

INFOLINK-EXP



PriceSpider

Induction

TOPICS TO COVER

First unit

- What's PriceSpider?
- PriceSpider Products
 - Where to buy?
 - Brand monitor
 - Prowl
- Product identification
 - SKU vs. UPC
 - VPNs and Location-based price strategizing
- WebCrawler vs WebScrapping

Second unit

- Web Interaction
 - Web Browsers
 - HTML
 - CSS
 - JS
 - SQL and Database Manegement Systems
 - URLs and HTTP
 - Inspect feature and source code
- WebCrawler example
- AI:Natural Language Processing

Installing Git and using GitHub

- Installing Git
- Cloning Repos

WHAT'S PRICESPIDER?

PriceSpider is a software platform that provides businesses with real-time competitive pricing and product information for a wide range of products. It is designed to help companies optimize their pricing strategies, increase their sales and profits, and enhance their overall market competitiveness.

PriceSpider's technology crawls the internet and collects data from various sources such as retailer websites, marketplaces, and social media platforms. The platform then analyzes this data using advanced algorithms and machine learning techniques to provide businesses with insights into market trends, consumer behavior, and competitor pricing.

In addition to providing real-time competitive pricing and product information, PriceSpider also offers a range of other features, including dynamic pricing tools, brand protection services, and analytics and reporting capabilities. These features help businesses to make data-driven decisions and optimize their pricing strategies for maximum profitability.

EXAMPLE

Let's say you run an online retail business that sells electronics, including smartphones. You want to optimize your pricing strategy to stay competitive in the market and maximize your profits.

With PriceSpider, you can track the prices of smartphones from your competitors, as well as the prices of the same models on different marketplaces, such as Amazon or eBay. You can also track the prices of the same smartphone models sold by different retailers, both online and offline.

Based on this data, PriceSpider can give you insights into the most popular smartphone models, the average selling price, and the range of prices across different channels. You can then use this information to adjust your pricing strategy and offer more competitive prices to attract more customers and increase your sales.

PriceSpider also offers dynamic pricing tools that can automatically adjust your prices based on real-time market conditions, such as changes in competitor pricing or demand. This can help you stay ahead of the competition and improve your margins.

In summary, PriceSpider can help you gather and analyze real-time competitive pricing and product information, which can be used to optimize your pricing strategy and increase your sales and profits.

PRODUCTS



- Offers MAP price monitoring.
- Map pricing, or Minimum Advertised Price, is a pricing policy that some manufacturers or brands put in place to control the price at which their products are advertised or sold by retailers.
- Under a MAP policy, the manufacturer or brand sets a minimum price that retailers must advertise their products at or above. This means that retailers are not allowed to advertise or sell the products below the set MAP price, but they are free to sell the products at a higher price.
- PriceSpider offers a MAP monitoring and enforcement tool as part of its suite of e-commerce solutions. This tool enables manufacturers and brands to track and monitor the advertised prices of their products across different online channels, and enforce their MAP policy by identifying any violations and taking appropriate actions.
- The MAP monitoring and enforcement tool also provides manufacturers and brands with real-time insights into market trends, competitive pricing, and other relevant data that can help them optimize their pricing strategy and improve their sales performance. By ensuring that their products are being sold at the minimum advertised price, manufacturers and brands can protect their brand image, maintain healthy profit margins, and build stronger relationships with their retail partners.

WHERE TO BUY

This tool helps brands and manufacturers to provide accurate and up-to-date information on where their products are available for purchase across different online and offline retailers.

Advertising: PriceSpider also offers various advertising solutions, including paid search, display ads, and social media ads, to help brands and retailers drive traffic and sales to their online stores.

BRAND MONITOR

- This tool enables brands to monitor their online presence and reputation, analyze consumer feedback and sentiment, and track product performance.
- This product uses advanced analytics and AI to help retailers and brands optimize their online product pages and improve conversion rates.
- This product provides retailers with real-time pricing data, competitive analysis, and insights into market trends and consumer behavior.

HOW IS IT WORKS?

PriceSpider uses a combination of advanced data collection, machine learning, and natural language processing techniques to gather and analyze product data from a variety of sources, such as online marketplaces, retailer websites, and social media platforms.

PriceSpider's algorithm employs techniques such as web crawling, data extraction, and data normalization to collect and clean the product data from these sources. The algorithm then uses machine learning techniques to analyze this data and identify patterns and trends in the market, such as changes in consumer behavior, pricing trends, and product availability.

The algorithm also employs natural language processing (NLP) techniques to analyze product descriptions and user reviews to gain insights into consumer sentiment and preferences. This allows PriceSpider to provide businesses with a more holistic view of the market and helps them to make more informed pricing decisions.

Overall, the algorithm behind PriceSpider is complex and sophisticated, and it relies on advanced technologies to gather, analyze, and interpret product data in real-time.

UPC VS. SKU

In summary, UPCs are used to identify products universally and are mainly used for scanning at the point of sale, while SKUs are used by retailers to manage their inventory and track sales.

Local Pricing Strategy

The basic package, which costs \$7.99 per month in the US, will run you about \$5 there, when converted to US dollars at recent exchange rates, a May report by the research-firm Bernstein showed.

YOUR CLIENTS ARE OUR CLIENTS....



