

2015 天猫双11全球狂欢节

2015 TMALL 11.11 GLOBAL SHOPPING FESTIVAL



“剁手”狂狂欢节交易额达133亿元，其中无线占比70.88%。



北京

Beijing

上海

Shanghai

广州

Guangzhou

深圳

Shenzhen

杭州

Hangzhou



北京
Beijing

上海
Shanghai

广州
Guangzhou

深圳
Shenzhen

杭州
Hangzhou

量级

1亿 100万 0.1s 30万年

地址合并

上海市闵行区东川路800号 西区32幢 506



上海市闵行区东川路800号

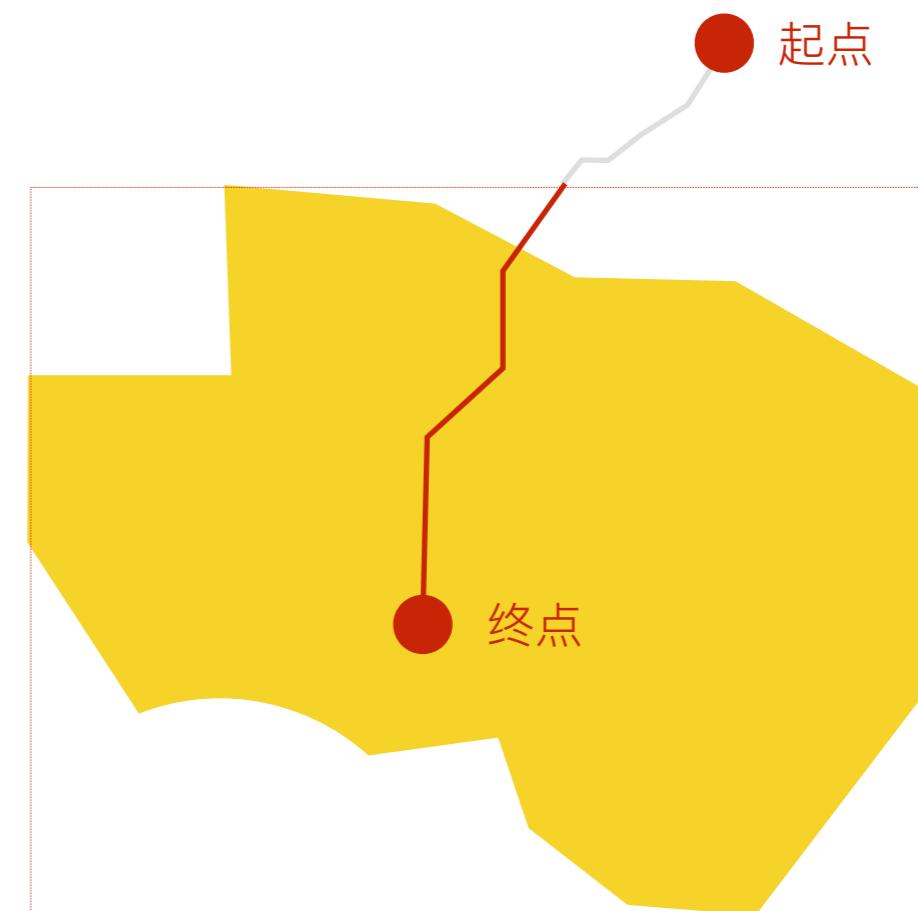
外省合并

300万

count随机

线路切割

线路简化



语言生态

crawler

Languages



spider

Languages



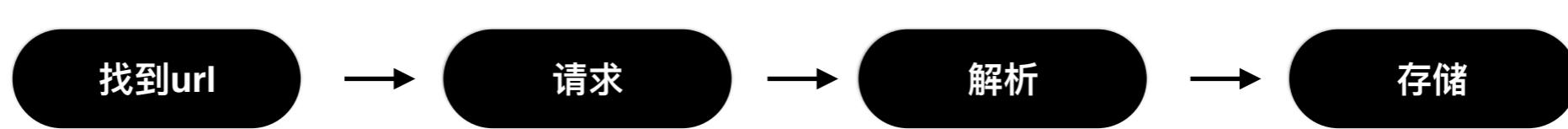
Node.js + JS + MongoDB / MySQL / Postgres

