

# 网络爬虫专题培训

费斌杰

为什么要学爬虫？

# 为什么要学爬虫？

把互联网变成你的数据库！

# 为什么要学爬虫？

## 天猫：抓取商品量价数据，及时发现趋势变化



¥2299.00

飞利浦空气净化器AC2880家用卧室智  
飞利浦上游专卖店

月成交 299笔

评价 1892



¥1499.00

小米空气净化器pro室内办公家用卧室  
小米官方旗舰店

月成交 2962笔

评价 1.5万

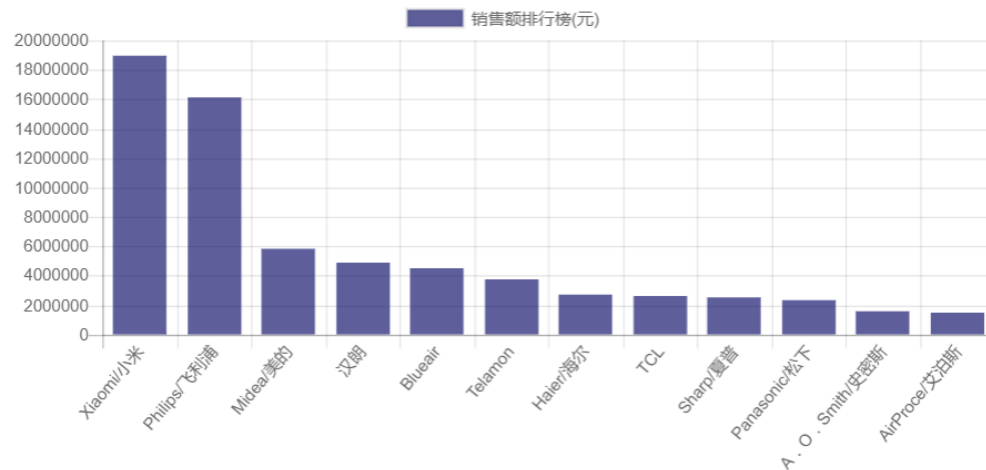


¥1490.00

TIPON/德国汉朗空气净化器除甲醛雾  
汉朗旗舰店

月成交 1249笔

评价 1.0万



¥1298.00

美的 KJ210G-C42空气净化器家用 卧室  
美的空气净化器旗舰店

月成交 177笔

评价 6203



¥1999.00

亚都空气净化器家用静音卧室除甲醛雾  
亚都智能电器旗舰店

月成交 235笔

评价 3921

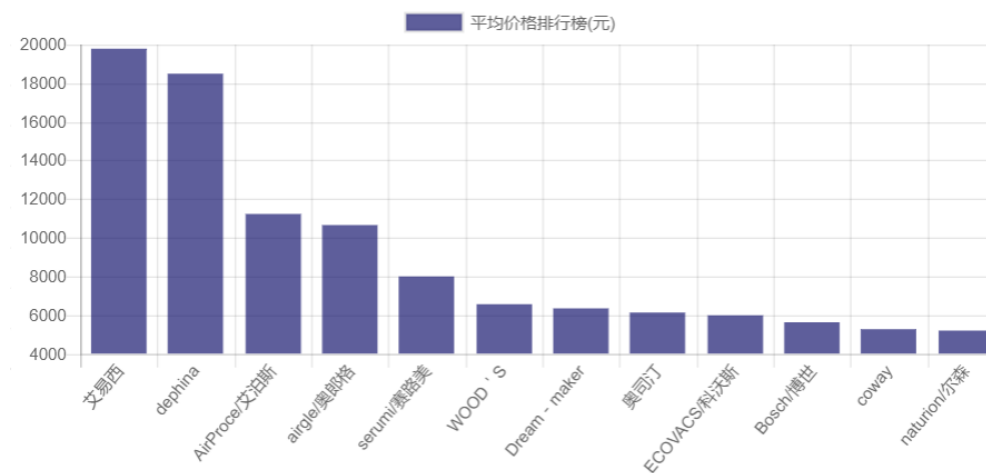


¥1299.00

海尔空气净化器家用除甲醛卧室氧吧负  
海尔海纳专卖店

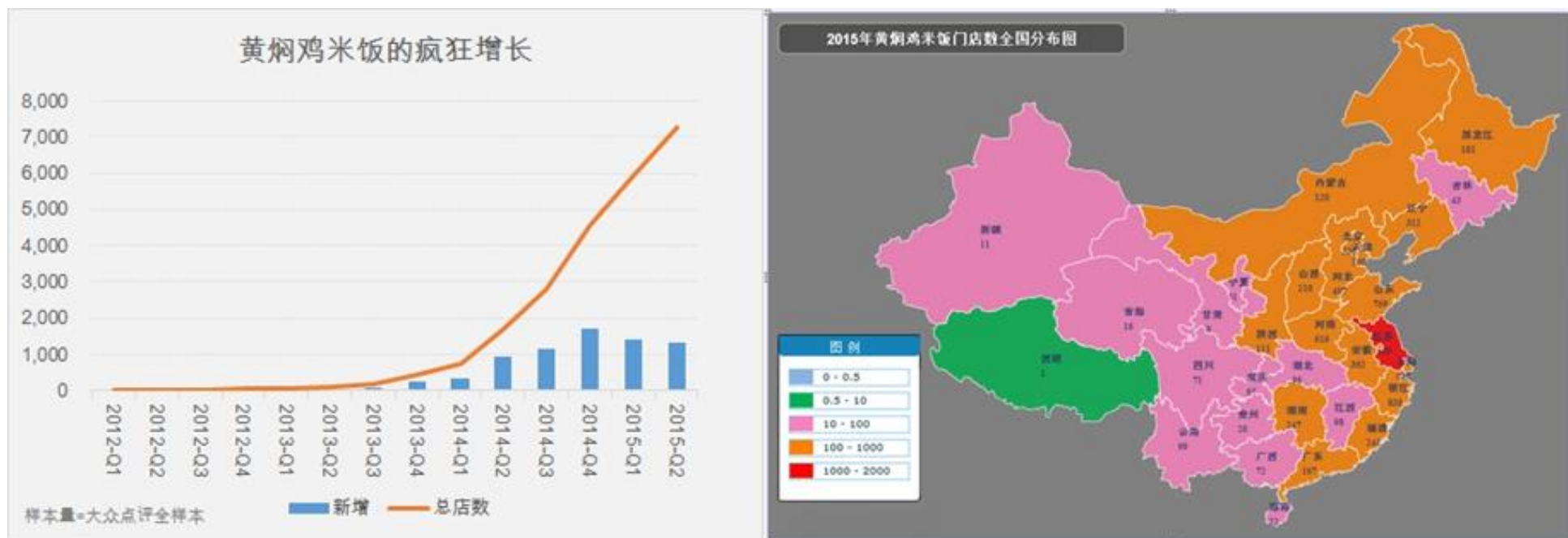
月成交 1266笔

评价 4604



# 为什么要学爬虫？

大众点评、美团网：抓取各种店面的开业情况以及用户消费和评价，了解周边变化的口味



# 为什么要学爬虫？

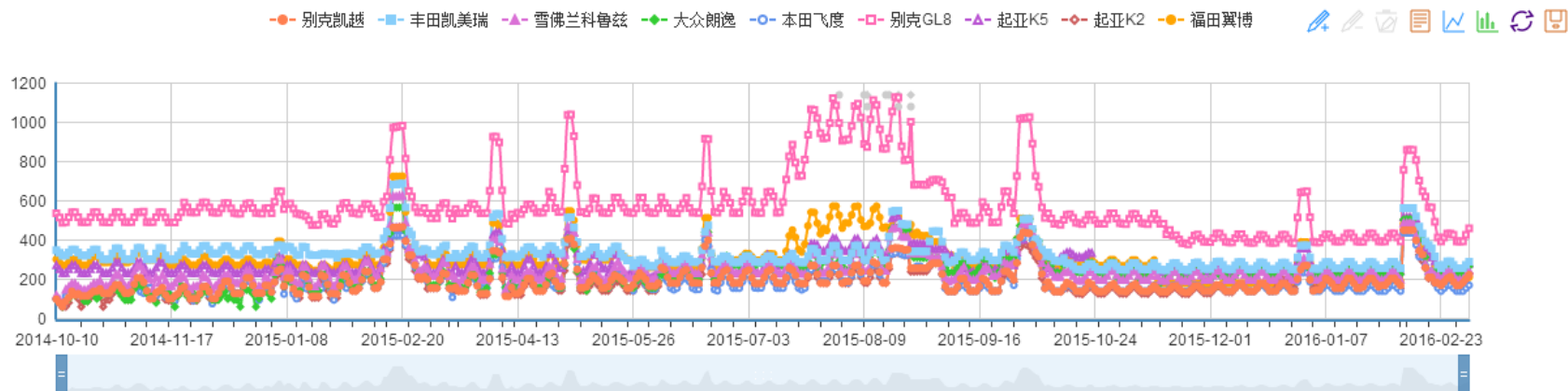
拉勾网、中华英才网：抓取各类职位信息，分析最热门的职业以及薪水



# 为什么要学爬虫？

## 神州租车、一嗨租车：跟踪租车价格及数量

北京地区各车型日均租赁价格变化趋势



# 为什么要学爬虫？

携程、去哪儿、12306：对航班及高铁等信息进行抓取，能从一个侧面反映经济情况





# 为什么要学爬虫？

把互联网变成你的数据库！

# 大数据产业解析

# 大数据产业全貌：什么是大数据

第一阶段

海量数据？

第二阶段

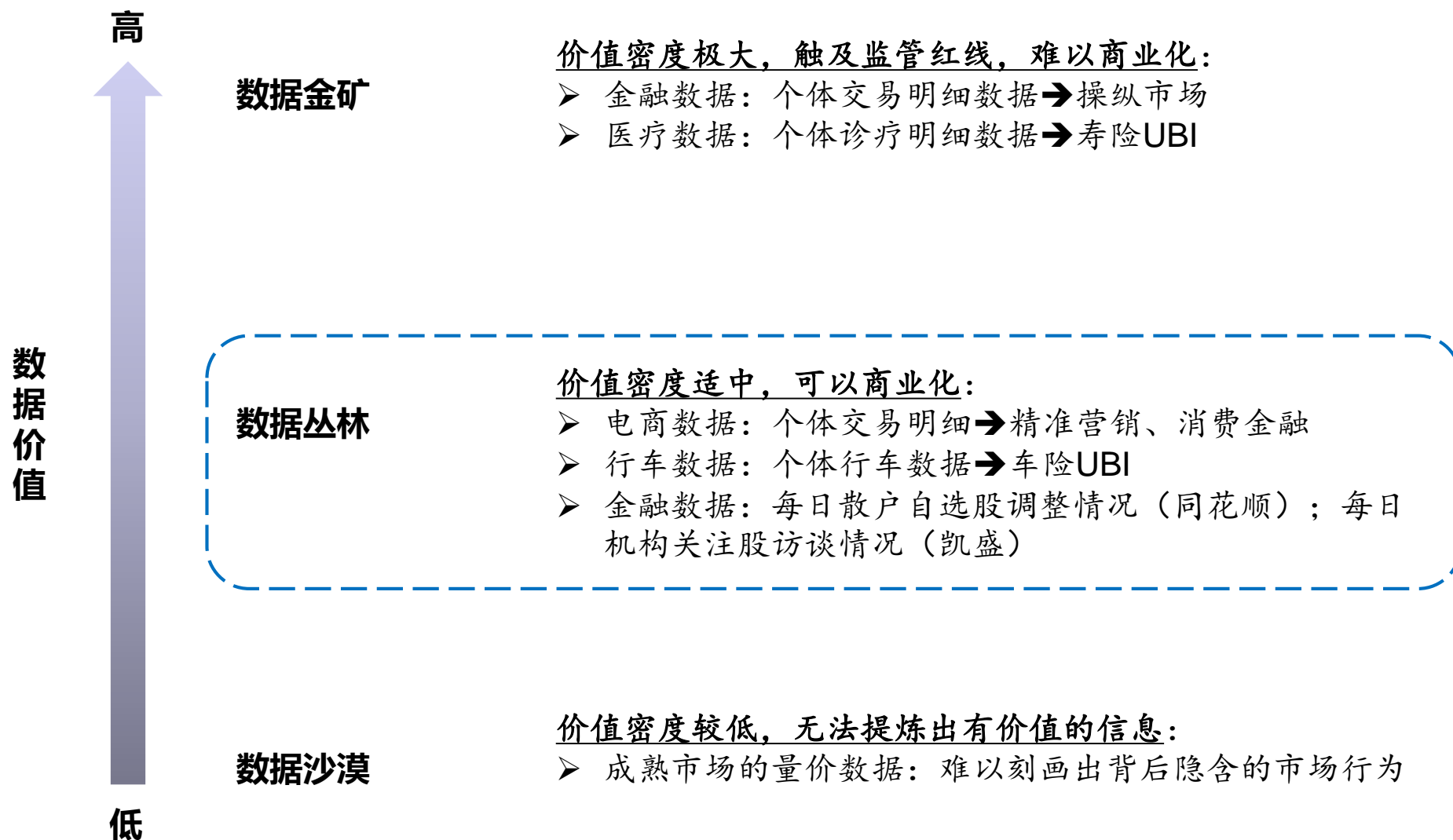
4V？

第三阶段

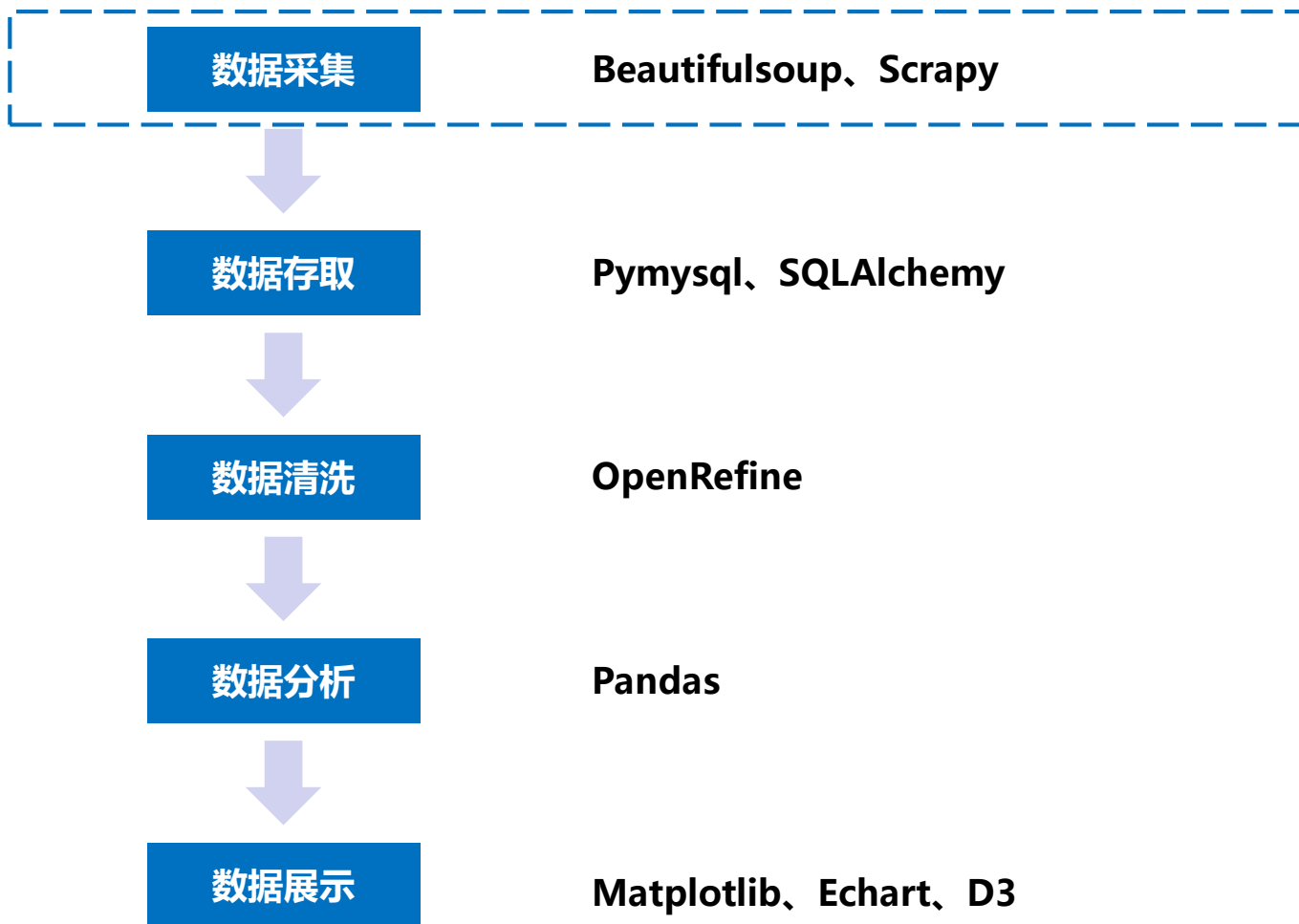
科技石油？

全量数据！

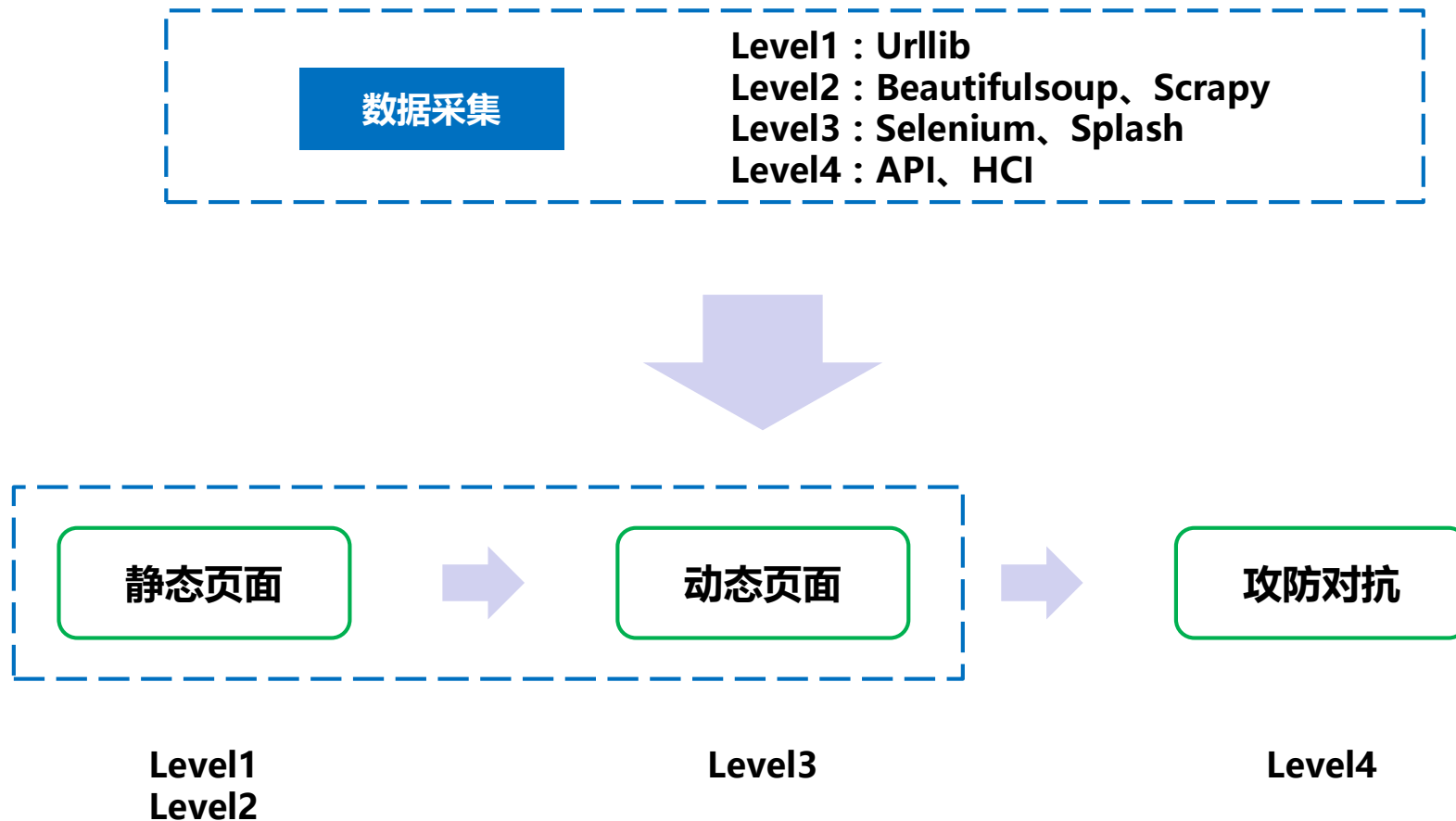
# 大数据产业全貌：三种大数据



# 大数据产业全貌：生产流水线

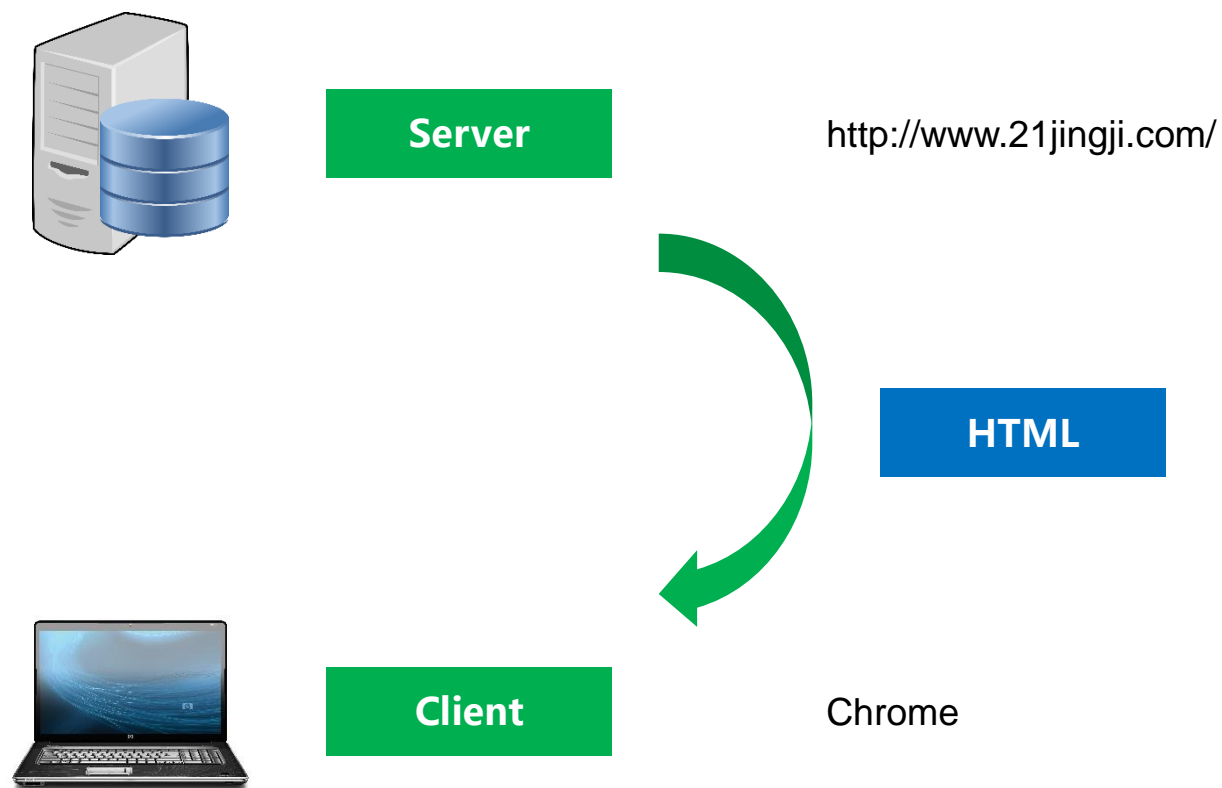


# 大数据产业全貌：爬虫进阶之路



# 基础技术知识

# 网络爬虫基础知识：什么是网页





# 网络爬虫基础知识：什么是网页



- 遏制日内过度投机 上期所出手“降温”螺纹钢

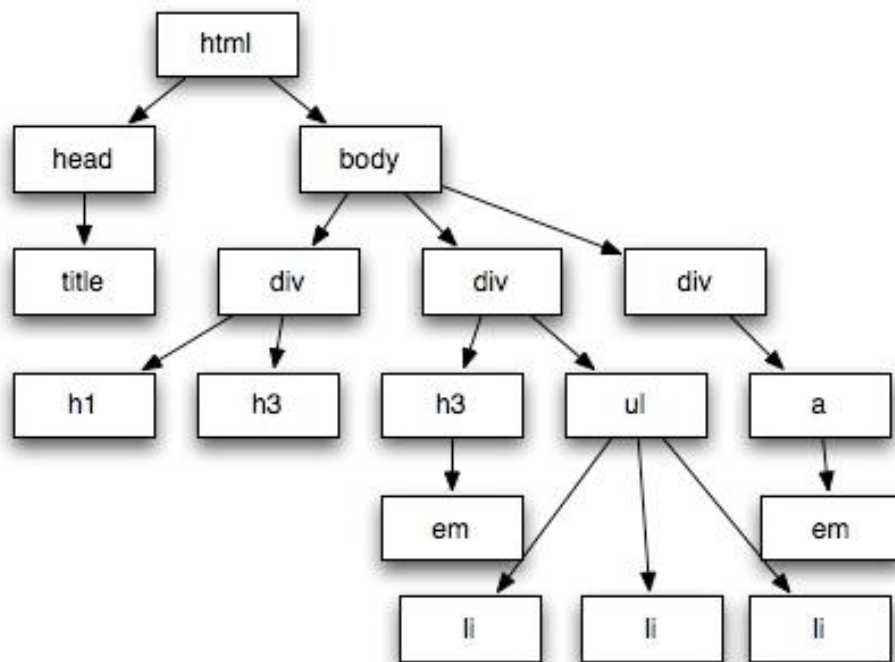
飙涨的螺纹钢，如今正面临多方面的调控压力。继中钢协公开表态后，8月11日上期所再次出手从期货市场“降温”。

- 一线城市供地面积已近去年全年 供求关系紧张将有所缓解

```
<body>
  <script language="JavaScript" type="text/JavaScript">...</script>
  <!--header-->
  <div id="head">...</div>
  <script type="text/javascript">...</script>
  <script type="text/javascript">...</script>
  <!--app下载广告-->
  <!-- <div style="font-size:12px; line-height:1.5em; background:#fff1db;
  border:solid 1px #f8c173; margin:0 10px 10px 10px; padding:5px 10px"><a
  href="http://www.21jingji.com/3g.html">下载21世纪经济报道APP客户端，阅读更方便。
  </a></div> -->
  <!--content-->
  <div id="slider" class="ui-slider">...</div>
  <div class="content">
    <div class="news_list">
      <a href="http://m.21jingji.com/article/20170808/herald/c0ed14318b9ac77dffe020934c72be9d.html">
        
        <div class="news_title">独家 | 监管征求意见：民营银行互联网贷款业务需核准</div>
