

SYBASE®

大数据时代的 列式数据库

卢东明

Sybase软件(中国)有限公司技术总监



新浪微博

Sybase卢东明

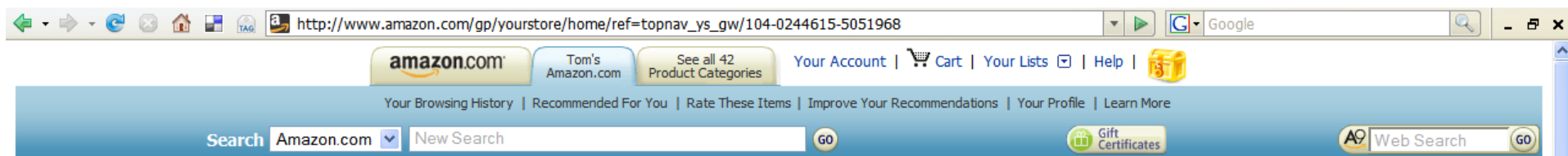


- 1992年加入Sybase China
- 1994-2006年在硅谷13年，效力Sybase公司，巴克莱资产管理公司(BGI)
- 2006年回到Sybase China任技术总监
- 新浪微博：Sybase卢东明
- 微群版主：“列式数据库”，“CEP”

Agenda

- 大数据有多“大”
- 大数据怎么“数”
- 大数据何以“据”

什么是Big Data：亚马逊新体验

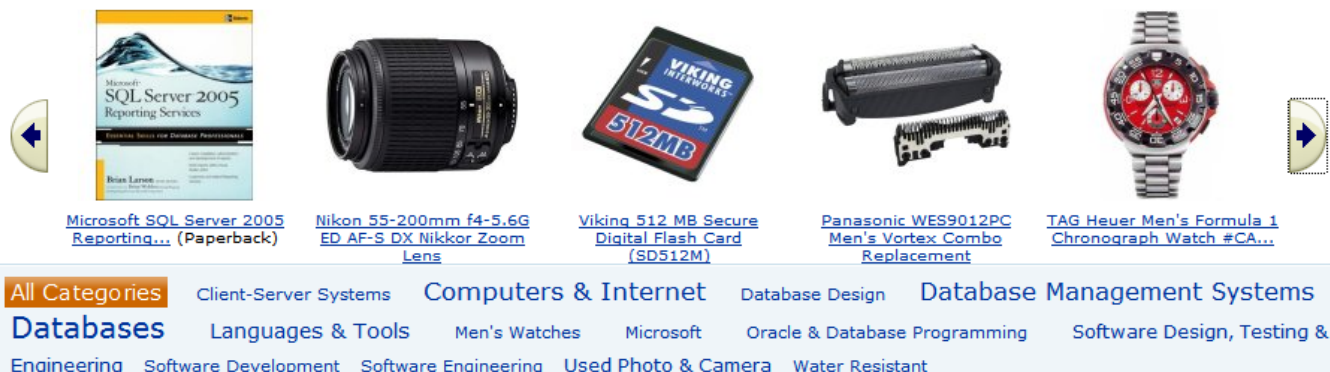


Tom, Welcome to Your Amazon.com™ (If you're not Tom Lu, [click here.](#))

1, 当日推荐

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations.](#)



Customers Who Searched for "panasonic es7056" Ultimately Bought

2, 其他搜索人后来买了那些东西



[Panasonic WES035P Vortex HydraClean Cleaning Cartridge an...](#)



[Panasonic ES7056S Vortex Triple Head HydraClean Men's Shaver](#)



[Panasonic WES9012PC Men's Vortex Combo Replacement](#)

[Your Shopping Cart](#) (0)
[Open & Recently Shipped Orders](#)

Your Lists
[Your Wish List](#)
[Your Gift List](#)
[Your Shopping List](#)

Your Community
[Your Amazon Friends](#)
[Your Interesting People](#)
[Your Reminders](#)

什么是Big Data: 亚马逊购物新体验

http://www.amazon.com/gp/yourstore/home/ref=topnav_ys_/104-0244615-5051968

From Your Wish List



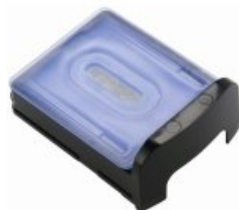
3, 希望购买列表

[Nikon SB-800 AF Speedlight for Nikon D50, D70s, D80, and...](#)

