

彩云天气

创新点、扩展性、技术架构

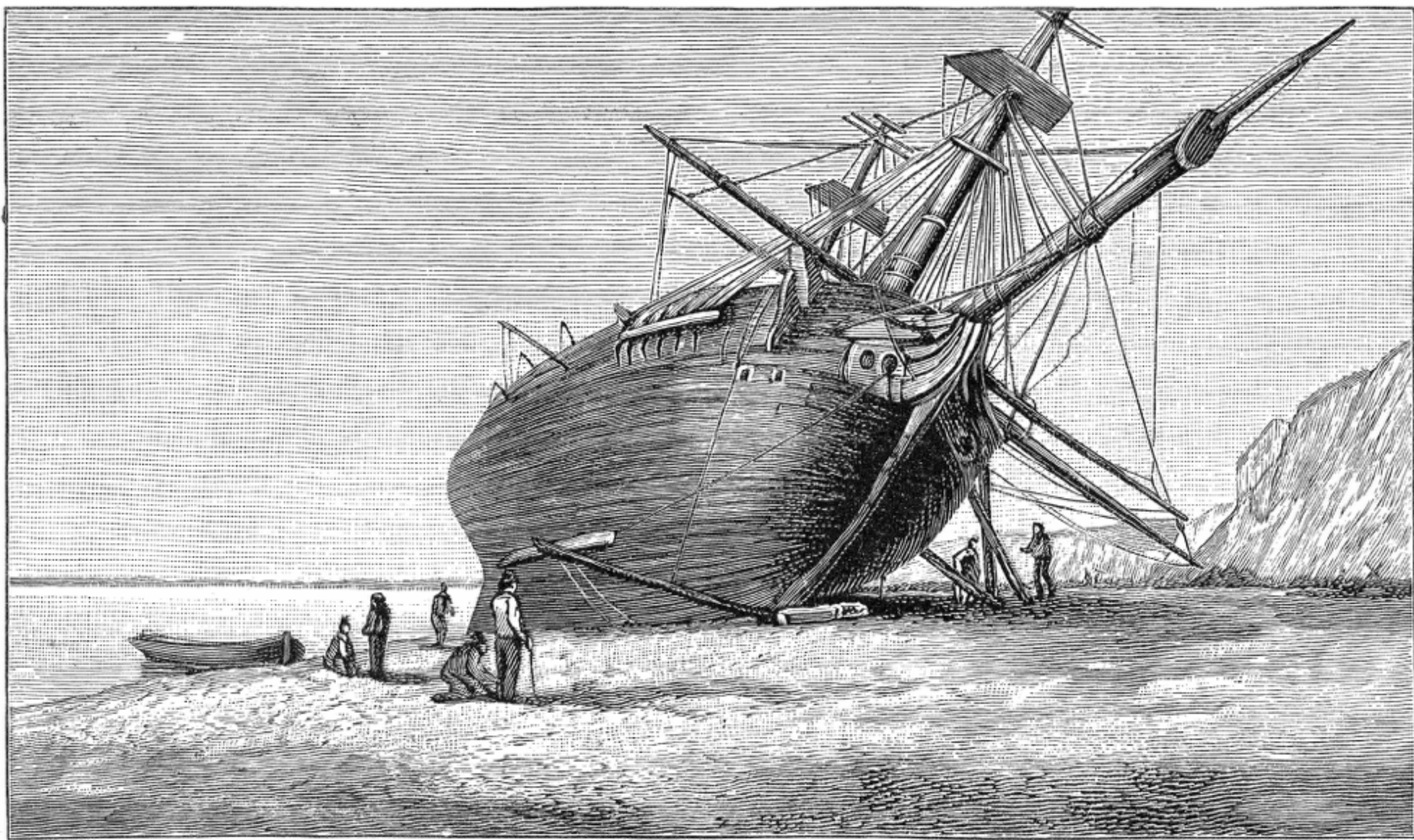
苑明理
2016年7月

目录

- 创新点：媒体变革与天气预报（商业创新）
- 扩展性：计算思维视角下的规模扩张（技术路线）
- 技术与架构：程序员的小技雕虫（技术架构）

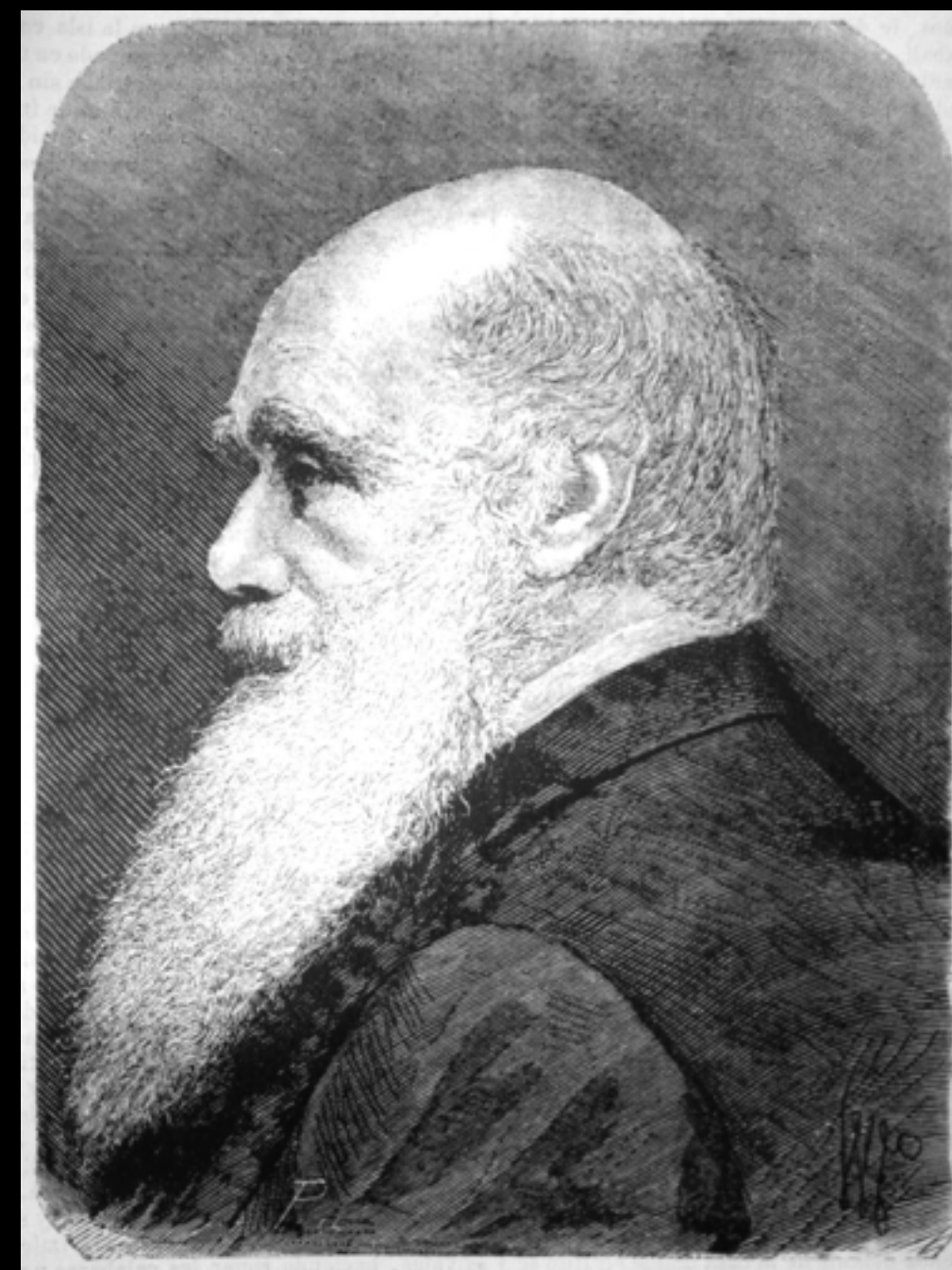
创新点

媒体变革与天气预报



THE BEAGLE LAID ASHORE, RIVER SANTA CRUZ

达尔文的 小猎犬号之旅





GREENWICH HOSPITAL COLLECTION

罗伯特·菲茨罗伊

天气观测网与预报

1861年8月1日
《泰晤士报》

THE WEATHER.
METEOROLOGICAL REPORTS.

