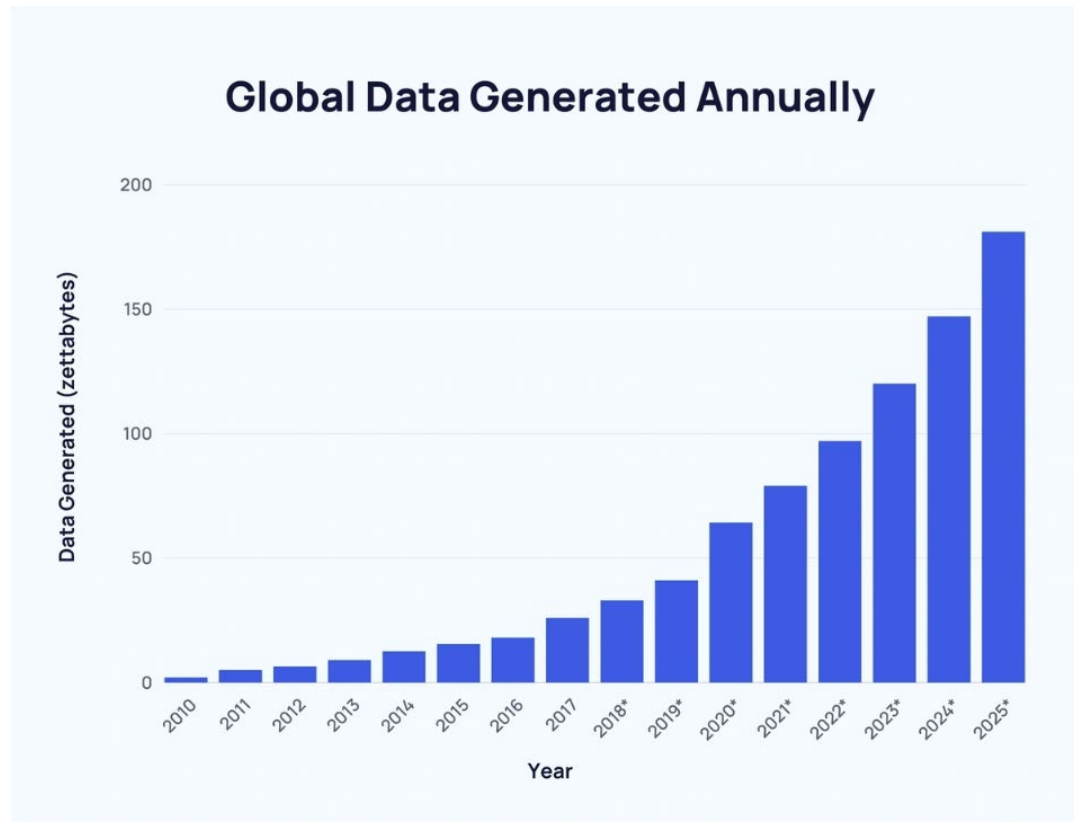# Future trends in Big Data

# Outline

1. The future of data storage

2. Analytics platform and machine learning

3. An increasing need for data scientists

4. Data privacy in the future

5. Fast data and Actionable data

Co-funded by
the European Union

# Future of data storage

# Exponential growth in data volume



**Global Data Generated Annually**

**Source**: www.explodingtopics.com

- **328 million TB** of data created each day.

- **120 zettabytes** generated in 2023

- Data volumes expected to increase **over 150%** by 2025

- Only **2%** of data produced in 2020 is stored until 2021.

- Data storage capacity is expected to grow **19.2%** from 2020 to 2025.