SONY®



NETGEAR



WRIGLEY



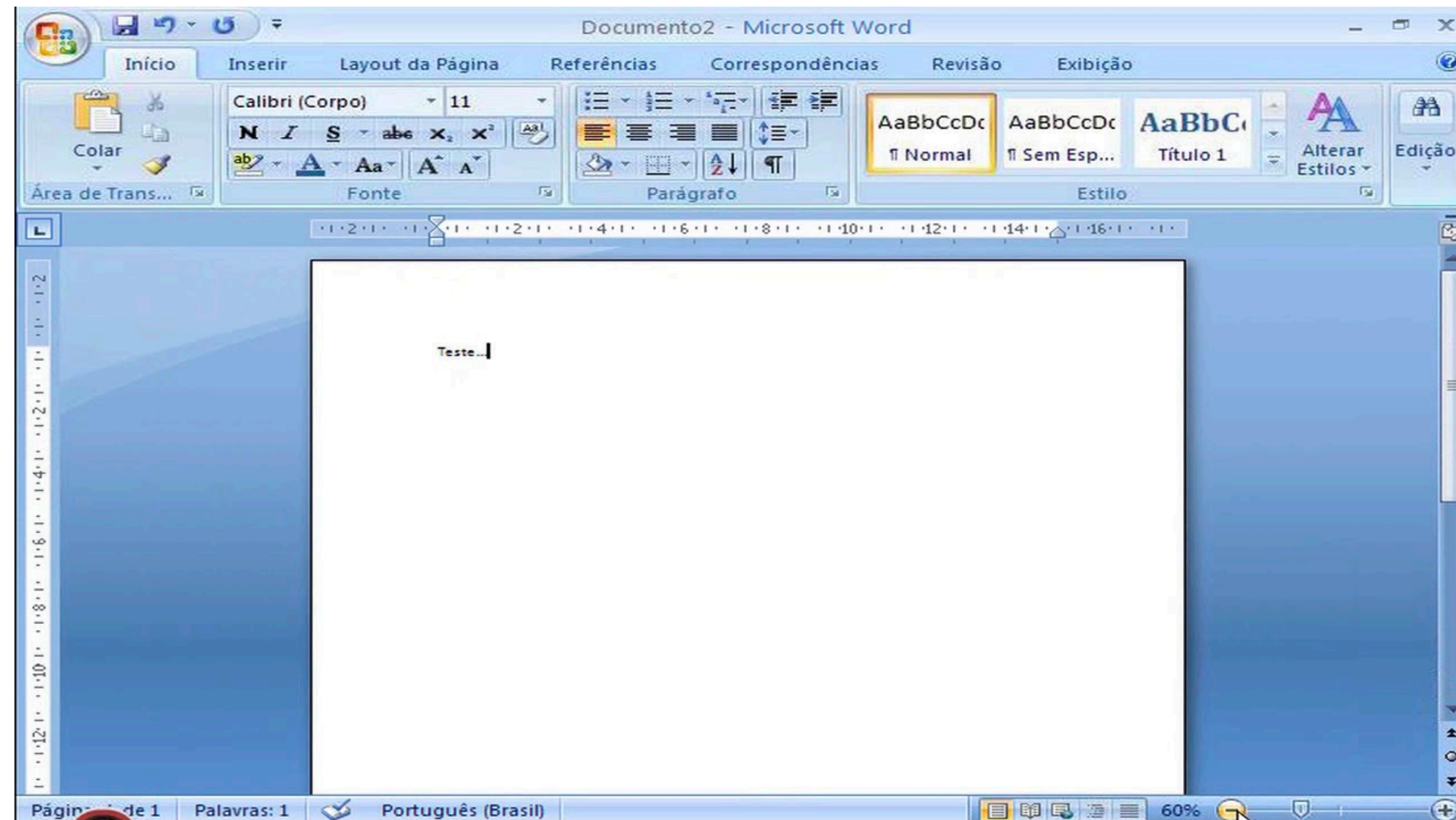
DURACELL®



SAMSUNG

2ND UNIT.- WEB INTERACTION

WHAT IS A PROGRAMMING LANGUAGE AND WHY IS NECESSARY?



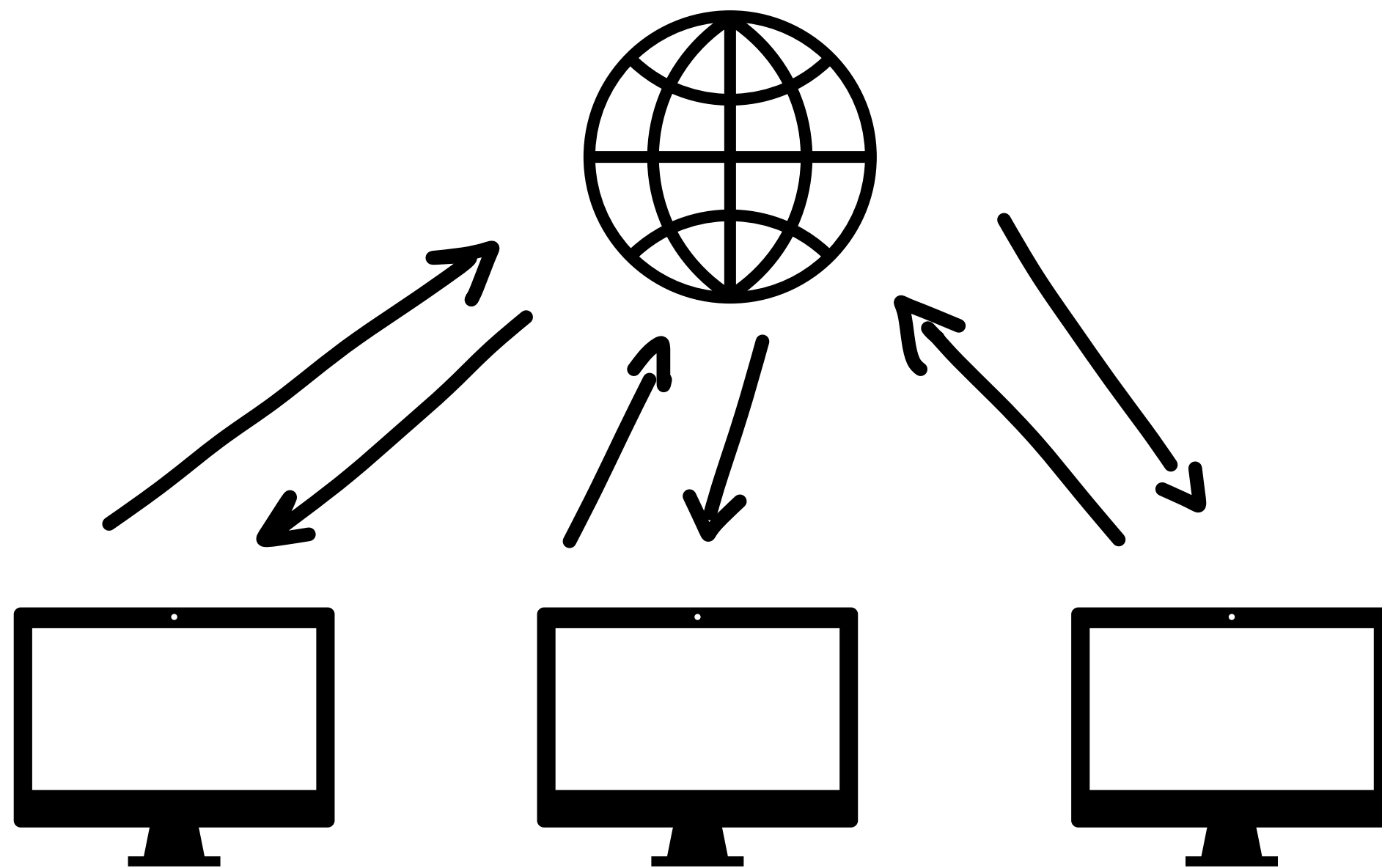
- ≈ HTML

- **A programming language is the way to speak to computers and how computers talk to other computers. They are high level languages make them easier to understand for humans.**
- **HTML is not OS dependent. Less resource intensive. Back in the day there weren't visual interfaces (they do exist). Text based avoid confusion in alike visual elements. It's easier to integrate CSS and JS in text based. WYSIWYG interfaces are useful in some contexts like design.**

1. Flexibility: Visual interfaces are great for specific tasks, such as designing a web page or creating a user interface. However, programming languages offer much greater flexibility, allowing developers to create custom solutions to unique problems.
2. Efficiency: While visual interfaces may help simplify certain tasks, they often require a lot of clicking and dragging, which can be slow and cumbersome. In contrast, experienced programmers can write complex algorithms quickly and efficiently.
3. Control: Visual interfaces often abstract away the underlying code, which can limit a developer's control over the application. This can be problematic if the developer needs to make specific changes or optimizations to the code.
4. Learning curve: While visual interfaces can be helpful for beginners, they can also create a steeper learning curve. In contrast, programming languages provide a consistent syntax and set of rules that can be learned and applied in a variety of contexts.
5. Scope: Visual interfaces are often geared towards specific types of tasks, such as creating user interfaces or designing web pages. In contrast, programming languages can be used to solve a wide range of problems, from data analysis to machine learning to building operating systems.