Wednesday, July 31, 8 to 9 a.m.	B.	E.	M.	D.	F.	C.	I.	S.
Nairn.. ..	29.54	57	56	W.S.W.	6	9	o.	3
Aberdeen ..	29.60	59	54	S.S.W.	5	1	b.	3
Leith	29.70	61	55	W.	3	5	c.	2
Berwick .. .	29.69	59	55	W.S.W.	4	4	c.	2
Ardrossan .	29.73	57	55	W.	5	4	c.	5
Portrush ..	29.72	57	54	S.W.	2	2	b.	2
Shields .. .	29.80	59	54	W.S.W.	4	5	o.	3
Galway .. .	29.83	65	62	W.	5	4	c.	4
Scarborough	29.86	59	56	W.	3	6	c.	2
Liverpool ..	29.91	61	56	S.W.	2	8	c.	2
Valentia .. .	29.97	62	60	S.W.	2	5	o.	3
Queenstown ..	29.88	61	59	W.	3	5	c.	2
Yarmouth.. .	30.05	61	59	W.	5	2	c.	3
London .. .	30.02	62	56	S.W.	3	2	b.	—
Dover.. ..	30.04	70	61	S.W.	3	7	o.	2
Portsmouth ..	30.01	61	59	W.	3	6	o.	2
Portland .. .	30.03	63	59	S.W.	3	2	c.	3
Plymouth.. .	30.00	62	59	W.	5	1	b.	4
Penzance .. .	30.04	61	60	S.W.	2	6	c.	3
Copenhagen ..	29.94	64	—	W.S.W.	2	6	c.	3
Helder .. .	29.99	63	—	W.S.W.	6	5	c.	3
Brest .. .	30.09	60	—	S.W.	2	6	c.	5
Bayonne .. .	30.13	68	—	—	—	9	m.	5
Lisbon .. .	30.18	70	—	N.N.W.	4	3	b.	2

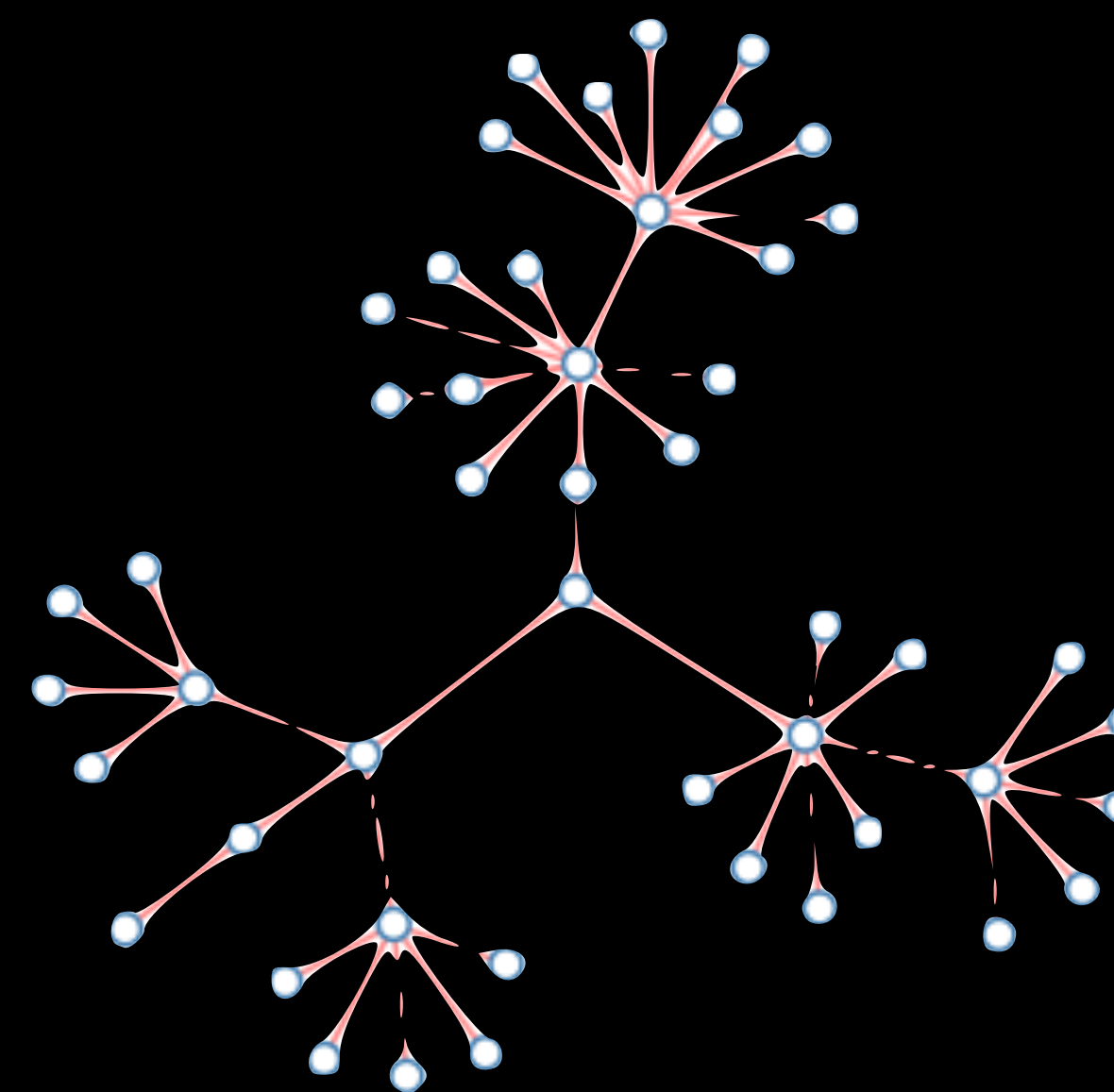
General weather probable during next two days in the—
 North—Moderate westerly wind; fine.
 West—Moderate south-westerly; fine.
 South—Fresh westerly; fine.

Explanation.

B. Barometer, corrected and reduced to 32° at mean sea level; each 10 feet of vertical rise causing about one-hundredth of an inch diminution, and each 10° above 32° causing nearly three-hundredths increase. E. Exposed thermometer in shade. M. Moistened bulb (for evaporation and dew-point). D. Direction of wind (true—two points left of magnetic). F. Force (1 to 12—estimated). C. Cloud (1 to 9). I. Initials:—b., blue sky; c., clouds (detached); f., fog; h., hail; l., lightning; m., misty (hazy); o., overcast (dull); r., rain; s., snow; t., thunder. S. Sea disturbance (1 to 9).