# Where is the data stored

## Where is the data stored?

● Consumer %    ● Enterprise %    ● Public cloud %



**Image Source**: www.itransition.com
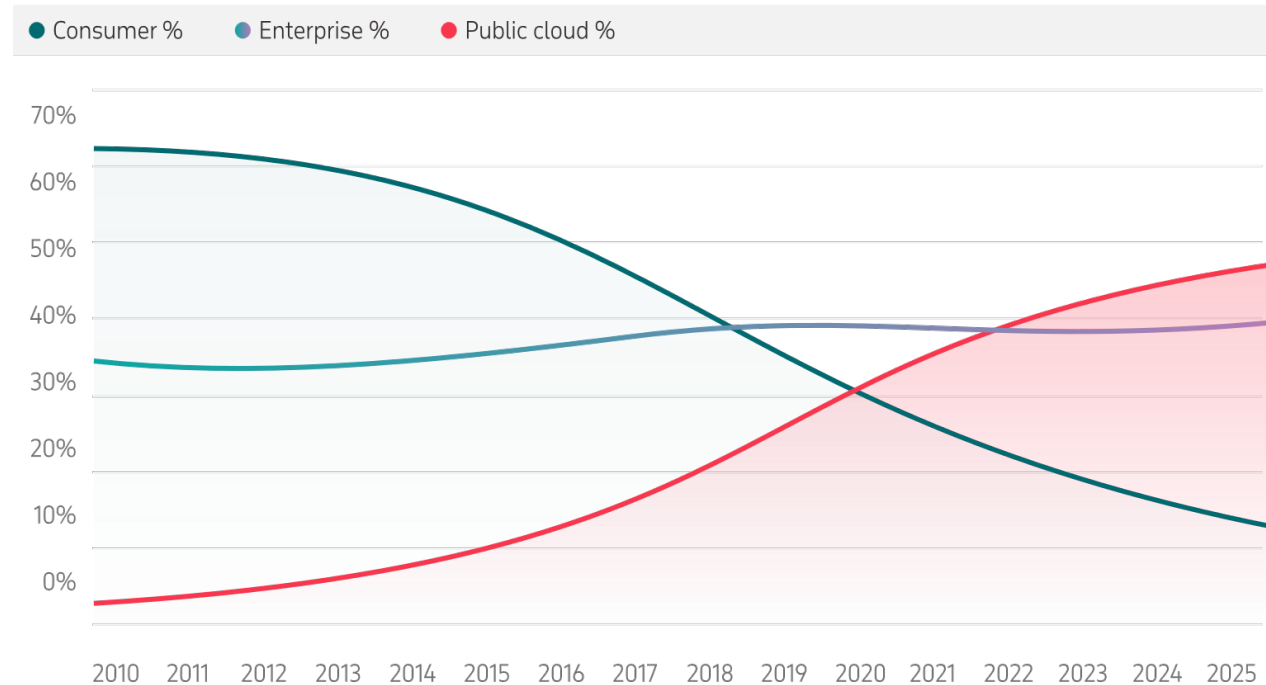
- Fewer data is being stored locally.

- **60%** of corporate data is stored in the cloud

- Cloud storage is expected to have a compound annual growth rate (CAGR) **of 22.3%**

- Data Center Storage market is expected to surpass **40 billion** USD.

- By 2023, the SSD market is expected to reach a value of around **80 billion** USD.