        <div class="news_date">3小时前</div>
      </a>
      <a href="http://m.21jingji.com/article/20170808/herald/d6aa726070fe5168a2143a41aff81494.html">
        
        <div class="news_title">战狼2登顶中国最高票房，北京文化暴涨56%，高管要套现了</div>
        <div class="news_date">1小时前</div>
      </a>
    </div>
  </div>
```

# 网络爬虫基础知识：HTML

**什么是html：**网页的主体是html代码，html代码是由一个个tag构成的，tag之间是树状结构，每个tag代表特定的功能。浏览器通过解析html代码，显示出图形化的操作界面。



# 网络爬虫基础知识：HTML

标签	描述
<a href="#"><u>&lt;!--...--&gt;</u></a>	定义注释。
<a href="#"><u>&lt;!DOCTYPE&gt;</u></a>	定义文档类型。
<a href="#"><u>&lt;a&gt;</u></a>	定义锚。
<a href="#"><u>&lt;abbr&gt;</u></a>	定义缩写。
<a href="#"><u>&lt;acronym&gt;</u></a>	定义只取首字母的缩写。
<a href="#"><u>&lt;address&gt;</u></a>	定义文档作者或拥有者的联系信息。
<a href="#"><u>&lt;applet&gt;</u></a>	不赞成使用。定义嵌入的 applet。
<a href="#"><u>&lt;area&gt;</u></a>	定义图像映射内部的区域。
<a href="#"><u>&lt;article&gt;</u></a>	定义文章。
<a href="#"><u>&lt;aside&gt;</u></a>	定义页面内容之外的内容。
<a href="#"><u>&lt;audio&gt;</u></a>	定义声音内容。
<a href="#"><u>&lt;b&gt;</u></a>	定义粗体字。
<a href="#"><u>&lt;base&gt;</u></a>	定义页面中所有链接的默认地址或默认目标。
<a href="#"><u>&lt;basefont&gt;</u></a>	不赞成使用。定义页面中文本的默认字体、颜色或尺寸。
<a href="#"><u>&lt;bdi&gt;</u></a>	定义文本的文本方向，使其脱离其周围文本的方向设置。
<a href="#"><u>&lt;bdo&gt;</u></a>	定义文字方向。
<a href="#"><u>&lt;big&gt;</u></a>	定义大号文本。
<a href="#"><u>&lt;blockquote&gt;</u></a>	定义长的引用。
<a href="#"><u>&lt;body&gt;</u></a>	定义文档的主体。
<a href="#"><u>&lt;br&gt;</u></a>	定义简单的折行。
<a href="#"><u>&lt;button&gt;</u></a>	定义按钮 (push button)。
<a href="#"><u>&lt;canvas&gt;</u></a>	定义图形。
<a href="#"><u>&lt;caption&gt;</u></a>	定义表格标题。
<a href="#"><u>&lt;center&gt;</u></a>	不赞成使用。定义居中文本。
<a href="#"><u>&lt;cite&gt;</u></a>	定义引用(citation)。

# 网络爬虫基础知识：HTML

标签	描述
<a href="#"><u>&lt;code&gt;</u></a>	定义计算机代码文本。
<a href="#"><u>&lt;col&gt;</u></a>	定义表格中一个或多个列的属性值。
<a href="#"><u>&lt;colgroup&gt;</u></a>	定义表格中供格式化的列组。
<a href="#"><u>&lt;command&gt;</u></a>	定义命令按钮。
<a href="#"><u>&lt;datalist&gt;</u></a>	定义下拉列表。
<a href="#"><u>&lt;dd&gt;</u></a>	定义定义列表中项目的描述。
<a href="#"><u>&lt;del&gt;</u></a>	定义被删除文本。
<a href="#"><u>&lt;details&gt;</u></a>	定义元素的细节。
<a href="#"><u>&lt;dir&gt;</u></a>	不赞成使用。定义目录列表。
