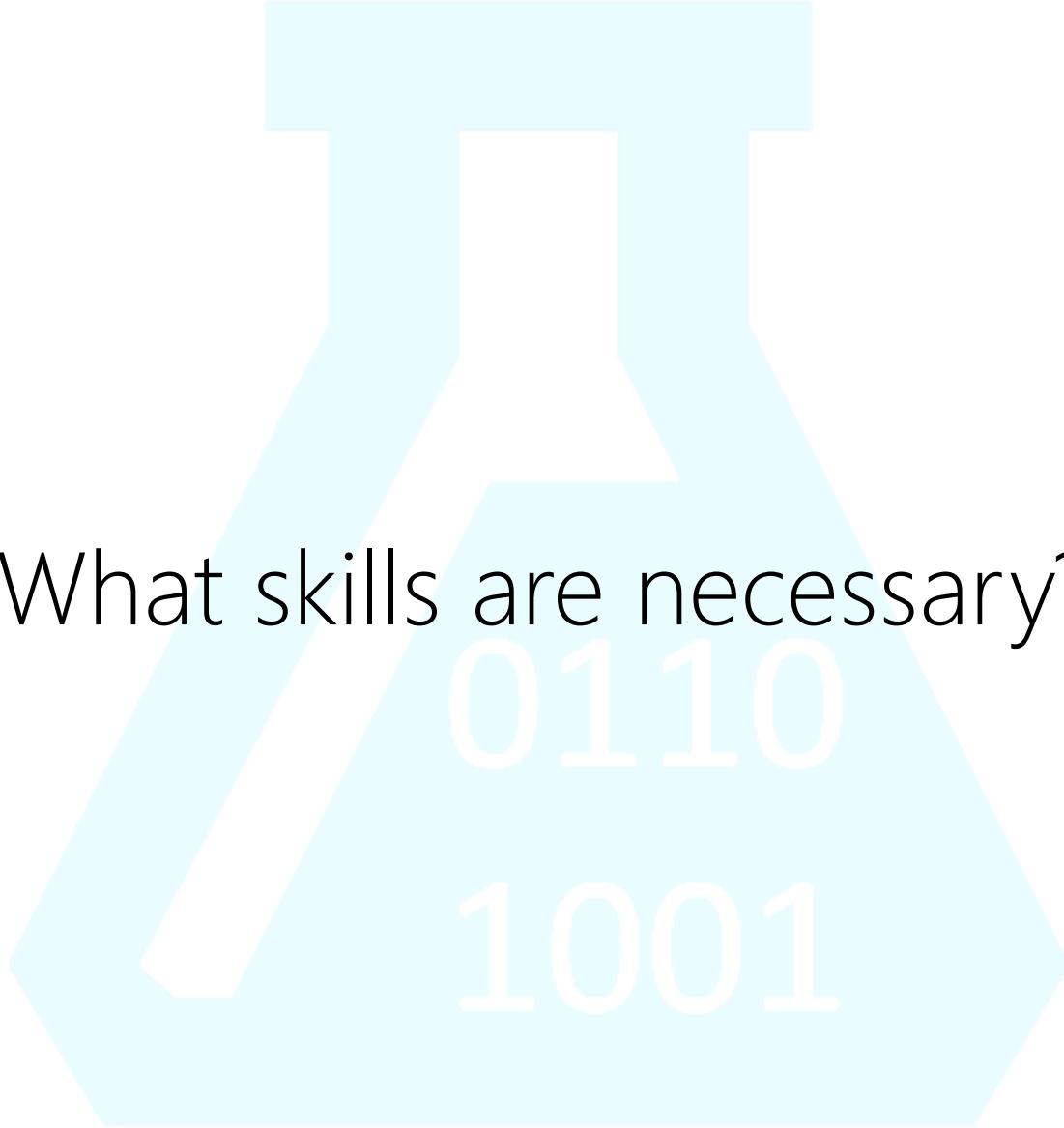
Performs data science

More than a scientist

More than an analyst

More than a developer





What skills are necessary?

0110

1001

Data Science Skills

Programming

Working with data

Descriptive statistics

Data visualization

Data Science Skills

Programming

Working with data

Descriptive statistics

Data visualization

Statistical modeling

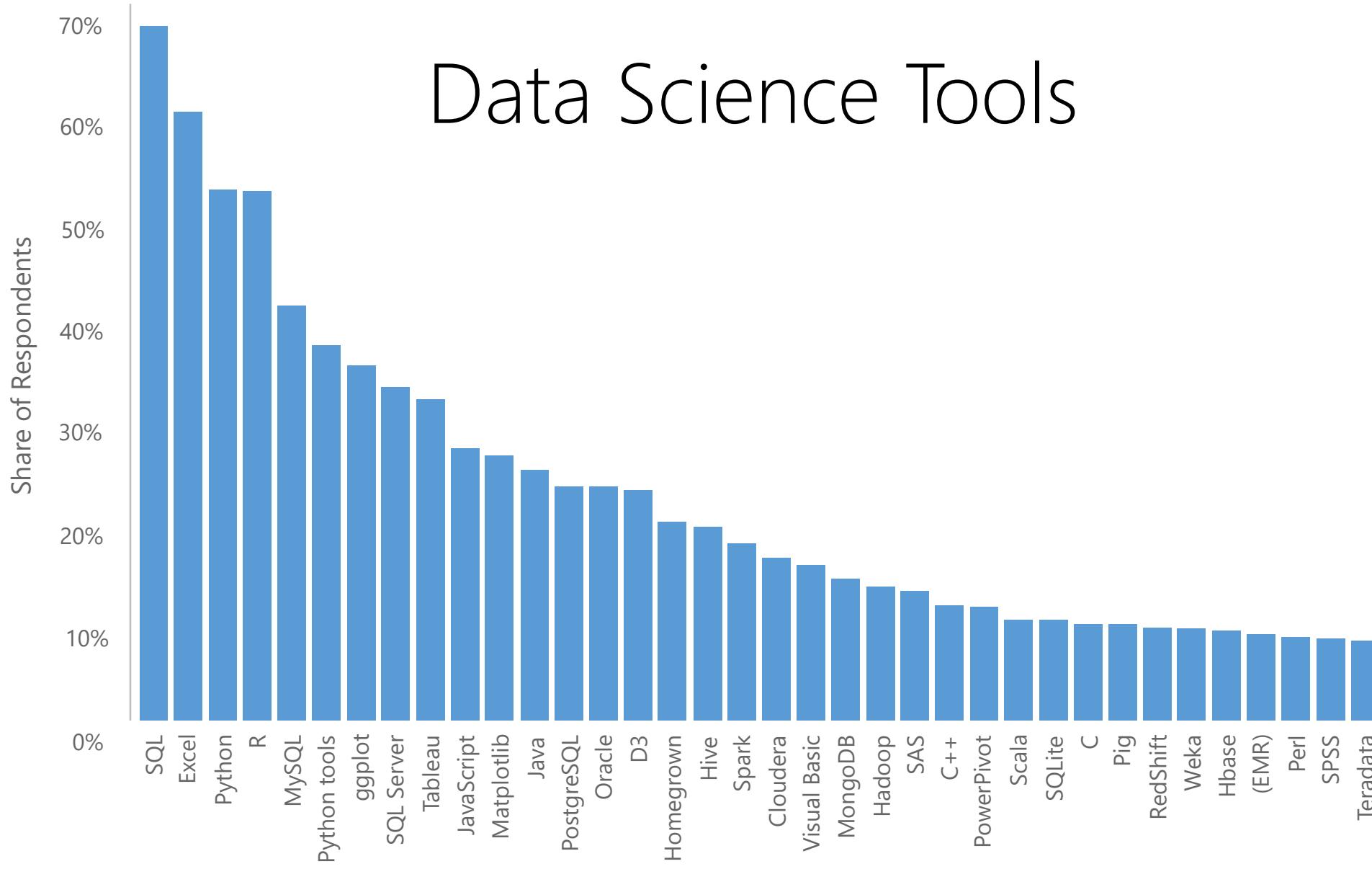
Handling Big Data

Machine learning

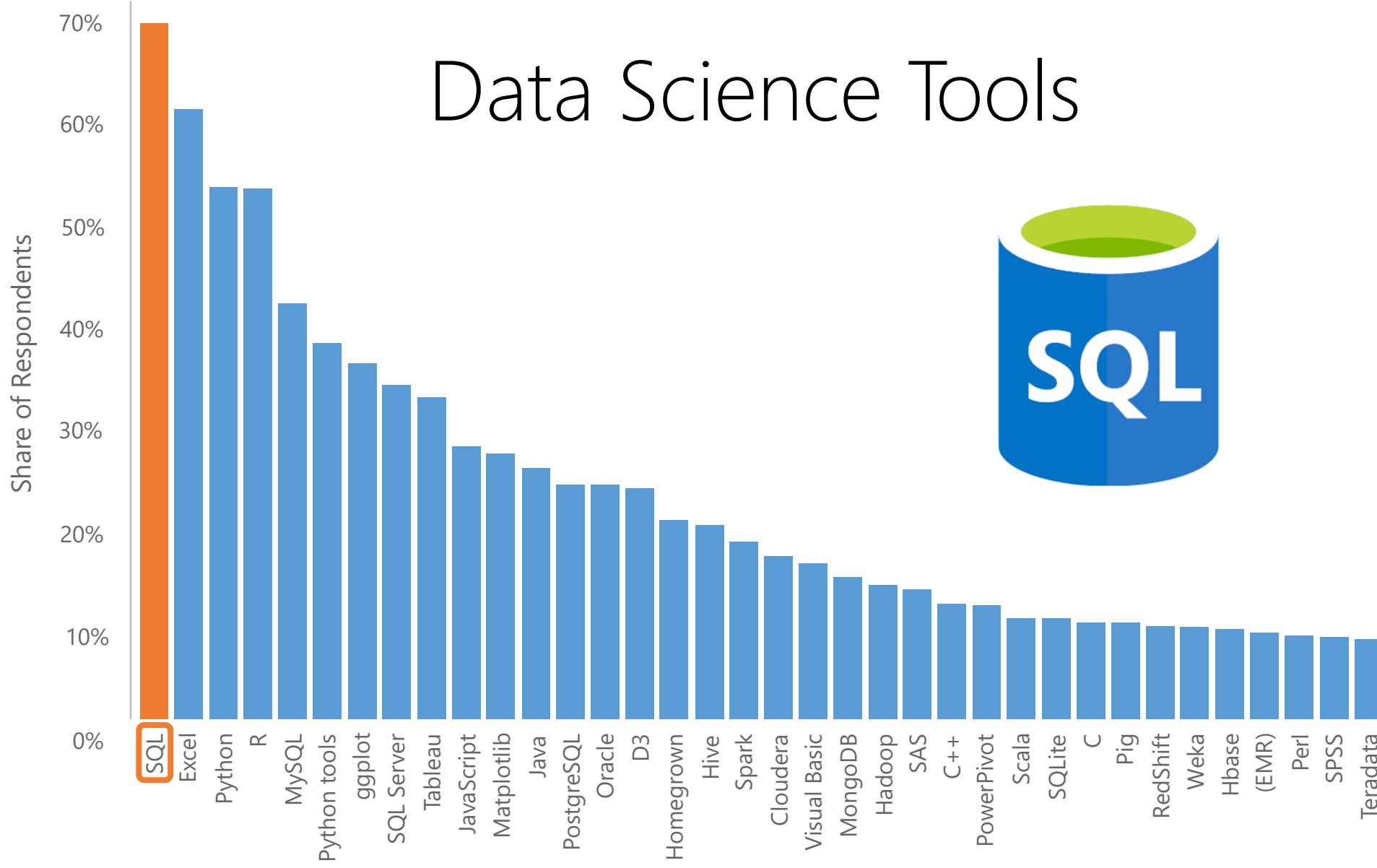
Deploying to production



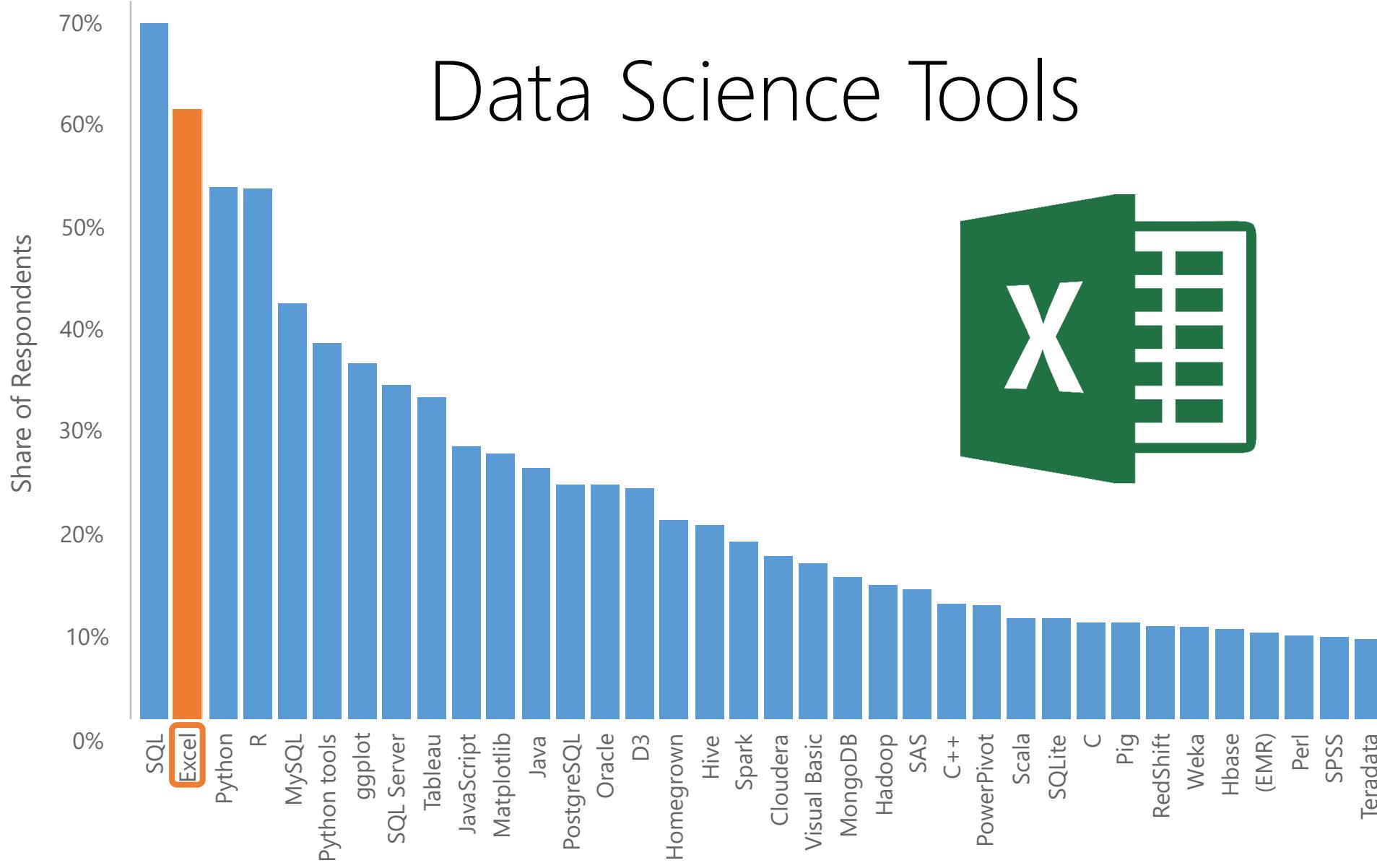
What tools are used?