乔治·考林

1951年1月11日



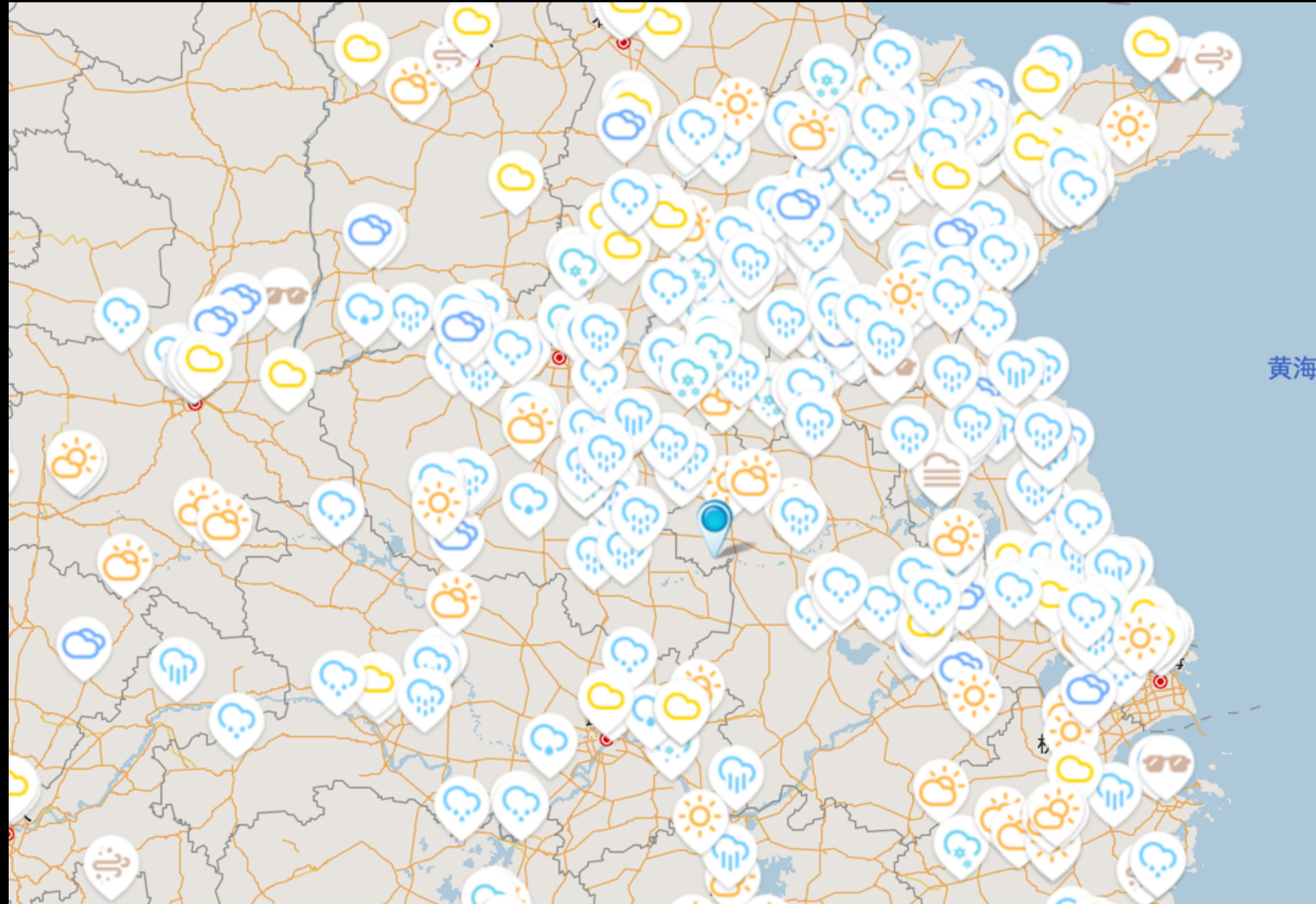
手机让我们可以捕捉到

每个人当下所在的时间和空间



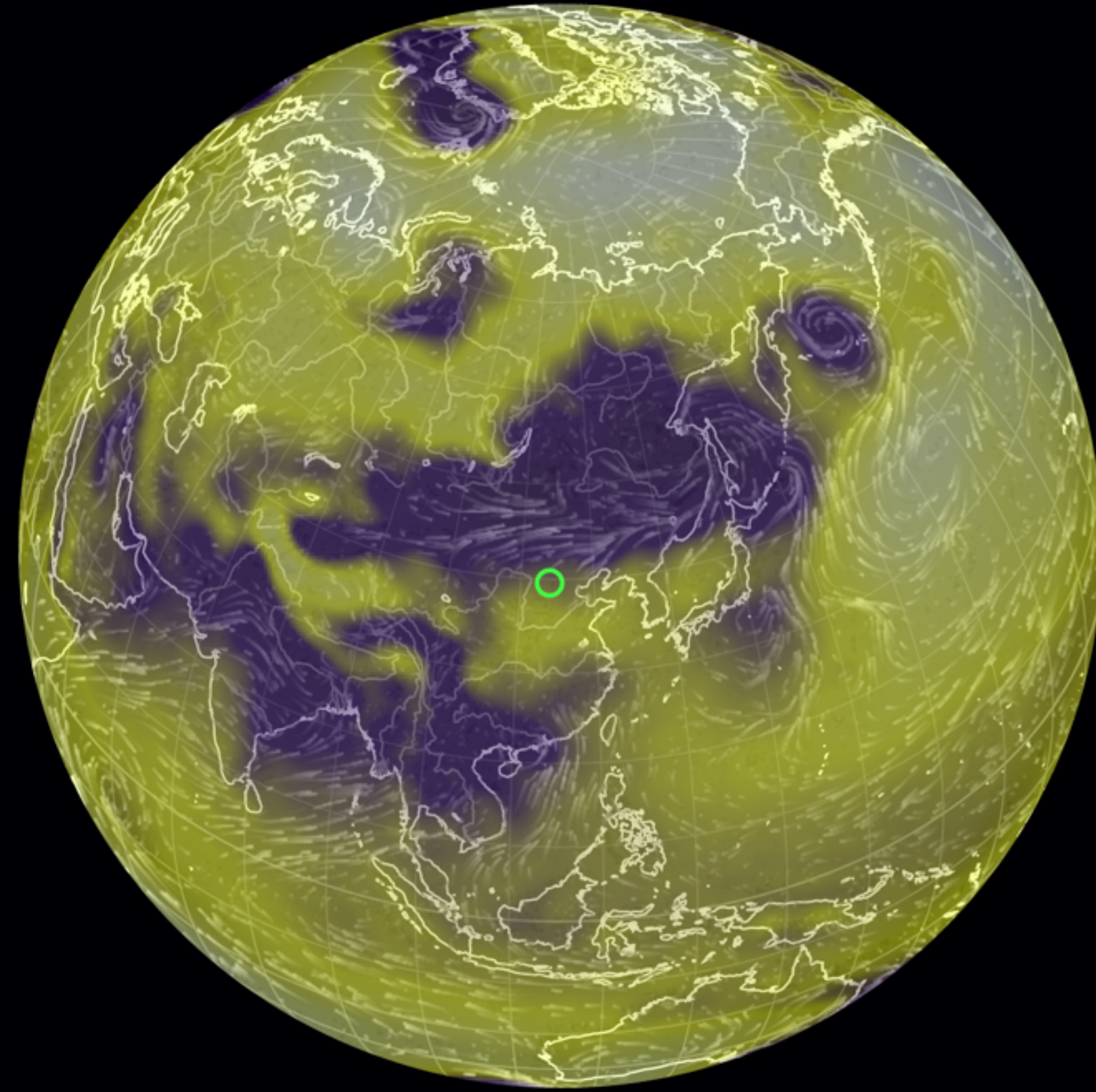
为公众提供每个人独有的个性化天气预报

手机让用户可以回传他本地的天气状况



用众包的方式来校准数据和预报

每一部手机都是小小气象站