THE INTERNET

- The web is a set of interconnected computers through a network. Even though, computers could be connected directly between them, usually they are connected through servers.



- They work under the request/response procedure. That is sending and receiving files. A special kind of file called web page. They use a standardized way of speaking between them to avoid ambiguity in the messages we ask and send.
- When a URL is entered in the browser this lookup for the IP direction through what is known as a DNS. The redirects the request to that IP and again the response to the original client in the browser. The web is surfed with HTTP, that is Hyper Text Transfer Protocol. Just a fancy name to say that we navigate it through links inside webpages. The way we tell computers that we are actually requesting a page is using HTML. This means we use hyper text and a markup language combined to develop web sites.

Web Navigation

It's the action of navigation a network of information resources in the World Wide Web (WWW), which is organized as hypertext. The user interface that is used to do so is called a web browser.

The Google logo, featuring the word "Google" in its characteristic multi-colored font (blue, red, yellow, blue, green, red).The Bing logo, featuring a green chevron icon followed by the word "Bing" in a green sans-serif font.The Yahoo! logo, featuring the word "yahoo!" in a bold, purple, lowercase sans-serif font.

Web browser

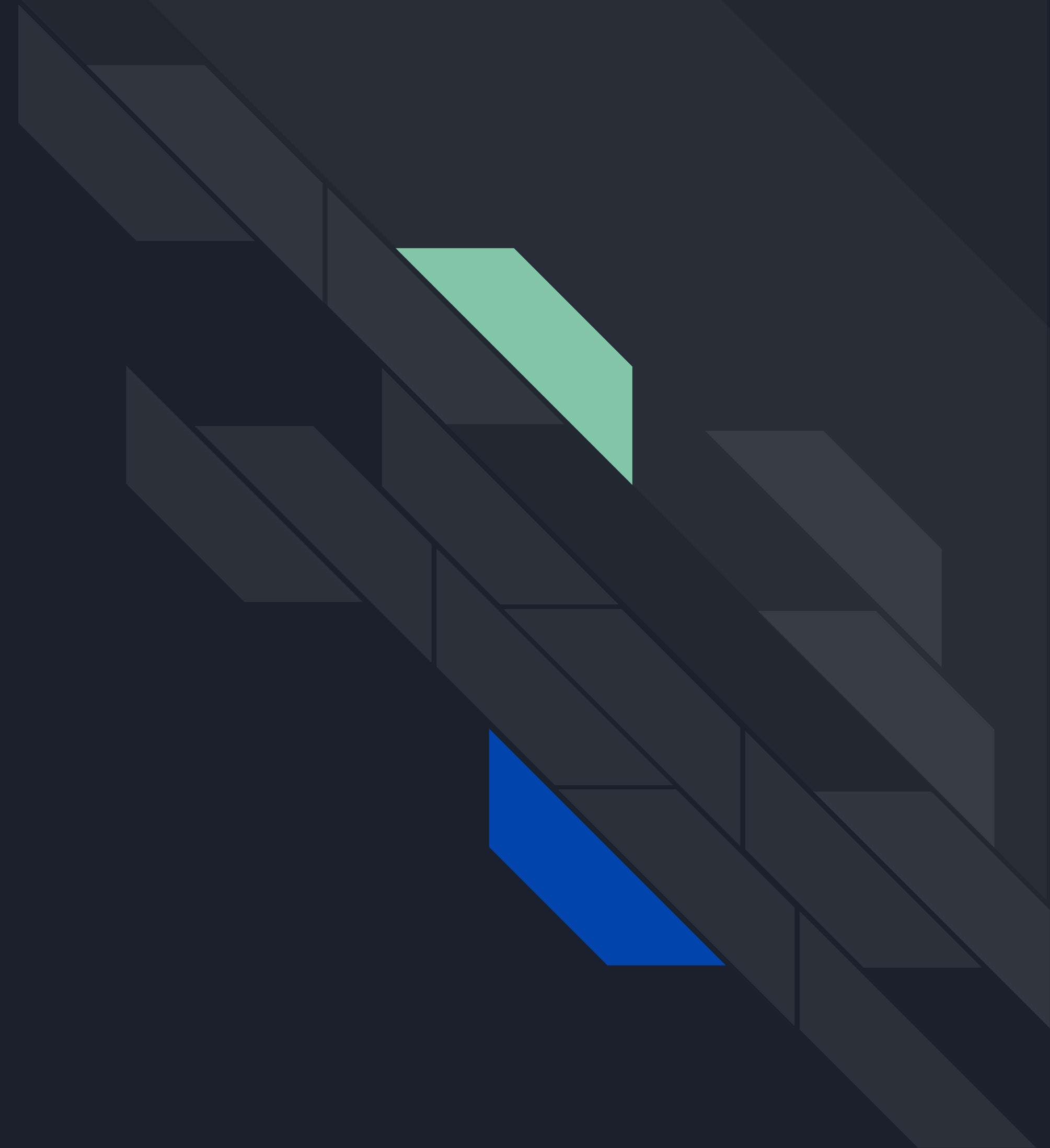
Is an application for accessing websites where the browser retrieves its files from a web server and displays the page from a particular site and displays the page on the user screen.

Its purpose is to fetch content from the WWW or from local storage.

This process begins when all URLs are retrieved using a Hypertext Transfer Protocol (HTTP), a set of rules for transfer data.



URL



What is a URL:

URL stands for *Uniform Resource Locator*. A URL is nothing more than the address of a given unique resource on the Web. A domain name is part of a URL. You can see the visual difference in the following examples

DOMAIN NAME

http://www.verisign.com/domain-names/online/index.shtml

URL

Authority

http://www.example.com:80/path/to/myfile.html?key1=value1&key2=value2#SomewhereInTheDocument

Scheme Domain Name Port Path to the file Parameters Anchor



URL

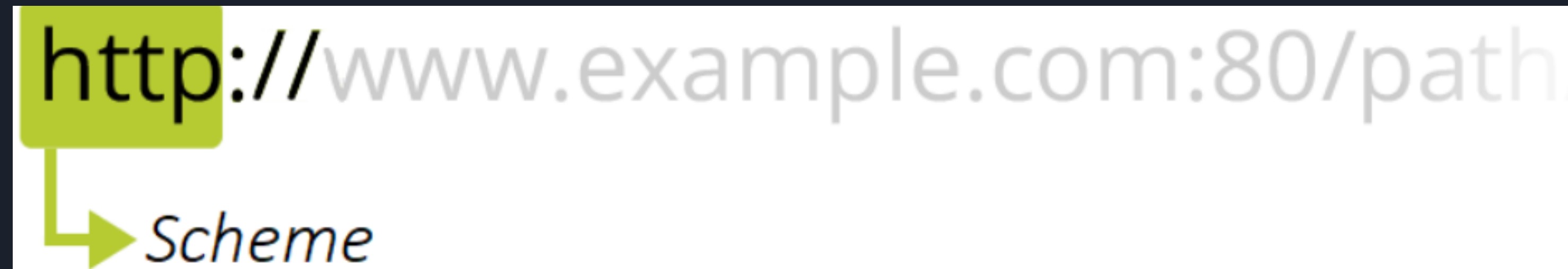
For computer networks and servers to “talk to one another,” on a language made up of numbers and letters called an IP address. Every device that connects to the internet has a unique IP address and looks like this:

22.231.113.64 or 3ffe:1900:4545:3:200:f8ff:fe21:67cf

To navigate easily around the web, typing in a long IP address isn’t ideal, or realistic, for an online user. This is the reason why domain names were created. You could consider the domain name as a “nickname” for the IP address.