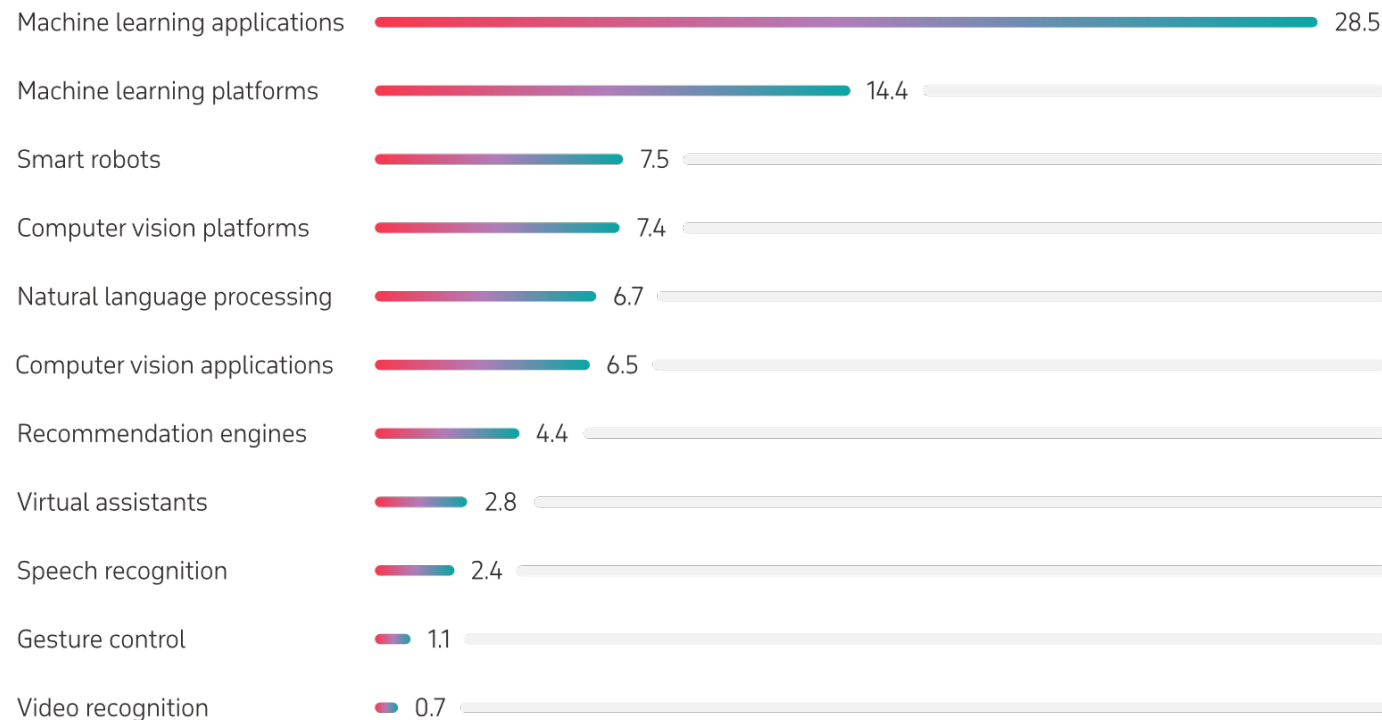# Analytics platforms and machine learning

# Global market spending for machine learning

## Machine learning tops AI dollars

AI funding worldwide cumulative through March 2019, by category | in billion U.S. dollars

| Category | Value |
|---|---|
| Machine learning applications | 28.5 |
| Machine learning platforms | 14.4 |
| Smart robots | 7.5 |
| Computer vision platforms | 7.4 |
| Natural language processing | 6.7 |
| Computer vision applications | 6.5 |
| Recommendation engines | 4.4 |
| Virtual assistants | 2.8 |
| Speech recognition | 2.4 |
| Gesture control | 1.1 |
| Video recognition | 0.7 |