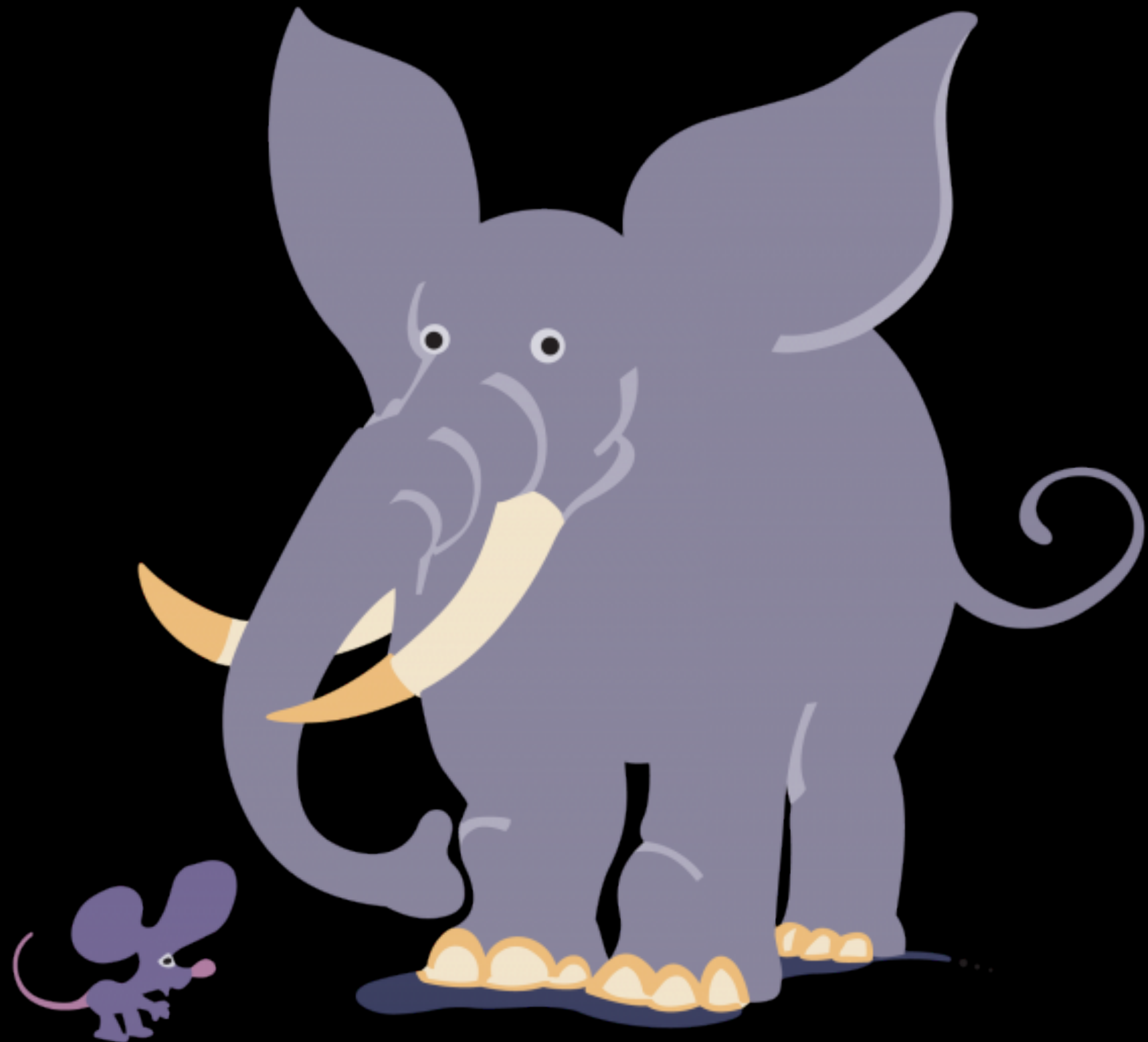
可以收集到大量数据
其数据量甚至可以超越传统的手段

扩展性

计算思维下的规模扩张

彩云为何可以用小的资源 去撬动中国气象市场的大局

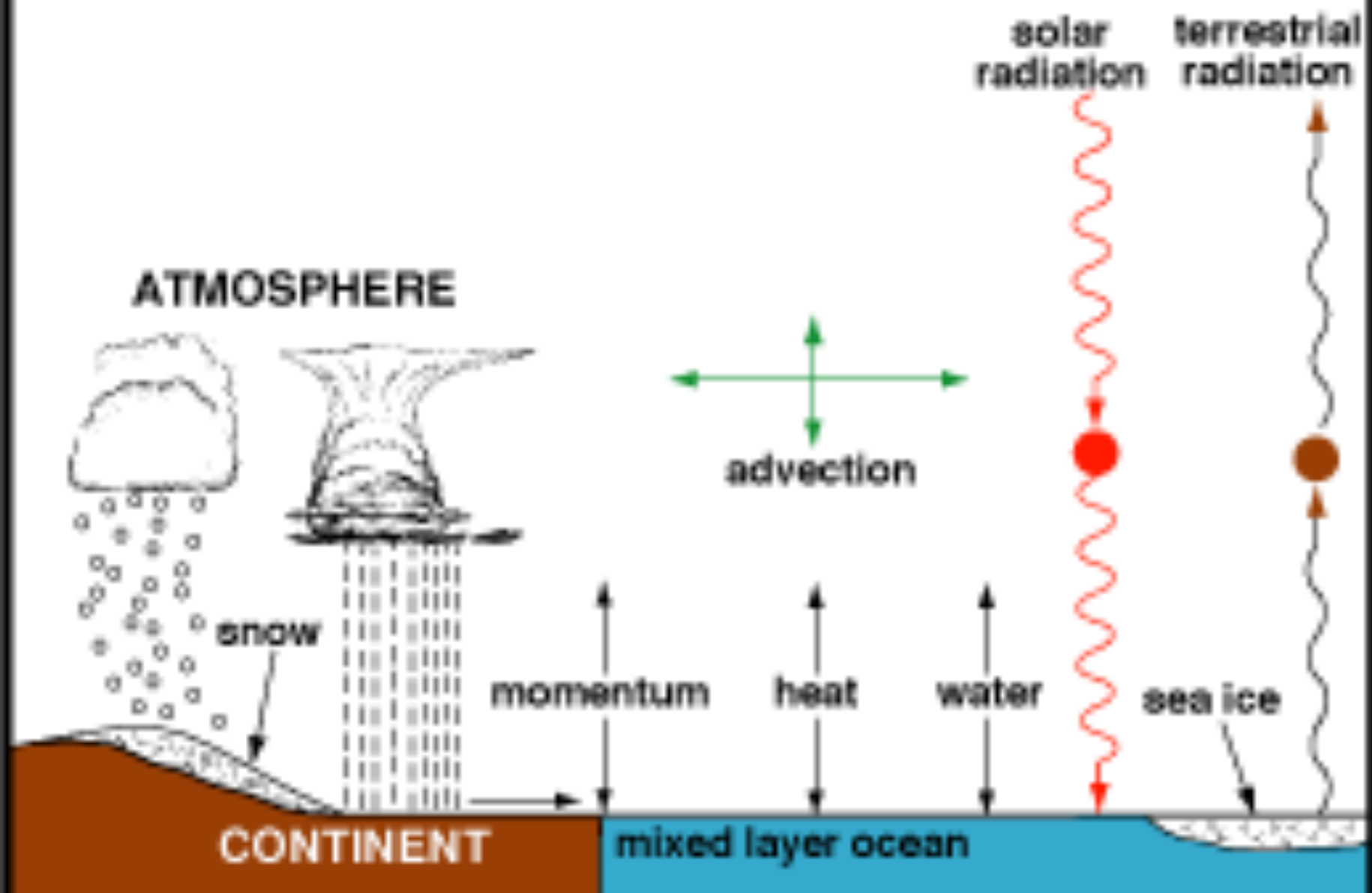
技术的角度如何解释？



Horizontal Grid
(Latitude-Longitude)

Vertical Grid
(Height or Pressure)