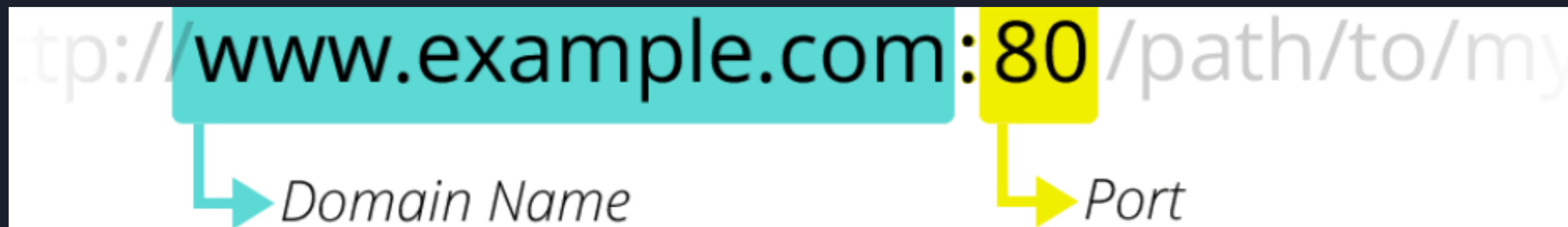
Anatomy of a URL

The first part of the URL is the *scheme*, which indicates the protocol that the browser must use to request the resource



The diagram shows a URL `http://www.example.com:80/path` where the `http` part is highlighted in a green box. A green arrow points from this box to the word *Scheme*.

the *authority*, which is separated from the scheme by the character pattern `://`. If present the authority includes both the *domain* (e.g. `www.example.com`) and the *port* (80), separated by a colon:



The diagram shows a URL `http://www.example.com:80/path/to/my` where `www.example.com` is highlighted in a teal box and `80` is highlighted in a yellow box. A teal arrow points from the teal box to the words *Domain Name*, and a yellow arrow points from the yellow box to the word *Port*.

Anatomy of a URL

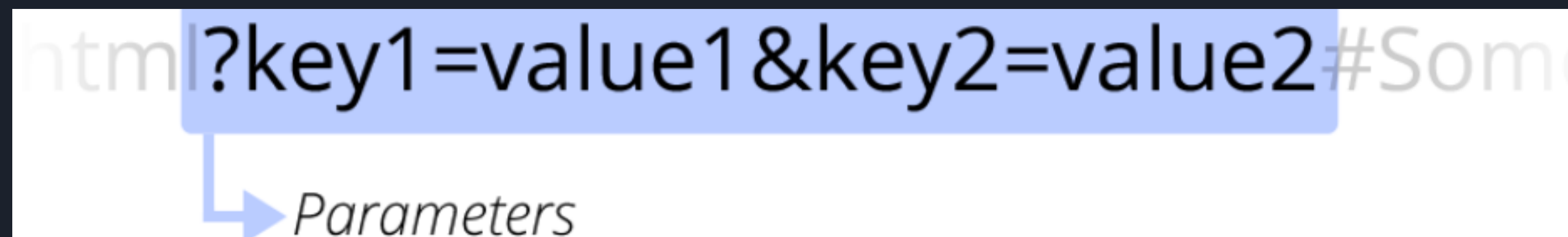
A path like this represented a physical file location on the Web server. Nowadays, it is mostly an abstraction handled by Web servers without any physical reality.



n:80/path/to/myfile.html?key1=value1

Path to resource

Those parameters are a list of key/value pairs separated with the & symbol. The Web server can use those parameters to do extra stuff before returning the resource.



html?key1=value1&key2=value2#Some

Parameters

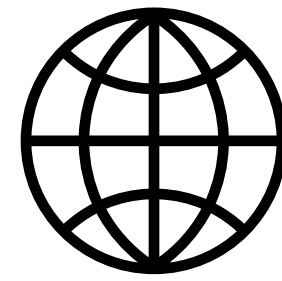
Anatomy of a URL

An anchor represents a sort of "bookmark" inside the resource, giving the browser the directions to show the content located at that "bookmarked" spot. On an HTML document, for example, the browser will scroll to the point where the anchor is defined;


lue2#SomewhereInTheDocument



- **The web is a set of interconnected computers.**



HTML




HTML stands for **HyperText Markup Language**. It is a standard markup language for web page creation. It allows the creation and structure of sections, paragraphs, and links using HTML elements (the building blocks of a web page) such as tags and attributes.



HTML has a lot of use cases, namely:

- **Web development.** Developers use HTML code to design how a browser displays web page elements, such as text, hyperlinks, and media files.
- **Internet navigation.** Users can easily navigate and insert links between related pages and websites as HTML is heavily used to embed hyperlinks.
- **Web documentation.** HTML makes it possible to organize and format documents, similarly to Microsoft Word.

It's also worth noting that HTML is not considered a programming language as it can't create dynamic functionality. It is now considered an official web standard. The [World Wide Web Consortium \(W3C\)](#) maintains and develops HTML specifications, along with providing regular updates.



The average website includes several different HTML pages. For instance, a home page, an about page, and a contact page would all have separate HTML files.

HTML documents are files that end with a .html or .htm extension. A web browser reads the HTML file and renders its content so that internet users can view it.

All HTML pages have a series of HTML elements, consisting of a set of tags and attributes.

HTML elements are the building blocks of a web page. A tag tells the web browser where an element begins and ends, whereas an attribute describes the characteristics of an element.




The three main parts of an element are:

- Opening tag – used to state where an element starts to take effect. The tag is wrapped with opening and closing angle brackets. For example, use the start tag `<p>` to create a paragraph.
- Content – this is the output that other users see.
- Closing tag – the same as the opening tag, but with a forward slash before the element name. For example, `</p>` to end a paragraph.

The combination of these three parts will create an HTML element:


```
<p>This is how you add a paragraph in HTML.</p>
```

Another critical part of an HTML element is its attribute, which has two sections – a name and attribute value. The name identifies the additional information that a user wants to add, while the attribute value gives further specifications.



For example, a style element adding the color purple and the font-family verdana will look like this:

```
<p style="color:purple;font-family:verdana">This is how you add a paragraph  
in HTML.</p>
```




Another attribute, the HTML class, is most important for development and programming. The class attribute adds style information that can work on different elements with the same class value.