<a href="#"><u>&lt;div&gt;</u></a>	定义文档中的节。
<a href="#"><u>&lt;dfn&gt;</u></a>	定义定义项目。
<a href="#"><u>&lt;dialog&gt;</u></a>	定义对话框或窗口。
<a href="#"><u>&lt;dl&gt;</u></a>	定义定义列表。
<a href="#"><u>&lt;dt&gt;</u></a>	定义定义列表中的项目。
<a href="#"><u>&lt;em&gt;</u></a>	定义强调文本。
<a href="#"><u>&lt;embed&gt;</u></a>	定义外部交互内容或插件。
<a href="#"><u>&lt;fieldset&gt;</u></a>	定义围绕表单中元素的边框。
<a href="#"><u>&lt;figcaption&gt;</u></a>	定义 <b>figure</b> 元素的标题。
<a href="#"><u>&lt;figure&gt;</u></a>	定义媒介内容的分组，以及它们的标题。
<a href="#"><u>&lt;font&gt;</u></a>	不赞成使用。定义文字的字体、尺寸和颜色。
<a href="#"><u>&lt;footer&gt;</u></a>	定义 <b>section</b> 或 <b>page</b> 的页脚。
<a href="#"><u>&lt;form&gt;</u></a>	定义供用户输入的 <b>HTML</b> 表单。
<a href="#"><u>&lt;frame&gt;</u></a>	定义框架集的窗口或框架。
<a href="#"><u>&lt;frameset&gt;</u></a>	定义框架集。
<a href="#"><u>&lt;h1&gt; to &lt;h6&gt;</u></a>	定义 <b>HTML</b> 标题。

# 网络爬虫基础知识：HTML

标签	描述
<a href="#"><u>&lt;head&gt;</u></a>	定义关于文档的信息。
<a href="#"><u>&lt;header&gt;</u></a>	定义 section 或 page 的页眉。
<a href="#"><u>&lt;hr&gt;</u></a>	定义水平线。
<a href="#"><u>&lt;html&gt;</u></a>	定义 HTML 文档。
<a href="#"><u>&lt;i&gt;</u></a>	定义斜体字。
<a href="#"><u>&lt;iframe&gt;</u></a>	定义内联框架。
<a href="#"><u>&lt;img&gt;</u></a>	定义图像。
<a href="#"><u>&lt;input&gt;</u></a>	定义输入控件。
<a href="#"><u>&lt;ins&gt;</u></a>	定义被插入文本。
<a href="#"><u>&lt;isindex&gt;</u></a>	不赞成使用。定义与文档相关的可搜索索引。
<a href="#"><u>&lt;kbd&gt;</u></a>	定义键盘文本。
<a href="#"><u>&lt;keygen&gt;</u></a>	定义生成密钥。
<a href="#"><u>&lt;label&gt;</u></a>	定义 input 元素的标注。
<a href="#"><u>&lt;legend&gt;</u></a>	定义 fieldset 元素的标题。
<a href="#"><u>&lt;li&gt;</u></a>	定义列表的项目。
<a href="#"><u>&lt;link&gt;</u></a>	定义文档与外部资源的关系。
<a href="#"><u>&lt;map&gt;</u></a>	定义图像映射。
<a href="#"><u>&lt;mark&gt;</u></a>	定义有记号的文本。
<a href="#"><u>&lt;menu&gt;</u></a>	定义命令的列表或菜单。
<a href="#"><u>&lt;menuitem&gt;</u></a>	定义用户可以从弹出菜单调用的命令/菜单项目。
<a href="#"><u>&lt;meta&gt;</u></a>	定义关于 HTML 文档的元信息。
<a href="#"><u>&lt;meter&gt;</u></a>	定义预定义范围内的度量。
<a href="#"><u>&lt;nav&gt;</u></a>	定义导航链接。
<a href="#"><u>&lt;noframes&gt;</u></a>	定义针对不支持框架的用户的替代内容。
<a href="#"><u>&lt;noscript&gt;</u></a>	定义针对不支持客户端脚本的用户的替代内容。