MEAN



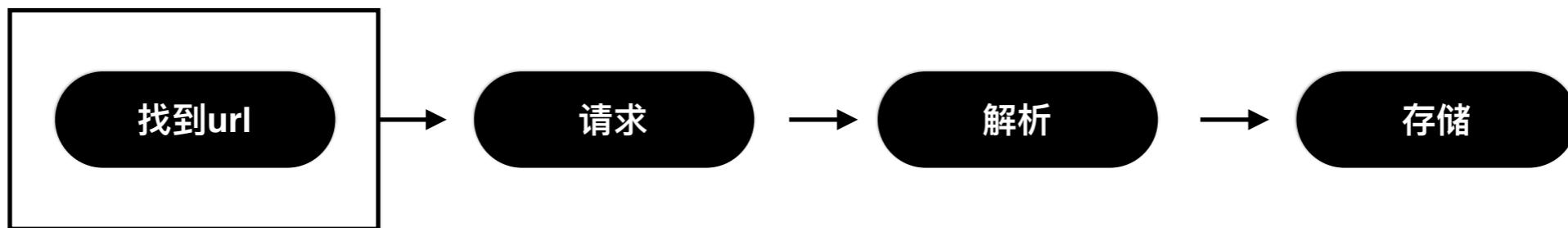
PERN

服务器

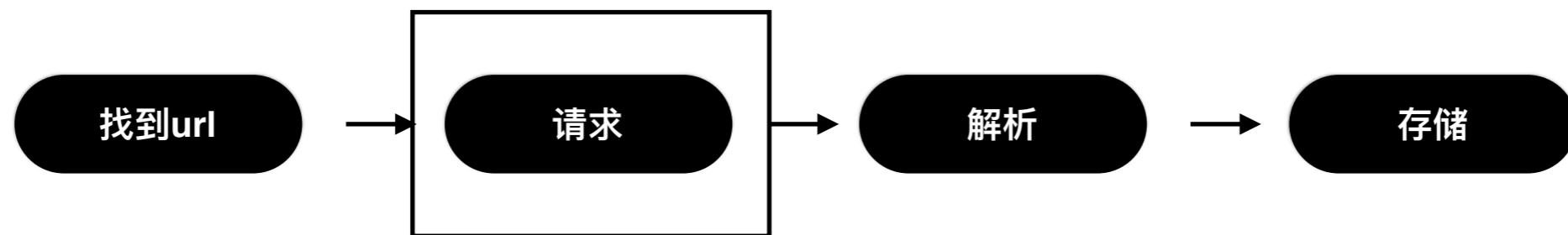


流水线的生命周期

—url对多table



自增id 翻页page
持久化的问题
dynamic



发起什么请求 **json | html | text**

请求怎么排队 **并发 | 排队**

请求怎么搞模拟

发出请求

请求的返回

URL Error 没网

HTTPError 服务器不给数据

1xx 临时返回

2xx 基本成功

3xx 转跳为主

4xx 基本挂了

5xx 服务器灭了

其他，被服务器发现是爬虫

请求的结构

User-Agent: 浏览器型号

Referer: 请求从哪来

Cookie: 模拟登陆

Content-Type: 请求体的类型(POST)