Data source: Venture Scanner, Statista estimates

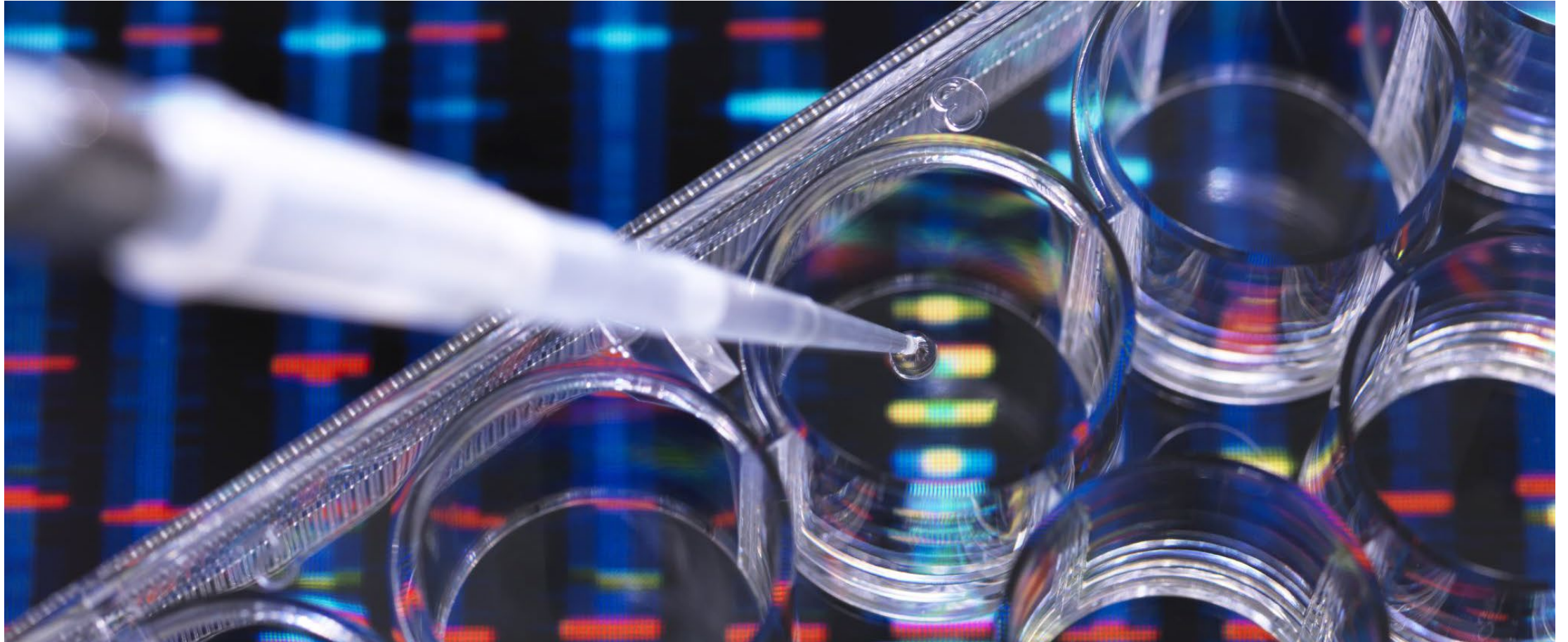**Image Source**: www.itransition.com

➤ **Increasing use of**

- Unsupervised ML algorithms
- Deep personalisation
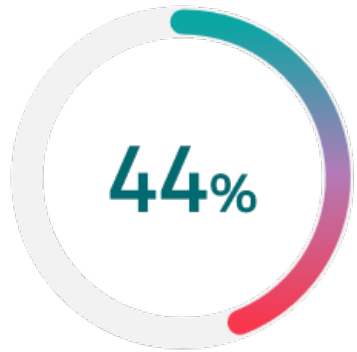- Cognitive services

➤ **More applications for**

- Emotion detection
- Self-drive cars
- Space exploration
- Medical science

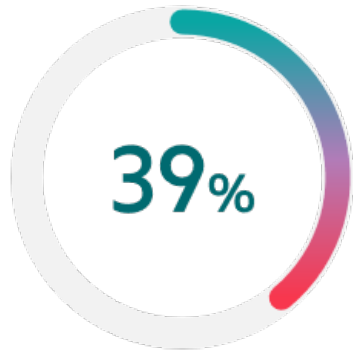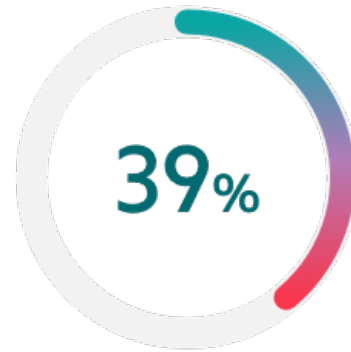# Growing need for data scientists in water sector
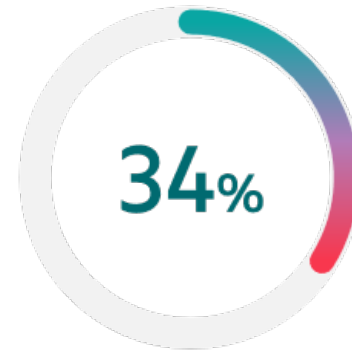
# A need for data scientists