For example, we will use the same style for a heading `<h1>` and a paragraph `<p>`. The style includes background color, text color, border, margin, and padding, under the class `.important`. To achieve the same style between `<h1>` and `<p>`, add `class="important"` after each start tag:

```
<html>
<head>
<style>
  .important {
    background-color: blue;
    color: white;
    border: 2px solid black;
    margin: 2px;
    padding: 2px;
  }
</style>
</head>
<body>

<h1 class="important">This is a heading</h1>
<p class="important">This is a paragraph.</p>

</body>
</html>
```




Most elements have an opening and a closing tag, but some elements do not need closing tags to work, such as empty elements. These elements do not use an end tag because they do not have content:

```


```

This image tag has two attributes – an src attribute, the image path, and an alt attribute, the descriptive text. However, it does not have content nor an end tag.



Lastly, every HTML document must start with a `<!DOCTYPE>` declaration to inform the web browser about the document type. With HTML5, the doctype HTML public declaration will be:

```
<!DOCTYPE html>
```



Currently, there are 142 HTML tags available that allow for the creation of various elements. Even though modern browsers no longer support some of these tags, learning all the different elements available is still beneficial. This section will discuss the most-used HTML tags and two main elements – block-level elements and inline elements.



Block-Level Elements

A block-level element takes up the entire width of a page. It always starts a new line in the document. For example, a heading element will be in a separate line from a paragraph element.

Every HTML page uses these three tags:

- `<html>` tag is the root element that defines the whole HTML document.
- `<head>` tag holds meta information such as the page's title and charset.
- `<body>` tag encloses all the content that appears on the page.

```
<html>
  <head>
    <!-- META INFORMATION -->
  </head>
  <body>
    <!-- PAGE CONTENT -->
  </body>
</html>
```

Other popular block-level tags include:

- Heading tags – these range from `<h1>` to `<h6>`, where heading h1 is largest in size, getting smaller as they move up to h6.
- Paragraph tags – are all enclosed by using the `<p>` tag.
- List tags – have different variations. Use the `` tag for an ordered list, and use `` for an unordered list. Then, enclose individual list items using the `` tag.



Inline Elements

An inline element formats the inner content of block-level elements, such as adding links and emphasized strings. Inline elements are most commonly used to format text without breaking the flow of the content.

For example, a `` tag would render an element in bold, whereas the `` tag would show it in italics. Hyperlinks are also inline elements that use an `<a>` tag and an href attribute to indicate the link's destination

```
<a href="https://example.com/">Click me!</a>
```

HTML is the primary markup language found on the internet.

Every HTML page has a series of elements that create the content structure of a web page or application.

HTML is a beginner-friendly language with plenty of support and is mainly used for static website pages. HTML works best together with CSS for the styling and JavaScript for the functionality.

[Reference guide to HTML](#)

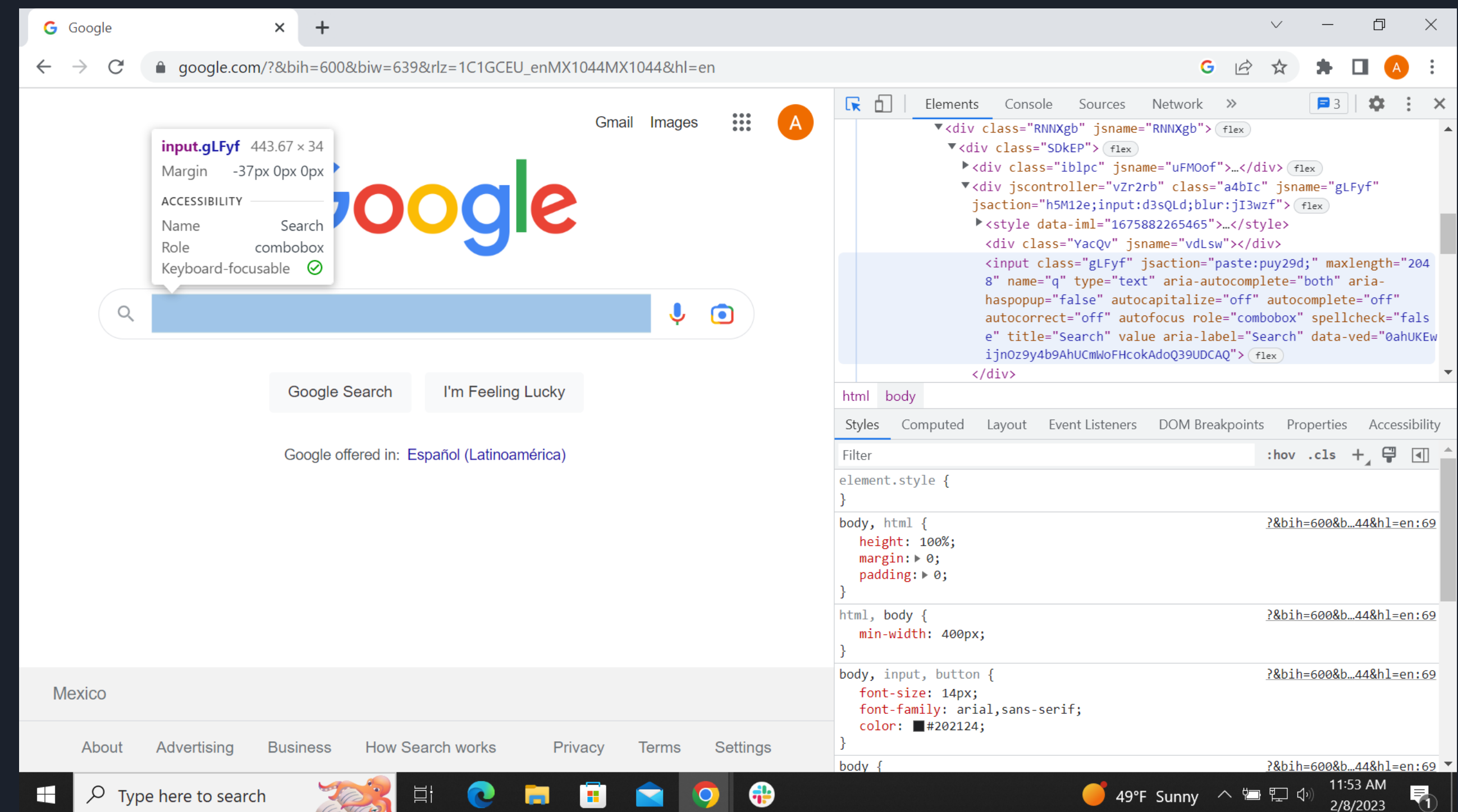
PAGE INSPECTION

Page Inspection

Enables you to get details about a page, design, elements, and source behind data displays. It is ideal for learning the data model behind a page.