反爬虫与反反爬虫

网站：

明杀404 / 验证码 && 暗杀

有一定流速控制：

3小时和1分钟检测

要素伪造与发现

假百度爬虫 user-agent与ip穿帮

IP 和 user-agent的绑定关系

Referer

X-forward-for

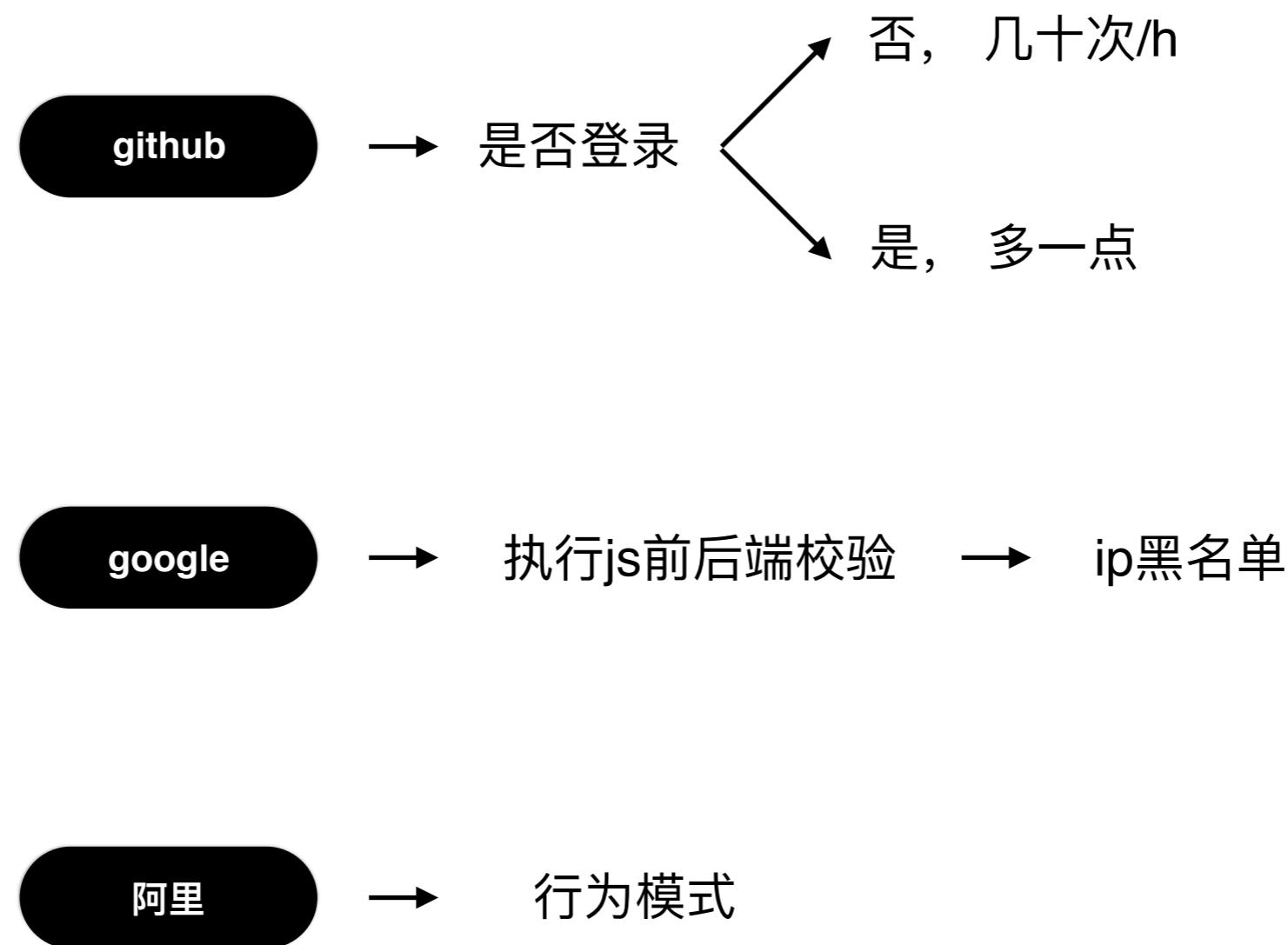
是否访问了js、html

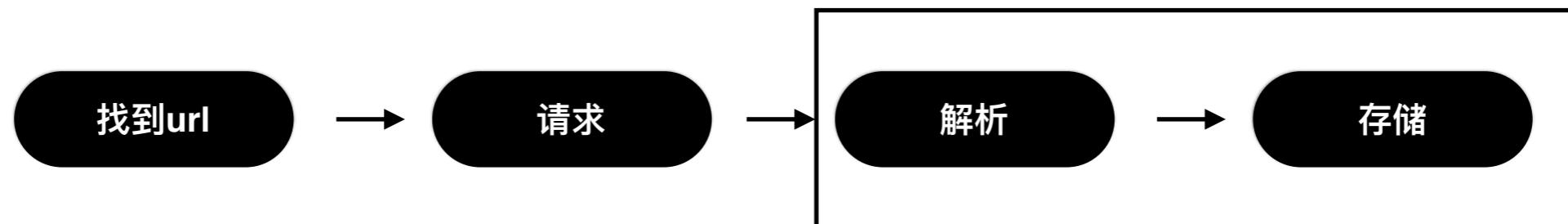
大招

绑定微信 + 限制访问次数

陷阱法

display:none法