Physical Processes in a Model



问题的规模

要求每分钟计算的格点数

国土东西距离 × 国土南北距离 × 高度上的层数

$5200 \times 5500 \times 32$

915200000

↑
亿





CRAY

CRAY

CRAY

CRAY

CRAY

CRAY

CRAY

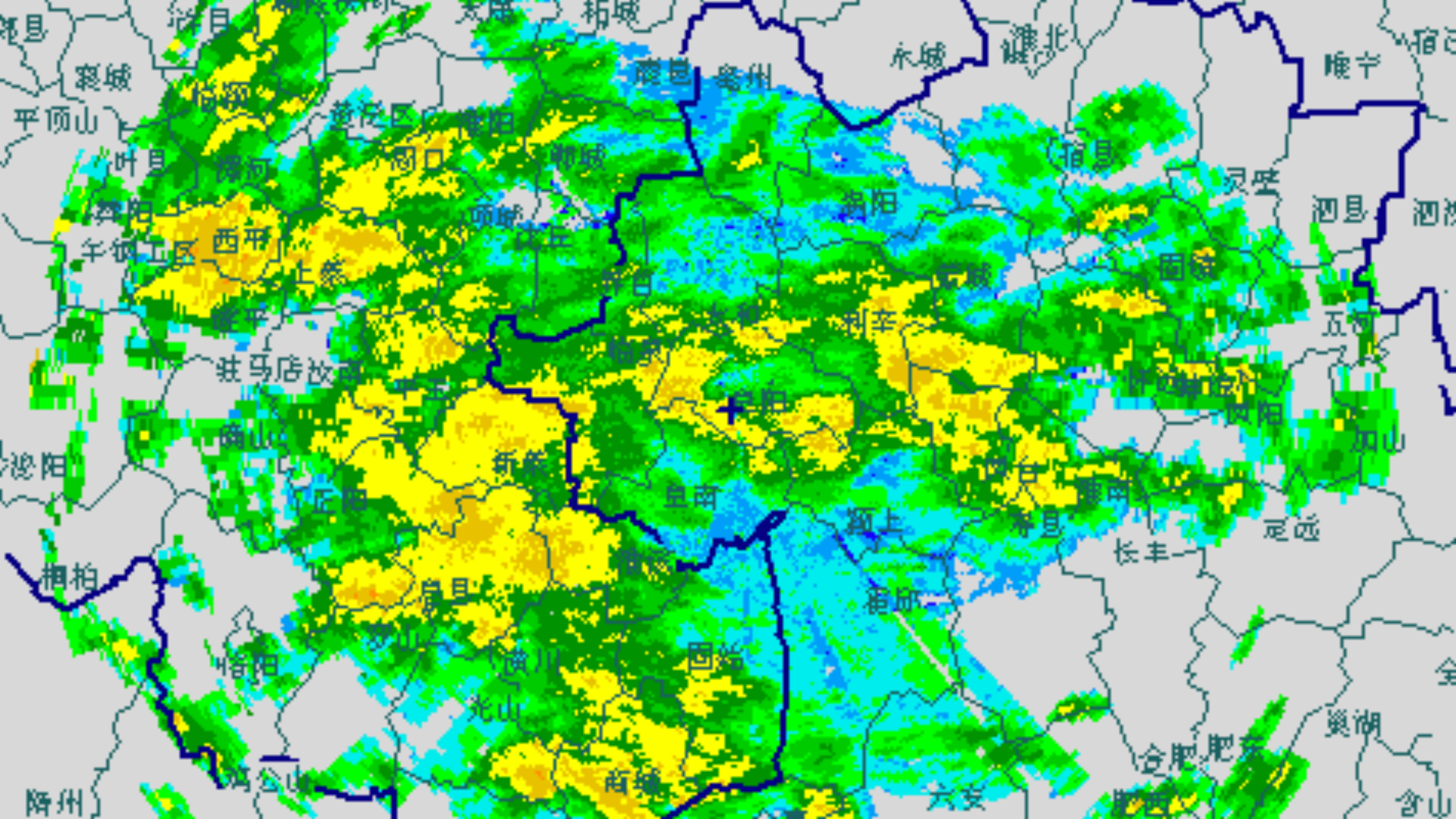
CRAY

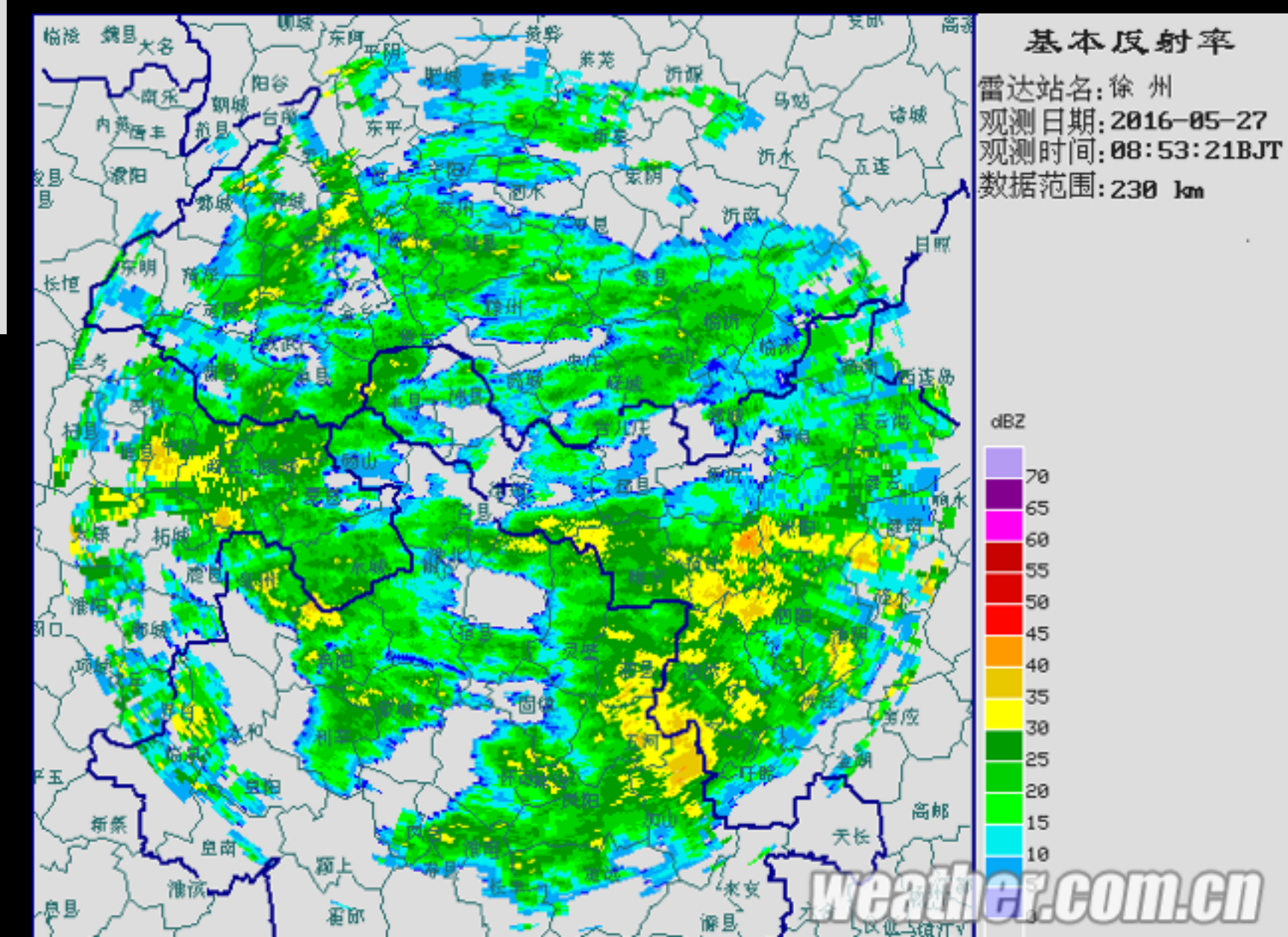
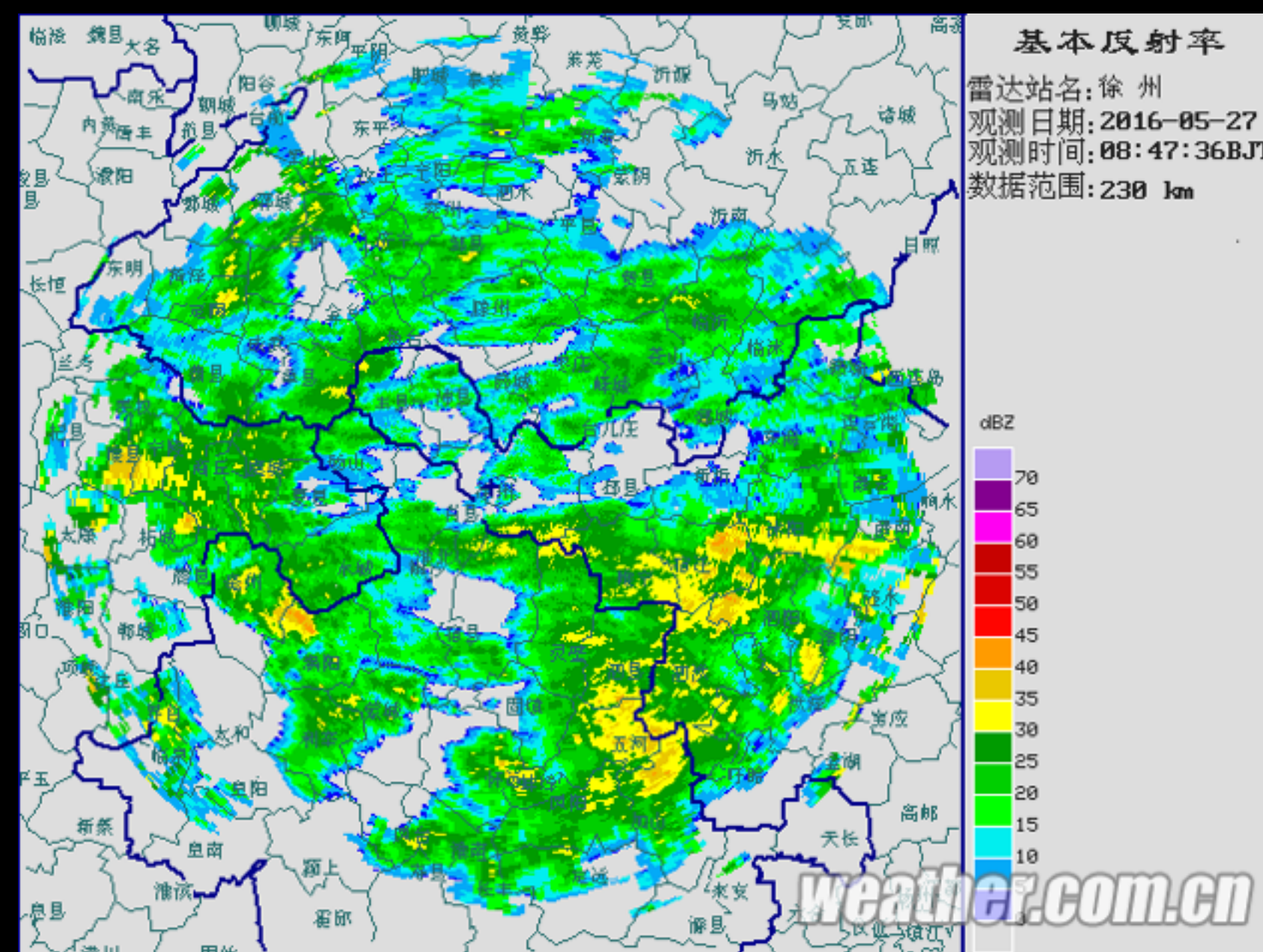
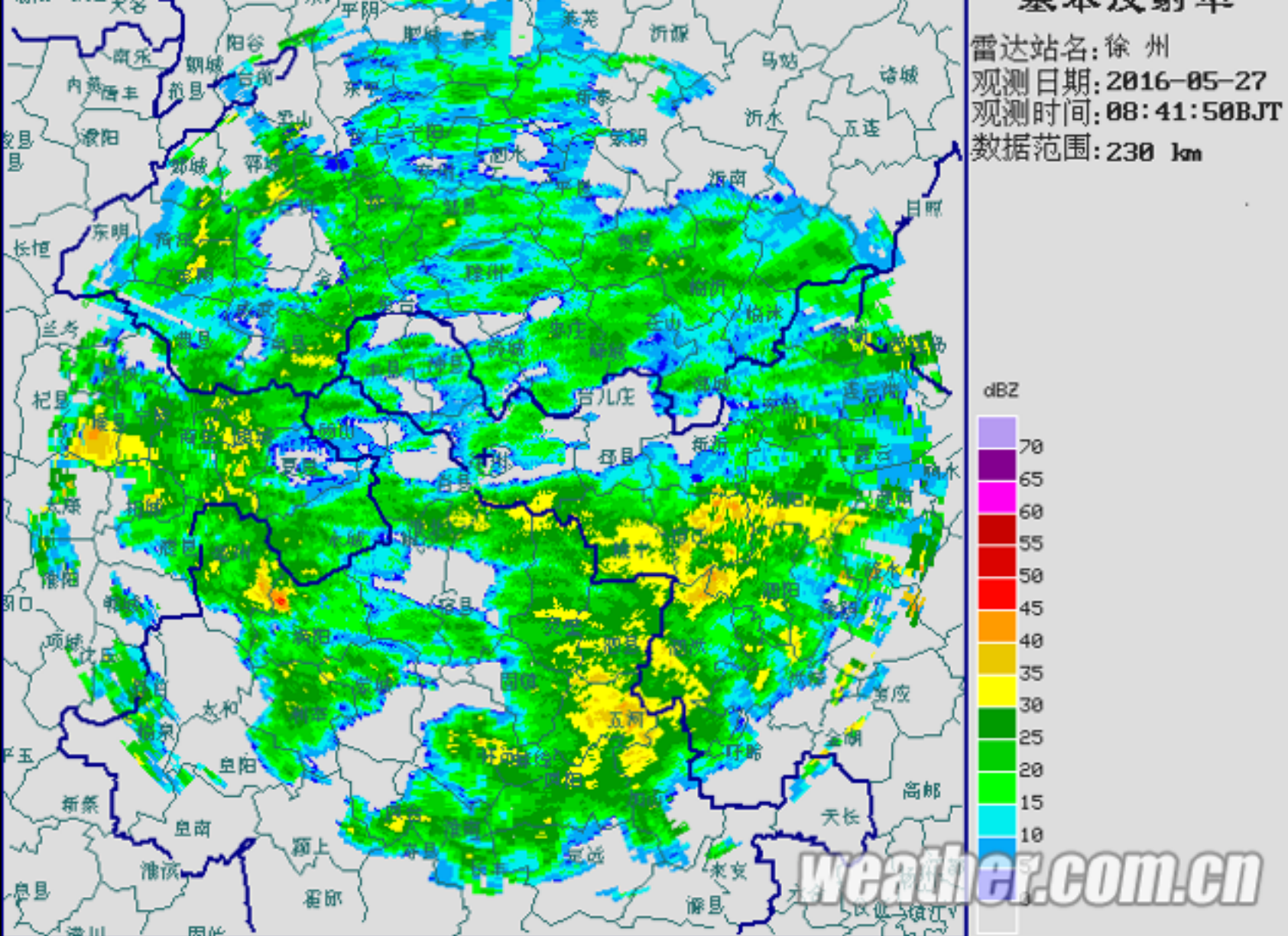
Amazon

