zyte | Web Scraping Maturity Model

A roadmap for success in web scraping

1 Planning & Design

2 Compliance

3 Building Quality

4 Scaling & Maintaining

What is the Web Scraping Maturity Model?

Levelling up the web scraping field

With over 13 years of experience helping customers worldwide extract web data at scale with pioneering technology and services, we've discovered some patterns that consistently yield effective web scraping projects.

We've found a relationship between a team's level of maturity in each of the four dimensions of web scraping, and the results those teams are getting in terms of return on investment, data accuracy, and consistency.

These four dimensions, and the four stages of maturity through them, make up what we call the Web Scraping Maturity Model. When understood, this model will serve as a formal playbook and benchmarking framework for data teams, allowing them to evaluate how they are performing against industry counterparts.

The model is meant to improve your ability to assess and evolve your own web scraping projects for maturity, reliability, and scalability, no matter the current size of the operation.

We built the model for you

By using this model as a reference, you can avoid the most common mistakes we see many companies making when they first reach out to us looking to fuel their businesses with quality data from the web.

Increase the chances of success of your new web scraping projects with Zyte's Web Scraping Maturity Model.

By increasing your web scraping maturity level, you'll gain greater access to new business opportunities

→

Your projects will yield better results from stronger business cases

Robust business cases for web scraping are essential for organizations because they help prioritize investments based on ROI, enhance the probability of project success through clear objectives and efficient communication, and serve as reusable templates, thus reducing the effort required for evaluating future project ideas.

You'll get more alignment between the stakeholders of your web scraping projects

A lot of time and money is lost due to miscommunications between the people that use the data and the people that extract it. When you learn how more mature organizations manage their web scraping structure, you'll get better at sourcing the right types of data, at the right frequency, from the places that matter.

Your product will be better! And your decisions too

Your digital products will improve tremendously thanks to the increased strength of your more mature web scraping projects. And access to fresh, reliable data will help your leaders make data-driven decisions across various departments, **a serious competitive advantage.**

You'll be more efficient at identifying and mitigating risks

Mature companies think about sustainability in the long term by performing constant compliance reviews on their web data feeds structure. They also de-risk investments by ensuring data can be collected in a sustainable and compliant manner. A strong compliance process helps **protect the operation and the brand by mitigating legal risk.**

Your spiders will be more aligned with your business strategy

It's easy to build a basic spider ad-hoc to get web data one time, from a single source.

It's much harder to build a consolidated and mature system (and culture) that is reliable, scalable, compliant, reusable, trackable, maintainable and functions to ensure that your company's strategy is realized.

You'll be more proactive on Quality Assurance

Mature data teams know that Quality is King, and this model will help you assess your current state of QA, and what you can do to improve your strategy to get the right data consistently. The model also provides insight into staying at the forefront of quality with the latest tools, techniques and strategies.
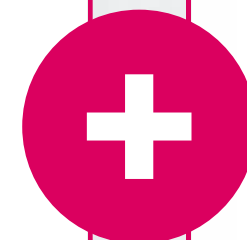
Who contributed to the model

The model is built according to findings from a research project that revealed the web scraping approaches, gaps, and successes of over 40 companies spanning a range of sizes and markets. We also gained insights from our engineers and project managers who have served over 5,000 customers and dealt with every conceivable type of web scraping project.

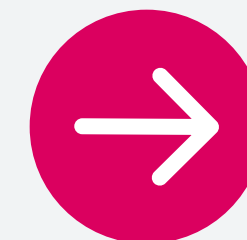## Web Data Extraction Professionals

Business & Strategy

- CEO, Founders
- Chief Data Officers
- Pricing Analysts
- Directors
- Product Managers