Source: O'Reilly 2015 Data Science Salary Survey

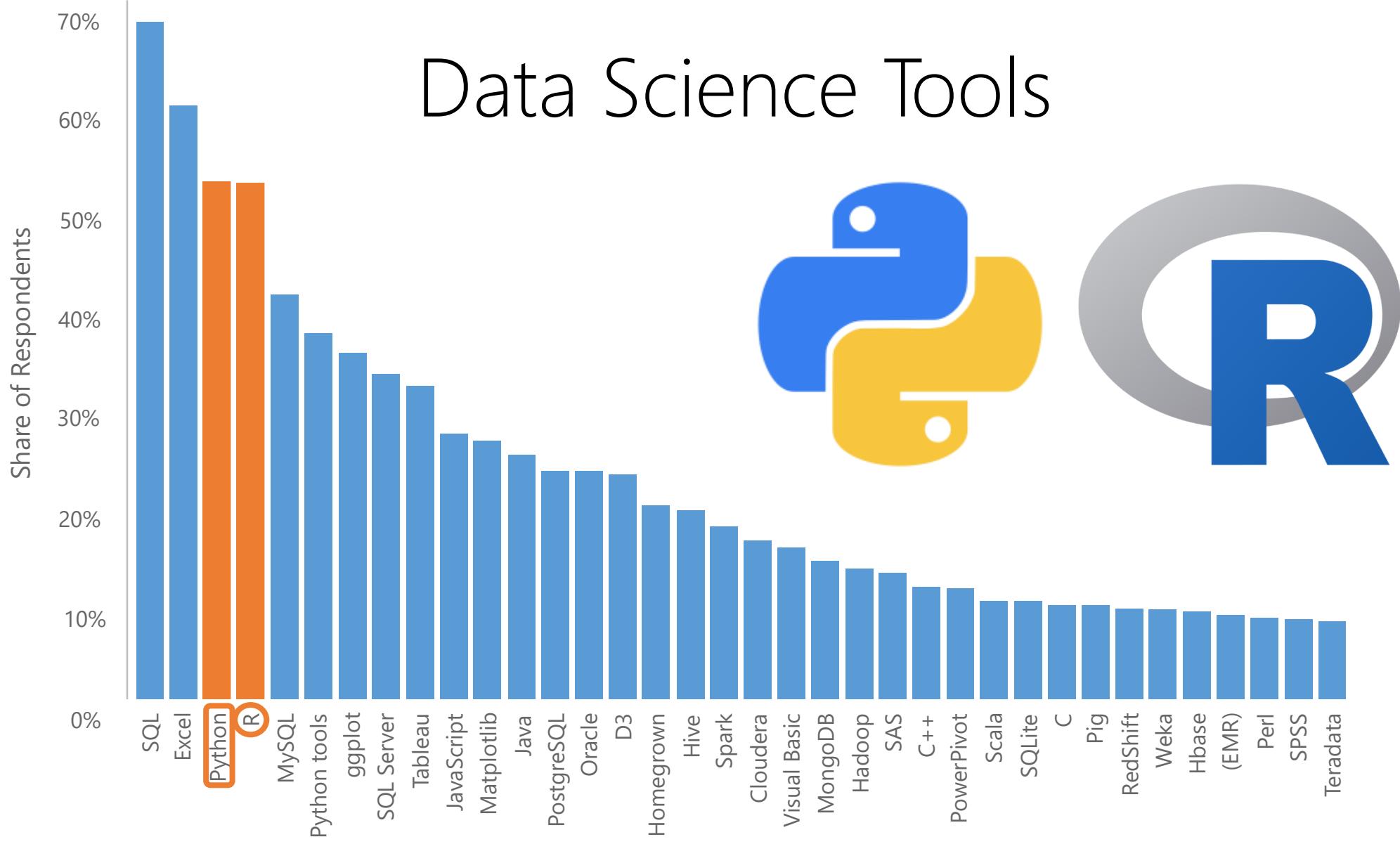


Source: O'Reilly 2015 Data Science Salary Survey

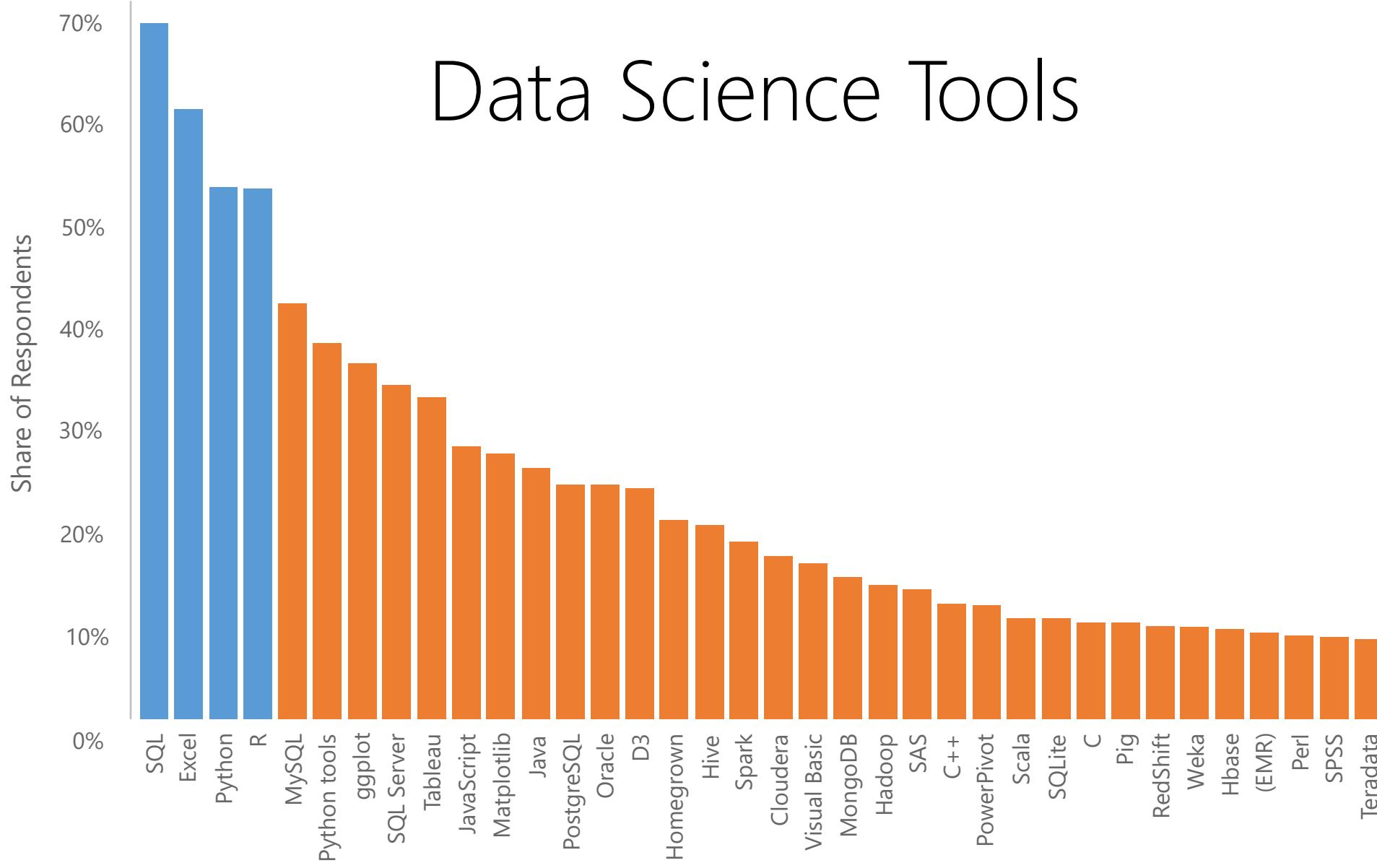


Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey



Source: O'Reilly 2015 Data Science Salary Survey

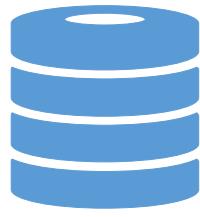


Source: O'Reilly 2015 Data Science Salary Survey



How is data science performed?