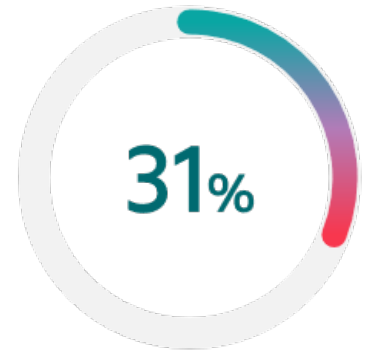| 44% | 39% | 39% | 34% | 31% |
|---|---|---|---|---|
| Big data/ Analytics | Cyber security | Artificial intelligence | Enterprise architecture | Business analysis |

**Image Source**: www.itransition.com

# Necessary skills in the future

- Data platforms and tools

- Programming languages

- Machine learning algorithms

- Data processing techniques

  ➢ Building data pipelines

  ➢ Manage ETL (Extract, Transform, Load) process

  ➢ Data preprocessing

# Data privacy in the future
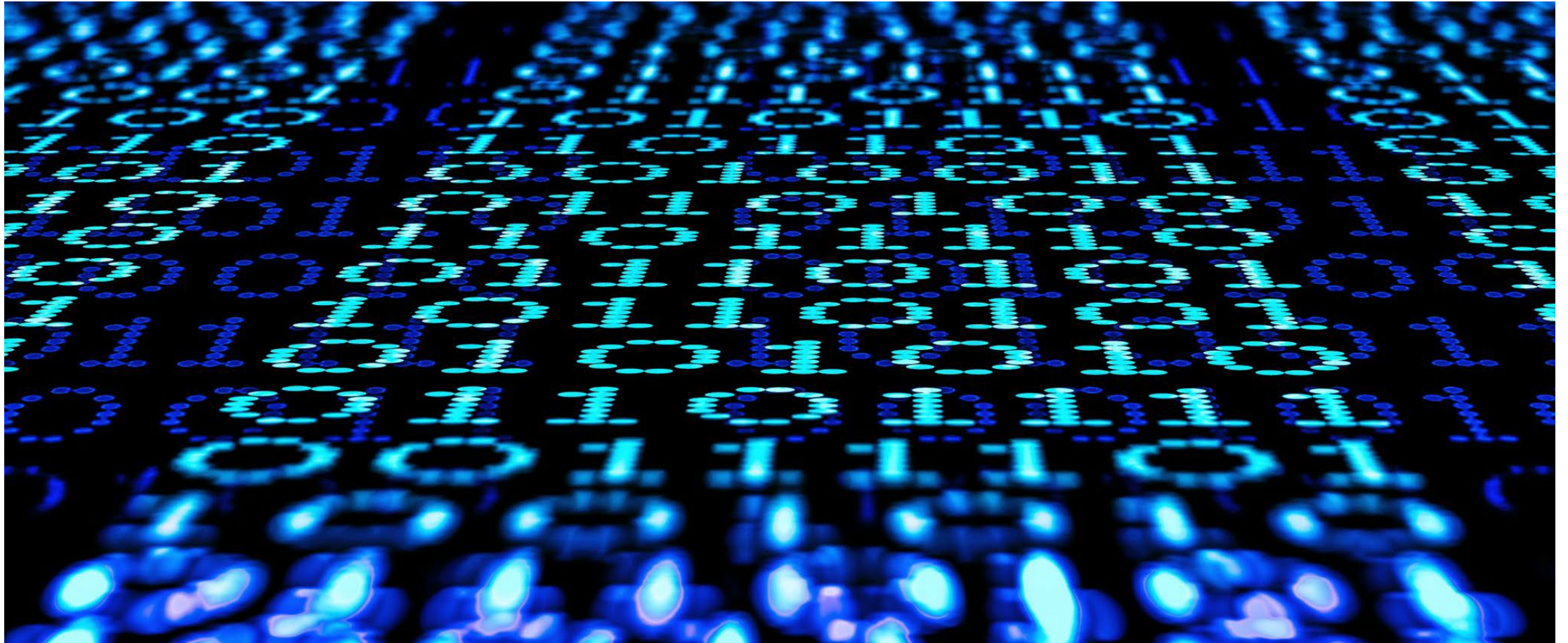
# How does it affect data privacy

- More data in the cloud  →  Less in your local hard drive

- Higher volumes of data in the shared platforms.

- A need for a higher degree of trust in our online storage resources.

- Can we store critical information online?

- How compromised are we in case of a data leak?

# Policies and systems

- Stricter Data Protection Regulations

- Emergence of Privacy-Preserving Technologies

- Privacy by Design

- Data Breach Preparedness and Incident Response

- Strengthened Encryption and Security Measures

- Increased User Control

# Education and awareness

- Ethical Data Practices

- Cross-Border Data Transfers

- Increased Privacy Education and Awareness

- Educate individuals, businesses, and policymakers about
  - privacy rights
  - Risks
  - best practices

- Foster a privacy-conscious culture and encourage responsible data handling

# Fast and actionable data

# Fast data and actionable data

- Bigdata doesn't just mean **big storage volumes** but also **big incoming data rate**.

- Processing data in batches results in a time lag and a subsequent loss of value of that data.

- Can data be processed as soon as it arrives?

- "**Fast and actionable data**" will replace big data in the near future.

# Big data versus Fast-Actionable data

## Bigdata

- *Is like looking for a needle in a haystack*

**Characterised by 3 V's**

- ➤ Massive Data **Volumes**

- ➤ Data generated at high **Velocities**

- ➤ Contains diverse **Variety** in the dataset.