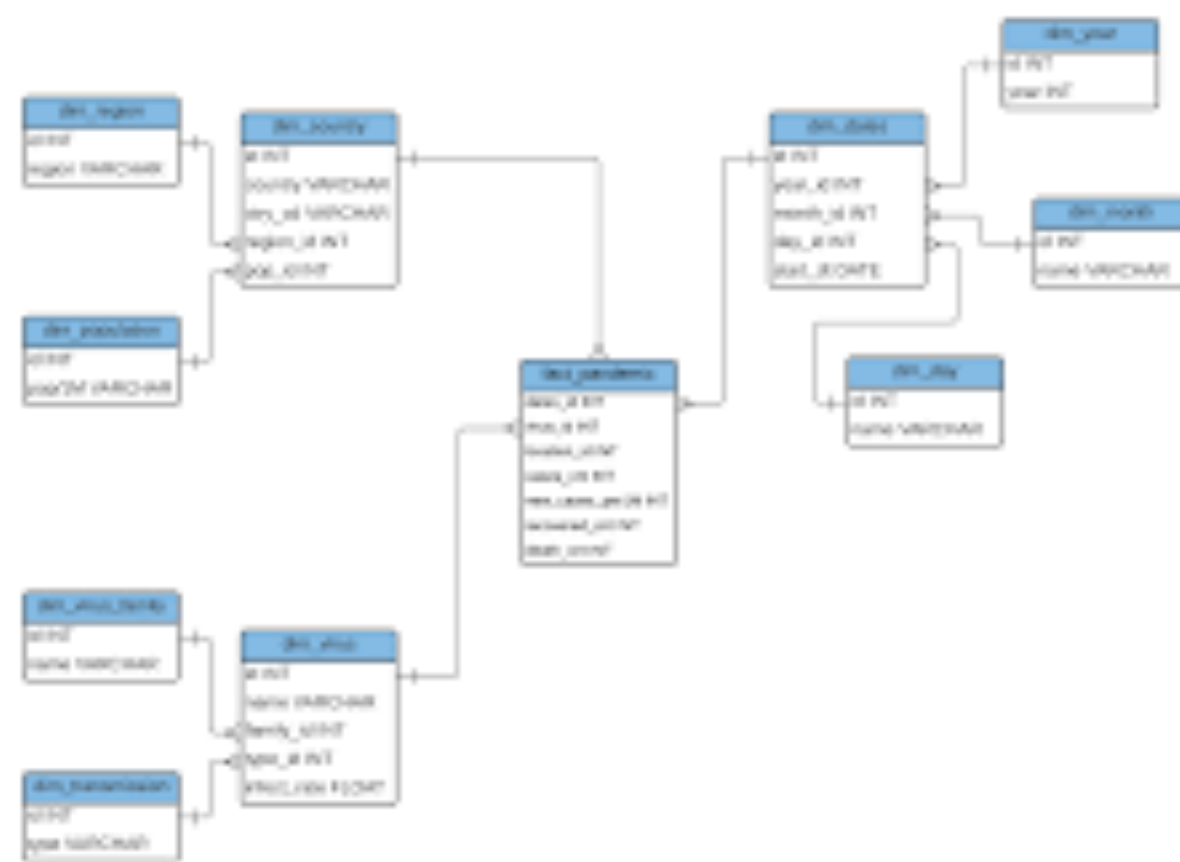
You can use Ctrl + Shift = I

And/Or F12



DDL modifies the schema of a database table

DML Manipulate data from the



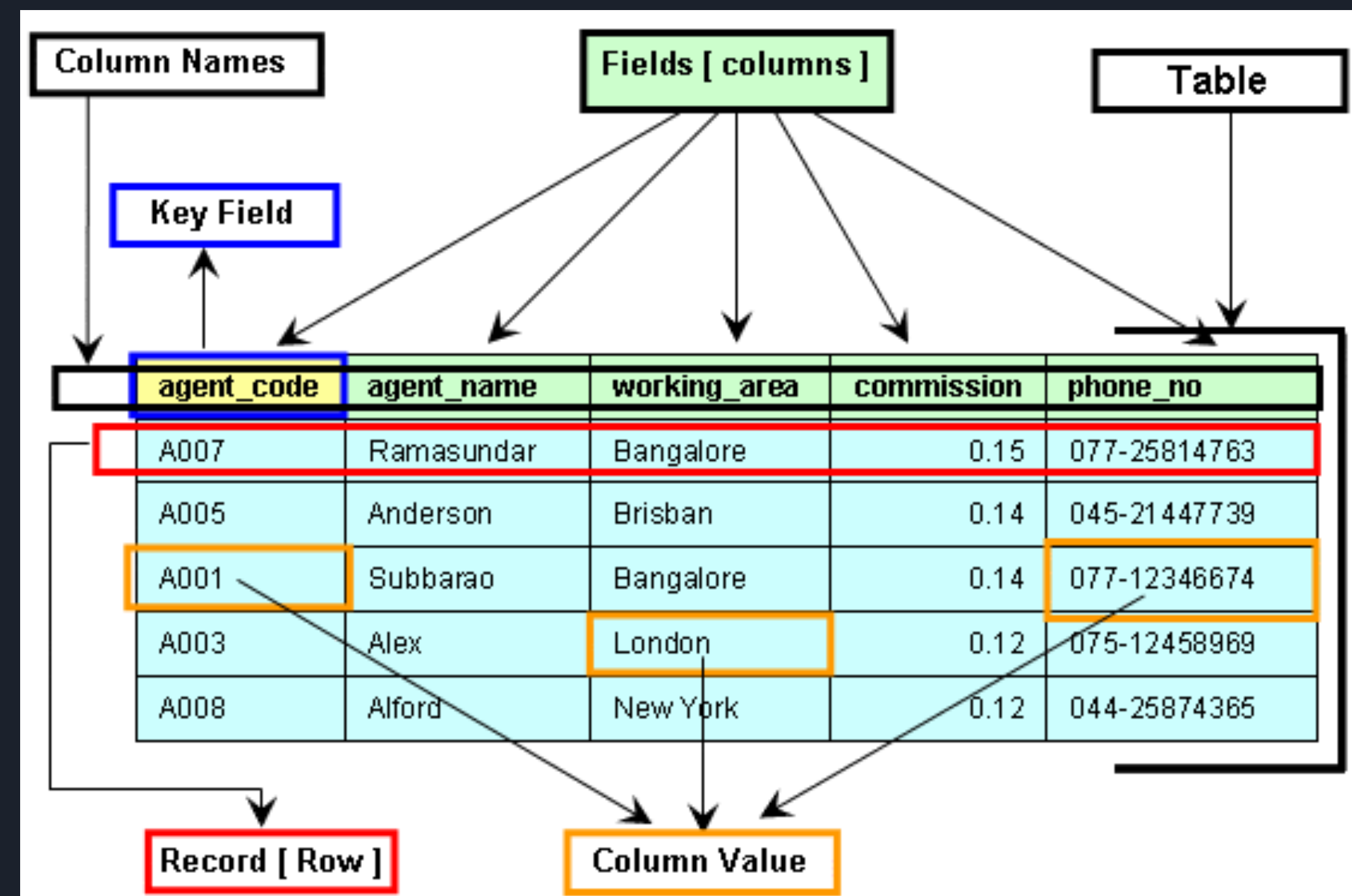
```
SELECT something FROM some table
      WHERE some conditions are satisfied
```

```
UPDATE some table SET something
    WHERE some conditions are satisfied
```

SQL

SQL Tables

Table is a database object which is comprised of rows and columns. Where data is logically organized in a row-and-column format similar to a spreadsheet. Each row represents a unique record and each column a field in the record.