The Data Science Process



Data

The Data Science Process

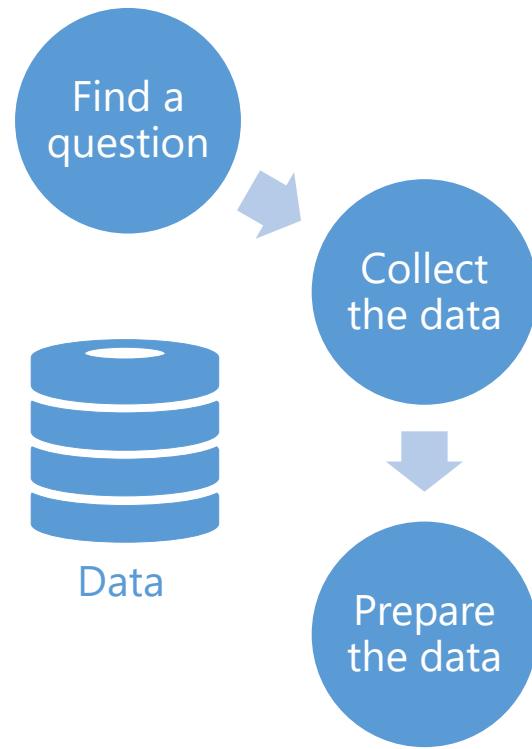


Data

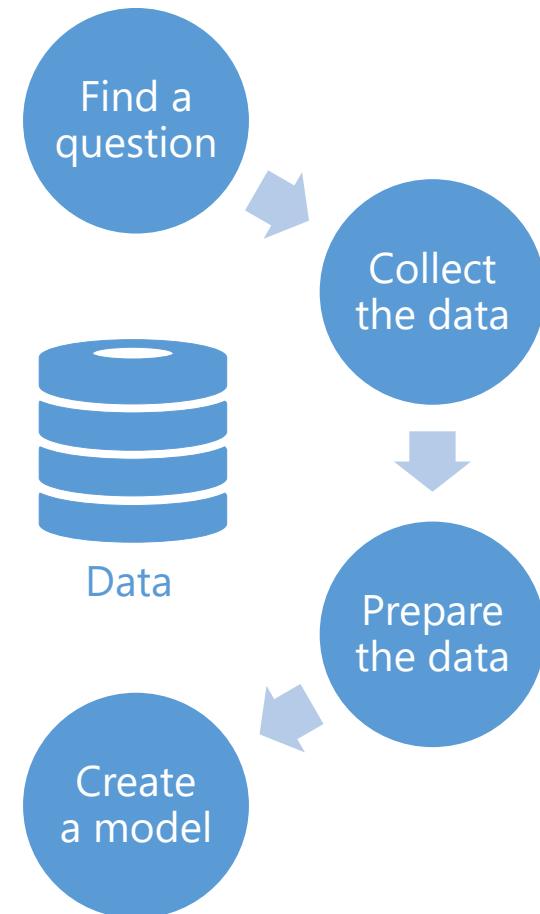
The Data Science Process



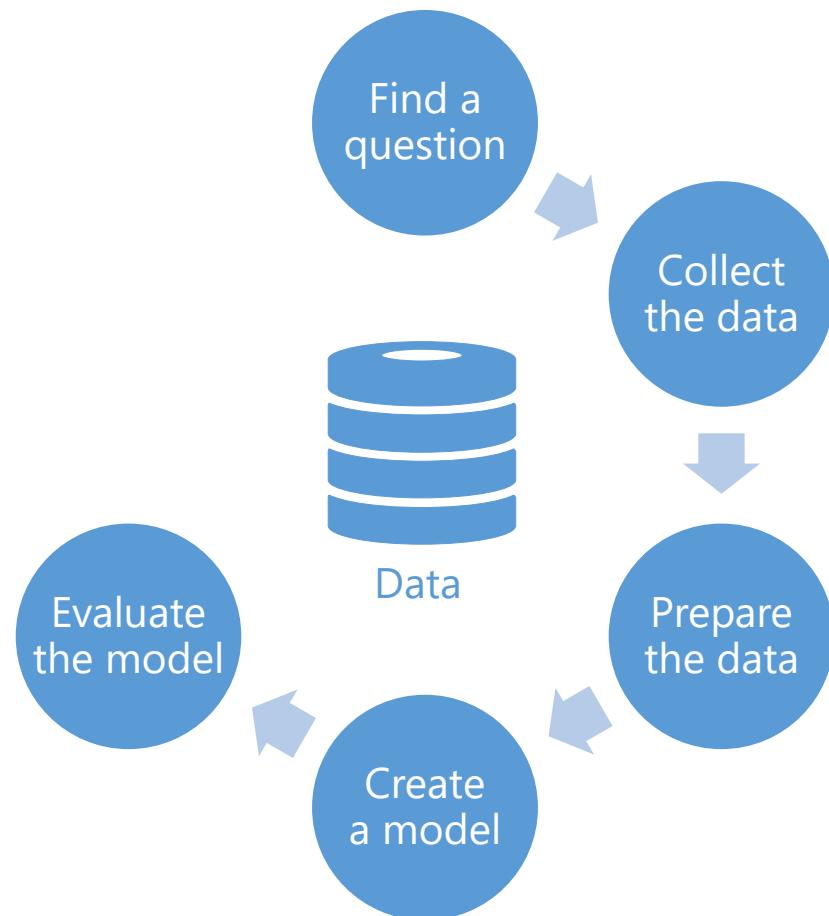
The Data Science Process



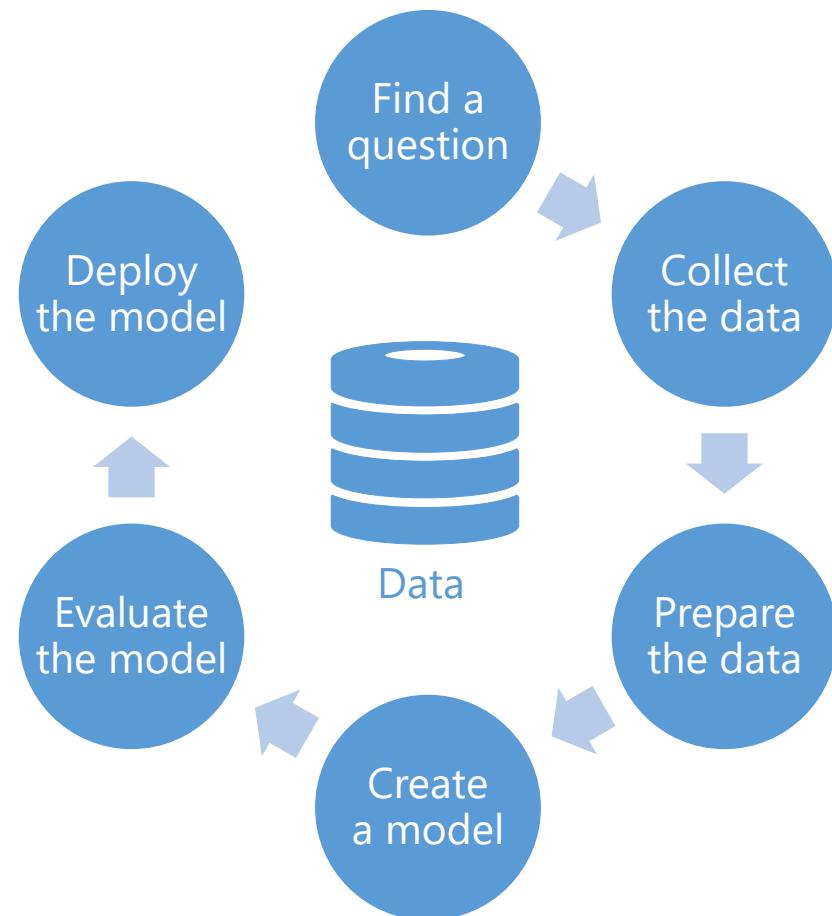
The Data Science Process



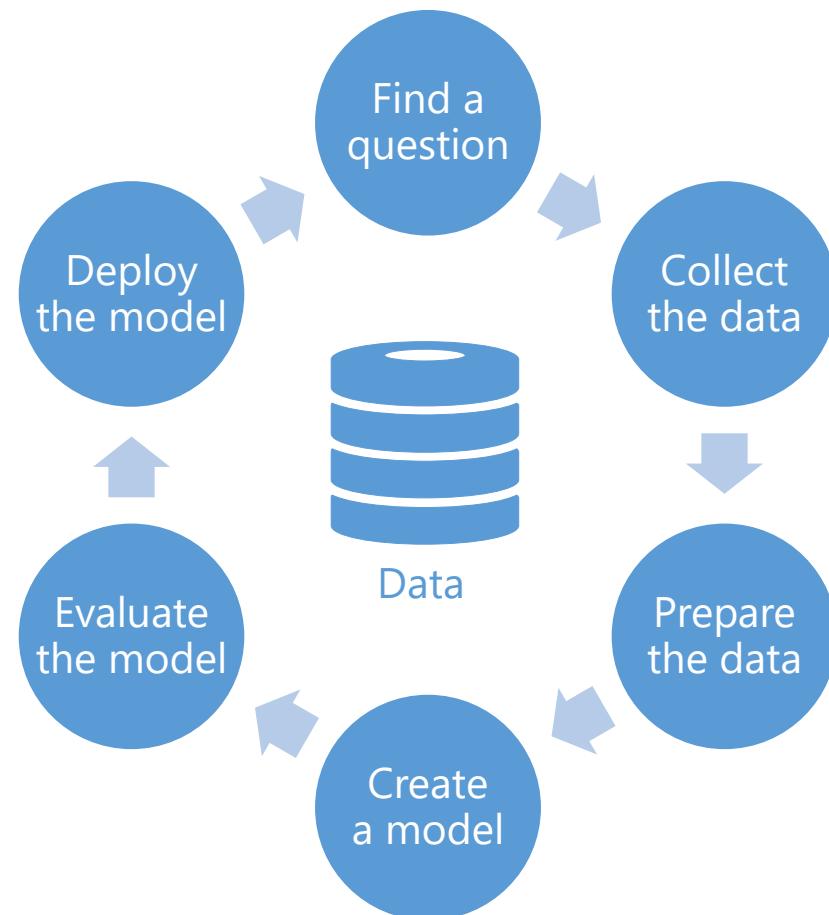
The Data Science Process



The Data Science Process

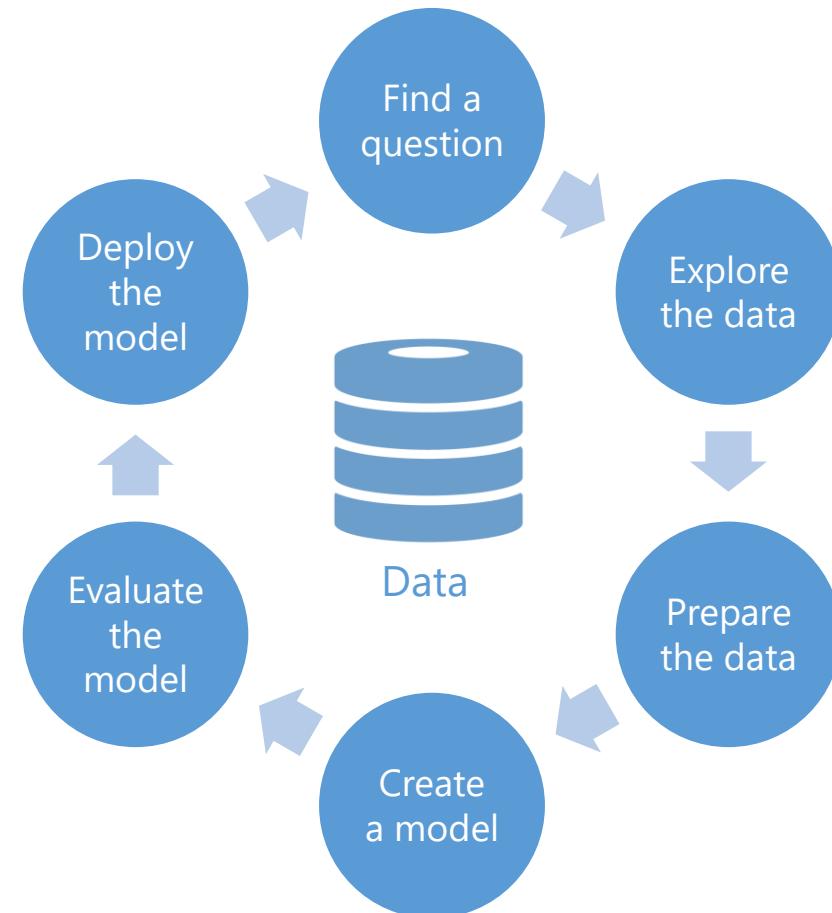


The Data Science Process



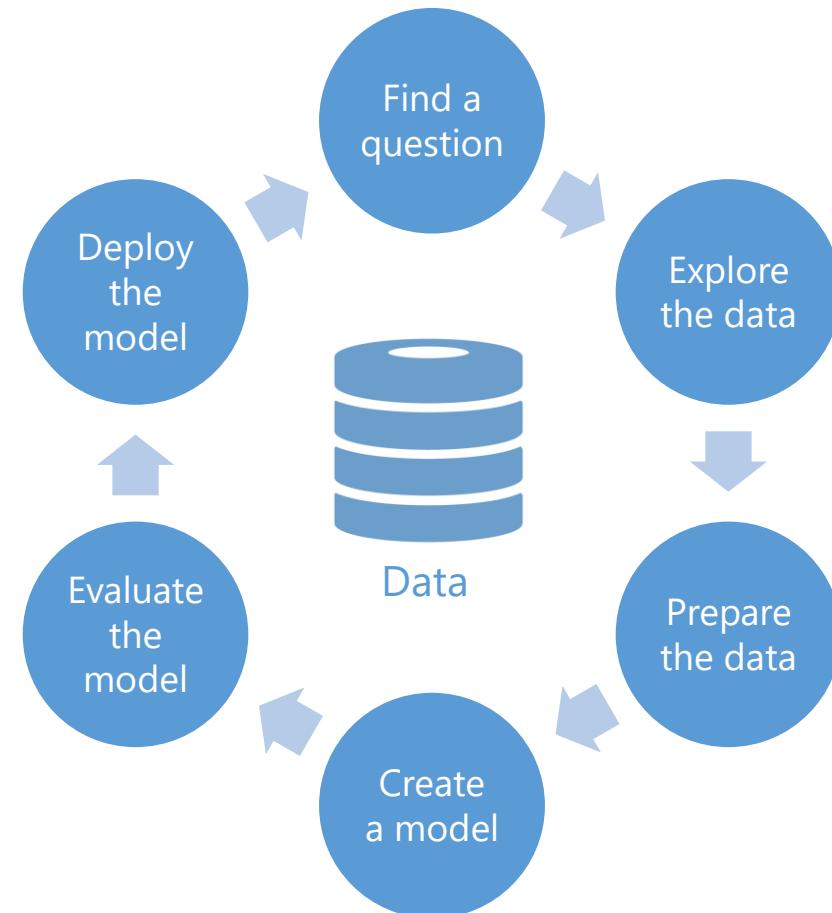
The Data Science Process

Iterative process



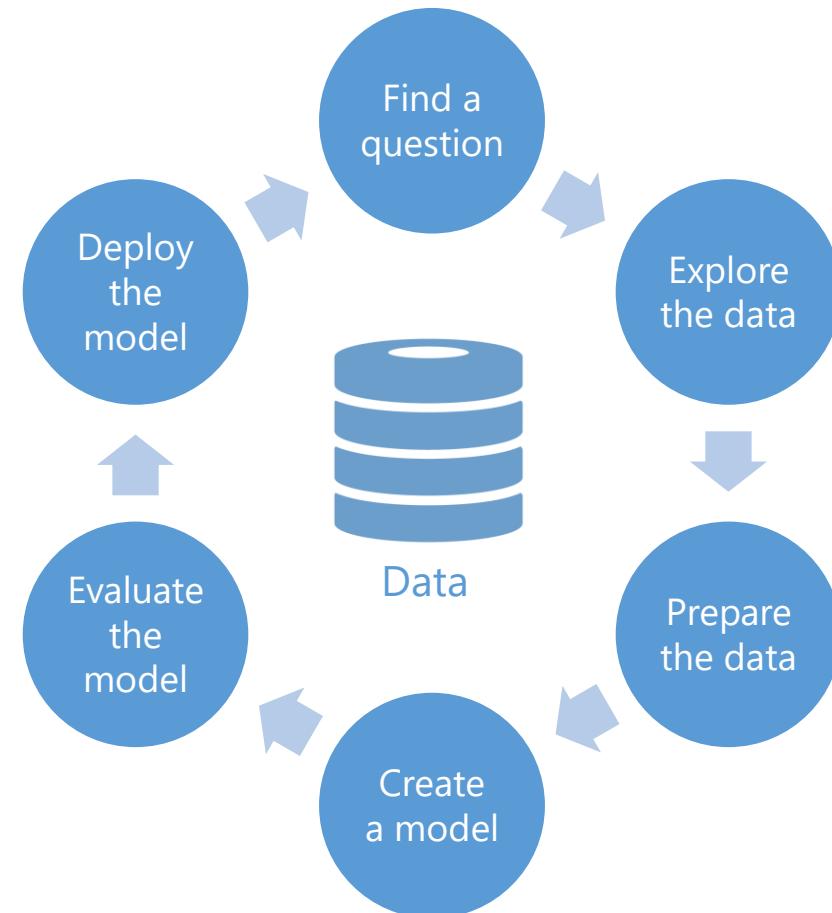
The Data Science Process