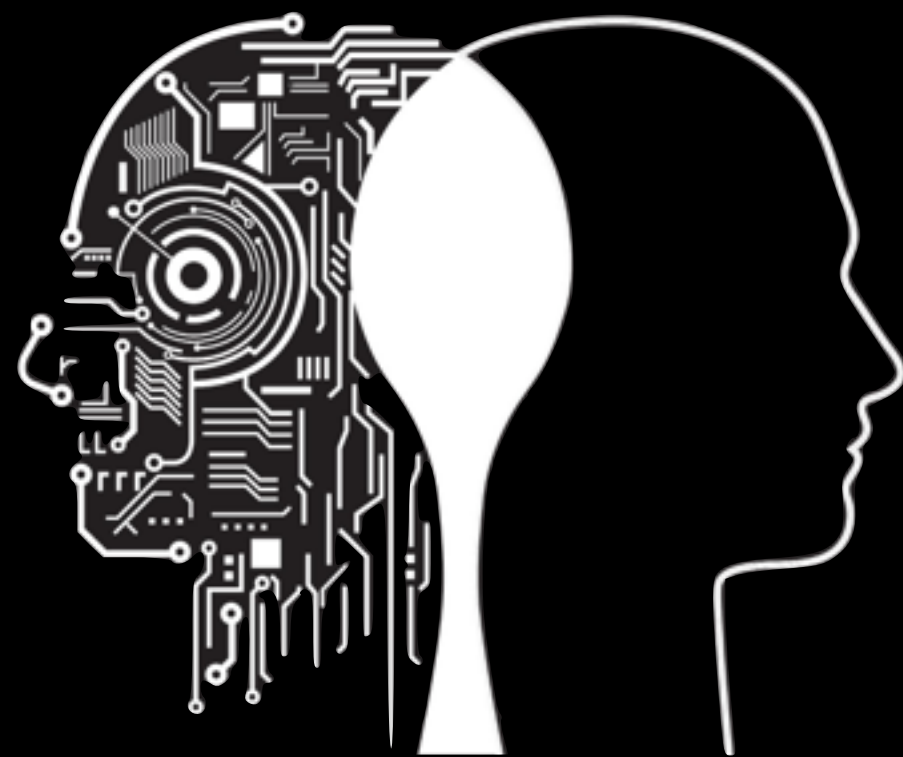
$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi$$
$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r})$$

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V(\mathbf{r}) \psi$$









动力学去追踪球面上的雨云移动
机器学习去捕捉雨的生成和消散

降三维问题为二维问题

915200000

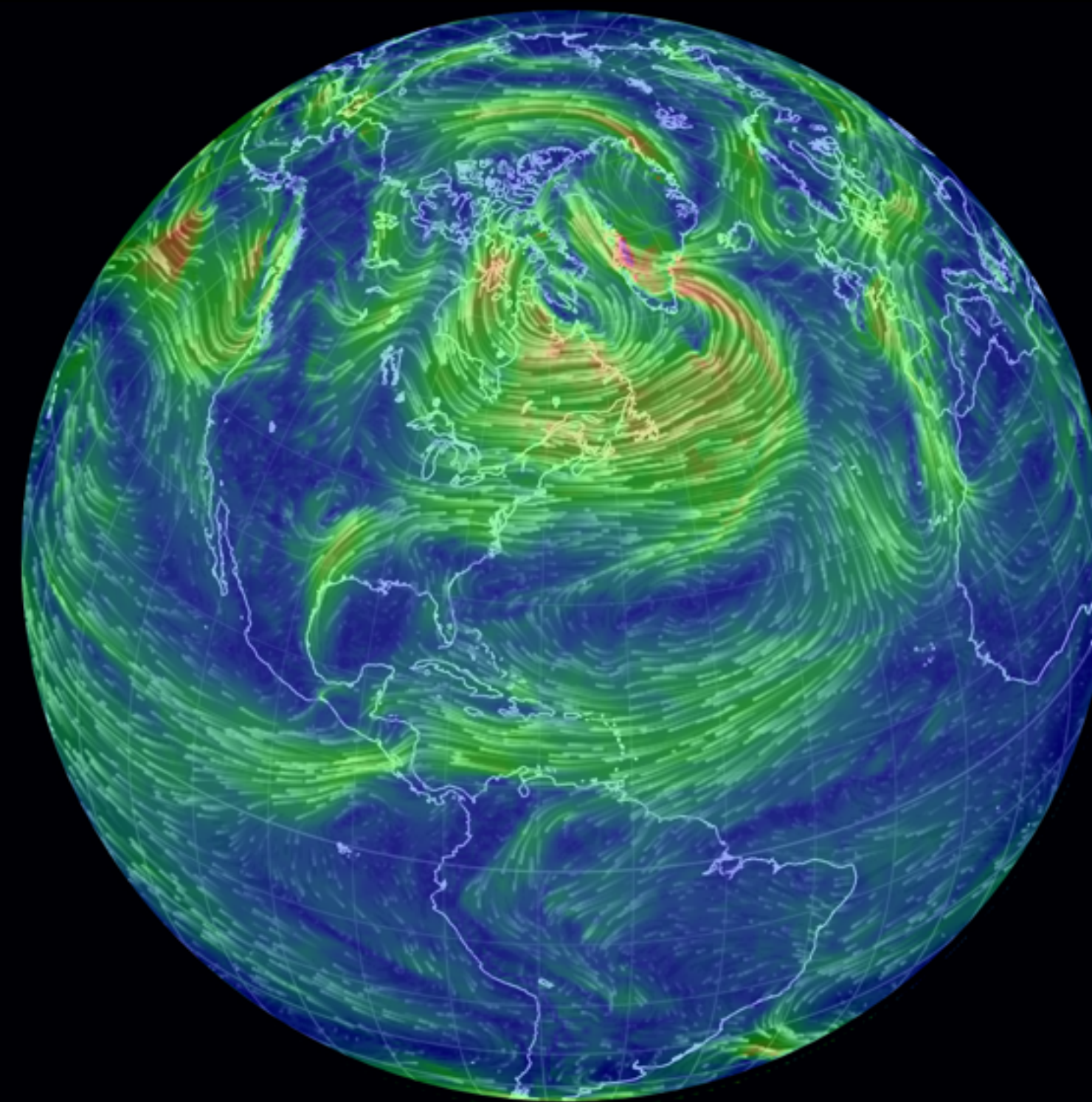


286000000

↑
千万



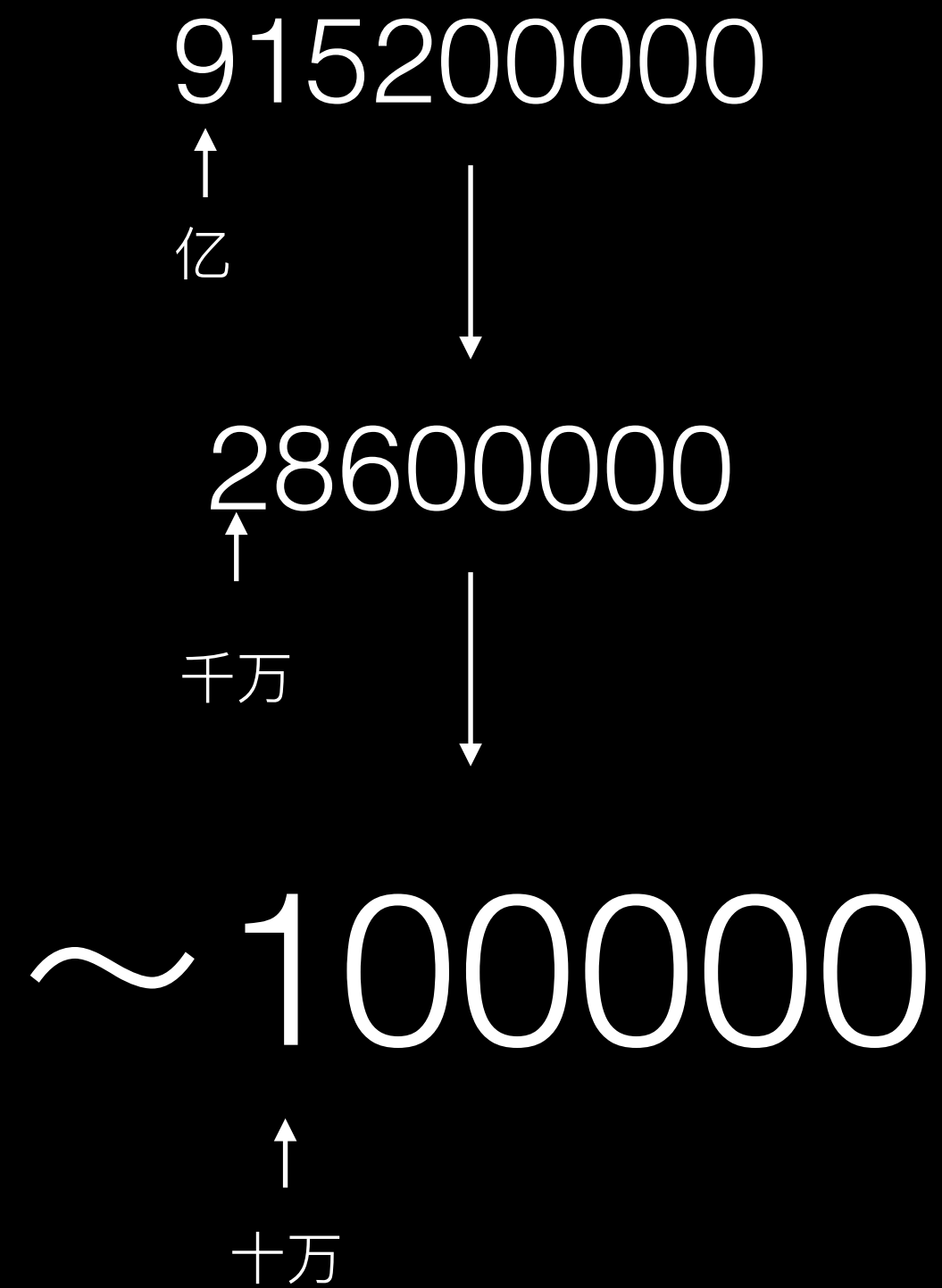
?