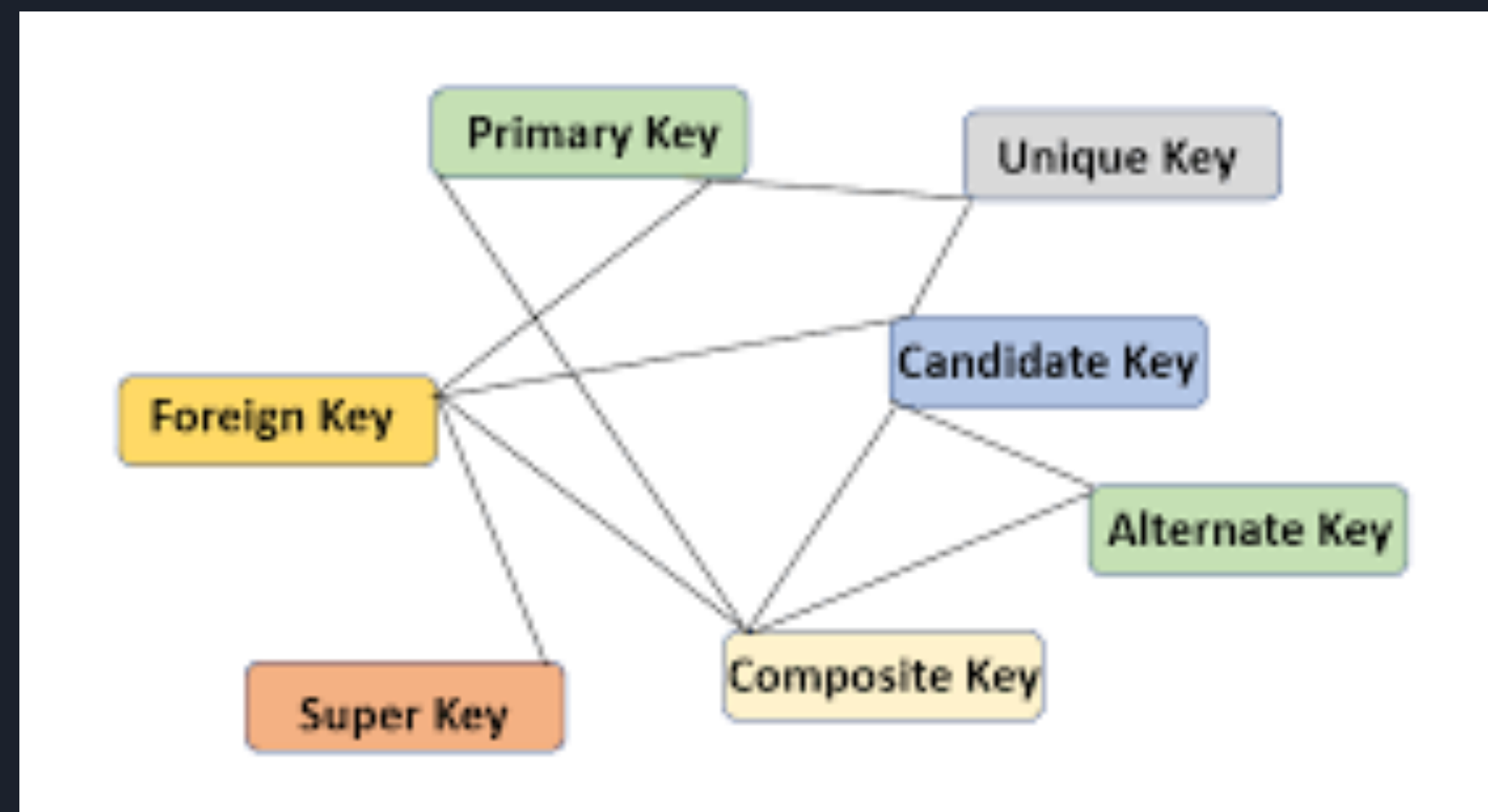


Elements of a table

- **Fields**
 - The information of a table store in some heads those are fields or columns shown vertically
- **Columns**
 - Each field or column has an individual name. A table cannot contain the same name of two different columns
- **Record (Row)**
 - All the columns in a table make a row> Each row contains all the information of individual topics
- **Column value**
 - The value of each field
- **Key Field**
 - Each table contain a field which can create a link with another one or more table is the key field

SQL key's

An SQL key is either a single column or group of columns that can uniquely identify rows in a table. They ensure that there are no rows with duplicate information. They also help in establishing a relationship between multiple tables in the database.



Candidate Key

Candidate key is a single key or a group of multiple keys that uniquely identify **rows** in a table.

Id	Name	Gender	City	Email	Dep_Id
1	Ajay	M	Delhi	ajay@gmail.com	1
2	Vijay	M	Mumbai	vijay@gmail.com	2
3	Radhika	F	Bhopal	radhika@gmail.com	1
4	Shikha	F	Jaipur	shikha@gmail.com	2
5	Hritik	M	Jaipur	hritik@gmail.com	2

5 rows in set (0.00 sec)

Primary key

Primary key is the Candidate key selected by the database administrator to uniquely identify tuples in a table. There can only be one and non-null attribute(s).

Alternate or Secondary Key

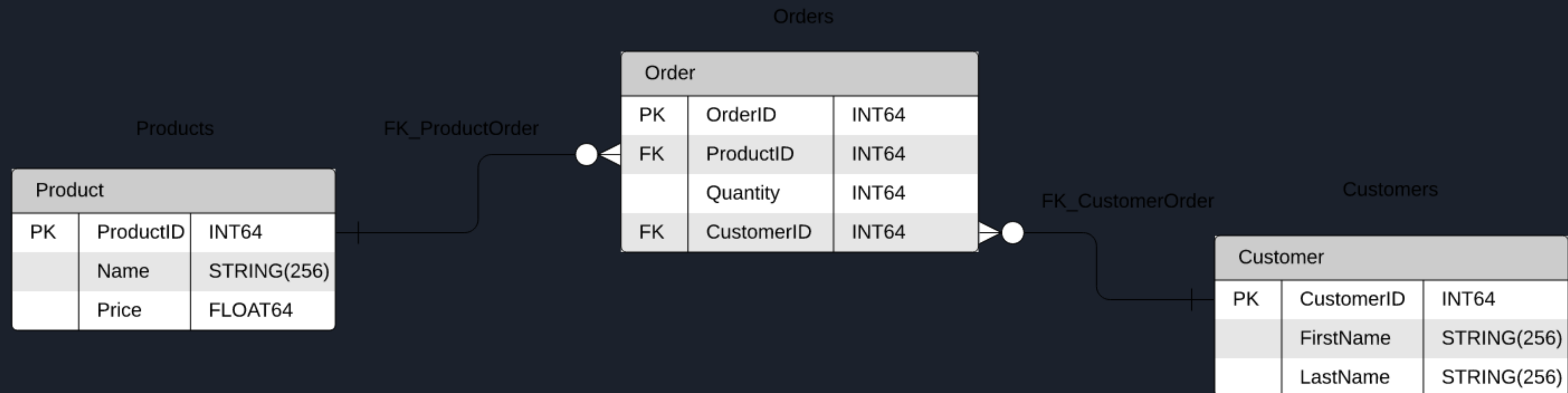
Alternate keys are those candidate keys which are not the Primary key.