> [View & edit all items in your Wish List](#) (1)

From Your Shopping List

4, 购买单



[Panasonic WES035P Vortex HydraClean Cleaning Cartridge and Filter \(2-Pack\)](#)

> [View & edit all items in your Shopping List](#) (2)



[Vortex HydraClean Shaver System](#)

5, 其他类似购物者还买了什么

Customers Who Bought Items in Your Recent History Also Bought:



[Ricoh DB-60 Lithium-ion Rechargeable Battery for the GR... \(Jun 17, 2003\)](#)

> [View & edit Your Browsing History](#)



[Nikon 55-200mm f/4-5.6G ED IF AF-S DX VR Zoom Nikkor Lens \(Mar 29, 2007\)](#)



[Ricoh GC-1 Leatherette Case for the GR Digital Camera \(Jun 17, 2003\)](#)

[Your Recommendations](#)
[Recommendations Based On Your Tags](#)

Rate items to improve your recommendations:



[Panasonic WES035P Vortex HydraClean Cleaning Cartridge and Filter \(2-Pack\)](#)

Rate It

X | ★★★★★

☐ Don't use to make recommendations

[Items You Own](#) (47)

[Items You've Rated](#) (10)

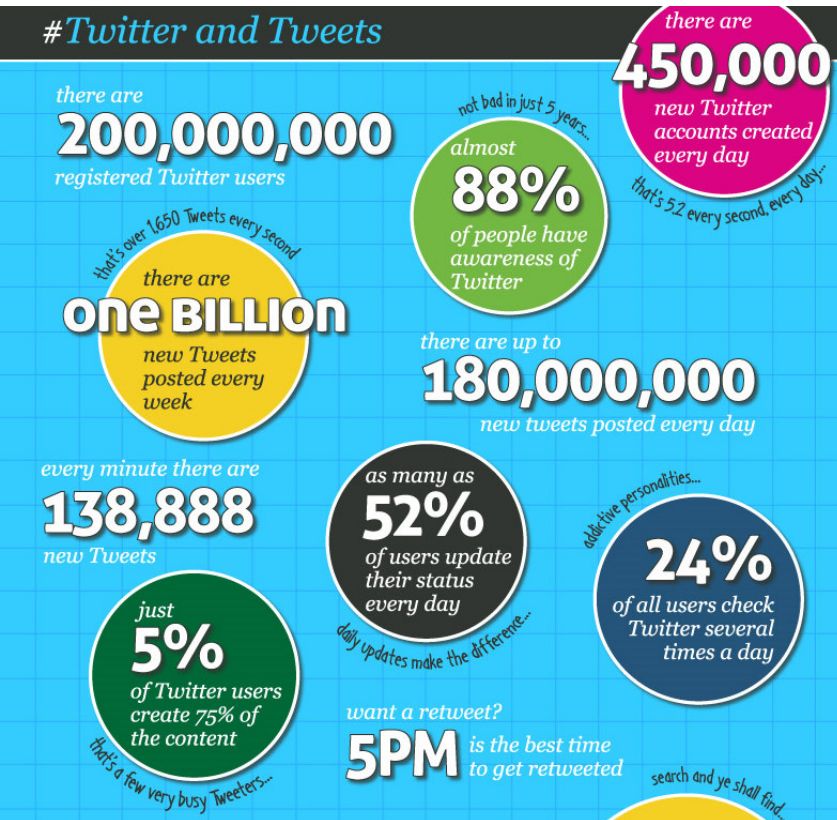
Search for items to rate:

Amazon.com

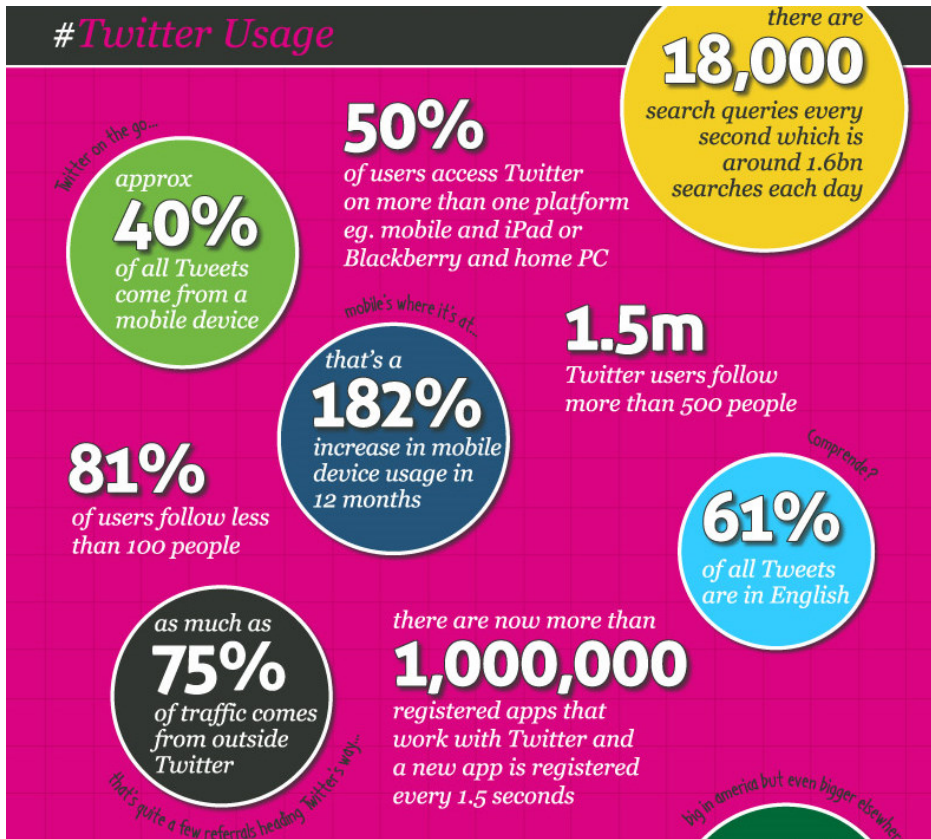
社交媒体 – 大数据的下一波挑战

Twitter 和 Tweets

- Twitter 有 **2,000,000,000** 注册用户
- 每天新注册用户 **450,000**
- 用户每周更新 **10 亿** 条 Tweets = 将近 **1.8 亿** 条每天 = **138888** 条每分 = 1650 条每秒
- 有将近 **52%** 的用户每天更新消息
- **5%** 的用户制造了 **75%** 的内容
- **24%** 的用户每天会查看好几次 Twitter
- 下午 **5** 点时发推最容易被转发
- Twitter 上每天有 **16 亿** 将搜索，每秒 **18000** 次



社交媒体 – 大数据的下一波挑战



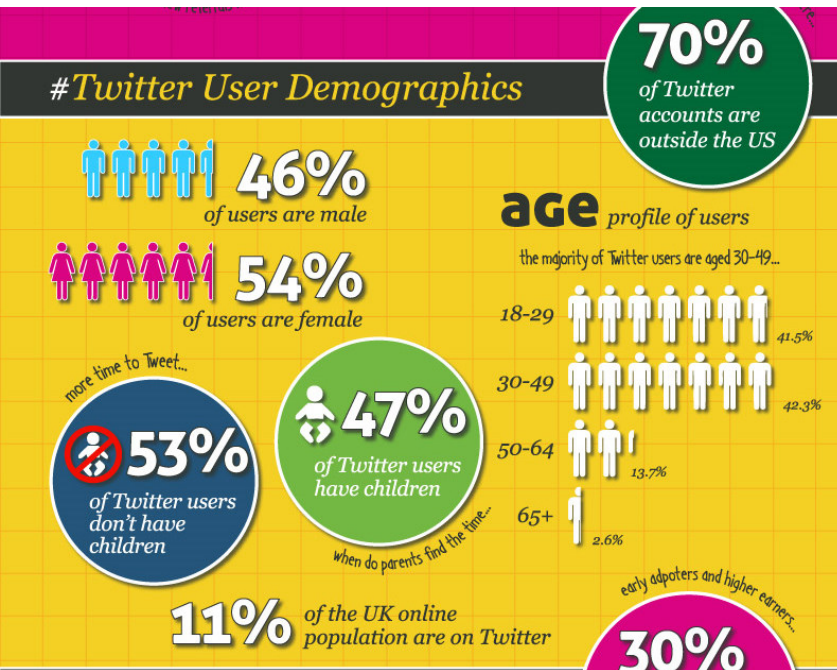
使用方式

- **40%** 的 Tweets 来自移动设备
- **50%** 的用户使用超过 1 个平台接入 Twitter，比如通过手机，iPad，家里电脑等
- **81%** 用户关注不到 **100** 人，**150 万** 用户关注超过 **500** 人
- Twitter 上注册的第三方应用有超过 **100 万** 个，每 **1.5 秒** 产生一个新应用
- **61%** 的 Tweets 是英语