# 网络爬虫基础知识：HTML

标签	描述
<a href="#"><u>&lt;object&gt;</u></a>	定义内嵌对象。
<a href="#"><u>&lt;ol&gt;</u></a>	定义有序列表。
<a href="#"><u>&lt;optgroup&gt;</u></a>	定义选择列表中相关选项的组合。
<a href="#"><u>&lt;option&gt;</u></a>	定义选择列表中的选项。
<a href="#"><u>&lt;output&gt;</u></a>	定义输出的一些类型。
<a href="#"><u>&lt;p&gt;</u></a>	定义段落。
<a href="#"><u>&lt;param&gt;</u></a>	定义对象的参数。
<a href="#"><u>&lt;pre&gt;</u></a>	定义预格式文本。
<a href="#"><u>&lt;progress&gt;</u></a>	定义任何类型的任务的进度。
<a href="#"><u>&lt;q&gt;</u></a>	定义短的引用。
<a href="#"><u>&lt;rp&gt;</u></a>	定义若浏览器不支持 ruby 元素显示的内容。
<a href="#"><u>&lt;rt&gt;</u></a>	定义 ruby 注释的解释。
<a href="#"><u>&lt;ruby&gt;</u></a>	定义 ruby 注释。
<a href="#"><u>&lt;s&gt;</u></a>	不赞成使用。定义加删除线的文本。
<a href="#"><u>&lt;samp&gt;</u></a>	定义计算机代码样本。
<a href="#"><u>&lt;script&gt;</u></a>	定义客户端脚本。
<a href="#"><u>&lt;section&gt;</u></a>	定义 section。
<a href="#"><u>&lt;select&gt;</u></a>	定义选择列表（下拉列表）。
<a href="#"><u>&lt;small&gt;</u></a>	定义小号文本。
<a href="#"><u>&lt;source&gt;</u></a>	定义媒介源。
<a href="#"><u>&lt;span&gt;</u></a>	定义文档中的节。
<a href="#"><u>&lt;strike&gt;</u></a>	不赞成使用。定义加删除线文本。
<a href="#"><u>&lt;strong&gt;</u></a>	定义强调文本。
<a href="#"><u>&lt;style&gt;</u></a>	定义文档的样式信息。
<a href="#"><u>&lt;sub&gt;</u></a>	定义下标文本。

# 网络爬虫基础知识：HTML

标签	描述
<a href="#"><code>&lt;summary&gt;</code></a>	为 <code>&lt;details&gt;</code> 元素定义可见的标题。
<a href="#"><code>&lt;sup&gt;</code></a>	定义上标文本。
<a href="#"><code>&lt;table&gt;</code></a>	定义表格。
<a href="#"><code>&lt;tbody&gt;</code></a>	定义表格中的主体内容。
<a href="#"><code>&lt;td&gt;</code></a>	定义表格中的单元。
<a href="#"><code>&lt;textarea&gt;</code></a>	定义多行的文本输入控件。
<a href="#"><code>&lt;tfoot&gt;</code></a>	定义表格中的表注内容（脚注）。
<a href="#"><code>&lt;th&gt;</code></a>	定义表格中的表头单元格。
<a href="#"><code>&lt;thead&gt;</code></a>	定义表格中的表头内容。
<a href="#"><code>&lt;time&gt;</code></a>	定义日期/时间。
<a href="#"><code>&lt;title&gt;</code></a>	定义文档的标题。
<a href="#"><code>&lt;tr&gt;</code></a>	定义表格中的行。
<a href="#"><code>&lt;track&gt;</code></a>	定义用在媒体播放器中的文本轨道。
<a href="#"><code>&lt;tt&gt;</code></a>	定义打字机文本。
<a href="#"><code>&lt;u&gt;</code></a>	不赞成使用。定义下划线文本。
<a href="#"><code>&lt;ul&gt;</code></a>	定义无序列表。
<a href="#"><code>&lt;var&gt;</code></a>	定义文本的变量部分。
<a href="#"><code>&lt;video&gt;</code></a>	定义视频。
<a href="#"><code>&lt;wbr&gt;</code></a>	定义可能的换行符。
<a href="#"><code>&lt;xmp&gt;</code></a>	不赞成使用。定义预格式文本。

# 网络爬虫基础知识：HTML

属性	描述
<a href="#">accesskey</a>	规定激活元素的快捷键。
<a href="#">class</a>	规定元素的一个或多个类名（引用样式表中的类）。
<a href="#">contenteditable</a>	规定元素内容是否可编辑。
<a href="#">contextmenu</a>	规定元素的上下文菜单。上下文菜单在用户点击元素时显示。
<a href="#">data-*</a>	用于存储页面或应用程序的私有定制数据。
<a href="#">dir</a>	规定元素中内容的文本方向。
<a href="#">draggable</a>	规定元素是否可拖动。
<a href="#">dropzone</a>	规定在拖动被拖动数据时是否进行复制、移动或链接。
<a href="#">hidden</a>	规定元素仍未或不再相关。
<a href="#">id</a>	规定元素的唯一 id。
<a href="#">lang</a>	规定元素内容的语言。
<a href="#">spellcheck</a>	规定是否对元素进行拼写和语法检查。
<a href="#">style</a>	规定元素的行内 CSS 样式。
<a href="#">tabindex</a>	规定元素的 tab 键次序。
<a href="#">title</a>	规定有关元素的额外信息。
<a href="#">translate</a>	规定是否应该翻译元素内容。

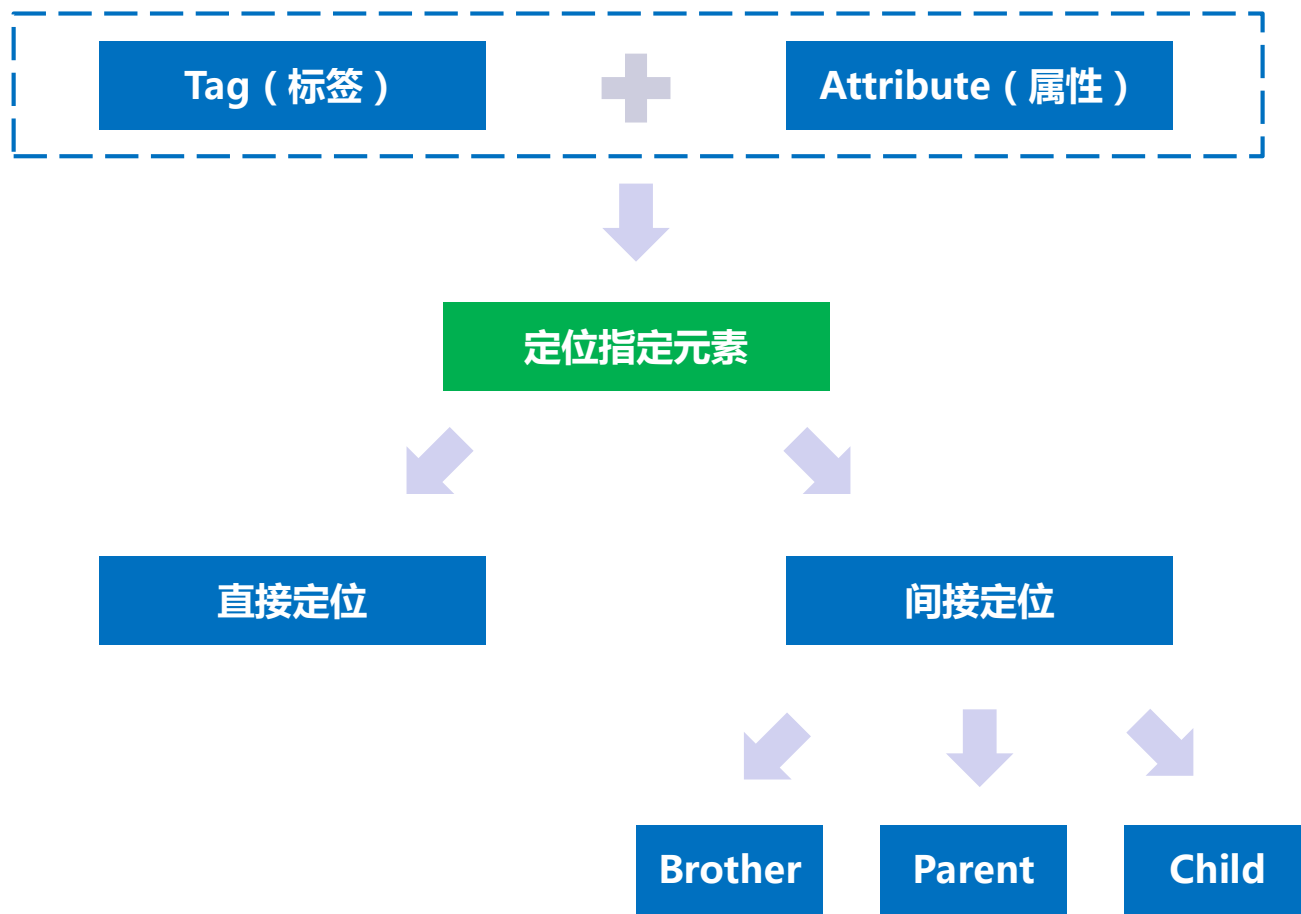
每个tag都有自己的属性（attribute），比如id、class、style，我们可以通过这些属性来定位指定tag。

```
<div class="news_date">1小时前</div>
```

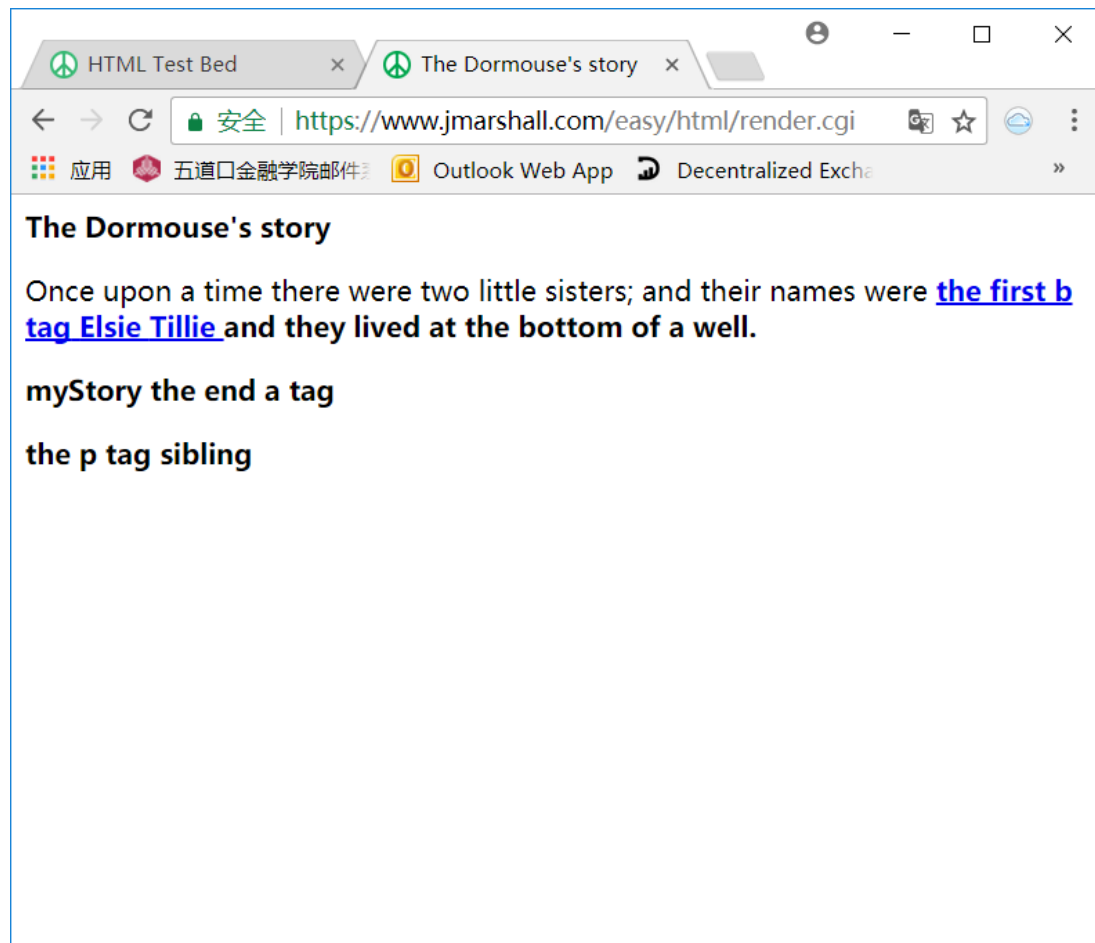
```
<div class="news_title">独家 | 监管征求意见：民营银行互联网贷款业务需核准</div>
```