Data & Software

- Engineers
- Data Scientists
- Data Analysts
- Developers
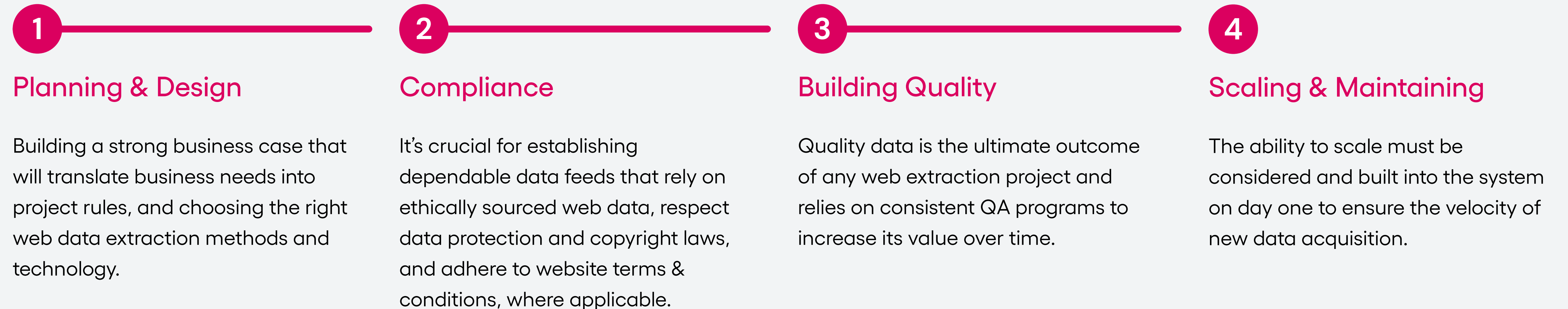- Quality Assurance

Market sectors

Sectors

- eCommerce & Retail
- Real Estate
- Marketing Intelligence
- Investment Management
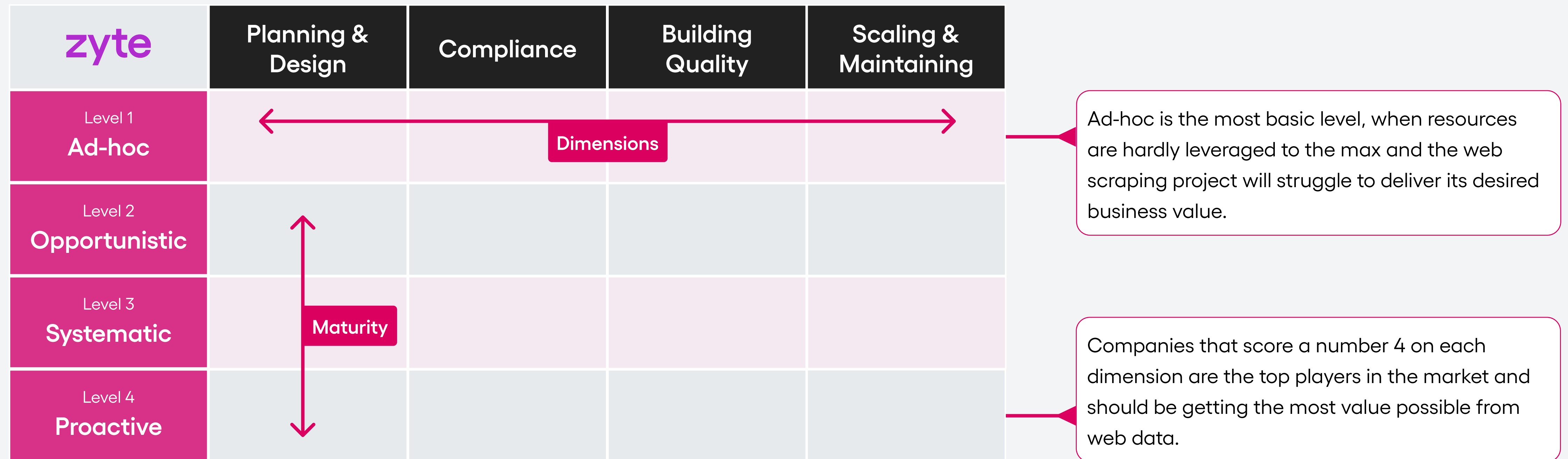- Trend Forecasting
- Publishing
- Insurance

The 4 dimensions of the model

Our research suggests there are four common dimensions that account for success in almost every web scraping project, that are most likely to impact a company's ability to get what it wants from web scraping — namely, consistent feeds of quality data from the web.