Primary Key				Alternate Key	
Id	Name	Gender	City	Email	Dep_Id
1	Ajay	M	Delhi	ajay@gmail.com	1
2	Vijay	M	Mumbai	vijay@gmail.com	2
3	Radhika	F	Bhopal	radhika@gmail.com	1
4	Shikha	F	Jaipur	shikha@gmail.com	2
5	Hritik	M	Jaipur	hritik@gmail.com	2

5 rows in set (0.00 sec)

Foreign key

Foreign key is an attribute which is a Primary key in a parent table, but is included as an attribute in another host table.





Super key

Super key is a single key or group of multiple keys that can uniquely identify **tuples** in a table.

ID	Name	Register_no
2026	Andrew	012917323
2025	John	123471904
2024	William	233490183



Composite Key

A Composite key is a candidate key or Primary key that consist of more than one attribute . Cases were no single attribute will have the property to uniquely identify tuples in a table.

```
mysql> select * from product;
```

Transaction_Id	Product_Id	Customer_Id	Product	Quantity
A1001	P1005	C9001	Smartphone	1
A1001	P2010	C9001	Screen guard	1
A1002	P2013	C9003	Smartwatch	1
A1003	P2010	C9010	Screen guard	2

```
4 rows in set (0.00 sec)
```

ANALOGY

Imagine that a company has a large number of employees, and the company needs to keep track of all their personal and employment information. The company could create a database to store this information, with each employee represented as a record in the database.

Just like how each person has a unique identifier such as a name or social security number, each record in the database would have a unique identifier, such as an employee ID. This ID would be used to uniquely identify each employee record in the database.

Now, just as people have different attributes such as age, gender, and job title, the employee records in the database would have different fields to store different pieces of information, such as name, address, phone number, job title, department, and salary.

SQL would be used to interact with the employee database. For example, an SQL query could be used to select all employees who work in a specific department, or to update the salary of all employees who meet certain criteria.

In this analogy, the employee database is like a huge filing cabinet, with each record representing a person, and SQL is like the language that the HR department uses to manage the information stored in the filing cabinet.





SQL (Structured Query Language) is a programming language that is used to manage and manipulate relational databases. It is used to create, modify, and delete databases, tables, and other database objects, as well as to insert, update, and retrieve data from the database.

SQL is used in a wide variety of applications, from small businesses managing customer data to large corporations managing complex financial systems. It is used by software developers, data analysts, and data scientists to work with databases.

SQL is a standard language, meaning that its syntax and commands are standardized across database platforms. However, there may be some variations in the way different database systems implement certain features or functions.

There are many different types of SQL commands, including SELECT, INSERT, UPDATE, and DELETE, as well as commands for creating and modifying tables, indexes, and views. SQL also includes functions for performing calculations, formatting data, and working with dates and times.

WHAT ABOUT THE RELATIONAL PART?

The "relational" part of relational databases refers to the way that different tables of data are related to each other. In a relational database, data is organized into one or more tables, with each table containing multiple rows (also known as records) and columns.

For example, if we go back to the employee database analogy, the company might have a separate table to store information about each department, with each row representing a different department and each column containing information about the department, such as its name, location, and manager.

To relate the two tables together, the employee table might have a column called "department ID" that is used to link each employee to their department. This column would contain the unique identifier for each department, which could be used to look up additional information about the department in the department table.

SQL includes commands for working with related data across multiple tables. For example, an SQL JOIN command can be used to combine data from two or more tables based on a common field, such as the department ID in the employee and department tables.

By using tables and establishing relationships between them, relational databases can efficiently store and manage large amounts of data while still allowing for complex queries and data analysis.

1. **CREATE TABLE:** This command is used to create a new table in the database. It specifies the table's name, the names and data types of its columns, and any constraints or rules for the data.

Example:

```
CREATE TABLE employees (  
    id INT PRIMARY KEY,  
    name VARCHAR(50),  
    department_id INT,  
    hire_date DATE,  
    salary DECIMAL(10,2)  
);
```

This command creates a new table called "employees" with columns for the employee ID, name, department ID, hire date, and salary.

2.INSERT INTO: This command is used to insert a new row of data into a table.

Example:

```
INSERT INTO employees (id, name, department_id, hire_date, salary)
VALUES (1, 'John Smith', 3, '2022-01-01', 50000.00);
```

This command inserts a new row of data into the "employees" table, with values for the employee ID, name, department ID, hire date, and salary.

3. SELECT: This command is used to retrieve data from one or more tables. It can be used to filter and sort the data, as well as to perform calculations and groupings.

Example:

```
SELECT name, department_id, salary
FROM employees
WHERE department_id = 3
ORDER BY salary DESC;
```

This command retrieves the name, department ID, and salary for all employees who work in department 3, and sorts them in descending order by salary.

4.UPDATE: This command is used to modify existing data in a table.

Example:

```
UPDATE employees  
SET salary = 55000.00  
WHERE id = 1;
```

This command updates the salary of the employee with ID 1 to 55000.00.

5.DELETE: This command is used to delete rows of data from a table.

Example:

```
DELETE FROM employees  
WHERE department_id = 3;
```

This command deletes all rows from the "employees" table where the department ID is 3.

SQL VS. EXCEL

Excel and SQL are two different tools used for different purposes.

Excel is a spreadsheet program used for organizing, analyzing, and manipulating data in a tabular format. It provides a graphical user interface that allows users to create spreadsheets with cells, rows, and columns. It can perform mathematical operations, create charts, graphs, and pivot tables. Excel is a standalone application that is installed on a user's computer.

SQL (Structured Query Language), on the other hand, is a programming language designed to manage and manipulate data in relational databases. SQL is used to create, modify, and retrieve data from databases. It is used to create, modify, and delete tables, as well as to insert, update, and delete data. SQL is often used in conjunction with database management systems such as MySQL, Oracle, and Microsoft SQL Server.

One key difference between Excel and SQL is that Excel is better suited for analyzing smaller data sets, while SQL is designed for working with larger data sets stored in a database. Excel is also more suitable for ad-hoc data analysis and quick calculations, while SQL is better suited for complex queries and data manipulation. Additionally, SQL offers more advanced functionalities such as data normalization, indexing, and data security that Excel does not provide.

ANALOGY

One possible analogy to illustrate the difference between Excel and SQL is that Excel is like a calculator or a small tool for performing quick calculations, while SQL is like a toolbox filled with a variety of tools that are used for complex construction and engineering tasks.

Just as a calculator can perform simple mathematical operations and provide quick answers, Excel can quickly organize, analyze, and manipulate small sets of data. However, just as a calculator has limited capabilities and cannot be used for more complex tasks, Excel also has its limitations and may not be the best tool for handling large or complex data sets.

On the other hand, SQL is like a toolbox filled with many tools designed for different tasks, such as hammers, saws, drills, and screwdrivers. SQL provides a set of tools for managing and manipulating large databases, just as a toolbox provides a set of tools for managing and manipulating different materials and construction projects. While SQL may require more knowledge and training to use effectively, it offers a much wider range of capabilities and can be used for complex data analysis and management.

WEB CRAWLERS

- EXAMPLE IN PYTHON USING BS