# 网络爬虫基础知识：HTML



# 网络爬虫基础知识：HTML



# 网络爬虫基础知识：HTML

```
<html>
  <head>
    <title>The Dormouse's story</title>
  </head>
  <body>
    <p class="title" name="dromouse">
      <b>The Dormouse's story</b>
    </p>
    <p class="story">
      Once upon a time there were two little sisters; and their names were
      <a class="mysis" href="http://example.com/elsie" id="link1">
        <b>the first b tag<b>
        Elsie
      </a>
      <a class="mysis" href="http://example.com/tillie" id="link2" myname="kong" >
        Tillie
      </a>and they lived at the bottom of a well.
    </p>
    <p class="story">
      myStory
      <a>the end a tag</a>
    </p>
    <a>the p tag sibling</a>
  </body>
</html>
```

# 网络爬虫基础知识：HTML

```
# 选择所有title标签
soup.select("title")
```

```
# 选择所有p标签中的第三个标签
soup.select("p:nth-of-type(3)")
```

```
# 选择body标签下的直接a子标签
soup.select("body > a")
```

```
# 选择id=link1后的所有兄弟节点标签
soup.select("#link1 ~ .mysis")
```

```
# 选择id=link1后的下一个兄弟节点标签
soup.select("#link1 + .mysis")
```

```
# 选择a标签，其类属性为mysis的标签
soup.select("a.mysis")
```

```
# 选择a标签，其id属性为link1的标签
soup.select("a#link1")
```

```
# 选择a标签，其属性中存在myname的所有标签
soup.select("a[myname]")
```

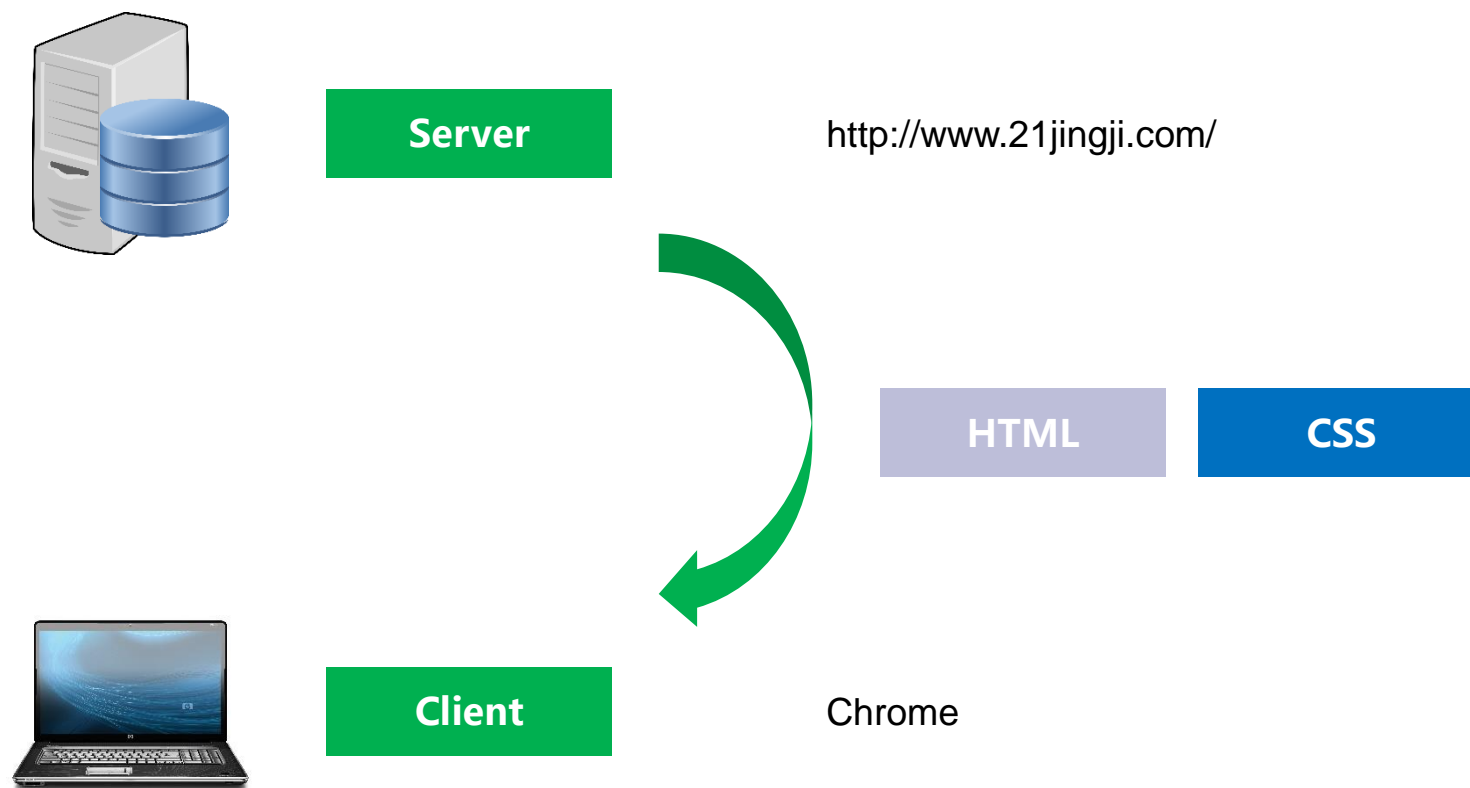
```
# 选择a标签，其属性href满足条件的所有标签
soup.select("a[href='http://example.com/elsie']")
```

```
# 选择a标签，其href属性以http开头
soup.select('a[href^="http"]')
```

```
# 选择a标签，其href属性以lacie结尾
soup.select('a[href$="elsie"]')
```

```
# 选择a标签，其href属性包含.com
soup.select('a[href*=".com"]')
```

# 网络爬虫基础知识：什么是网页



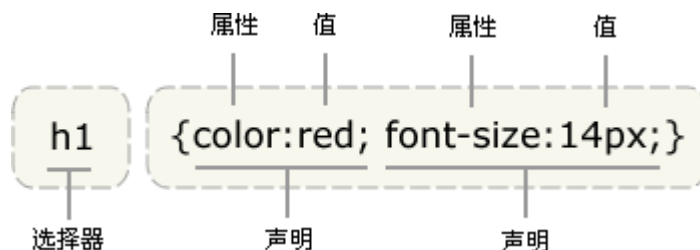
# 网络爬虫基础知识：CSS

什么是CSS：CSS 指层叠样式表 (Cascading Style Sheets)，简称“样式”。样式定义如何显示html元素。由于其多样性，样式可以当做筛选器来使用。

```
h1 { color: white;
      background: orange;
      border: 1px solid black;
      padding: 0 0 0 0;
      font-weight: bold;
    }
/* begin: seaside-theme */

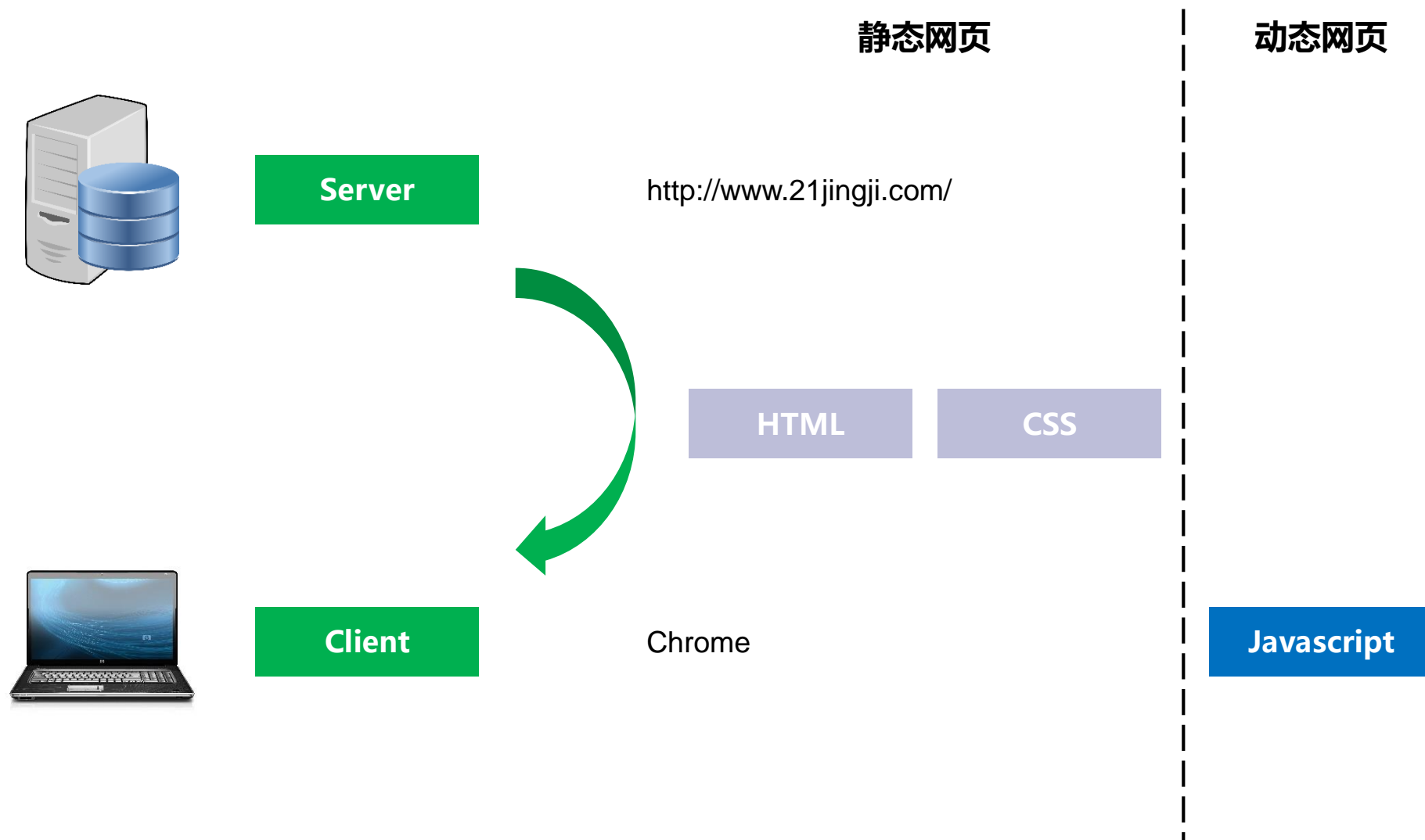
body {
  background-color: white;
  color: black;
  font-family: Arial, sans-serif;
  margin: 0 4px 0 0;
  border: 12px solid;
}
```