函数求值的方式

 积极求值：适合批处理计算

 惰性求值：适合响应式计算



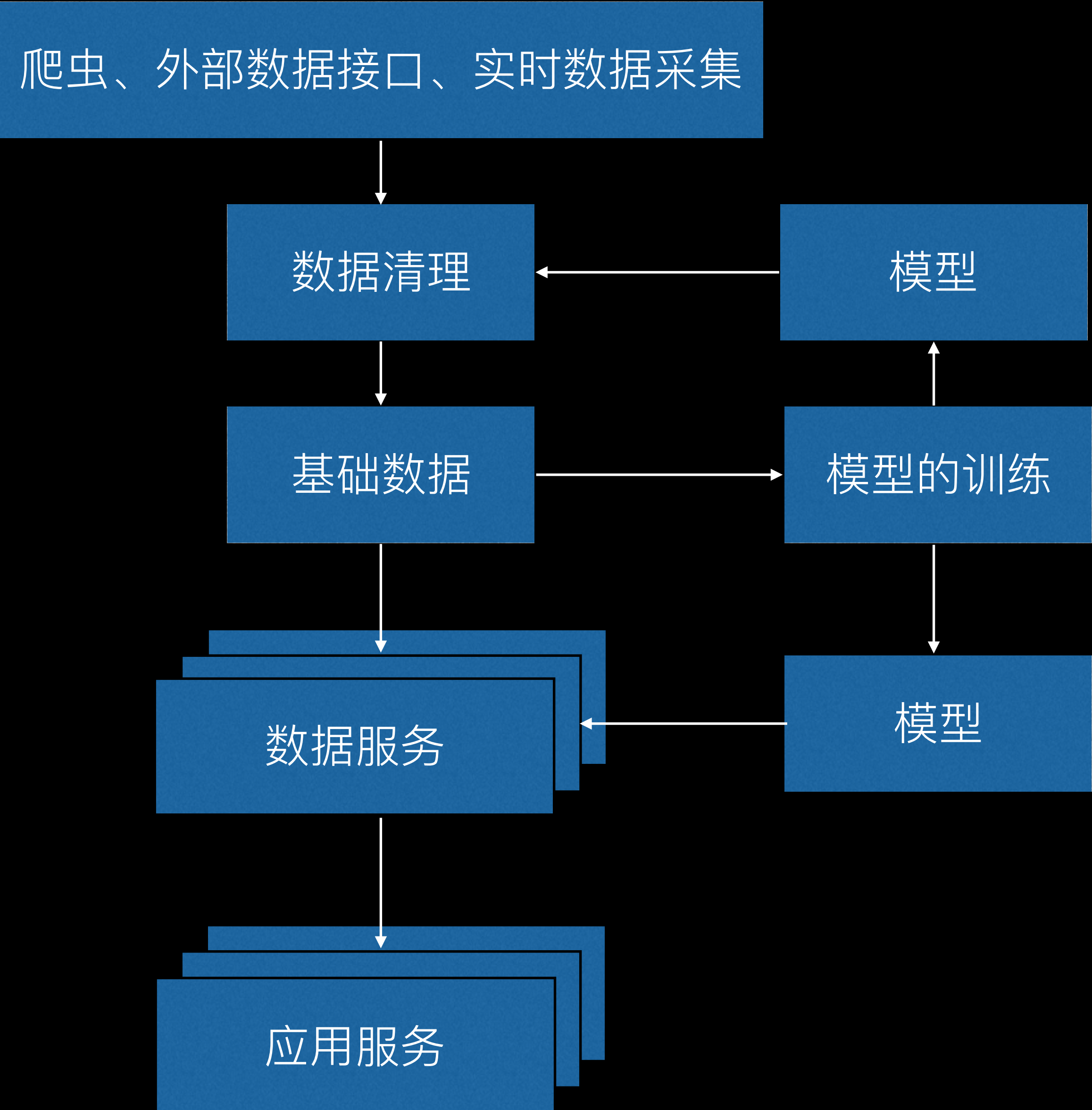
技术与架构

程序员的小技雕虫

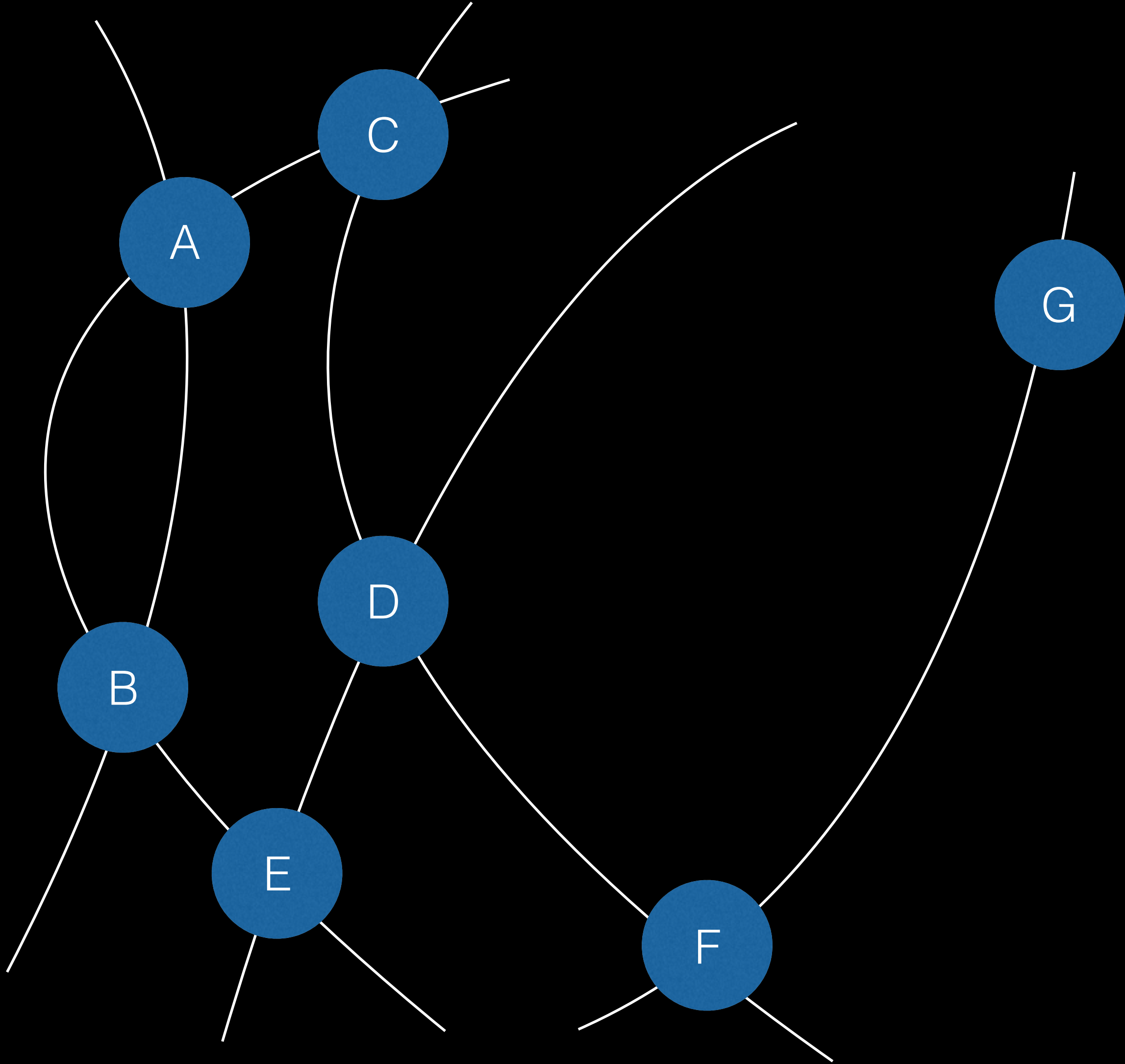
数据处理流程

流式数据系统：5 条数据流线

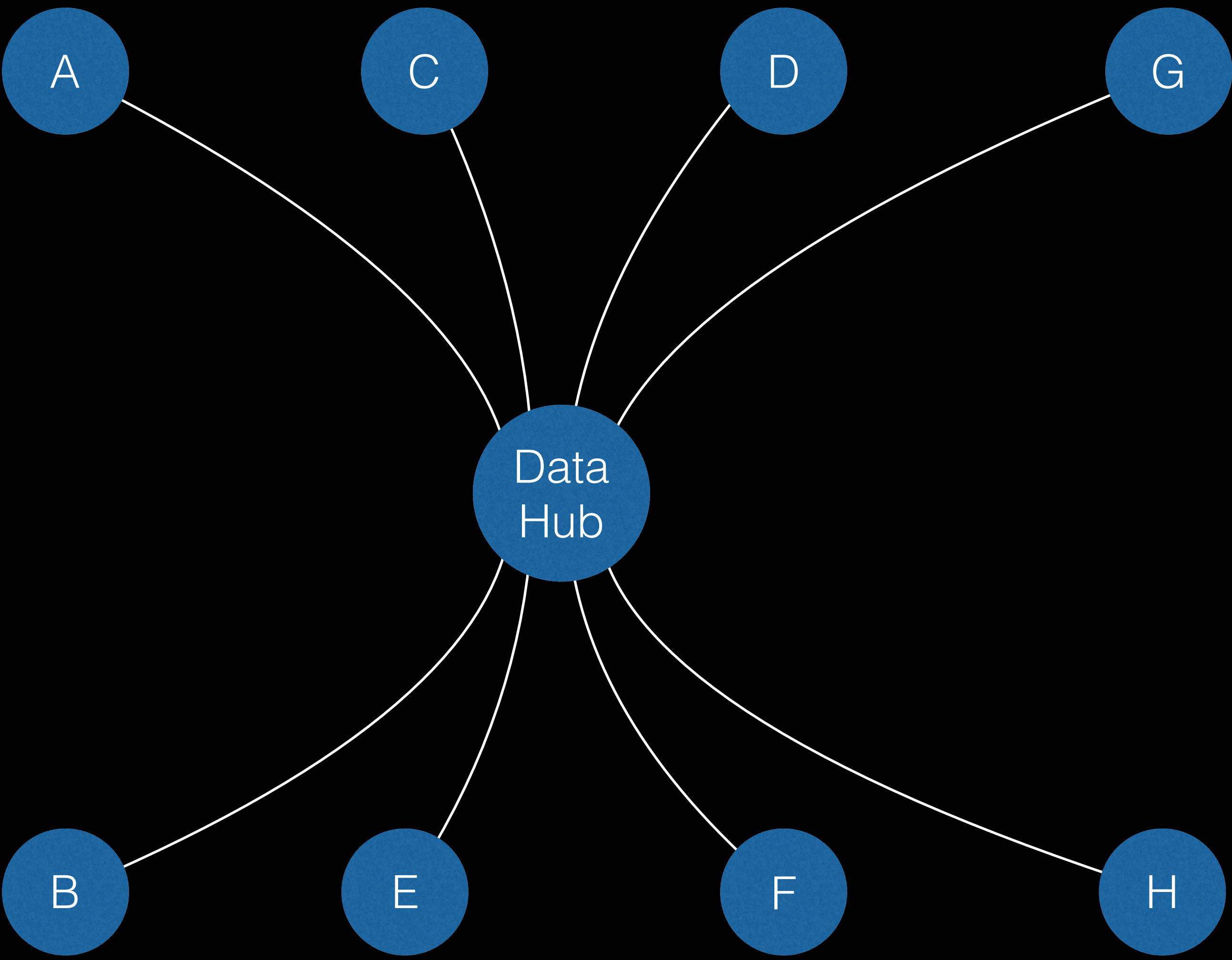
近实时数据系统：1 分钟延迟



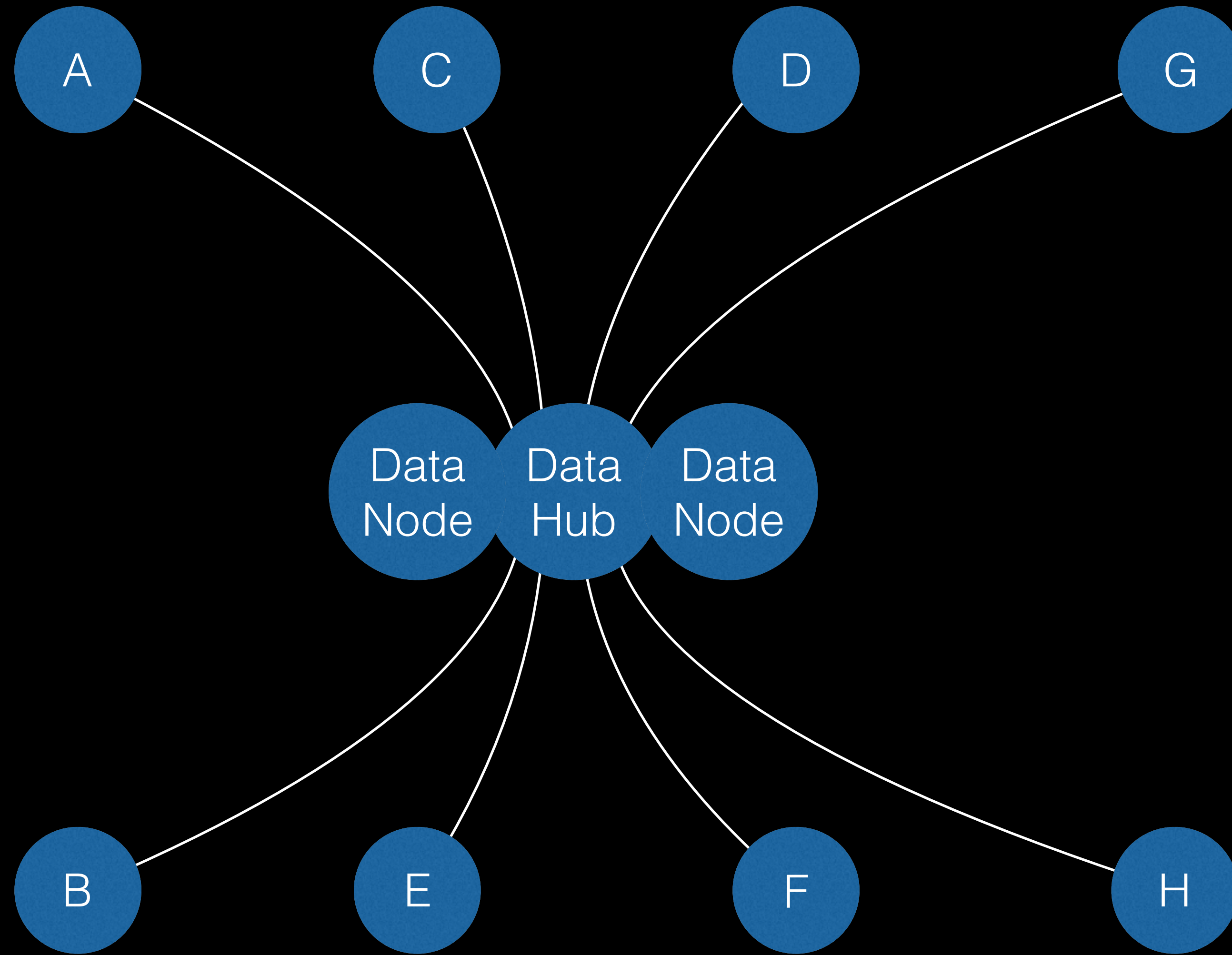
凌乱的数据流网



流网的有序化



分布式文件系统



编写高效数据服务的一些困难

快、快、快.....

程序处理速度快；数据的加载速度快；编程模型简单开发速度快

可用性

基础数据更新和模型更换时，能够不停机提供服务