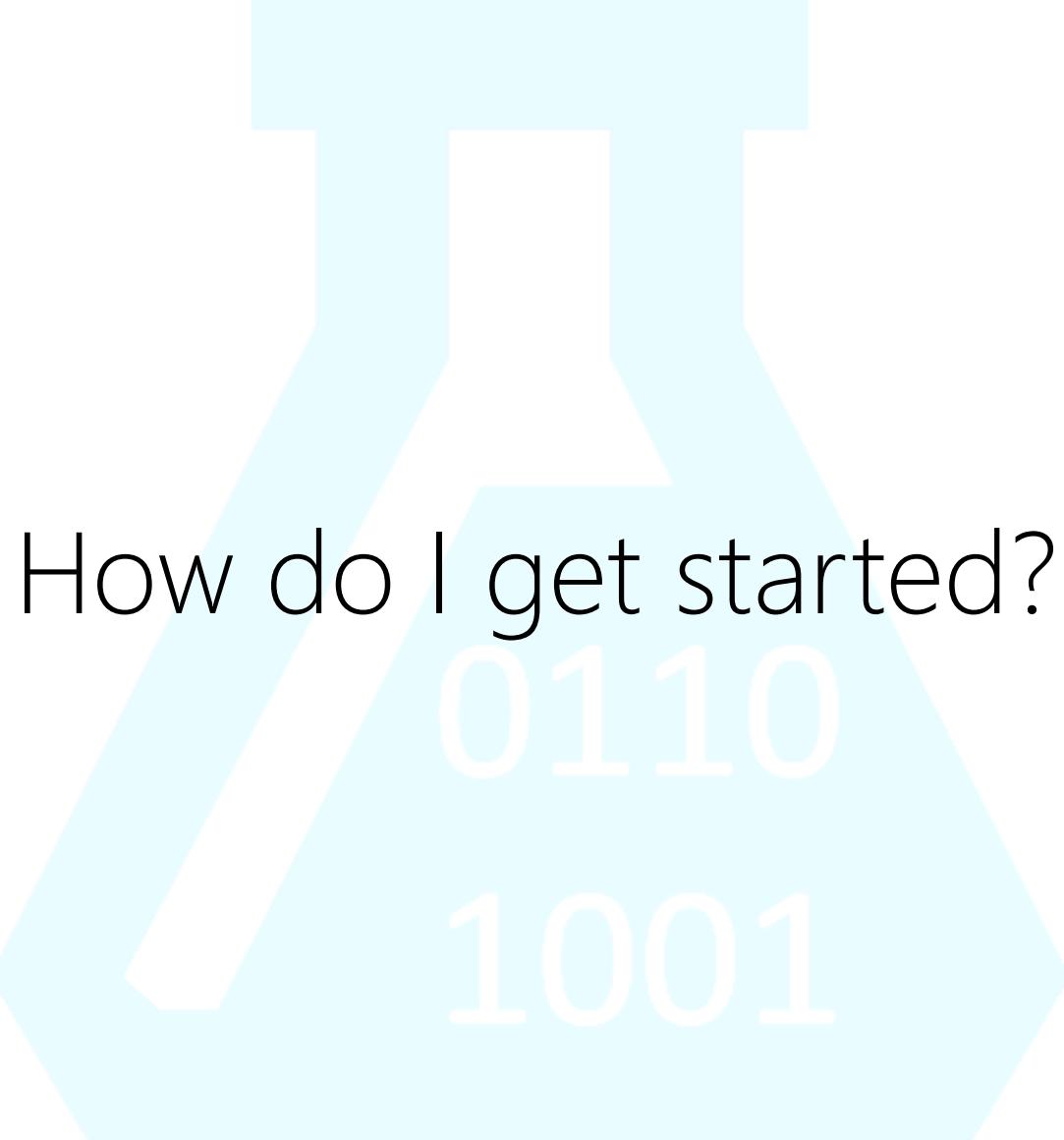
Iterative process
Non-sequential



The Data Science Process

Iterative process
Non-sequential
Early termination

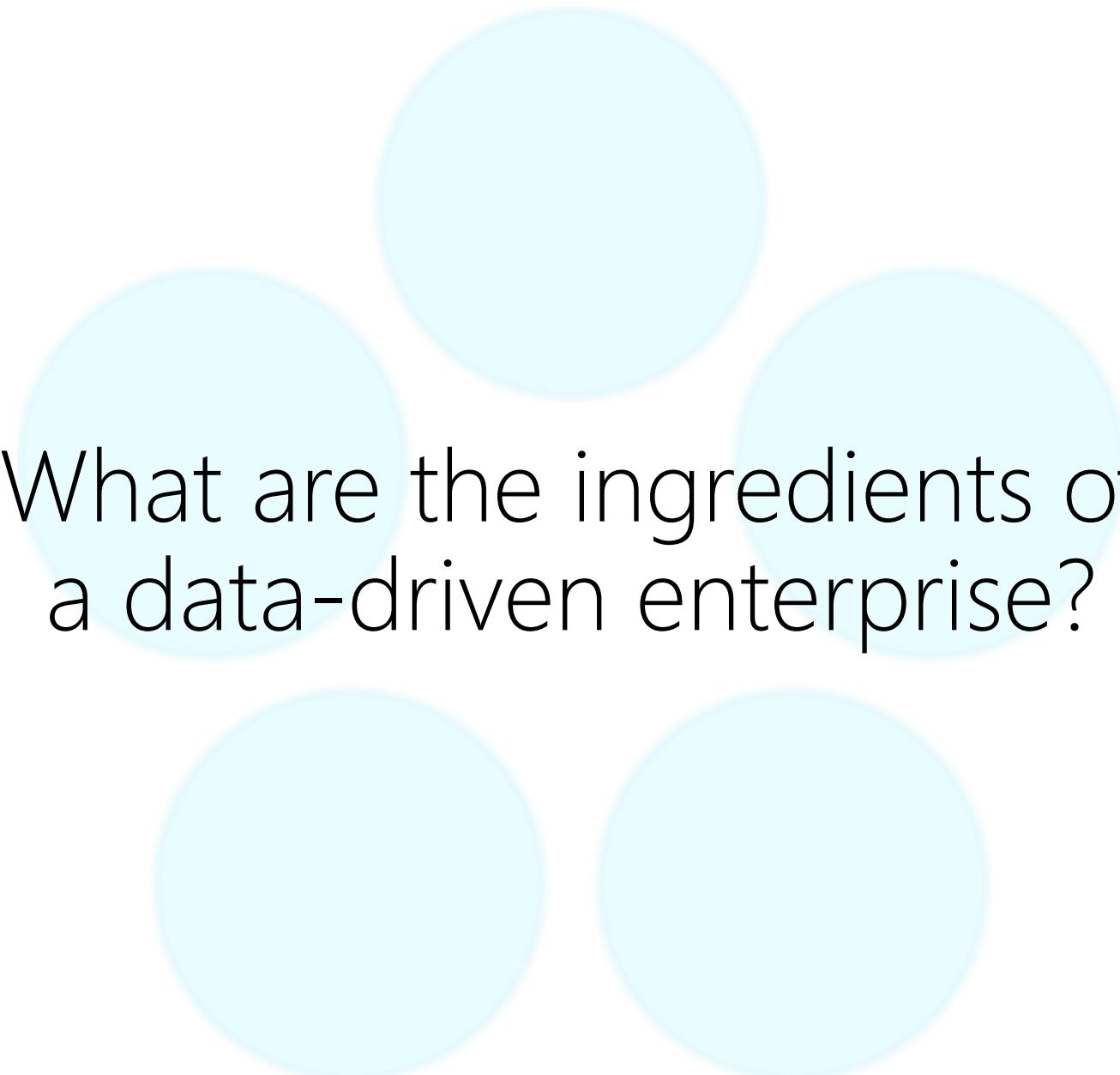




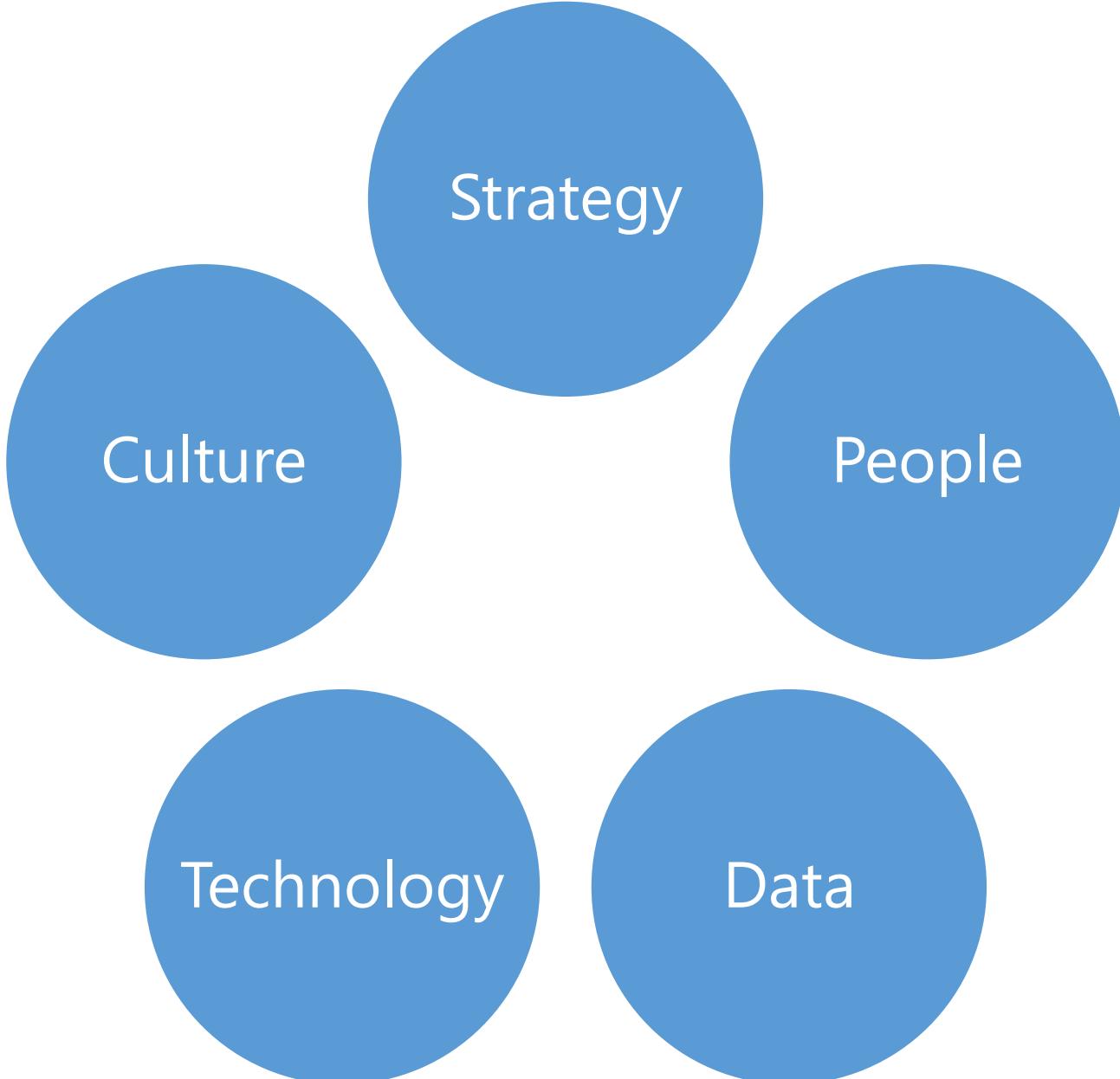
How do I get started?

0110

1001



What are the ingredients of
a data-driven enterprise?



Strategy

Culture

People

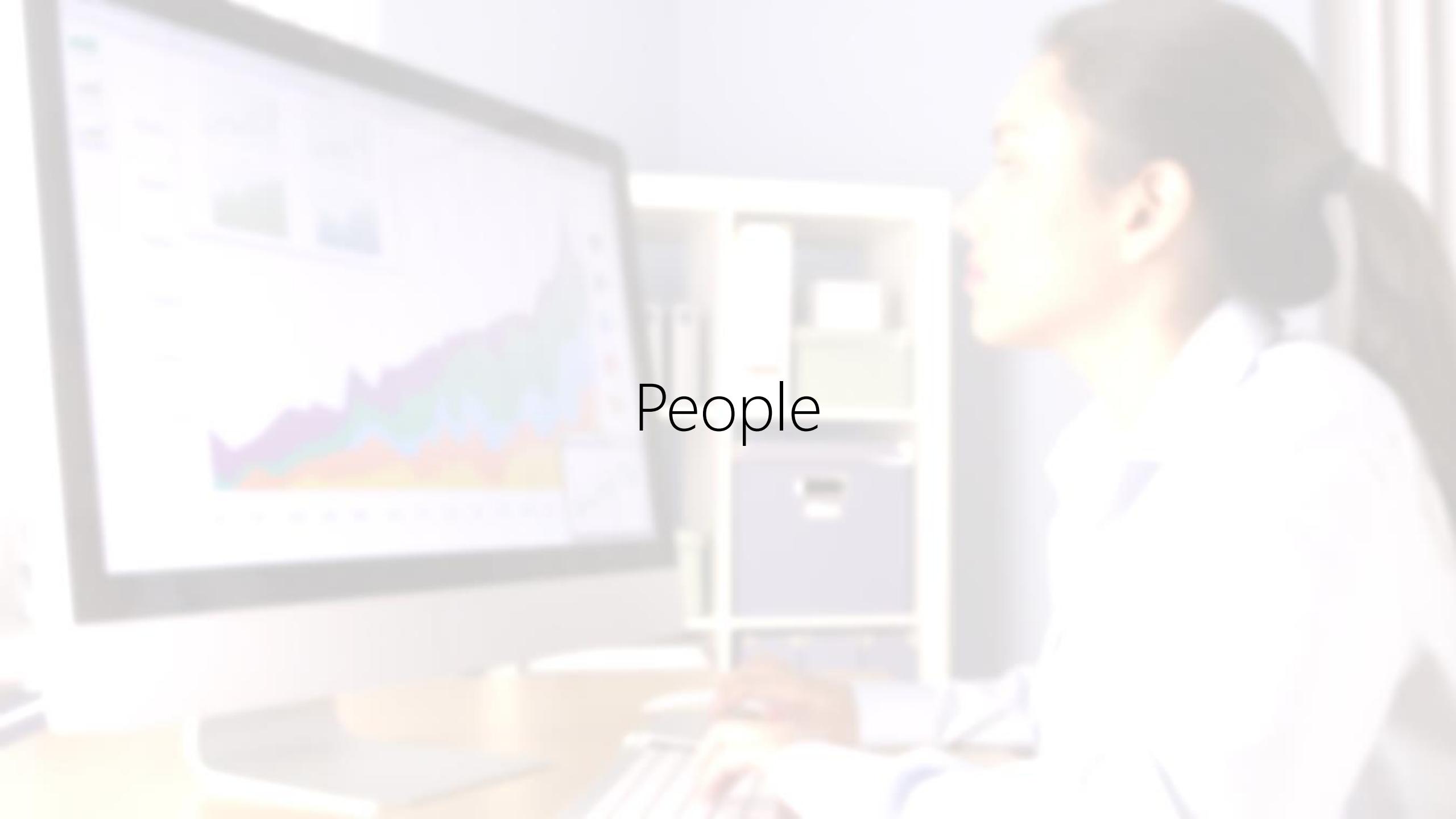
Technology

Data



Strategy



A man in a dark suit and white shirt is seated at a desk, looking intently at a large computer monitor. The monitor displays a complex, multi-colored bar chart with several peaks and troughs. The background is a blurred office environment.