## Fast and actionable data

- *Is like detecting the needle before it hits the haystack.*

**A subset of Big data, which also includes**

- ➤ Real-time processing

- ➤ Immediate action

- ➤ Targeted relevant insights

# Status quo

- Most current Big data platforms can handle and store large data volumes

- However, most fall behind in analysing the information in real-time

- To build a Fast data infrastructure, we need
  - An ultra-fast streaming system capable of delivering events as soon as they arrive
  - A data storage capable of processing each item as fast as it arrives

# Important feature of fast data systems

- Redundancy and scalability benefits

- Lean in-memory storage

- High per-node throughput

- Processing at ingestion time

- Handle concurrency and data divergence problems

- Should isolate operations from processing cores

# What does the future look like?

# Democratization of data

*The increasing accessibility and availability of data to a broader range of individuals and organizations. It empowers people with the ability to access, analyze, and utilize data for various purposes. Active developments are underway in the areas of*

- Open Data Initiative

- Data Visualization and Analytics Tools

- Crowdsourced and User-Generated Data

- Data Sharing Partnerships and Collaborations

- Personal Data Ownership and Control:

# No-code solution (NCS tools)

*Why are NCS tools gaining popularity?*

- Not everyone has to know how to code

- Enabling domain knowledge experts to generate value from big data

- Allow users to create applications and automate processes without writing code.

*Future trends and new developments*

- NCS platforms may expand further by integrating more complex and feature-rich applications.

- As the adoption of no-code platforms grows, there will be an increase in industry-specific templates and solutions

- Greater focus will be on enhancing collaboration and sharing capabilities

- Mobile app development is expected to become more accessible through no-code solutions.