CSS



CSS  
IS  
AWESOME

# 网络爬虫基础知识：什么是网页



# 网络爬虫基础知识：Javascript

**什么是Javascript:** Javascript是一种运行在浏览器中的解释型的编程语言，你在电脑、手机、平板上浏览的所有的网页，以及无数基于html5的手机App，其交互逻辑都是由JavaScript驱动的。一句话来说，任何动态网页都是离不开Javascript的。

```
<script type="text/javascript">
  var imgObj = null;

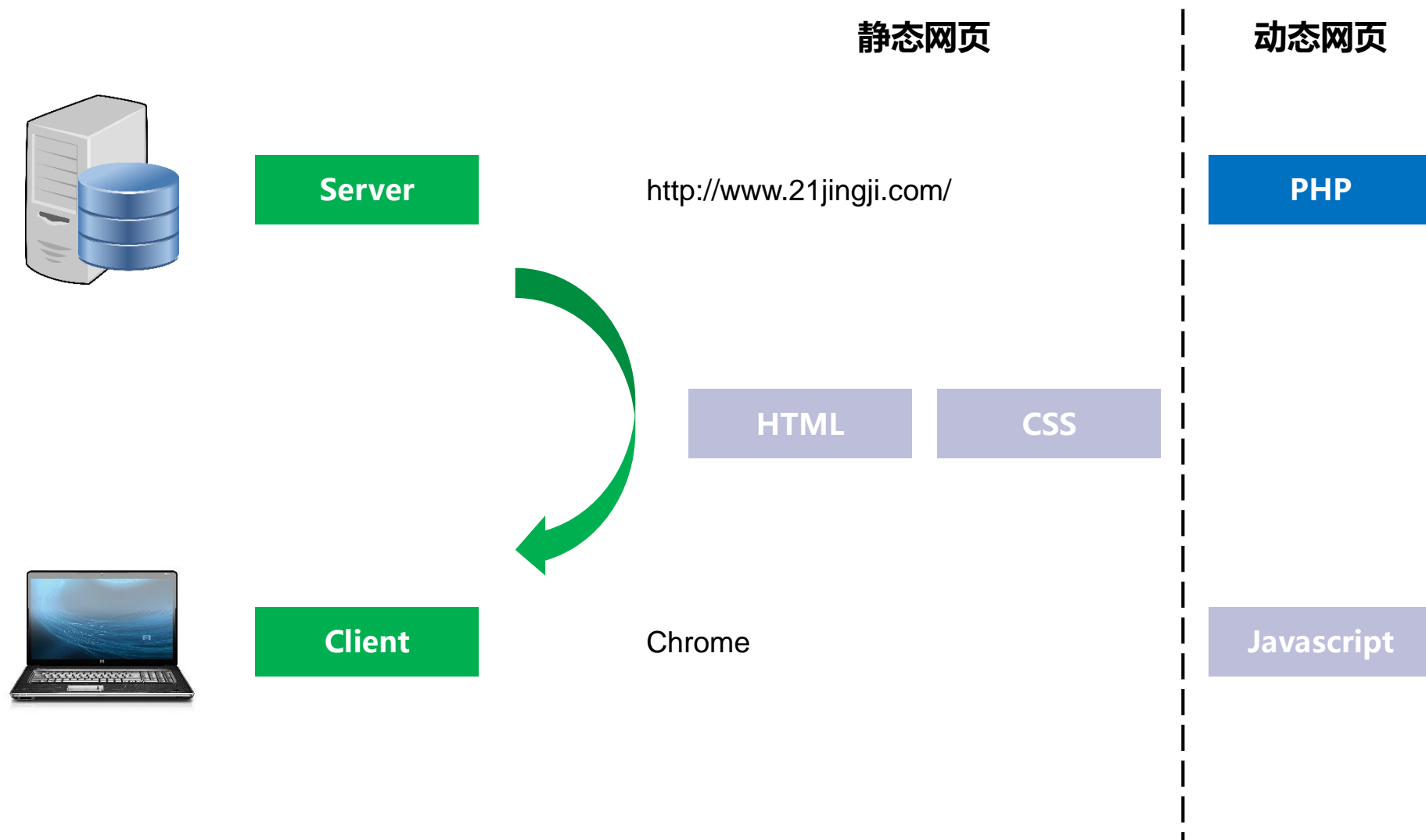
  function init(){
    imgObj = document.getElementById('myImage');
    imgObj.style.position= 'relative';
    imgObj.style.left = '0px';
  }

  window.onload =init;
</script>
```





# 网络爬虫基础知识：什么是网页



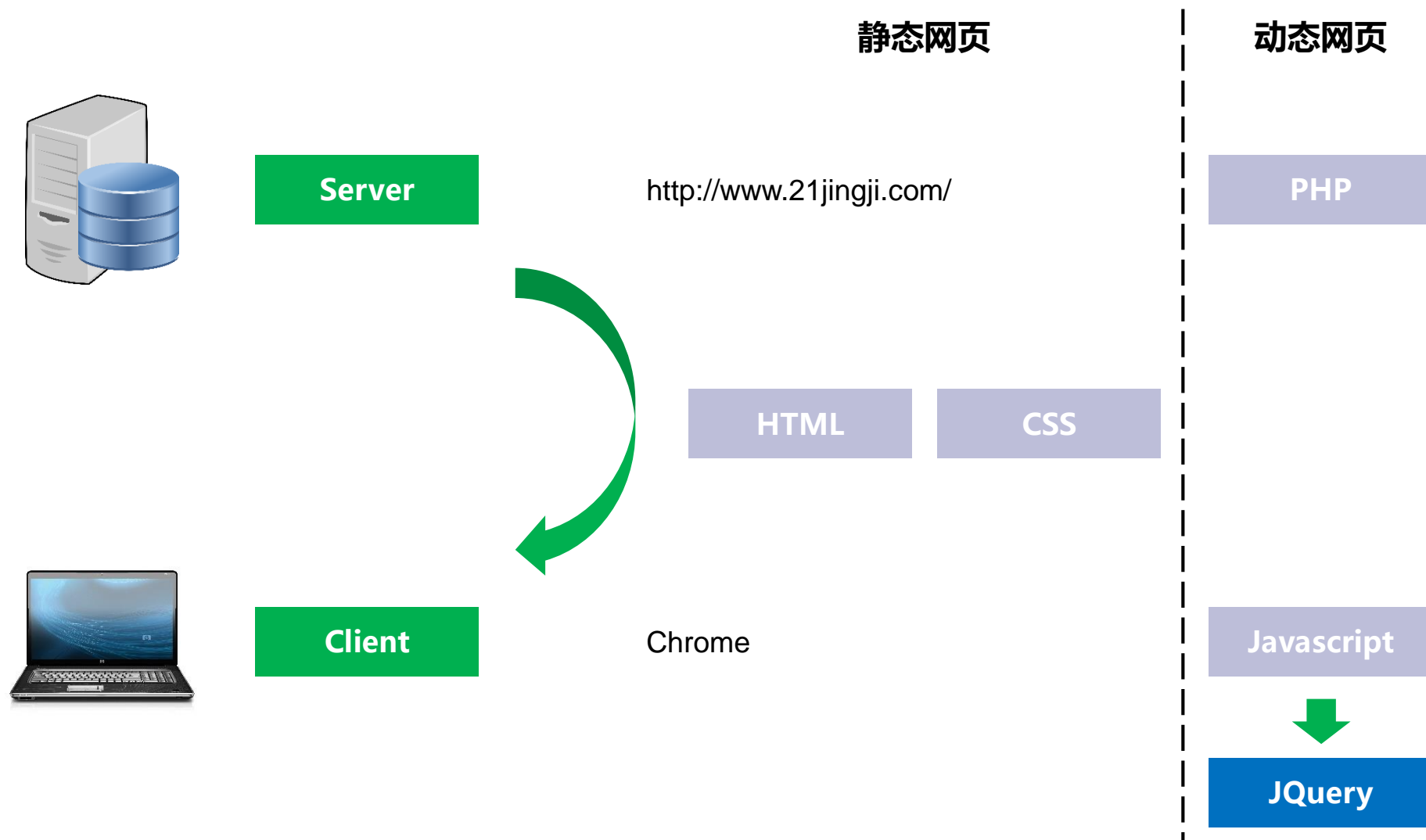
# 网络爬虫基础知识：PHP

什么是PHP：PHP全称为Hypertext Preprocessor，是一种开源的通用计算机脚本语言，尤其适用于网络开发并可嵌入html中使用。

```
<?php
    navi($username);
    $availableTags = array();
    $result = mysqli_query($con,"SELECT label
FROM `datacenter` WHERE 1 GROUP BY label");
    while($row = mysqli_fetch_array($result))
    {
        array_push($availableTags,$row['label']);
    }
?>
```



# 网络爬虫基础知识：什么是网页



# 网络爬虫基础知识：jQuery

什么是jQuery：jQuery是一个Javascript库，不仅提供丰富的交互功能，而且极大地简化了Javascript编程。

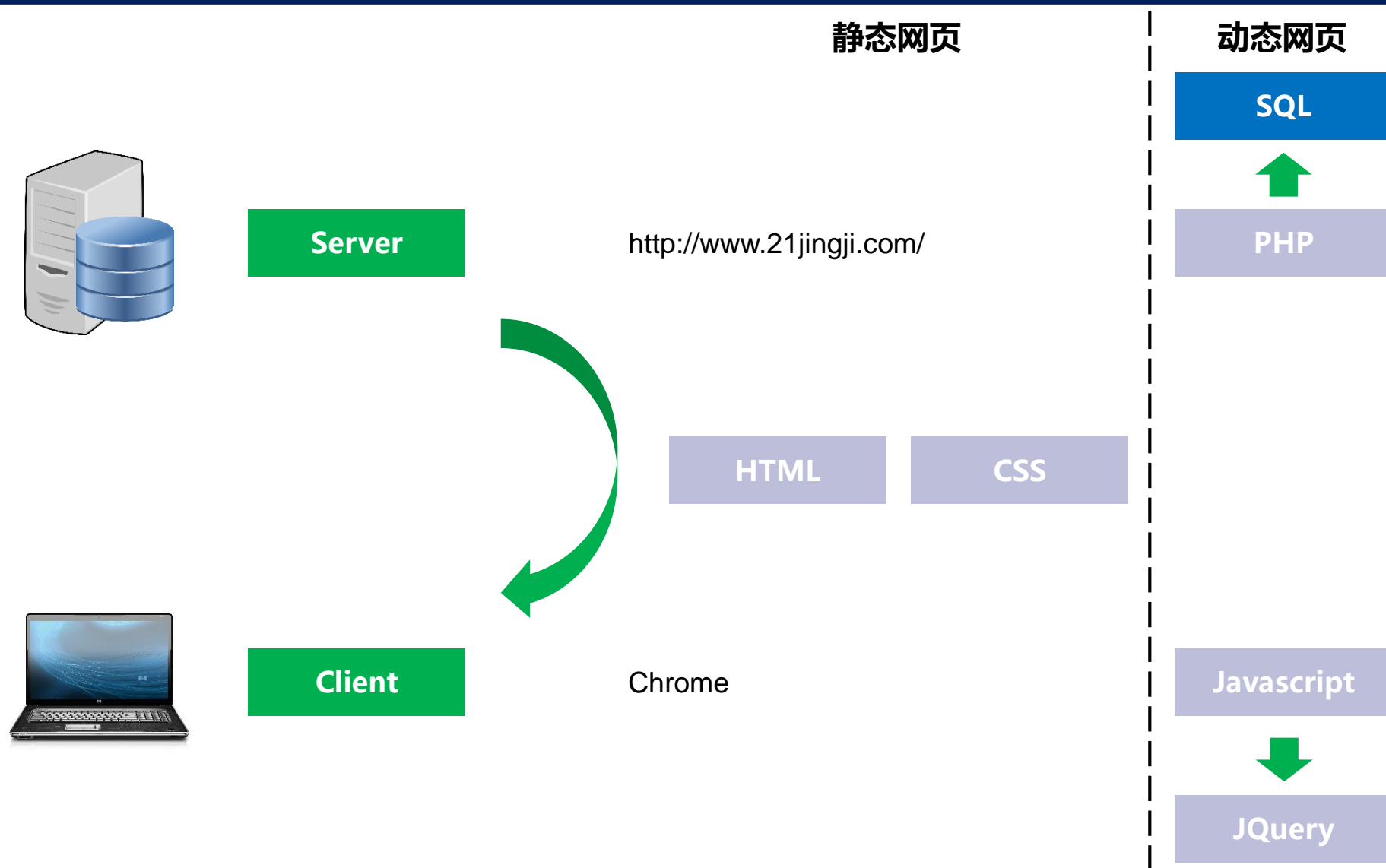
```
<html>
<head>
<script type="text/javascript"
src="jquery.js"></script>
<script type="text/javascript">
$(document).ready(function(){
    $("p").click(function(){
        $(this).hide();
    });
});
</script>
</head>

<body>
<p>If you click on me, I will disappear.</p>
</body>

</html>
```