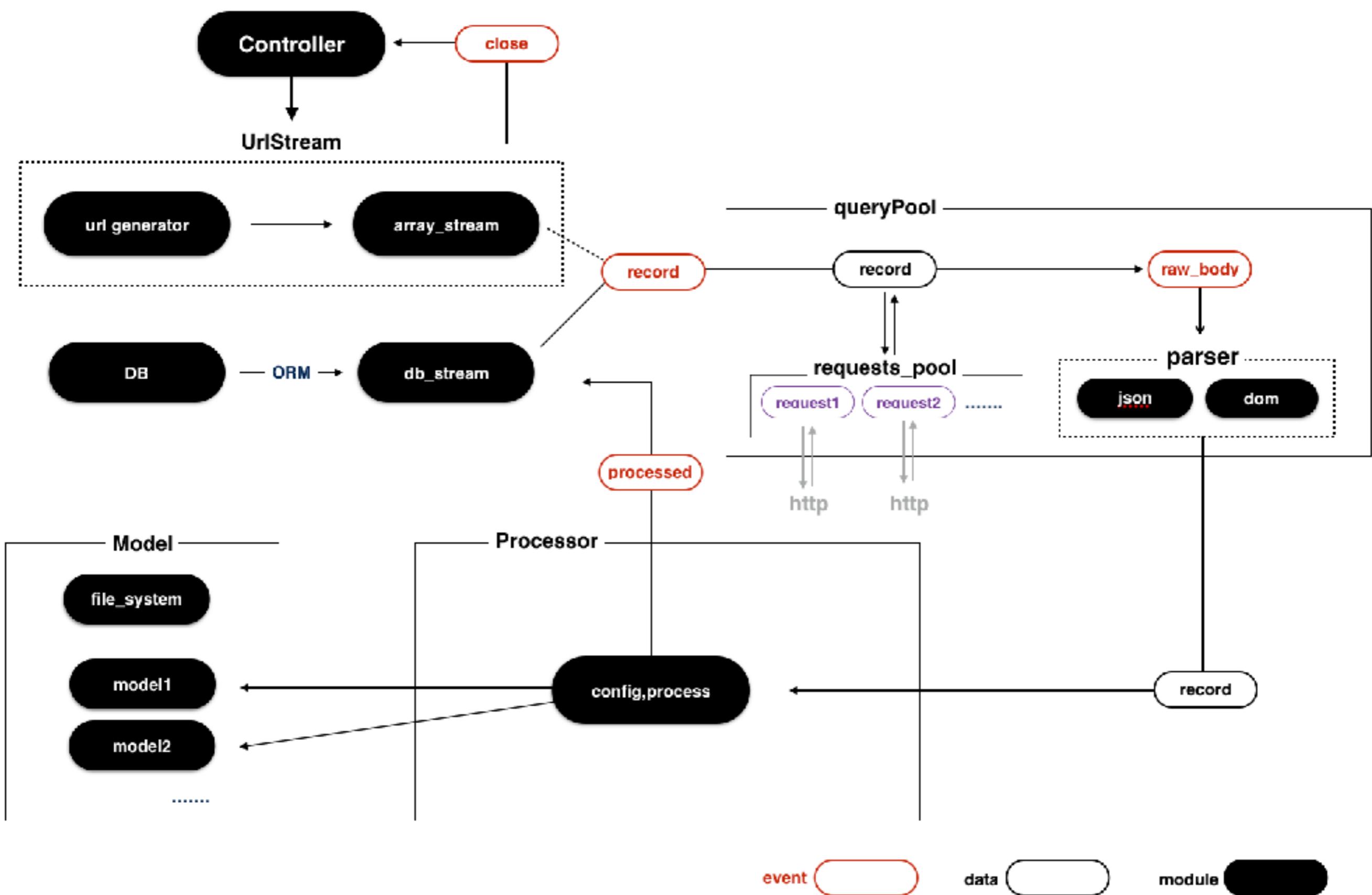


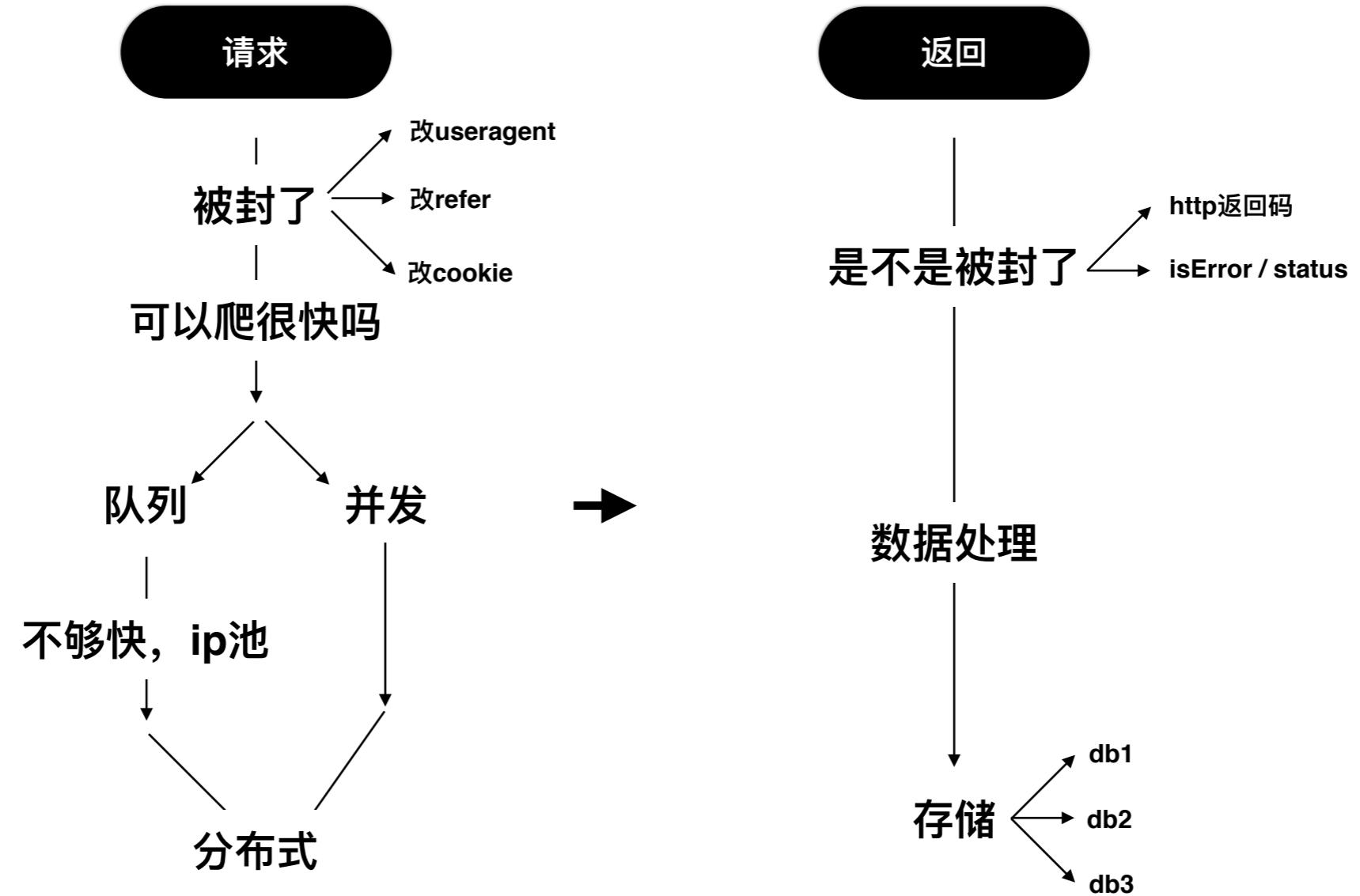


json / jsonp
html
text

postgres

Task 任务



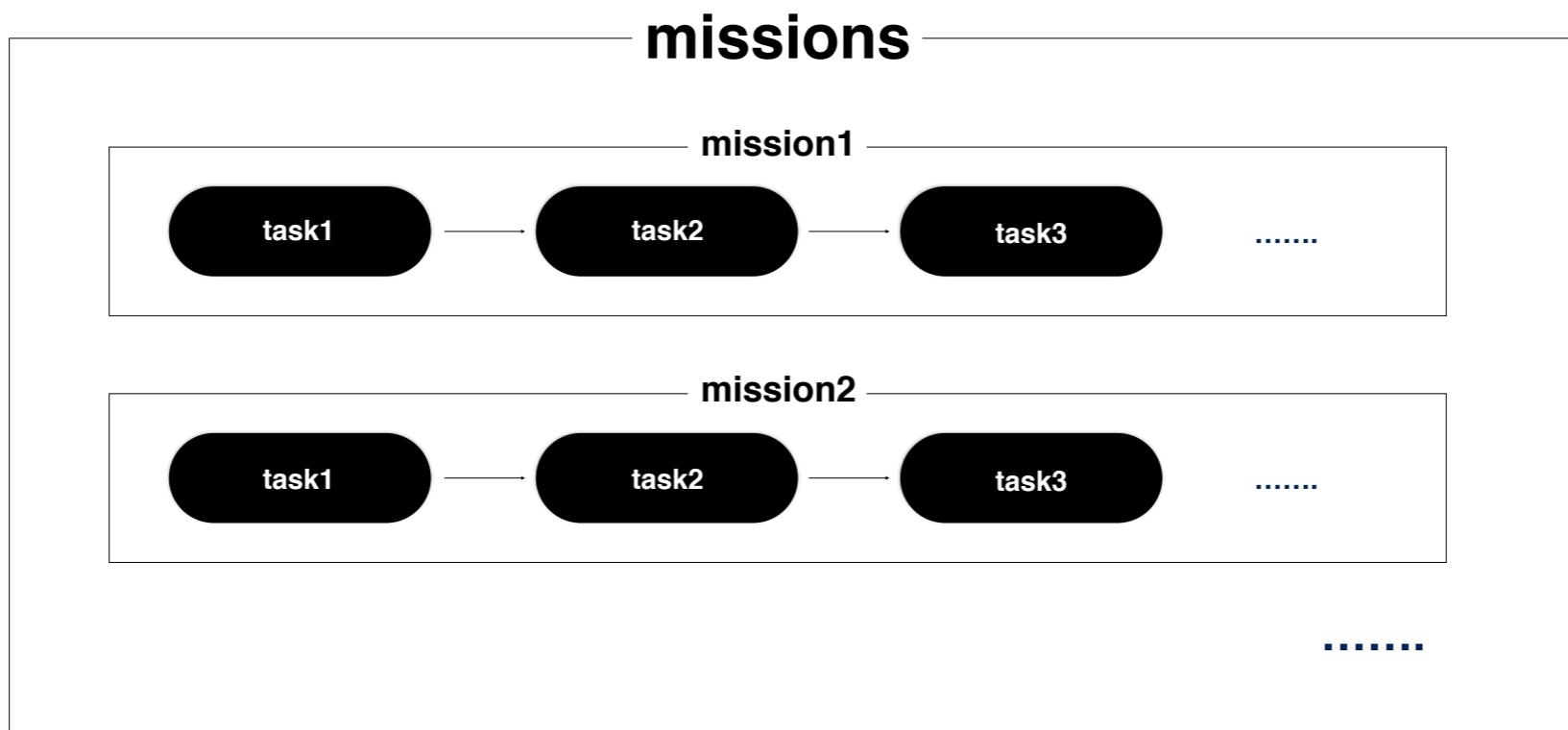


分布式的一些探索

js: 对象太弱

10台ECS -> 100台ECS

调度系统



可视化

keynote + qis / leaflet + echarts