People



Data



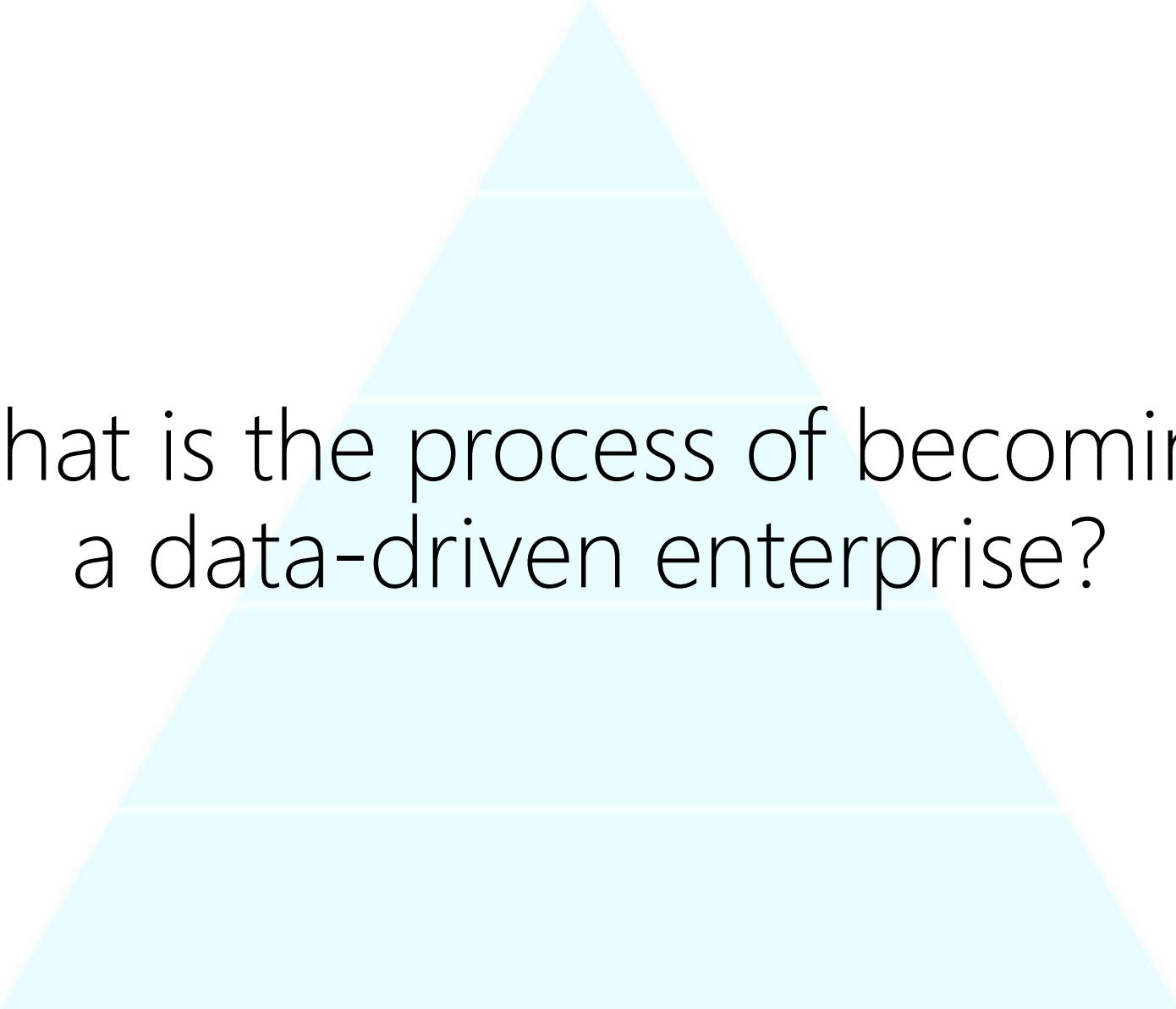
Technology



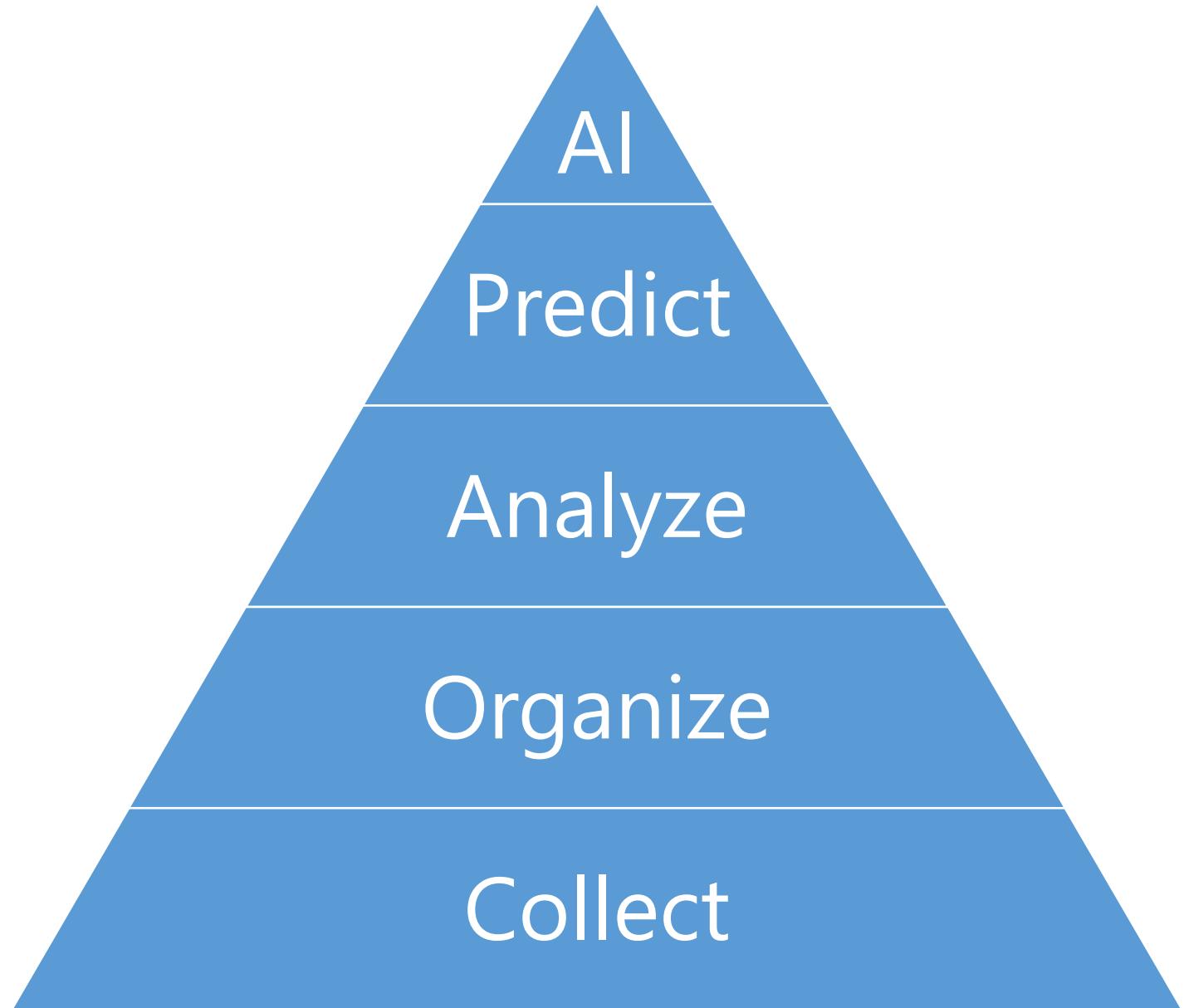
A blurred background image of a person working at a computer. The person is wearing a blue shirt and glasses. There are multiple computer monitors visible, displaying various screens. The word "Culture" is overlaid in large, black, sans-serif font in the center of the image.

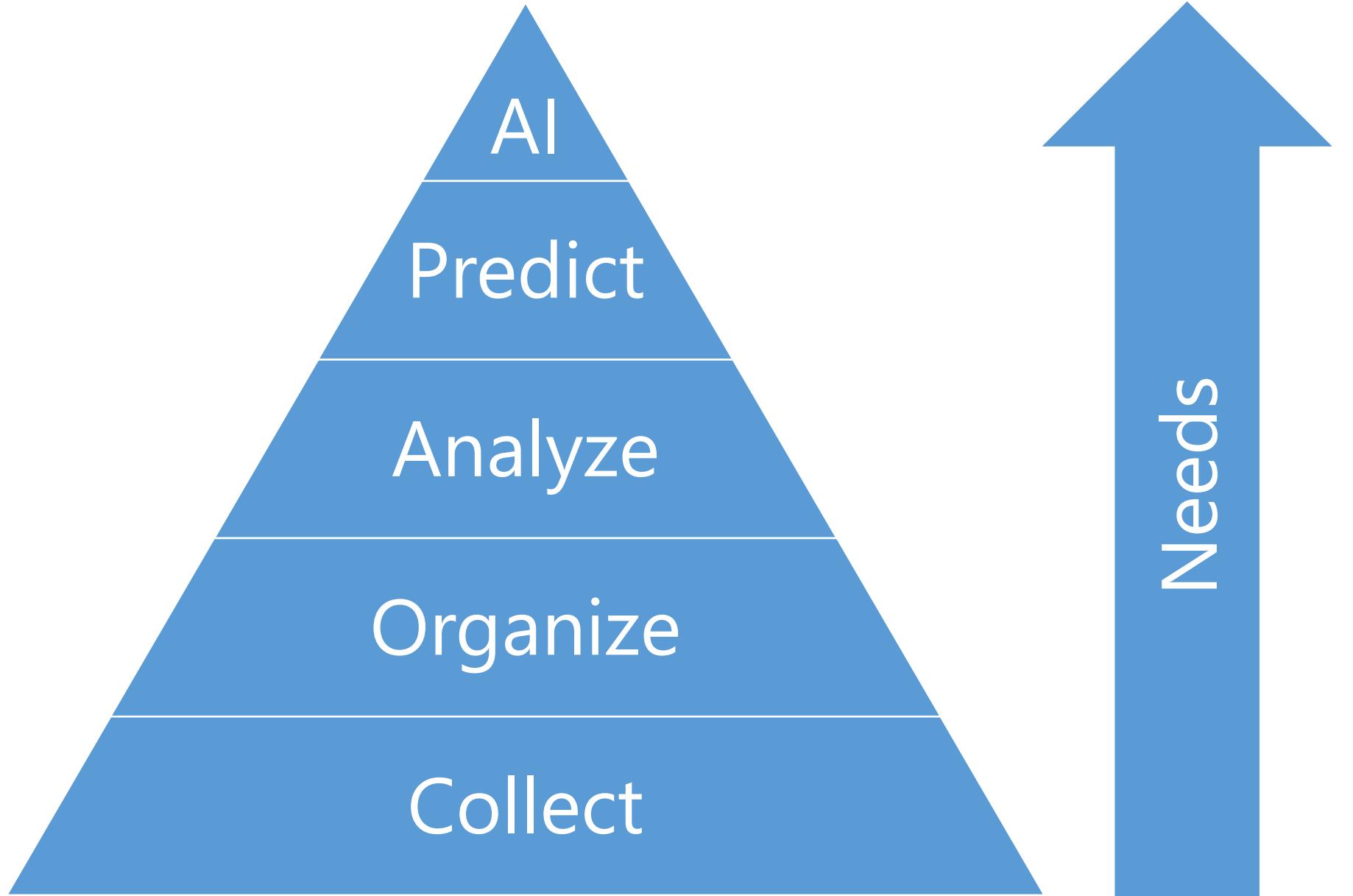
Culture





What is the process of becoming
a data-driven enterprise?