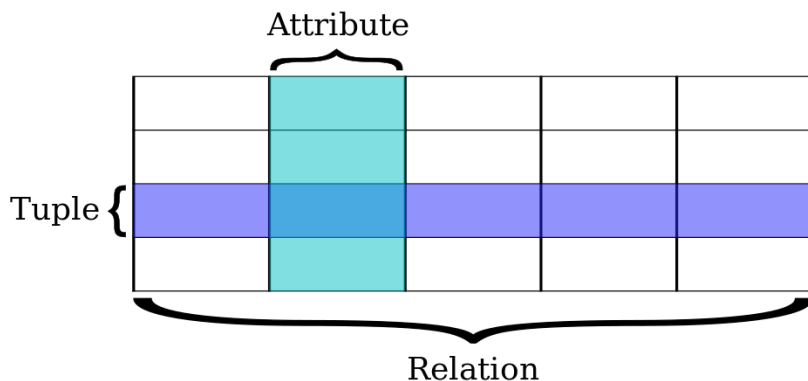
# 网络爬虫基础知识：什么是网页



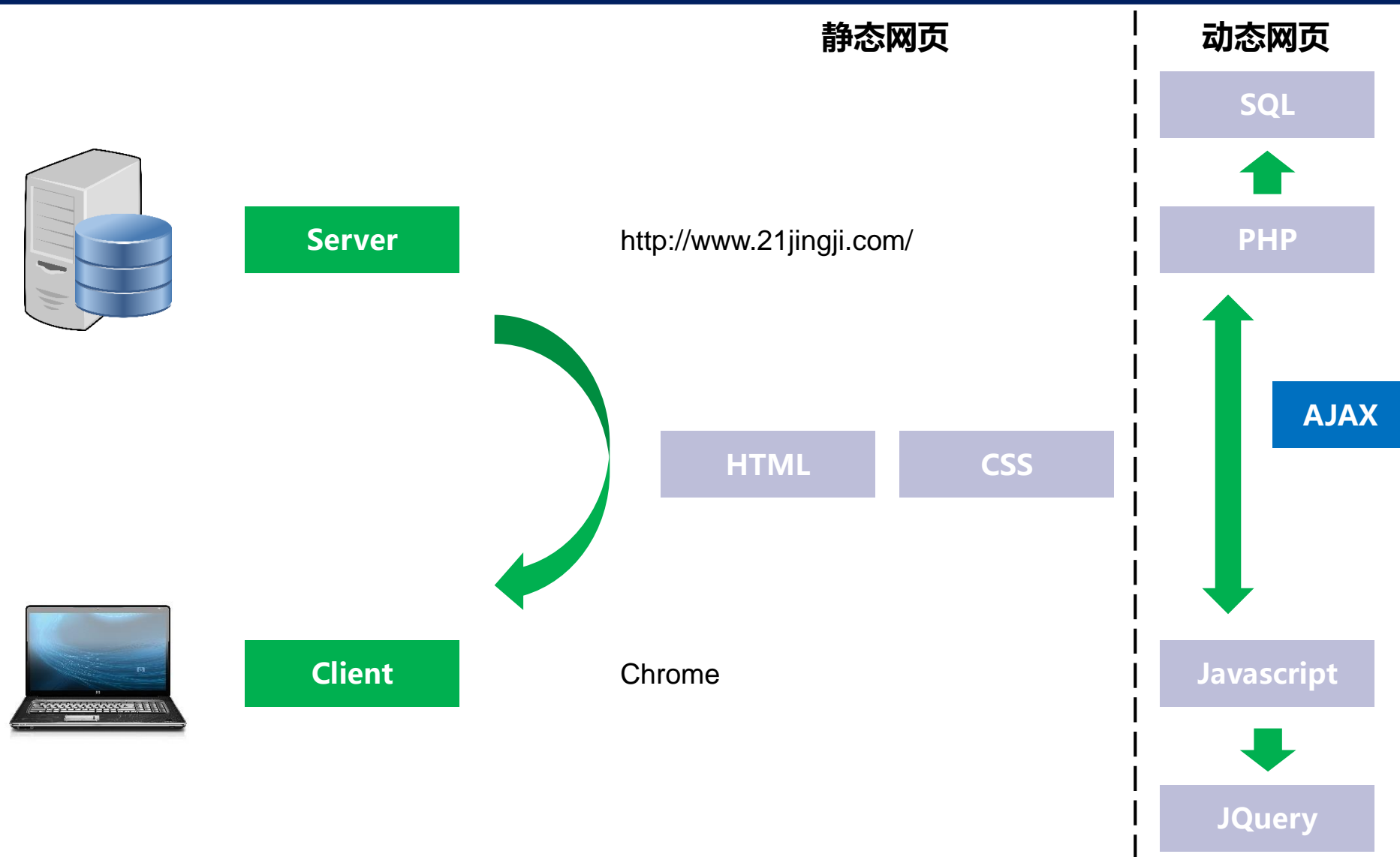
# 网络爬虫基础知识：SQL

数据库主要可以分为关系型数据库（SQL）与非关系型数据库（NoSQL），前者的典型代表是MySQL，后者的代表是MongoDB。关系型数据库中每一列称为Attribute，每一行称为Tuple，表与表之间可以通过指定字段进行JOIN操作。SQL的基础操作分为四块，即SELECT、INSERT、UPDATE、DELETE，分别对应选择、插入、修改、删除这四项功能。除此之外，SQL拥有种类丰富的Aggregate Function，需要搭配GROUP BY使用。

DBMS	Database Model	Score
Oracle	Relational DBMS	1439.72
MySQL	Relational DBMS	1272.45
Microsoft SQL Server	Relational DBMS	1177.48
MongoDB	Document store	267.24
PostgreSQL	Relational DBMS	262.34
DB2	Relational DBMS	202.42
Microsoft Access	Relational DBMS	140.54
Cassandra	Wide column store	107.08
SQLite	Relational DBMS	99.56
Redis	Key-value store	99.21



# 网络爬虫基础知识：什么是网页



# 网络爬虫基础知识：AJAX

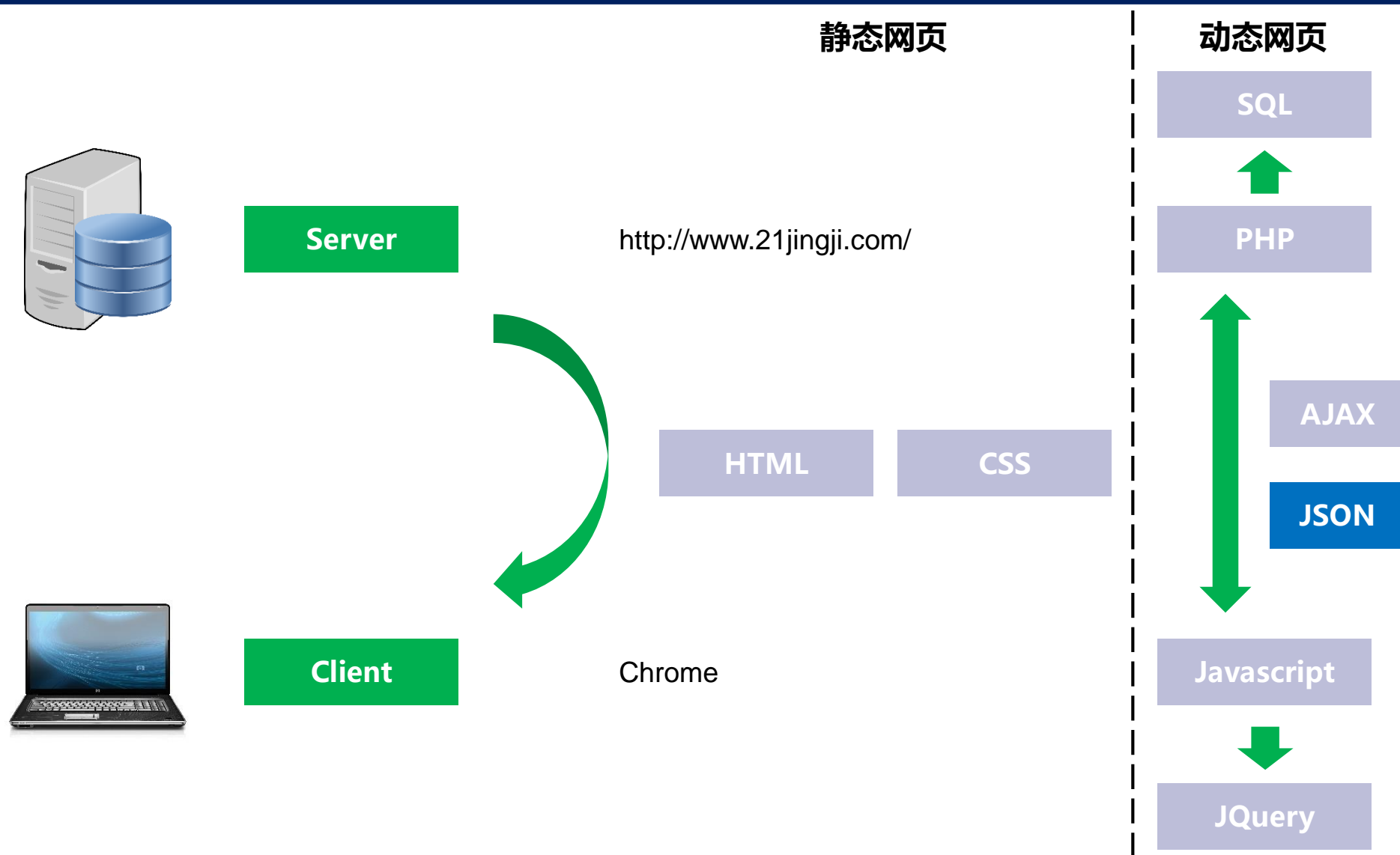
什么是AJAX：AJAX即“Asynchronous JavaScript and XML”，是基于Javascript之上的高级功能，重点在“Asynchronous”，即异步。当你访问的网页可以在不刷新页面的情况下动态加载内容，那它就使用了AJAX技术。

```
$(document).on('click', '#update', function () {  
    event.preventDefault();  
    var label = $("#label").val();  
    var DID = <?php echo json_encode($DID) ?>;  
    var html = $.ajax({  
        type: "POST",  
        url: "update_meta.php",  
        data: "label="+label+"&DID="+DID,  
        async: false  
    }).responseText;  
  
    alert(html);  
});
```