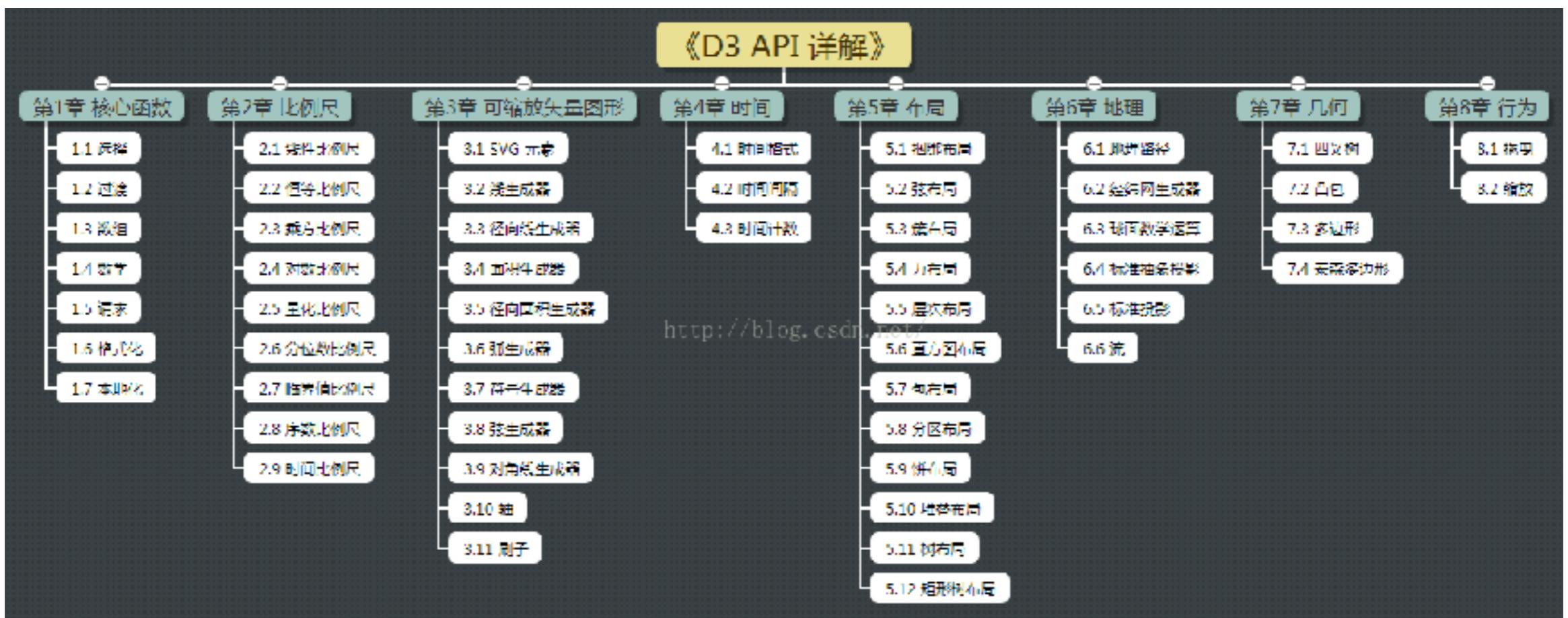
很实用的库

echarts
g2

leaflet
d3

three.js
stack.gl

D3.js



Leaflet.js

底图

底图

mapdata

simplyfy 简化边界
reduce 去掉小边界
+bBox:
+cp:中心点
+shortName:短名

svg

boundary

loadSVG(id) type
focu: 聚焦到某个位置
clear():及时清理场景\City\District
style: 形成某种风格
: boundary的交互