可扩展 加机器

在服务做扩展的时候，数据流水线要能简单顺畅地与新节点衔接

可扩展 加进程

基础数据规模比较大，如果 worker 数量多，内存会出现瓶颈

快、快、快

6 分钟：雷达的数据更新周期

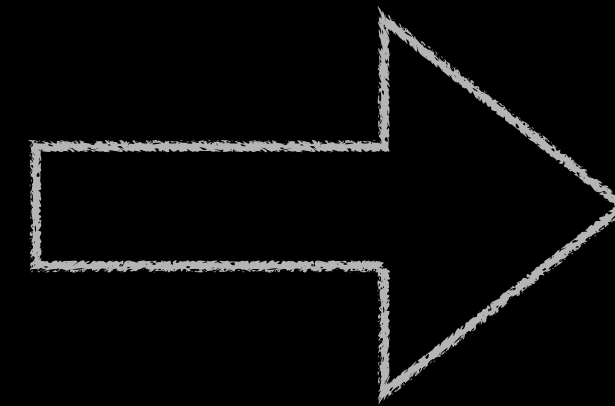
200 余部：实时更新的雷达数量

60 秒钟：抓取、分析、对外服务全部完成

程序处理最快的数据结构

加载最快的数据结构

编程最简单的数据结构



数组

进程间共享大数据块

mmap 是很老的技术，属于 POSIX 标准的一部分

mmap 在 numpy 里有非常易用的接口

案例

插值； 人工神经网络预测； 解算微分方程

8 核 16 G 的典型云服务器

20 个 Worker 的数据服务

单机 1000 ~ 4000 的 QPS

谢谢