# 网络爬虫基础知识：什么是网页



# 网络爬虫基础知识：JSON

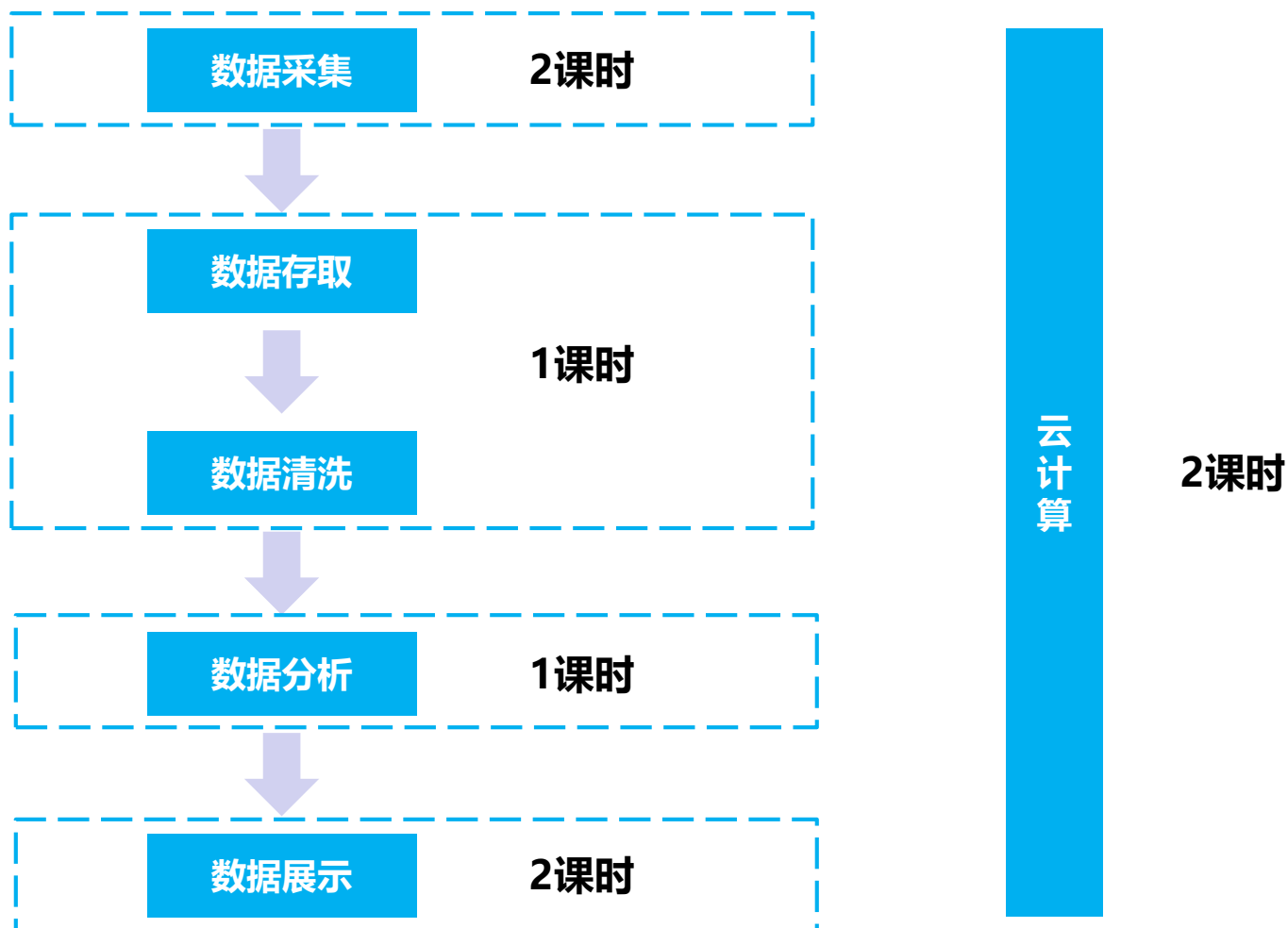
**什么是JSON：**JSON的全称是JavaScript 对象表示法（JavaScript Object Notation）。JSON可以将 JavaScript 对象中表示的一组数据转换为字符串，然后就可以在网络或者程序之间轻松地传递这个字符串，并在需要的时候将它还原为各编程语言所支持的数据格式。

```
<script type="text/javascript">
var JSONObject= {
    "name":"Bill Gates",
    "street":"Fifth Avenue New York 666",
    "age":56,
    "phone":"555 1234567"
};
document.getElementById("jname").innerHTML=JSONObject.name
document.getElementById("jage").innerHTML=JSONObject.age
document.getElementById("jstreet").innerHTML=JSONObject.street
document.getElementById("jphone").innerHTML=JSONObject.phone
</script>
```



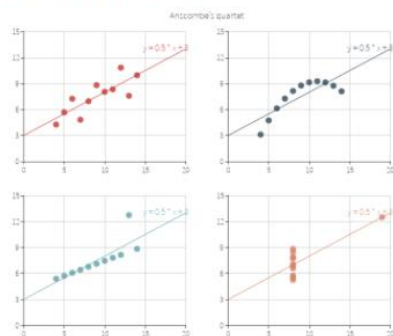
# 大数据系列课程体系

# 大数据系列课程体系

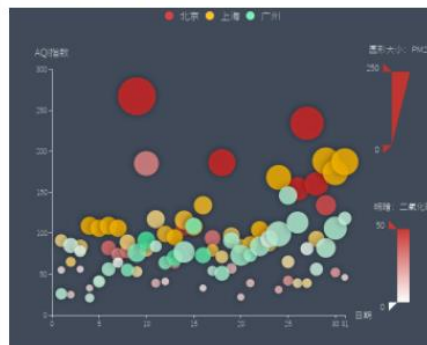


# 做大数据时代的梵高：ECharts

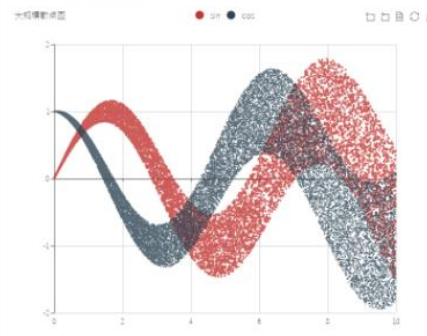
Anscombe's quartet



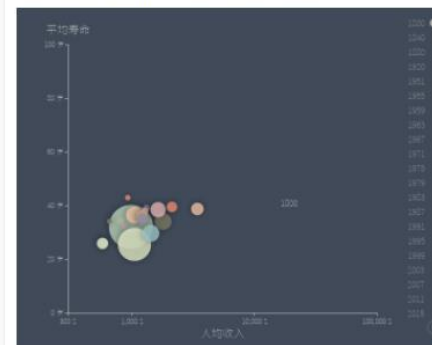
Scatter Aqi Color



大规模散点图



各国人均寿命与GDP关系演变



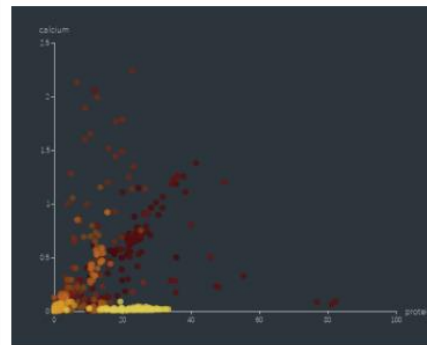
Scatter Map Brush



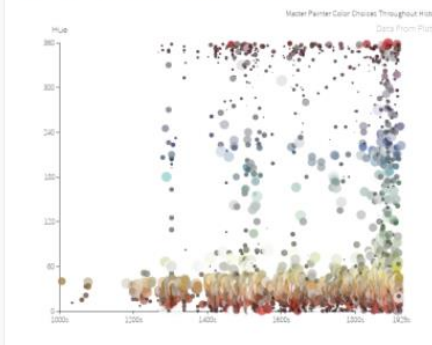
全国主要城市空气质量



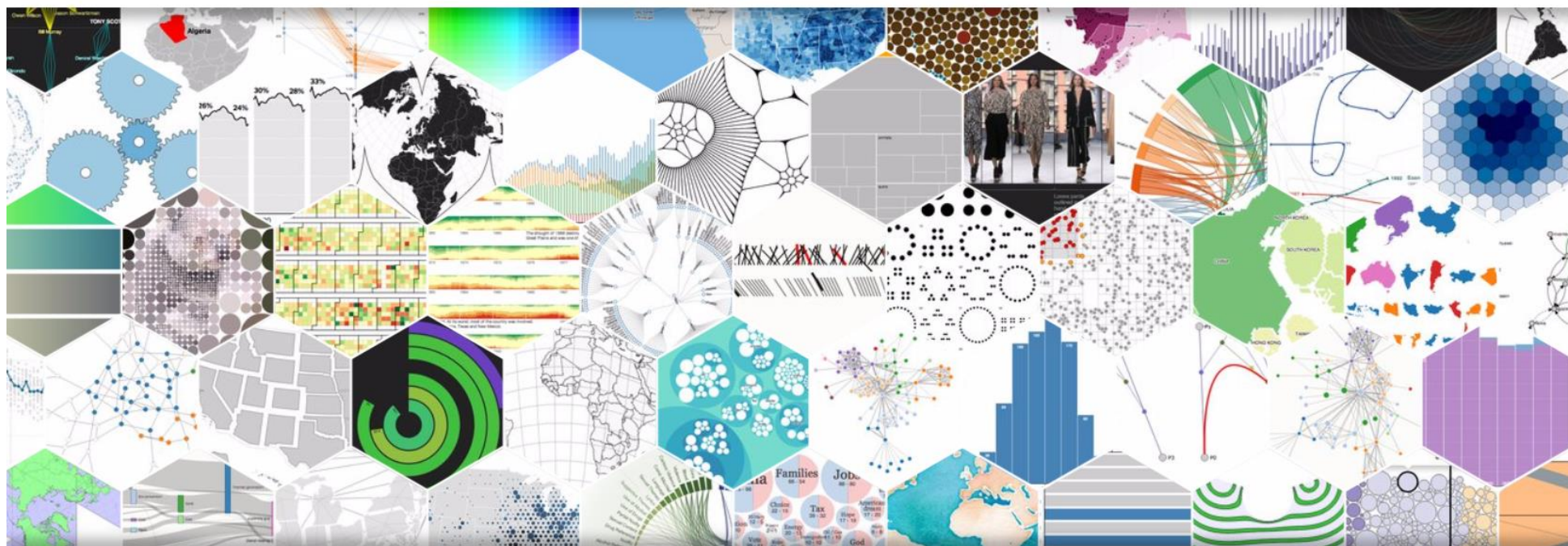
Scatter Nutrients



Master Painter Color Choices Througho...



## 做大数据时代的梵高：D3



# 做大数据时代的梵高：D3

```
<script>
var svg = d3.select("body").append("svg")
    .attr("width", width)
    .attr("height", height)
    .append("g")
    .attr("transform", "translate(" + 2 * margin + "," + margin + ")");

d3.json("marimekko.json", function(error, data) {
    if (error) throw error;
    var xtick = svg.selectAll(".x")
        .data(x.ticks(10))
        .enter().append("g")
        .attr("class", "x")
        .attr("transform", function(d) { return "translate(" + x(d) + ","
+ y(1) + ")"; });
</script>
```

## 本次培训需要用到的基础软件



**Python编译环境**



**PyCharm开发IDE**



**Apache+PHP+MySQL**



**开始实操！**