roads

tile型

地图服务

地图

layer: 到某tiles index/next/

卫星
普通
黑色
白色
路网

dom

floattag

svg

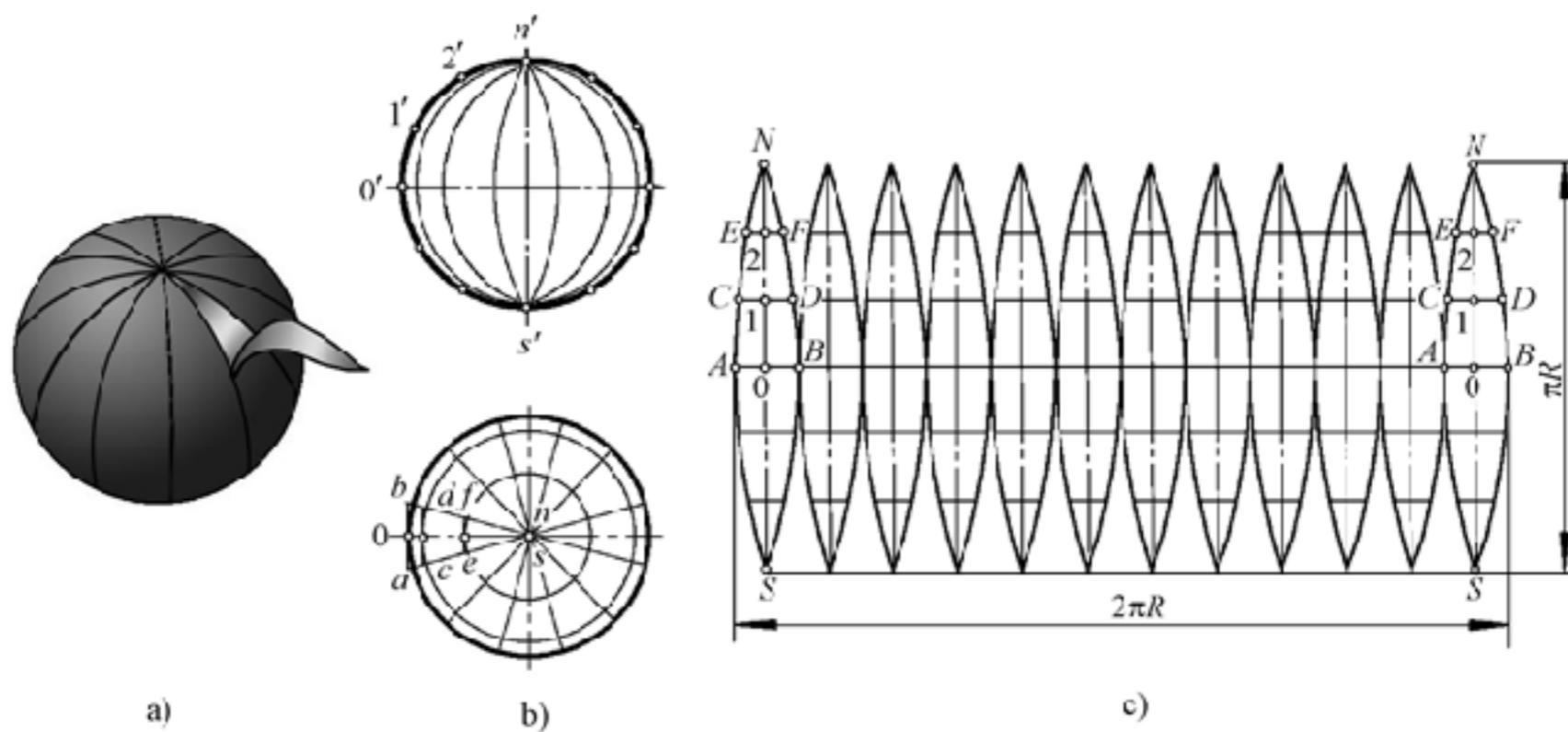
点图

路径图

canvas

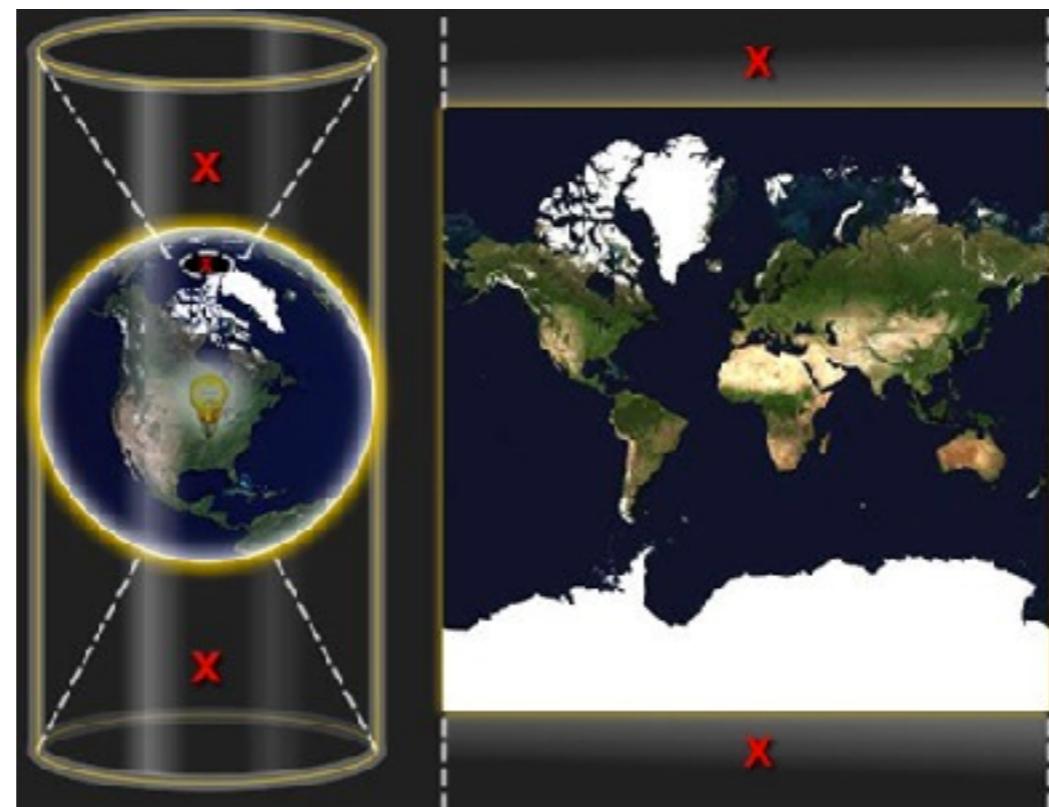
球的展开

拓扑不对等 球面几何学 头发的旋



mercator投影

修正 | web mercator | 火星 | 百度



$$\begin{aligned}x &= r^* \text{lat} \\y &= r^* \tan(\text{lng})\end{aligned}$$



$$\begin{aligned}x &= r^* \text{lat} \\y &= r^* \ln(\tan(\text{lng}/2 + \pi/2))\end{aligned}$$