1. Collect



Collect

1. Collect

Transactions

Logging

Digitization



Collect

1. Collect

Transactions

Logging

Digitization

Telemetry

Experiments

External data



Collect

2. Organize



Organize



Collect

2. Organize

Transform

Clean

Store



Organize



Collect

2. Organize

Transform

Clean

Store

Data ETL

Data Warehouse

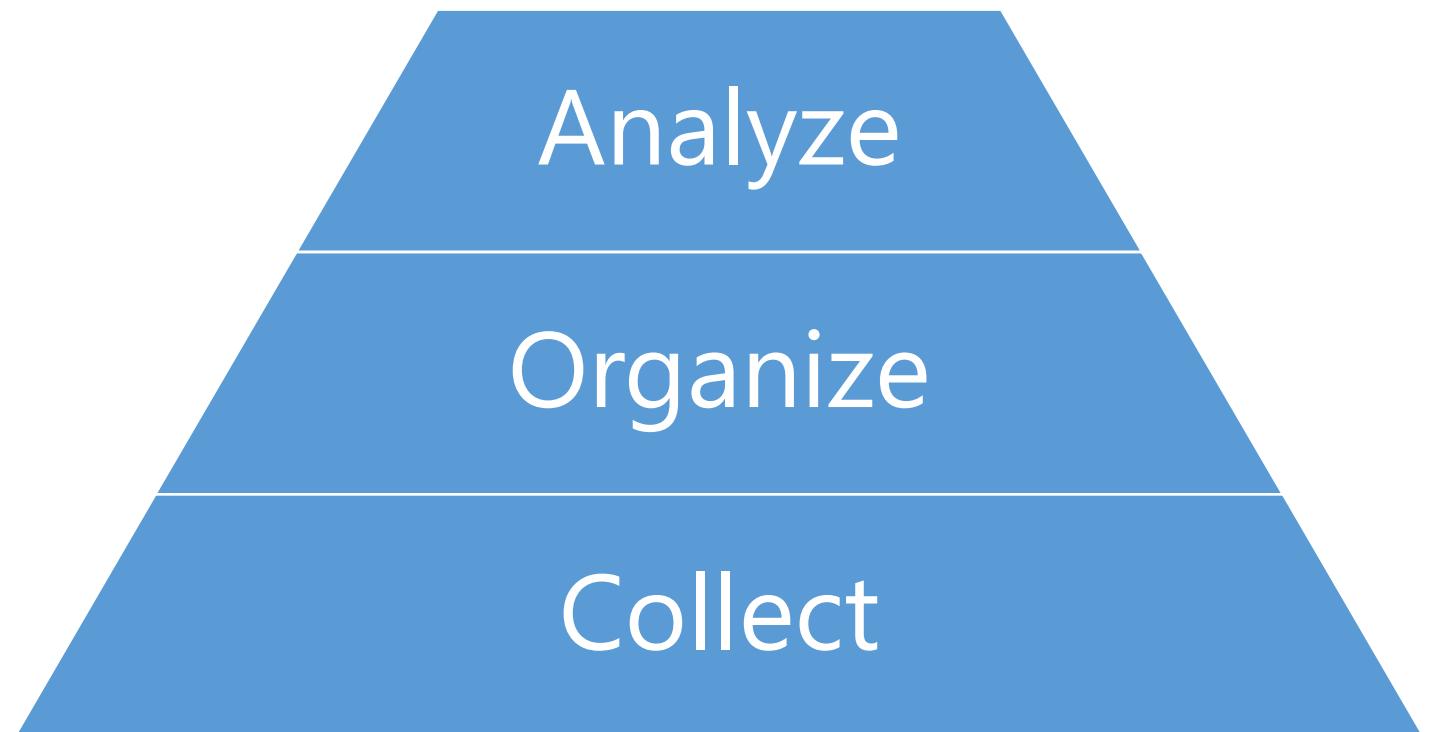
Data Lake



Organize

Collect

3. Analyze

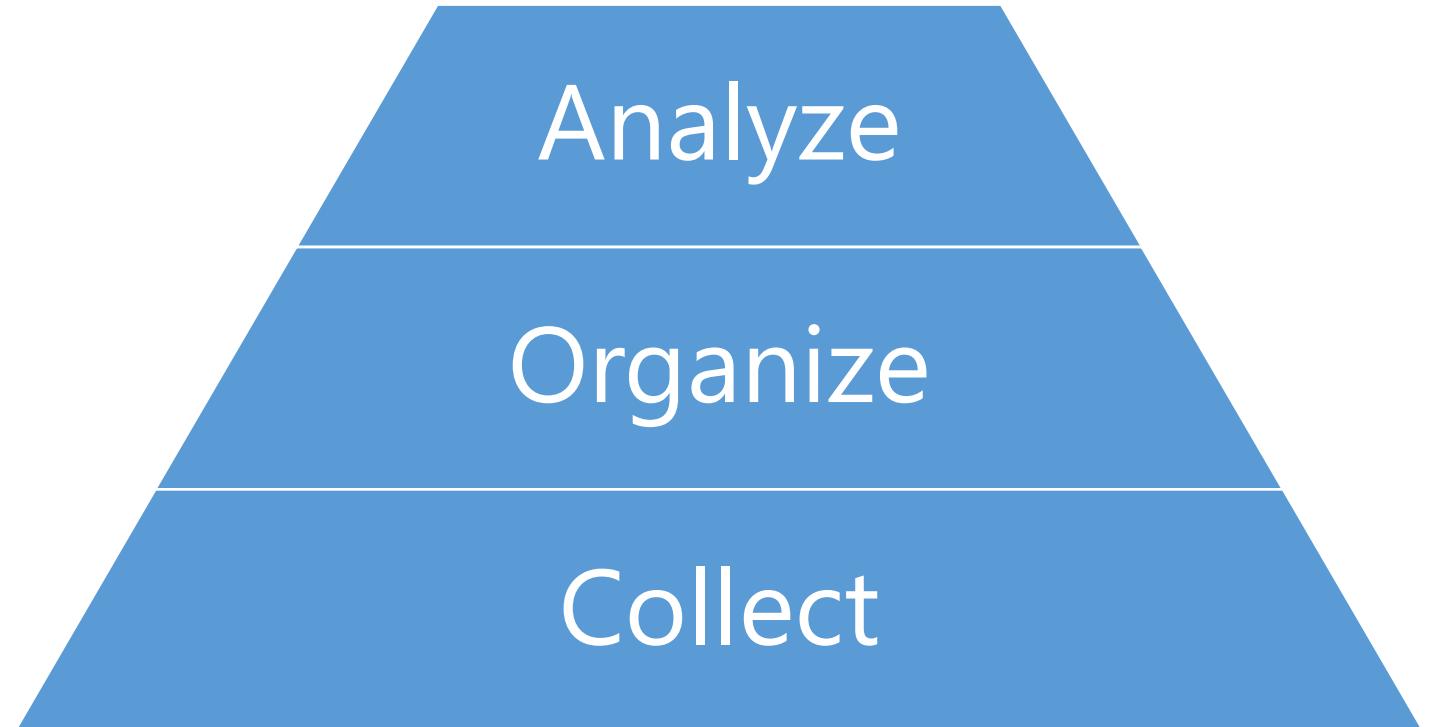


3. Analyze

Reports

Dashboards

KPI monitors



3. Analyze

Reports

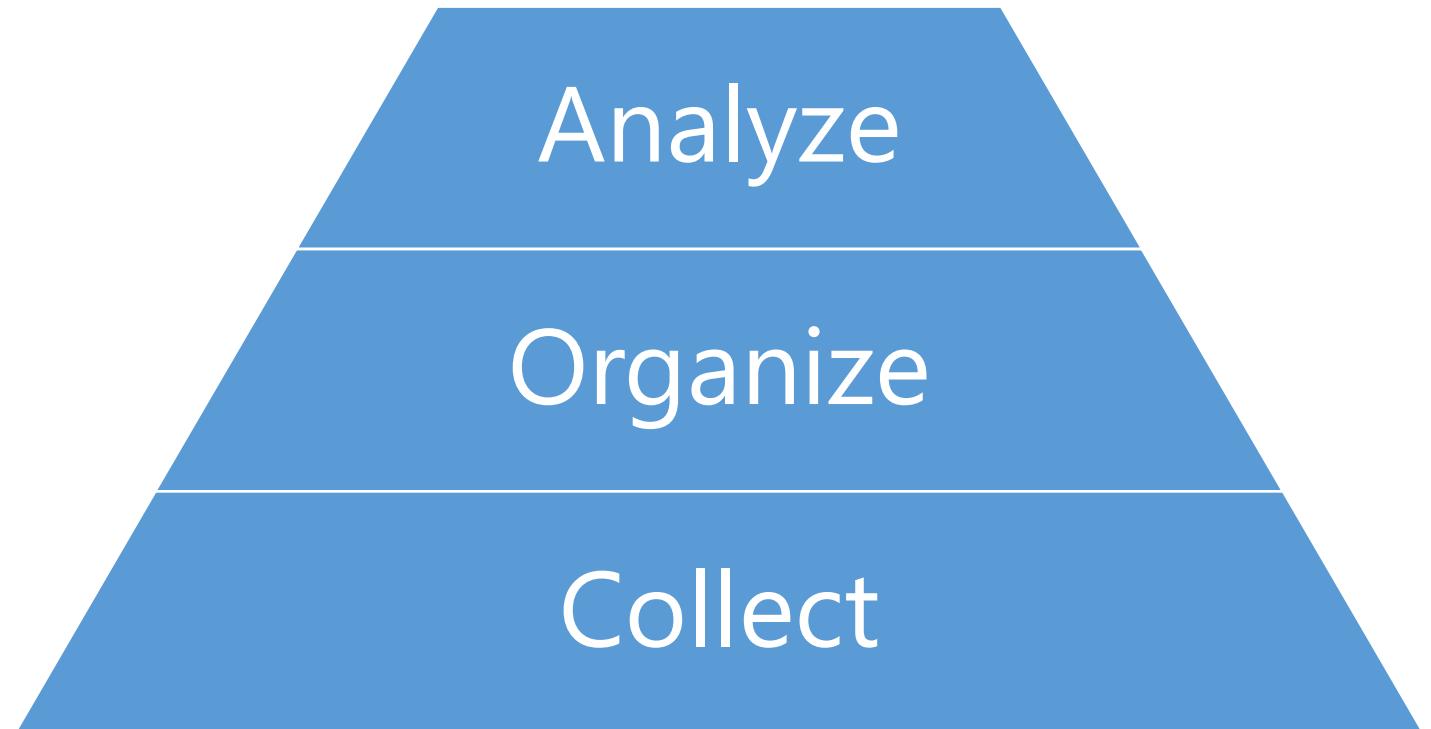
Dashboards

KPI monitors

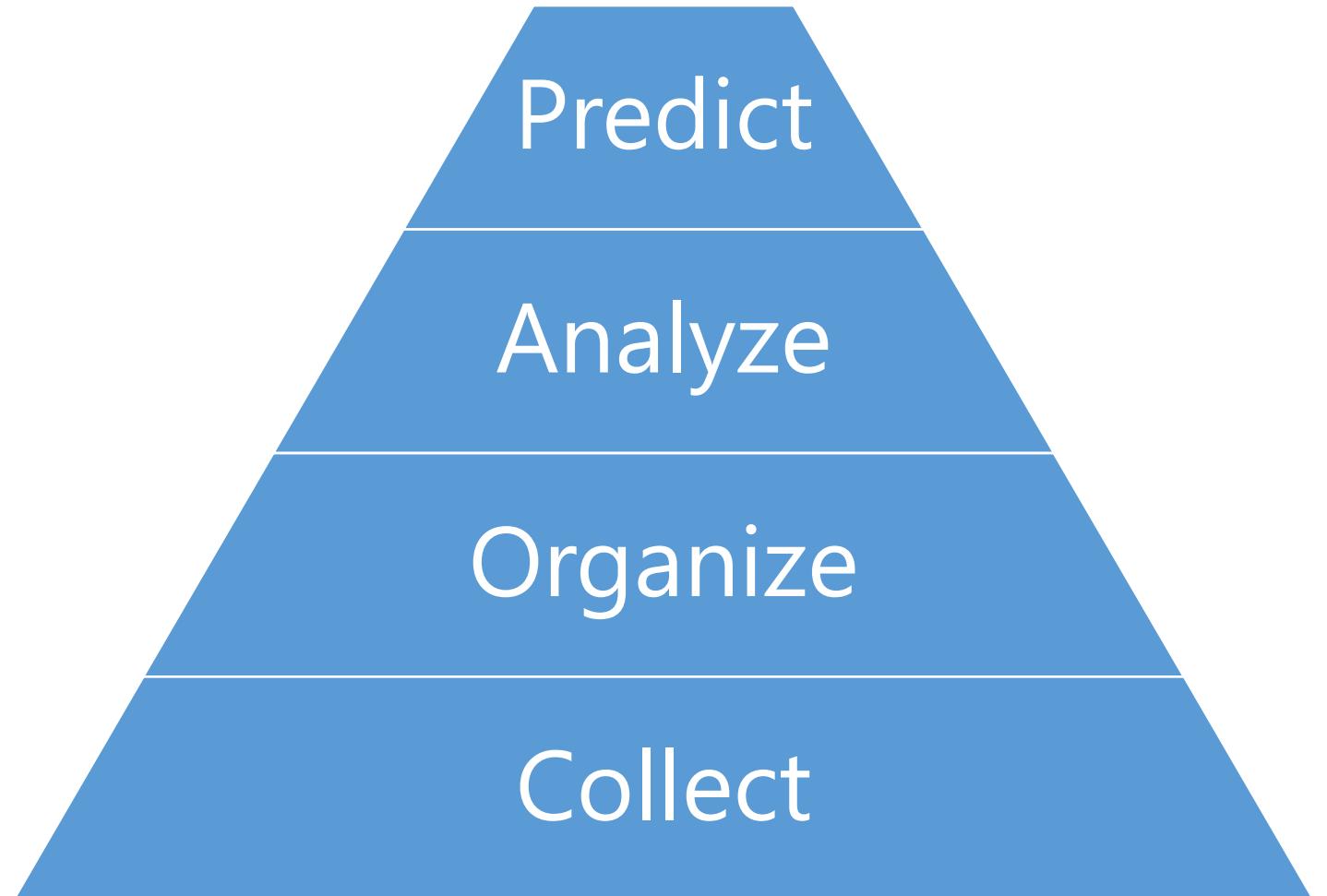
Data mining

Descriptive analytics

Diagnostic analytics

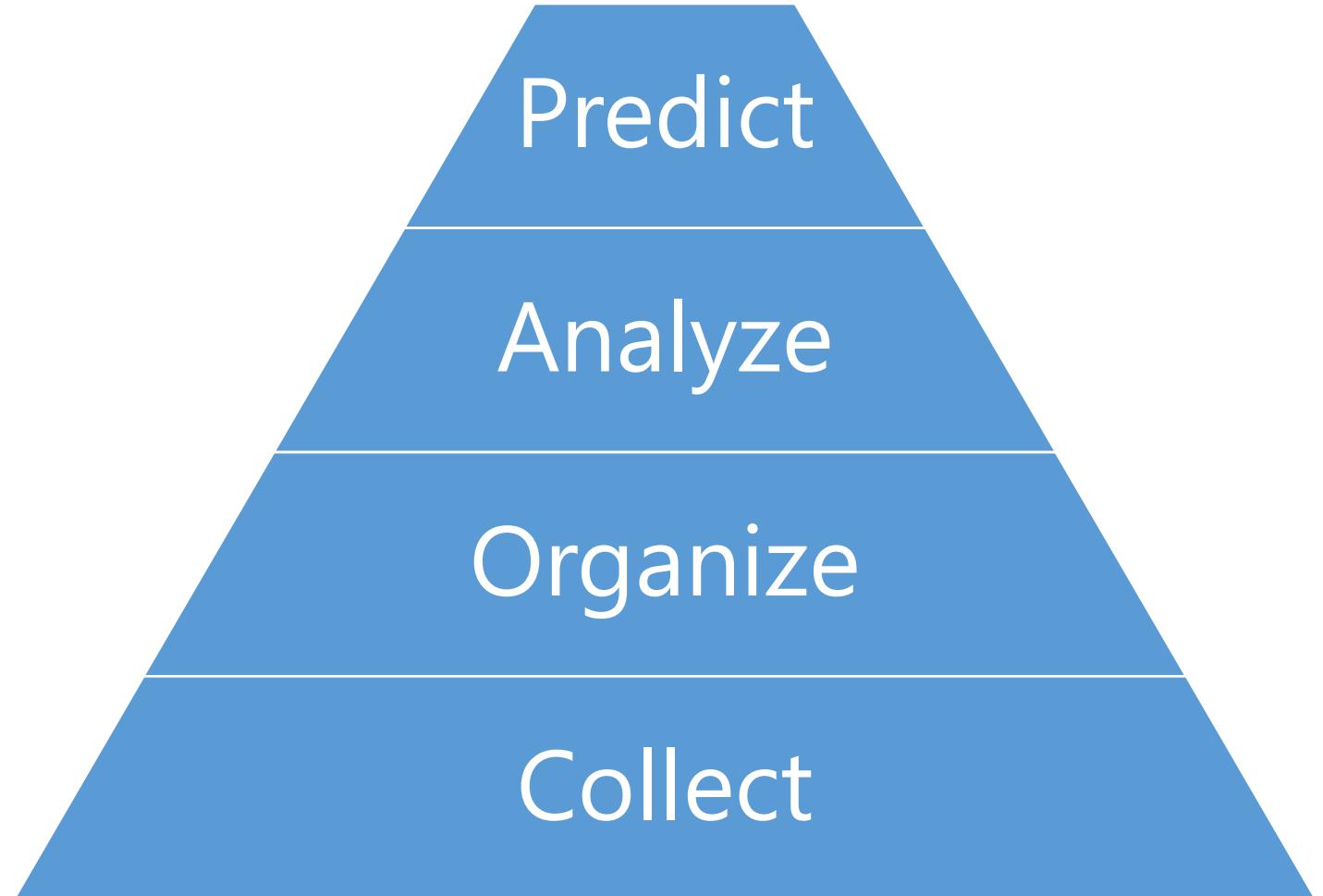


4. Predict

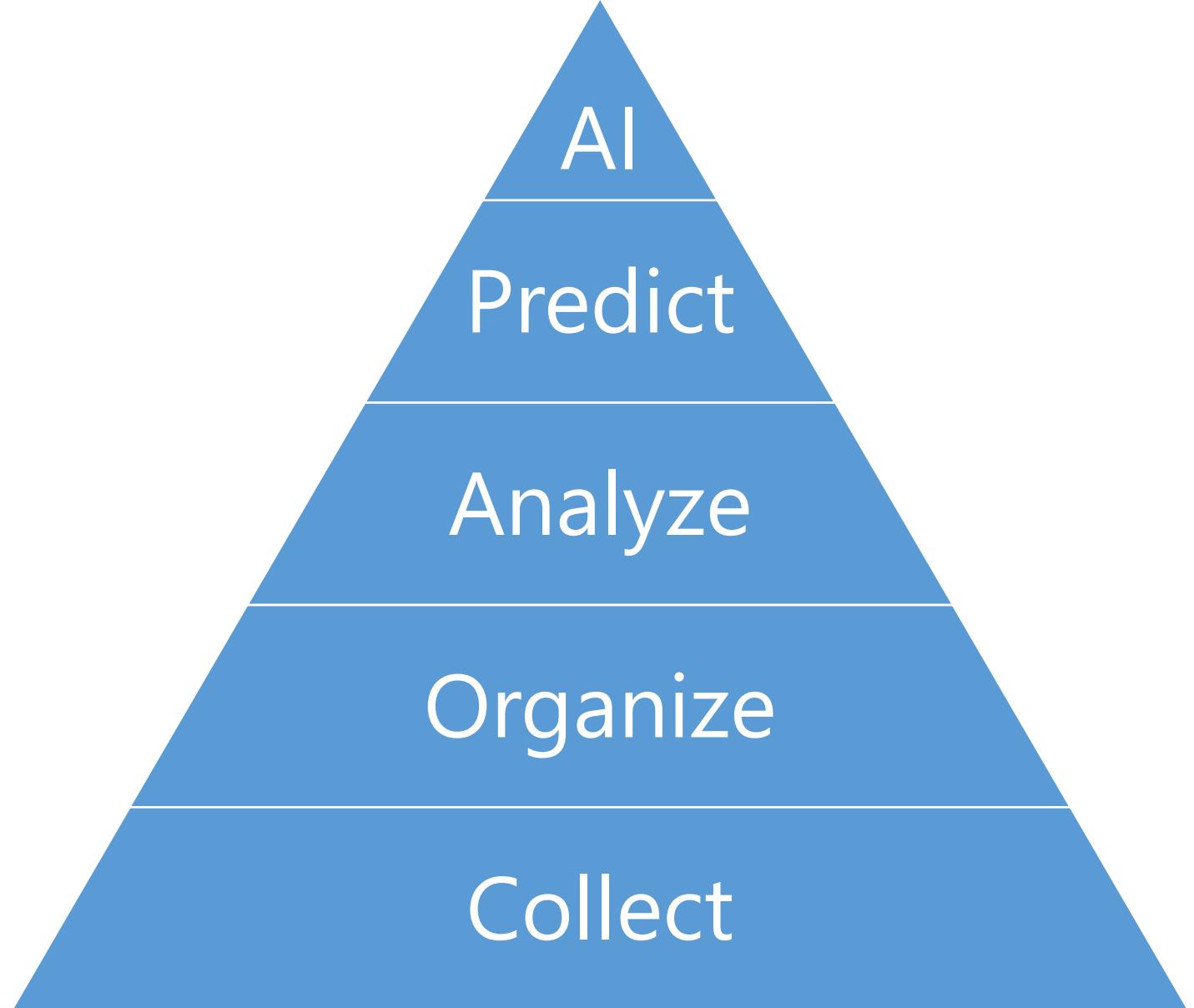


4. Predict

Predictive analytics
Prescriptive analytics
Machine learning



5. Automate

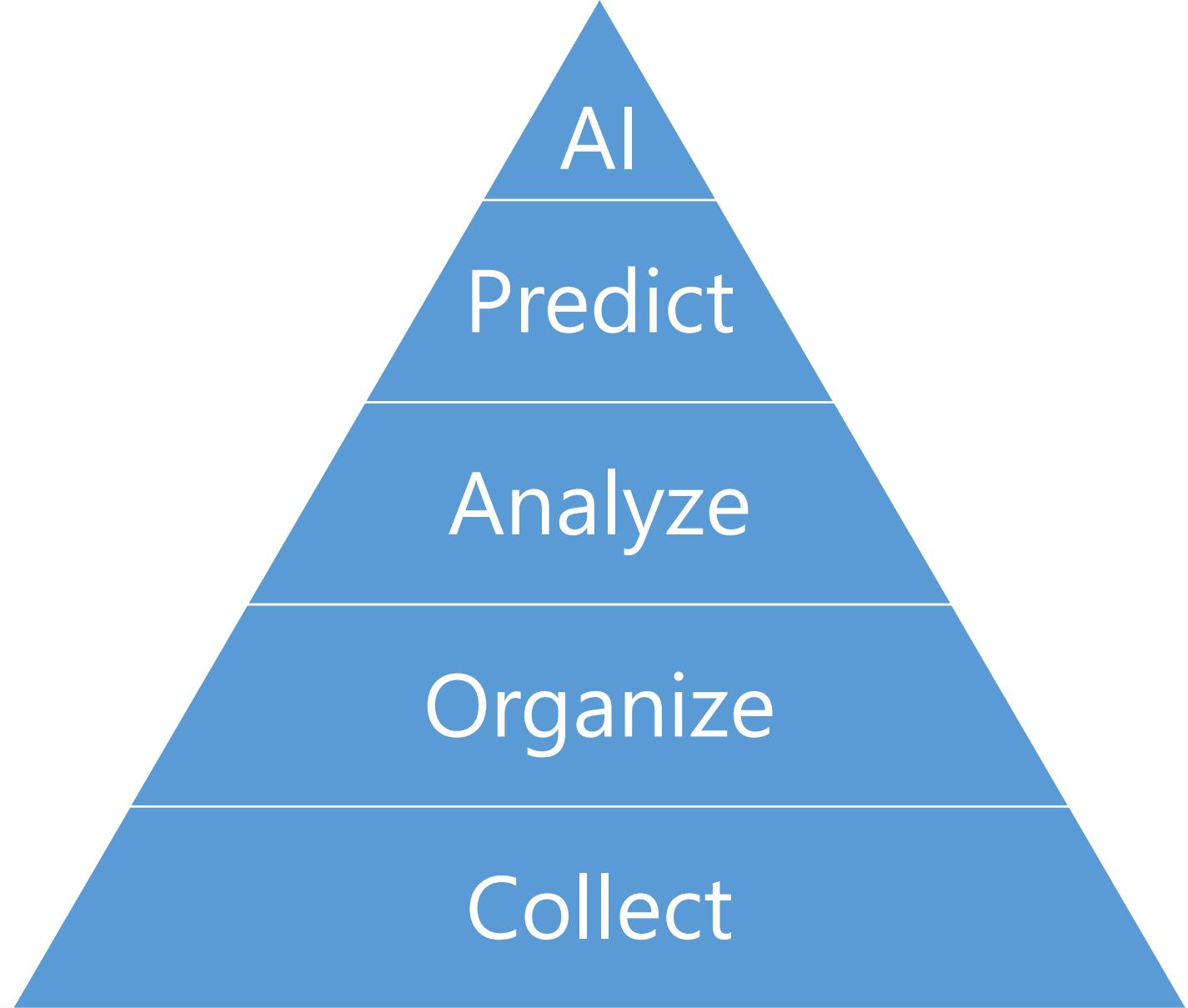


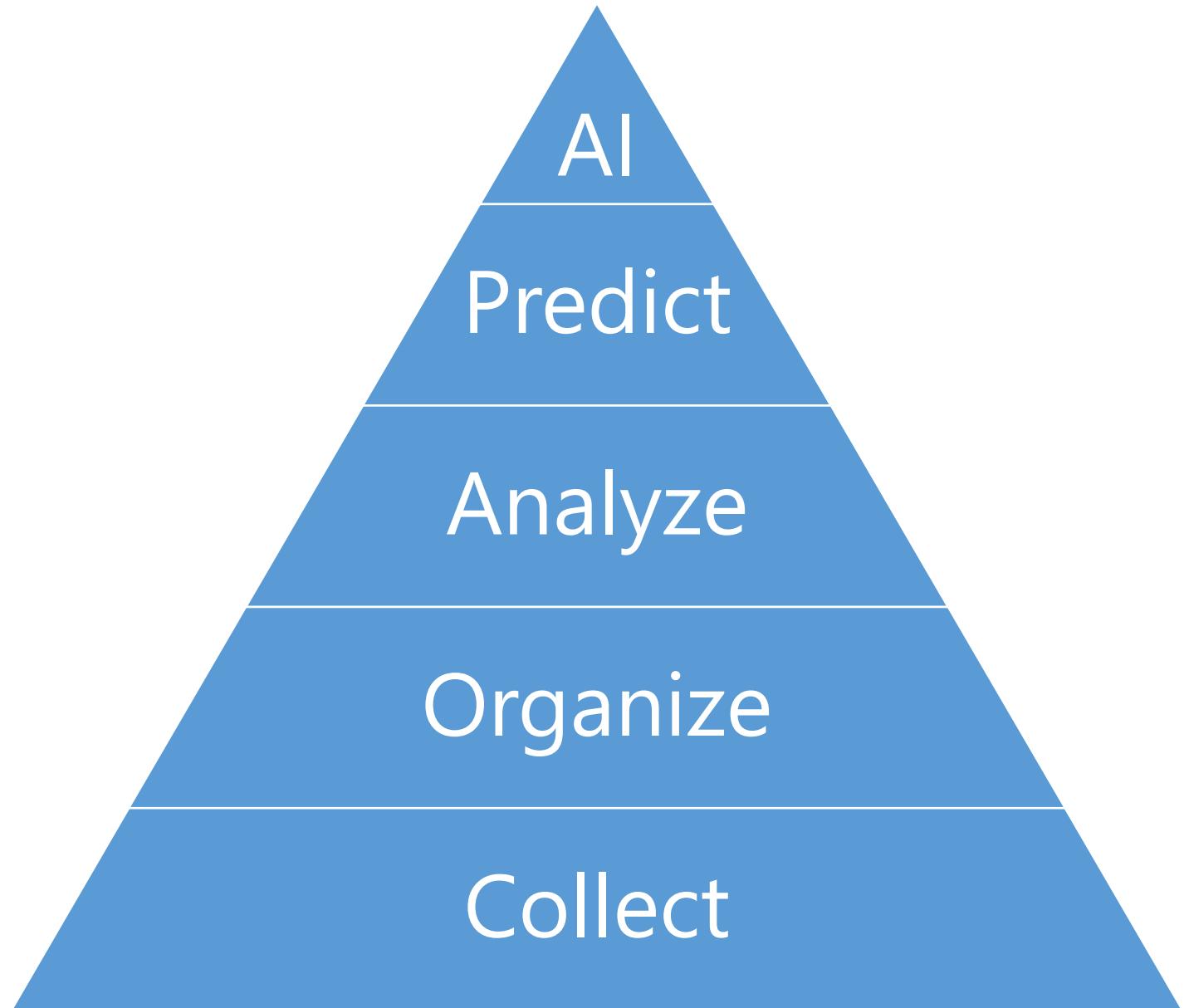
5. Automate

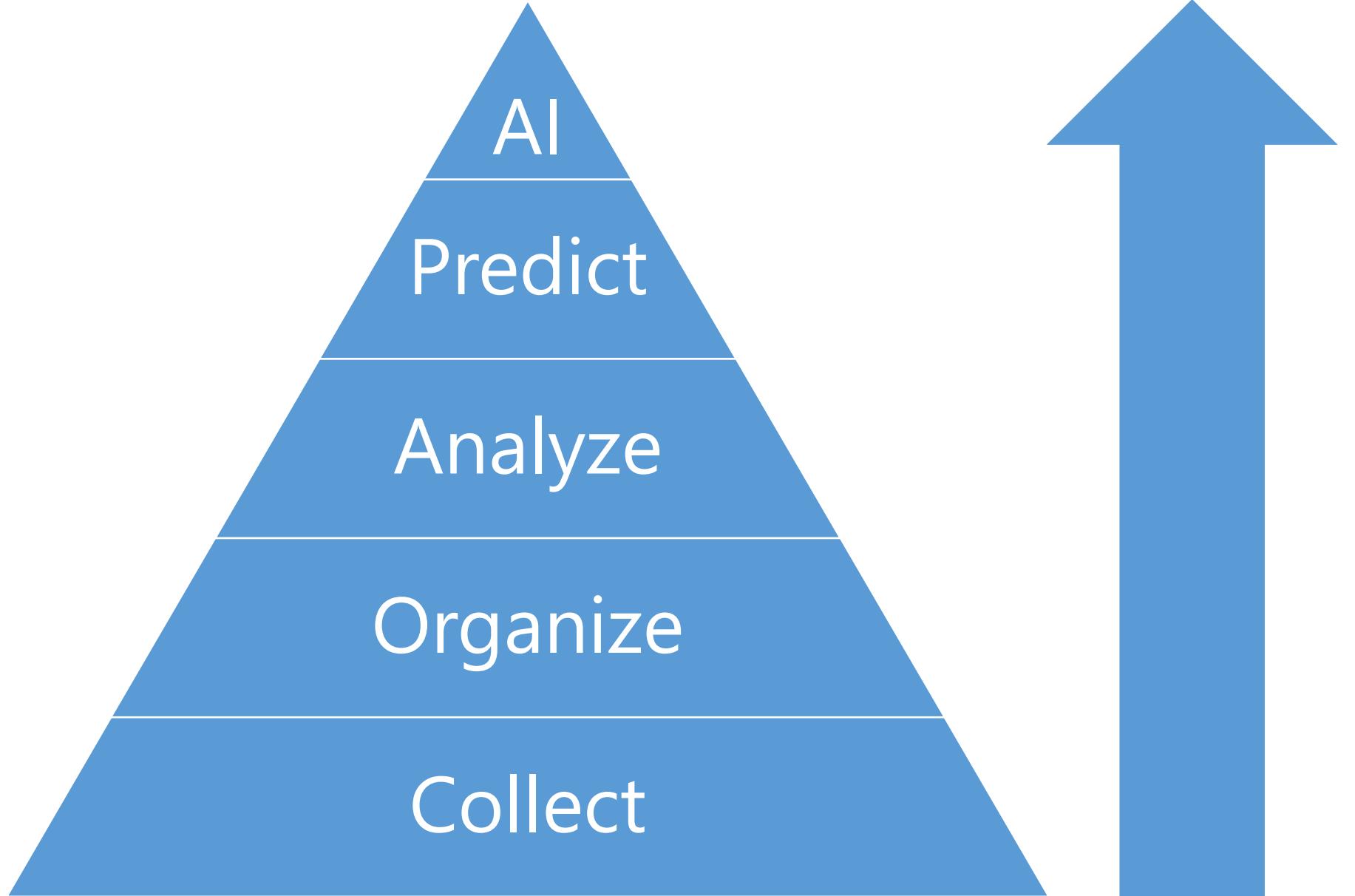
Artificial intelligence

Reinforcement learning

Deep learning







Advice for Success

- Get buy-in from leadership
- Focus on low-hanging fruit
- Don't silo data science teams
- Democratize your data

Advice for Success

Get buy-in from leadership