1

Planning & Design

Building a strong business case that will translate business needs into project rules, and choosing the right web data extraction methods and technology.

2

Compliance

It's crucial for establishing dependable data feeds that rely on ethically sourced web data, respect data protection and copyright laws, and adhere to website terms & conditions, where applicable.

3

Building Quality

Quality data is the ultimate outcome of any web extraction project and relies on consistent QA programs to increase its value over time.

4

Scaling & Maintaining

The ability to scale must be considered and built into the system on day one to ensure the velocity of new data acquisition.

How to read the model

The four columns represent the four dimensions of web scraping projects, and the four rows stand for a maturity score from 1 to 4, level 1 being Ad-hoc, and level 4 Proactive.

zyte	Planning & Design	Compliance	Building Quality	Scaling & Maintaining
Level 1 **Ad-hoc**			Dimensions	
Level 2 **Opportunistic**				
Level 3 **Systematic**	Maturity			
Level 4 **Proactive**				

Ad-hoc is the most basic level, when resources are hardly leveraged to the max and the web scraping project will struggle to deliver its desired business value.

Companies that score a number 4 on each dimension are the top players in the market and should be getting the most value possible from web data.

1 Planning & Design

The most common reason for Web Scraping projects being canceled, is the business benefit not being realized.

The main concerns can be:

- Are the resources being invested appropriate to the expected value?

- What are the potential opportunity costs?

- What are the lifetime costs?

- What does success look like?

2 Compliance

With ever-increasing legislation and case law, compliance can be a moving target.

Some ways to test your compliance preparedness are to ask yourself the following questions:

- Are we scraping copyrighted material? Have we completed an assessment of our use case?

- Are we scraping personal data? If so, how are we complying with the applicable data protection laws?

- Are we required to accept a website's terms and conditions? If so, what do the terms say about scraping?

3 Building **Quality**

Quality must be defined by the data end user, and not by what is easily measurable.

- Are your data pipeline development processes repeatable and predictable?

- Is quality slapped on after the fact?

- How quickly can new developers start adding value?

4 Scaling & **Maintaining**

Like Quality, the ability to scale must be considered and built on day one to ensure a high velocity of new data acquisition.

- Is the monitoring appropriate?

- Is there a dedicated monitoring and maintenance team?

- Are fixes centralized or on a per-website basis?