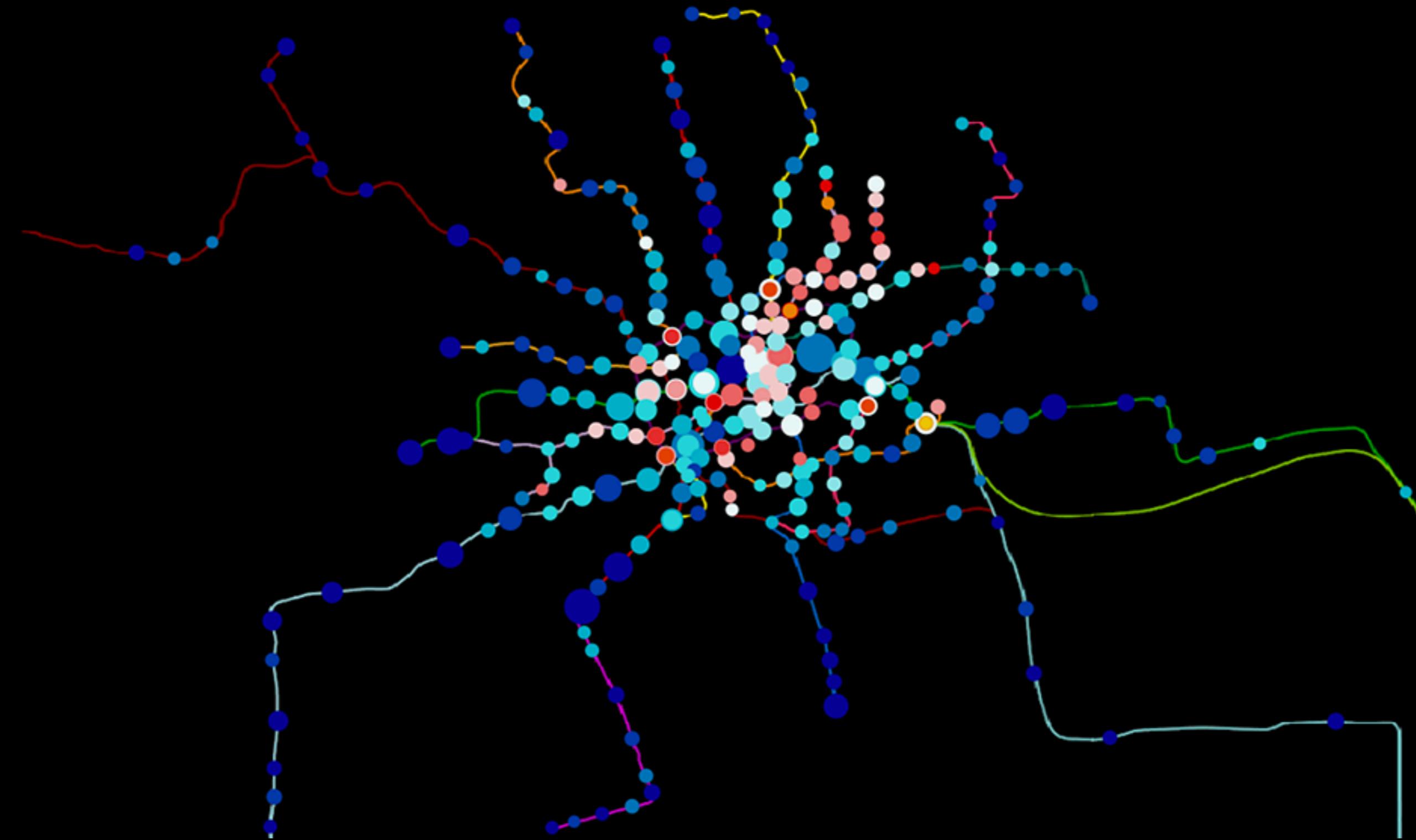


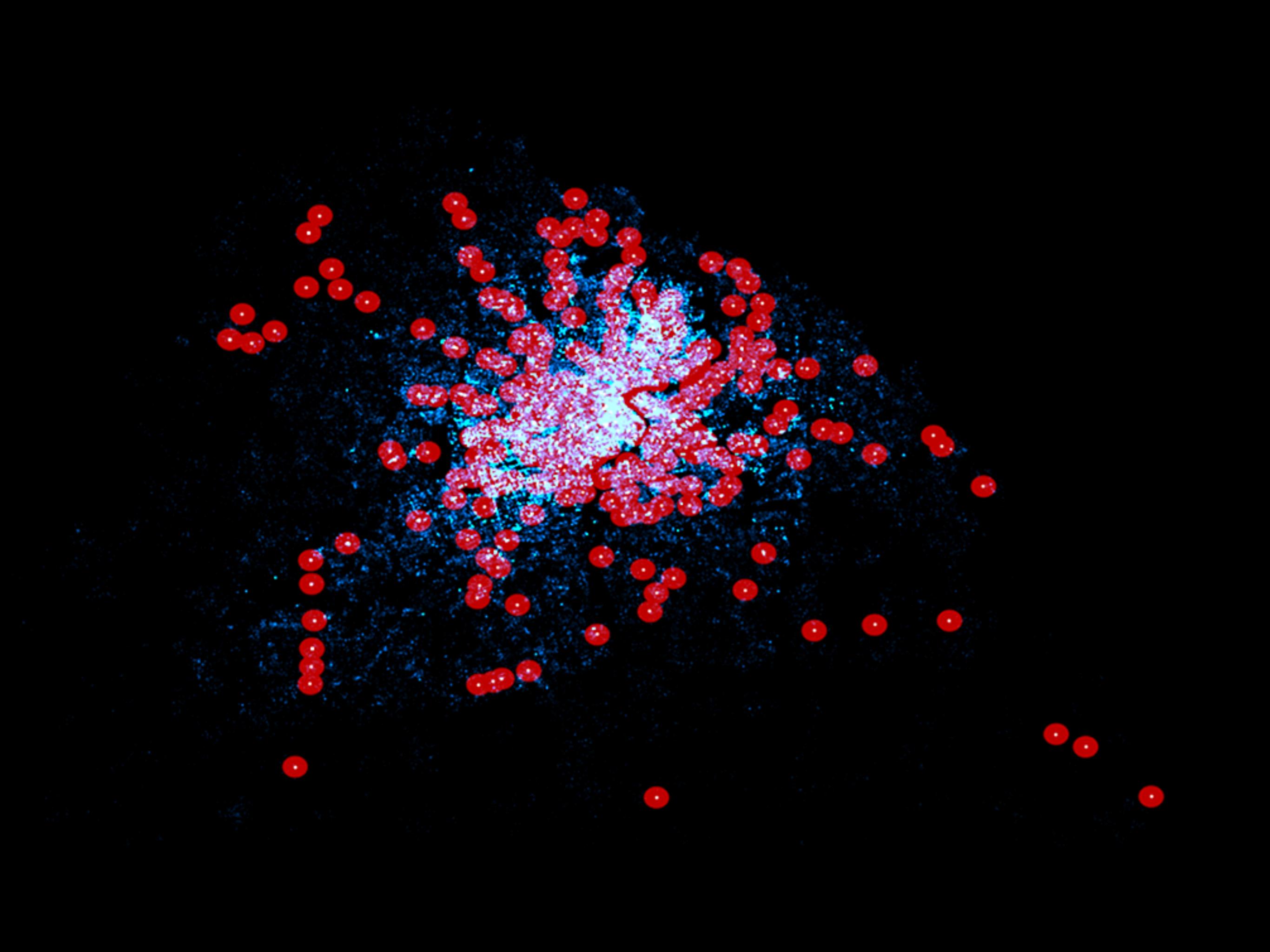
上海市黄浦区小东门街道中山东二路505号十
六铺码头

上海市闵行区吴泾镇吴泾镇市民体验中心上海氯碱化工股份有限公司集体宿舍

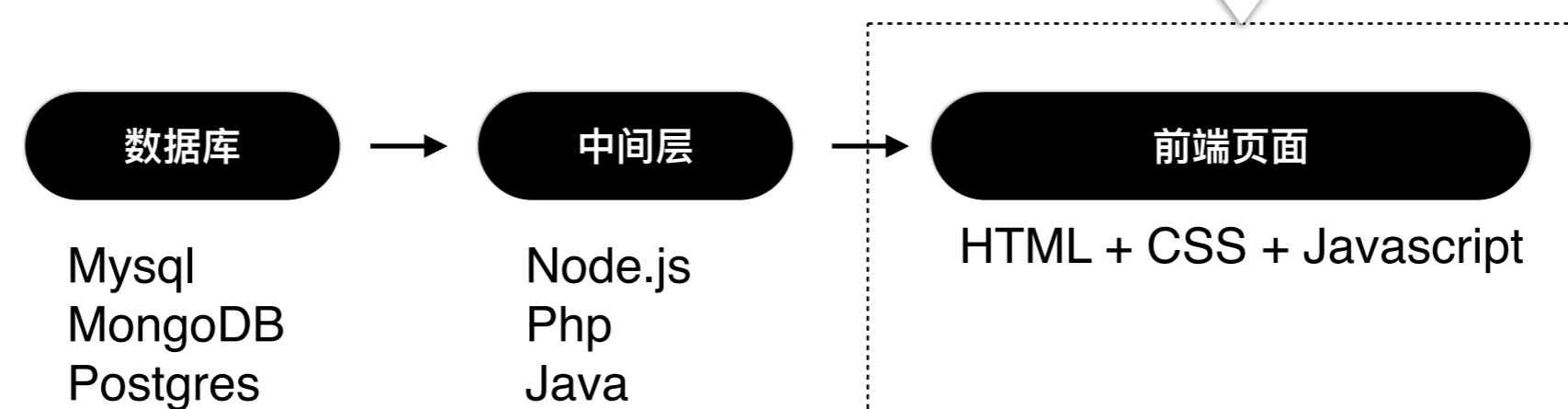