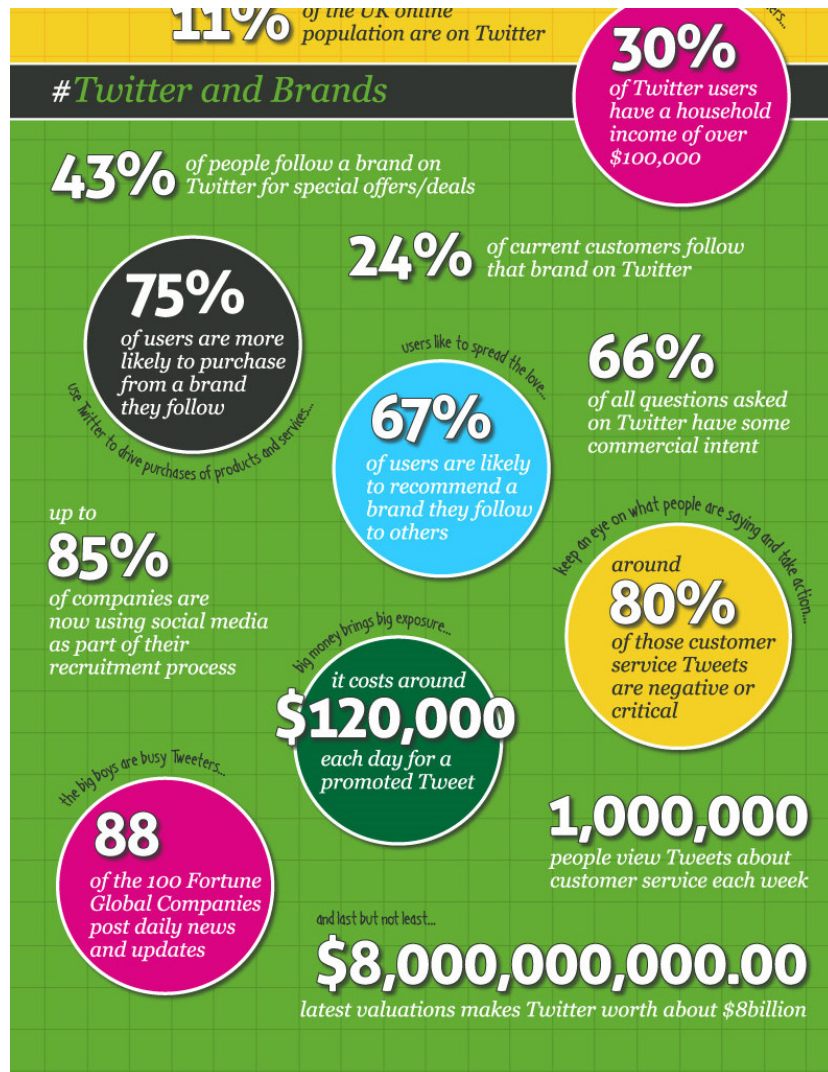
社交媒体 – 大数据的下一波挑战

用户分布

- **70%** 的 Twitter 用户是非美国用户
- **46%** 的用户是男性，**54%** 的用户是女性
- **53%** Twitter 用户没有小孩，有小孩的用户比例是 **47%**
- 英国网民 **11%** 使用 Twitter
- **30%** Twitter 用户的家庭收入超过 10 万美元