zyte | Web Scraping Maturity Model
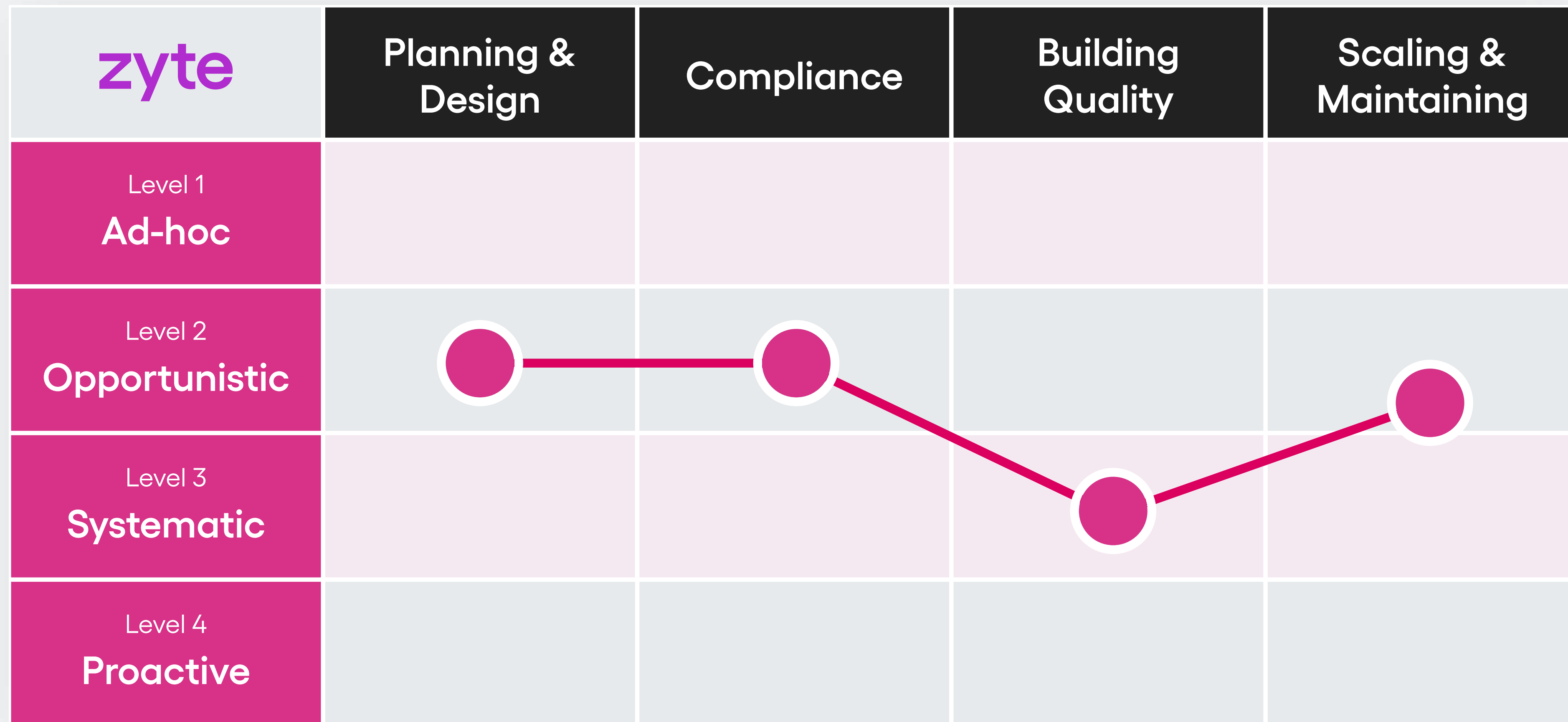
Discover your team's maturity level

How do you compare? Take the self-assessment survey on our website and receive your own scores by email.

Start the test →

1. Visit the **website**

2. Perform **the test**

3. Receive **your results** by email and **share with your team**

4. Book a **session with a Zyte specialist** for further insights

The current average web scraping maturity of the companies interviewed

zyte	Planning & Design	Compliance	Building Quality	Scaling & Maintaining
Level 1 **Ad-hoc**				
Level 2 **Opportunistic**	●	●		●
Level 3 **Systematic**			●	
Level 4 **Proactive**				

At Zyte we turn websites into data with industry-leading technology and services

Our solutions include:

Data Extraction Service

Let our web scraping experts build and manage the bespoke data extraction solution to satisfy your business needs.

Zyte API

Never get blocked again with Zyte proxies and smart browser tech all rolled into one powerful, lean, and ultra-reliable API.

Automatic Extraction powered by AI

Instantly access accurate web data through our user-friendly interface of various Extraction APIs, and save time getting the data you need.

Zyte API Enterprise

The best of Zyte technical expertise with Developer to Developer consultation, and access to our legal expertise for top challenges. Suited for operations willing to scale.

What are your web data needs?

Talk to a specialist →

Unlock the web data you need

Get clean, valuable data with web scraping tools
and services that drive your business forward.

Cuil Greine House, Ballincollig Commercial

Park Link Road, Ballincollig, Co. Cork, Ireland