用好软件





WebGIS | WebGL | Canvas



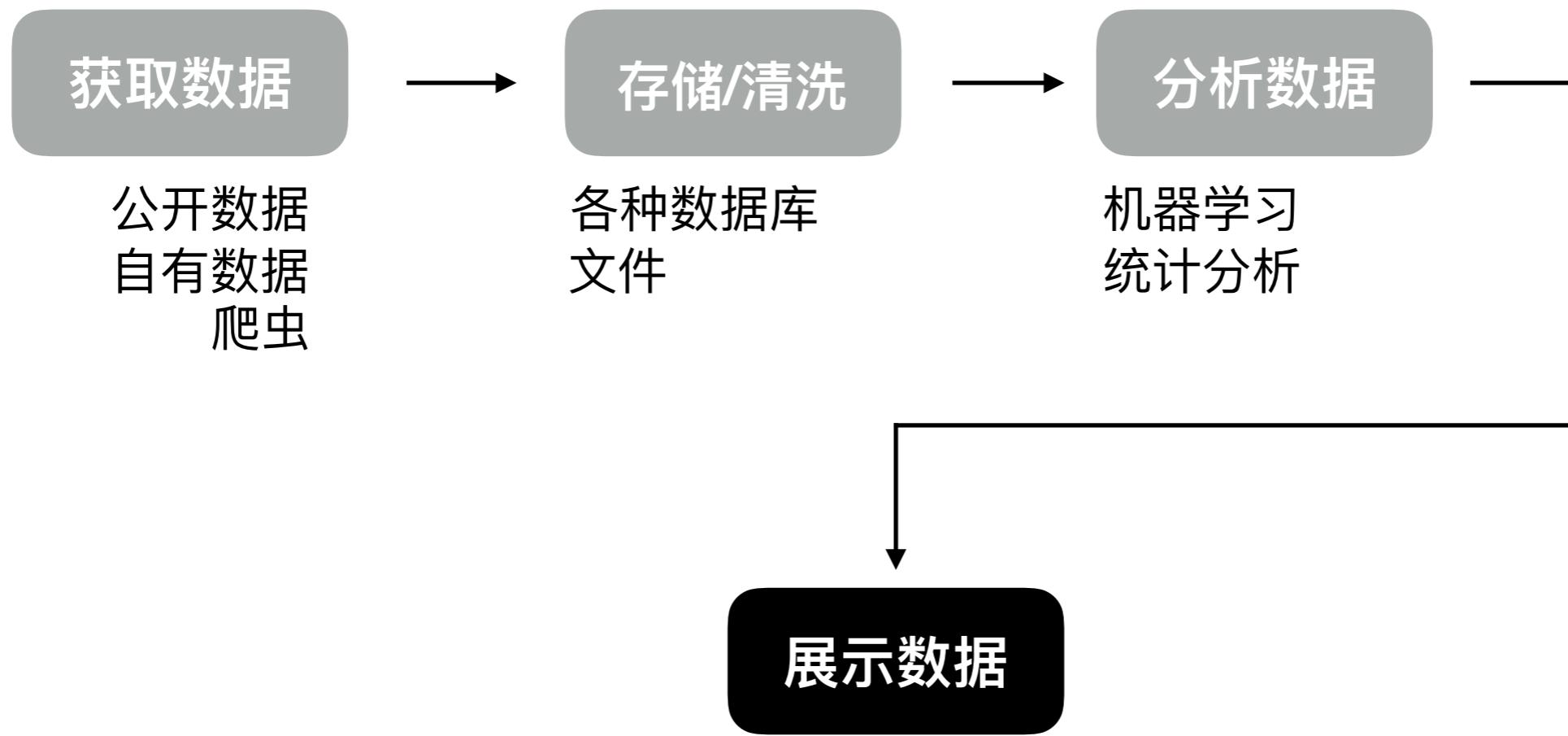


chart **graph** **map** 数据新闻
数据产品
数据大屏
数据分析



<https://zhuanlan.zhihu.com/za-zzz>