Focus on low-hanging fruit
Don't silo data science teams
Democratize your data

Embrace smart failure
Focus on feedback
Embed data collection
Avoid the Observer Effect



Where to Go Next?

0110

1001

Where to Go Next

Data Camp: <https://www.datacamp.com>

Pluralsight: <https://www.pluralsight.com>

Coursera: <https://www.coursera.org>

Pluralsight Courses

Data Science: The Big Picture

Data Science with R

Exploratory Data Analysis with R

Data Visualization with R (3-part)

Deep Learning: The Big Picture



<https://www.pluralsight.com/authors/matthew-renze>

News

2017-08-25 - Invitation to Speak at Devoxx Morocco

Very excited to announce that I've been invited to give a keynote in Casablanca at [Devoxx Morocco](#) in November. My keynote presentation will be on [Artificial Intelligence](#).



2017-08-16 - Invitation to Speak at Microsoft Ignite

I've been invited to speak at [Microsoft Ignite](#) in Orlando, Florida in September. This will be my first time speaking at Ignite. Talks will include both Data Science and Machine Learning with R.



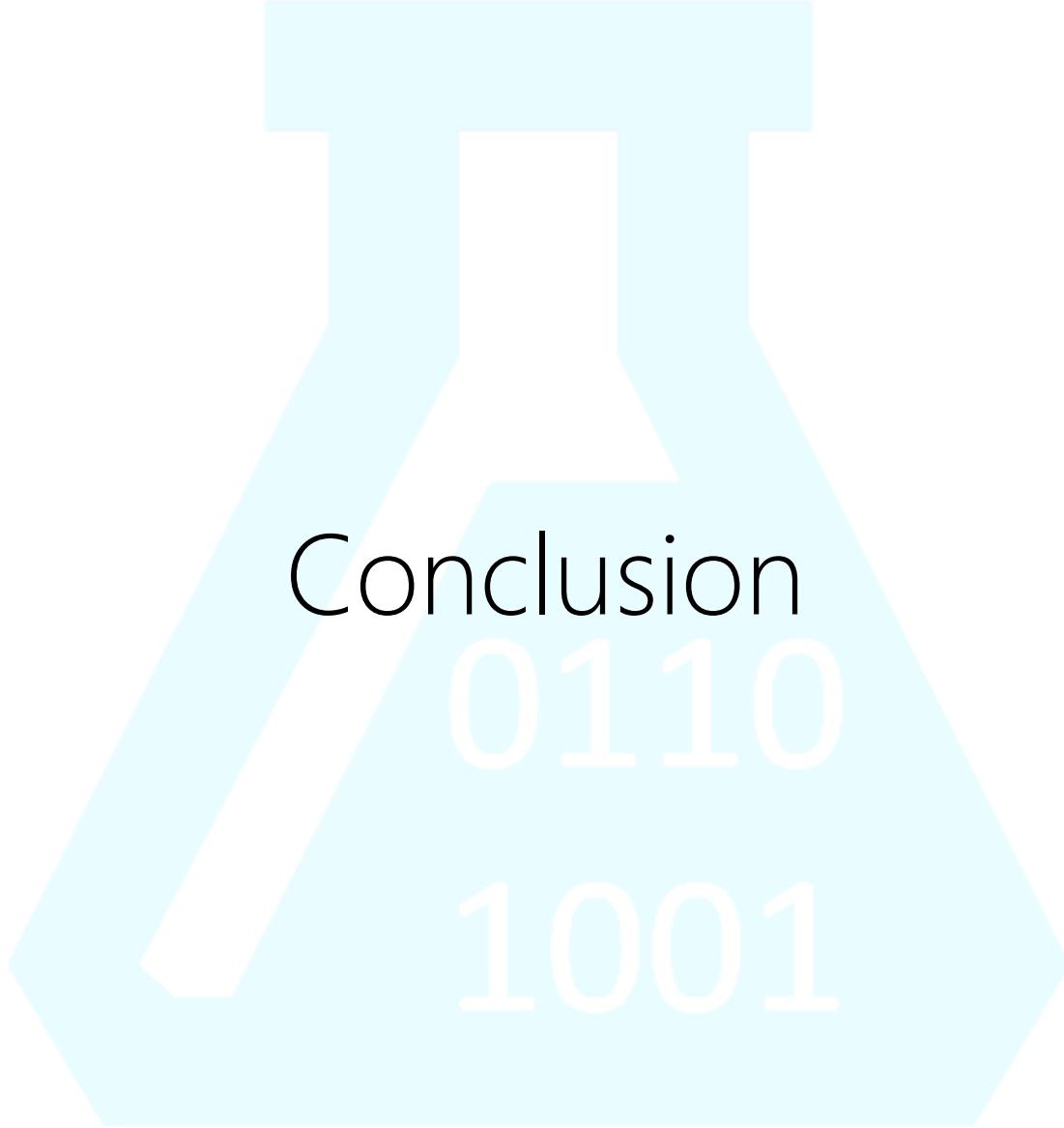
Matthew is a data science consultant, author for [Pluralsight](#), international public speaker, a [Microsoft MVP](#), [ASPIInsider](#), and open-source software contributor.