# Standardizing definitions for data
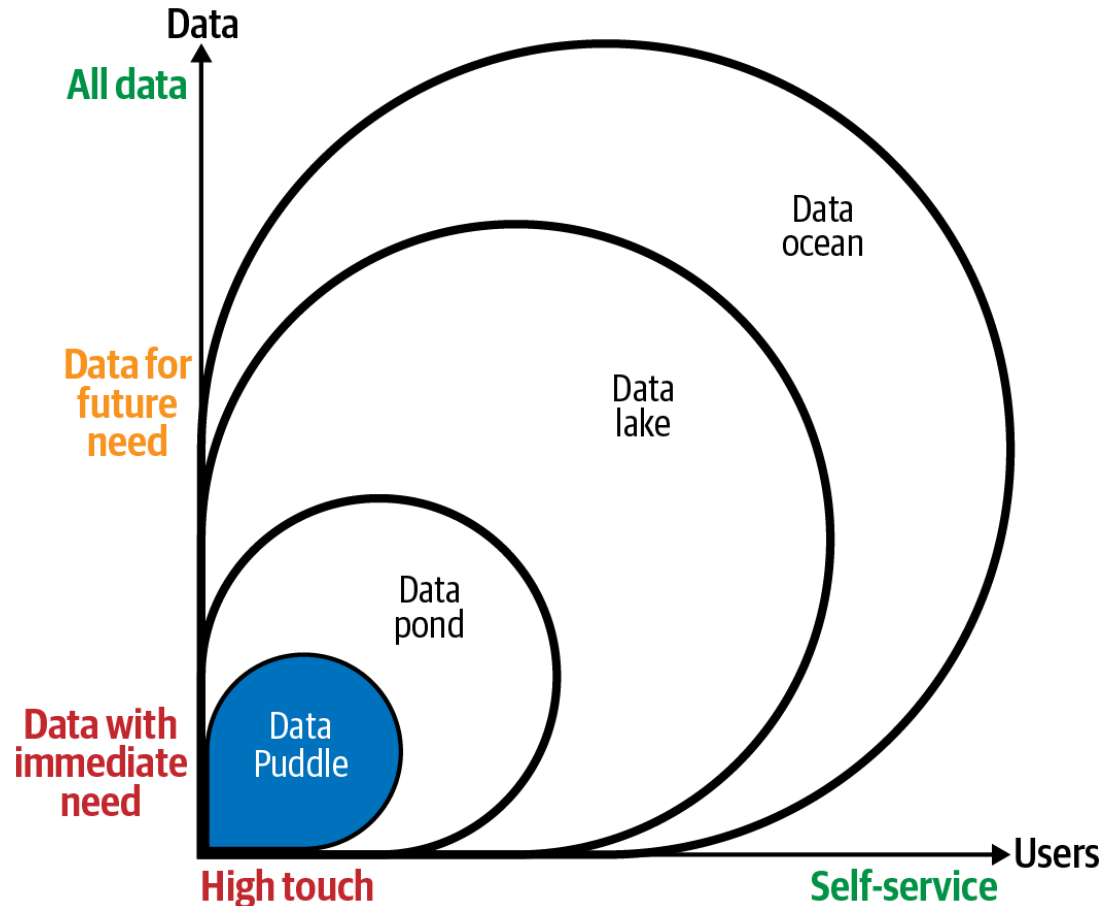


- Can there be a universally accepted definition of data types?

- From puddle → pond → lake → ocean

- Structured, unstructured, or hybrid?

- How to effectively classify data?

- Data types tailored to the water sector.

# Data mesh

*Data Mesh is an emerging approach to data architecture that aims to decentralise data ownership, improve data quality, and foster data-driven collaboration across organisations. A relatively new concept, with developments underway in the areas of*

- Interoperability and collaboration across multiple stakeholders in the Data Mesh.

- Data security and privacy in Data Mesh architectures.

- Emergence of dedicated tools and platforms that can support Data Mesh architecture.

- Specialized roles, such as data product managers, data domain owners, and data stewards, will be established to oversee the development, operation, and governance of data products and domains.

- Possibility to integrate Data Mesh into existing technologies and architectures (such as Kubernetes, microservices, and serverless computing)

**Data Mesh will continue to grow, facilitating knowledge sharing, case studies, and lessons learned**

# Thank you