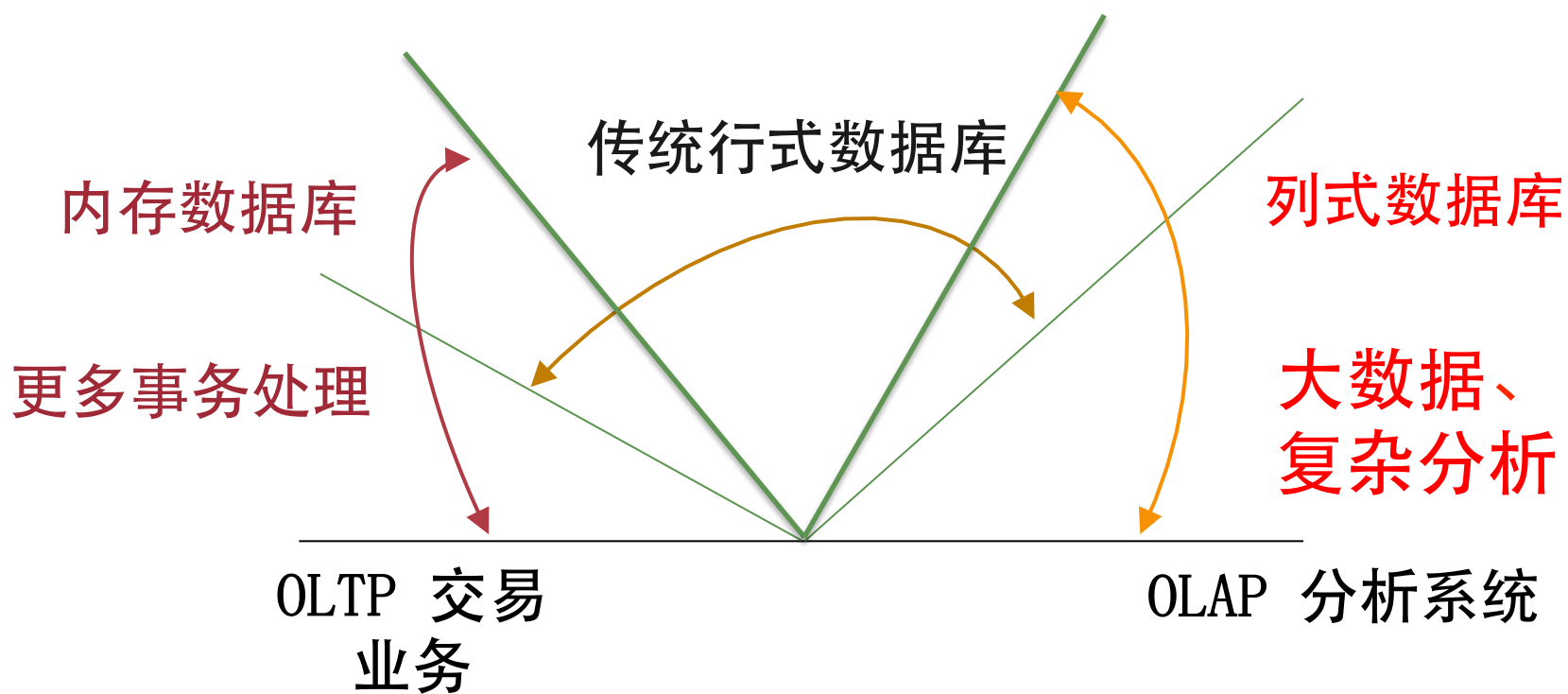
社交媒体 – 大数据的下一波挑战



Twitter 和品牌

- **43%** 的 Twitter 用户关注一个品牌（为了获取打折等信息）
- **75%** 的 Twitter 用户有可能从自己关注的品牌那里购买商品
- **67%** 的用户愿意将自己关注的品牌帐户推荐给其他人
- 一条推广的 Tweet 一天大概要花费 **120000 美元**
- 财富 100 强公司里有 **88** 家每天在 Twitter 上更新新闻
- Twitter 估值达到了 **80 亿**美元

数据库市场的细分格局



列式数据库的优秀案例

- ❖ 美国税务局(IRS): 全美国所有报税人7年报税记录(及原始文档)(15亿条记录):
 - ❖ 2006年审计数量为2000年的3倍
 - ❖ 2006年审计中收缴税款创下记录, 达到 592 亿美元, 上升 75%!
 - ❖ 190TB数据仓库 (压缩后数据量)
 - ❖ 六到八周加载一年的数据 [15 到 20TB] -- 四个小时!
- ❖ 花旗银行: Sybase IQ保存6年所有交易记录, HR数据及文档
- ❖ 美国证监会(SEC): 分析数据仓库以及非结构化数据(灾难备份点)
- ❖ 美国航空公司(American Airlines): 使用Sybase IQ节约了5百万美金
- ❖ ComScore: 全球第一大互联网监测公司: 1800亿URL
- ❖ 中国农业银行, 国家电网, 铁道部, 中国移动



SYBASE®

列式数据库
揭开云计算时代

SYBASE®



北京站 · 2012年4月18~20日
www.qconbeijing.com (11月启动)

QCon杭州站官网和资料
www.qconhangzhou.com

全球企业开发大会

INTERNATIONAL
SOFTWARE DEVELOPMENT
CONFERENCE