2017-08-14 - Dev on Fire Interview

Feedback

Very important to me!
What did you like?
What could I improve?

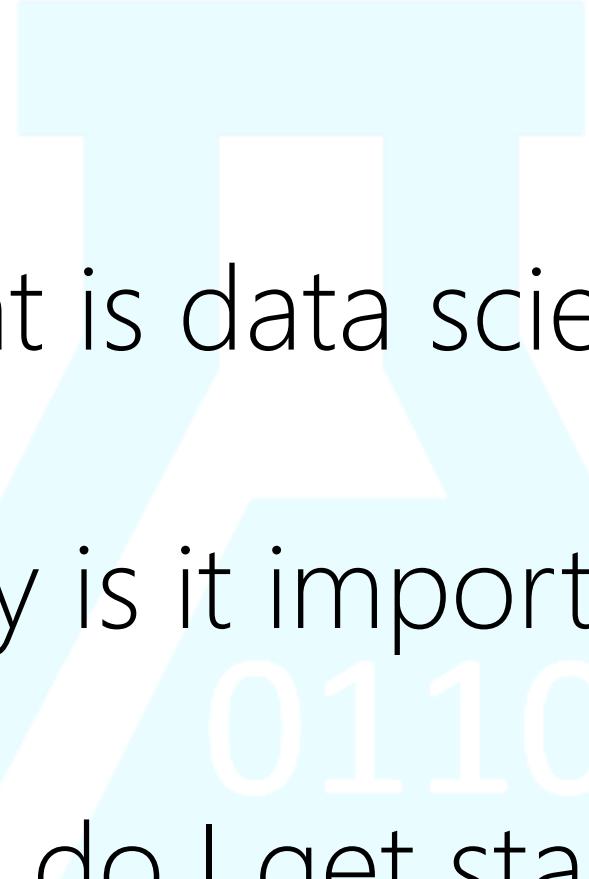




Conclusion

0110

1001



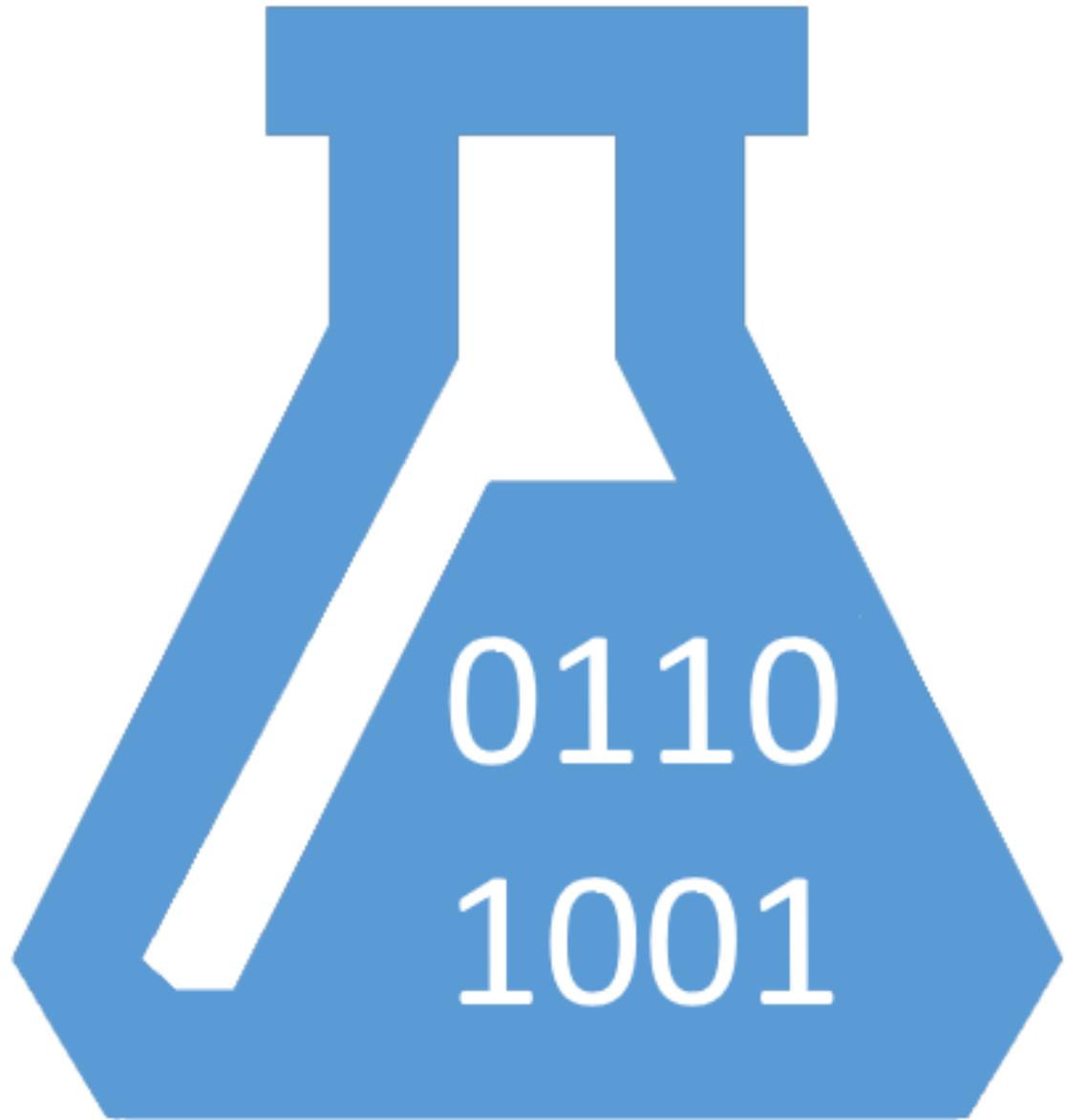
What is data science?

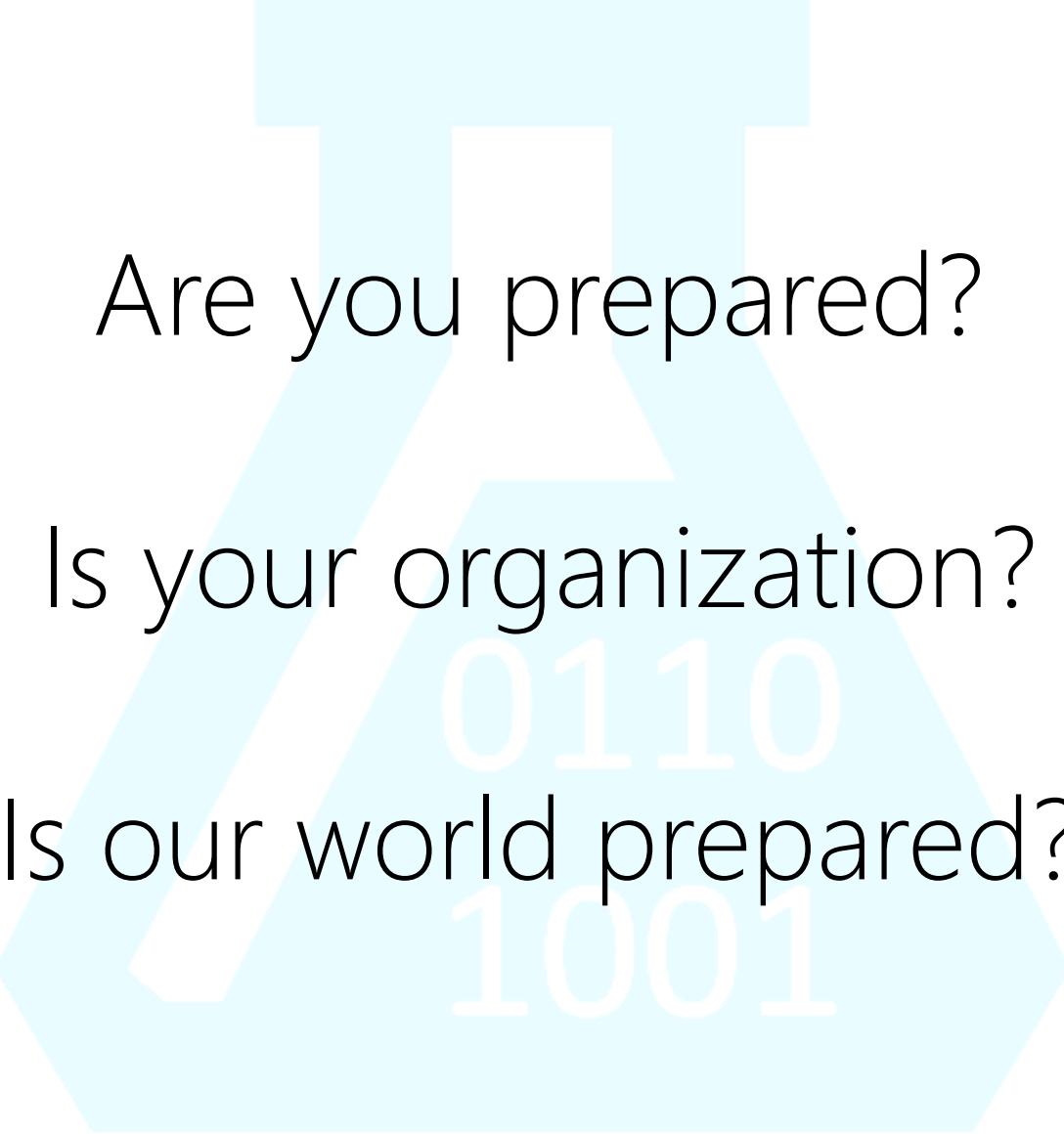


Why is it important?



How do I get started?

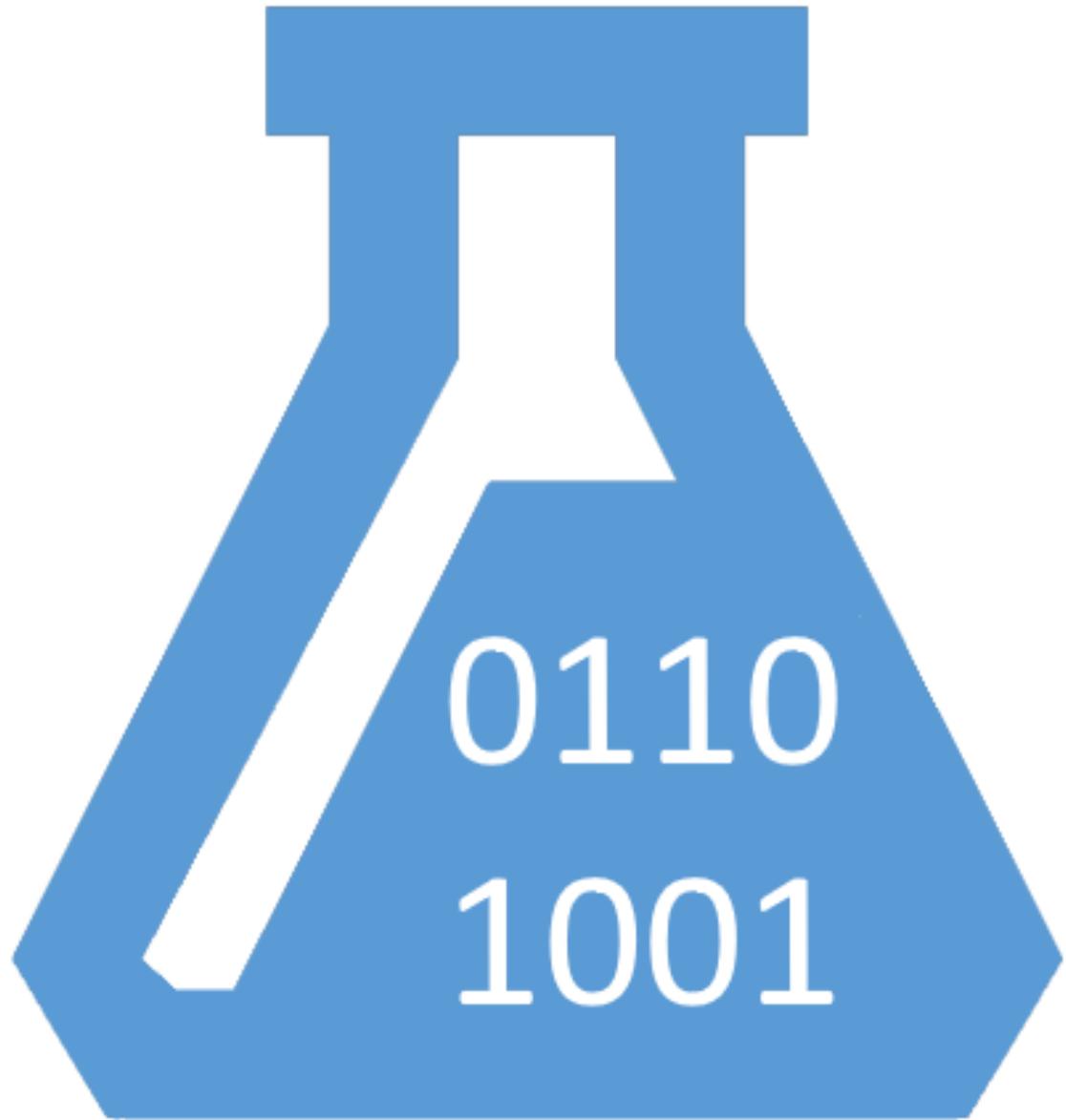




Are you prepared?

Is your organization?

Is our world prepared?



Thank You!

Matthew Renze
Data Science Consultant
Renze Consulting

Twitter: [@matthewrenze](https://twitter.com/matthewrenze)
Email: info@matthewrenze.com
Website: www.matthewrenze.com

