



Yahoo! Product Intelligence MicroStrategy World 2008

Amr A. Awadallah, PhD
VP Product Intelligence Engineering
Jan 2008

About Yahoo! and it's purpose.

Yahoo! who?

Yahoo! is the world's **largest global online network** of integrated services and is one of the most trafficked Internet destinations worldwide. For more than **12 years**, Yahoo! has been changing the way people **communicate** with each other, **conduct transactions**, and **access/share/create** information.

Our Purpose:

Yahoo! **powers** and **delights** our communities of **users, advertisers, and publishers** – all of us united in creating **indispensable** experiences, and fueled by **trust**.



Yahoo! Fast Facts

Our Name:

Yahoo! = **Y**et **A**nother **H**ierarchical **O**fficious **O**racle!
Dictionary definition= "rude, unsophisticated, uncouth"

Employees:

14,000 Yahoos worldwide

Global presence:

Operations in over 20 markets and regions around the world
Available in over 20 languages to **477M users**.

Milestones:

Founded in 1994
Incorporated in 1995
Public in 1996

Proud to be:

Fortune 500 Company
Fortune '100 Best Company to Work For'



Yahoo! Pillars and Strategic Objectives

Three Strategic Pillars:

- **Insights** will be leveraged to deliver **10x relevant** experiences
- **Open**, allow publishers to use our content/services, and vice versa
- **Partner-of-Choice**

Three Strategic Objectives:

- Online **starting point** for most **consumers**
- **Platforms** that attract the most **developers and publishers**
- A **must buy** for most **advertisers**

- We believe the **focus on relevance as a measure** will create a unifying focus to our work and drive increased value in everything we do.
- We are building the largest **content, services, and advertising exchange**.
- **Example Strategic Partners:** eBay, AT&T, Comcast, Newspaper Consortium, Bebo, WebMD, Cars.com, Forbes.com, Ziff Davis Media, DivX, Hearst.



Product Intelligence Engineering (PIE)

Continuously **generate** and
leverage **insights** to maximize
sustainable value created
through interactions within the
Yahoo! ecosystem



















Why Measure?

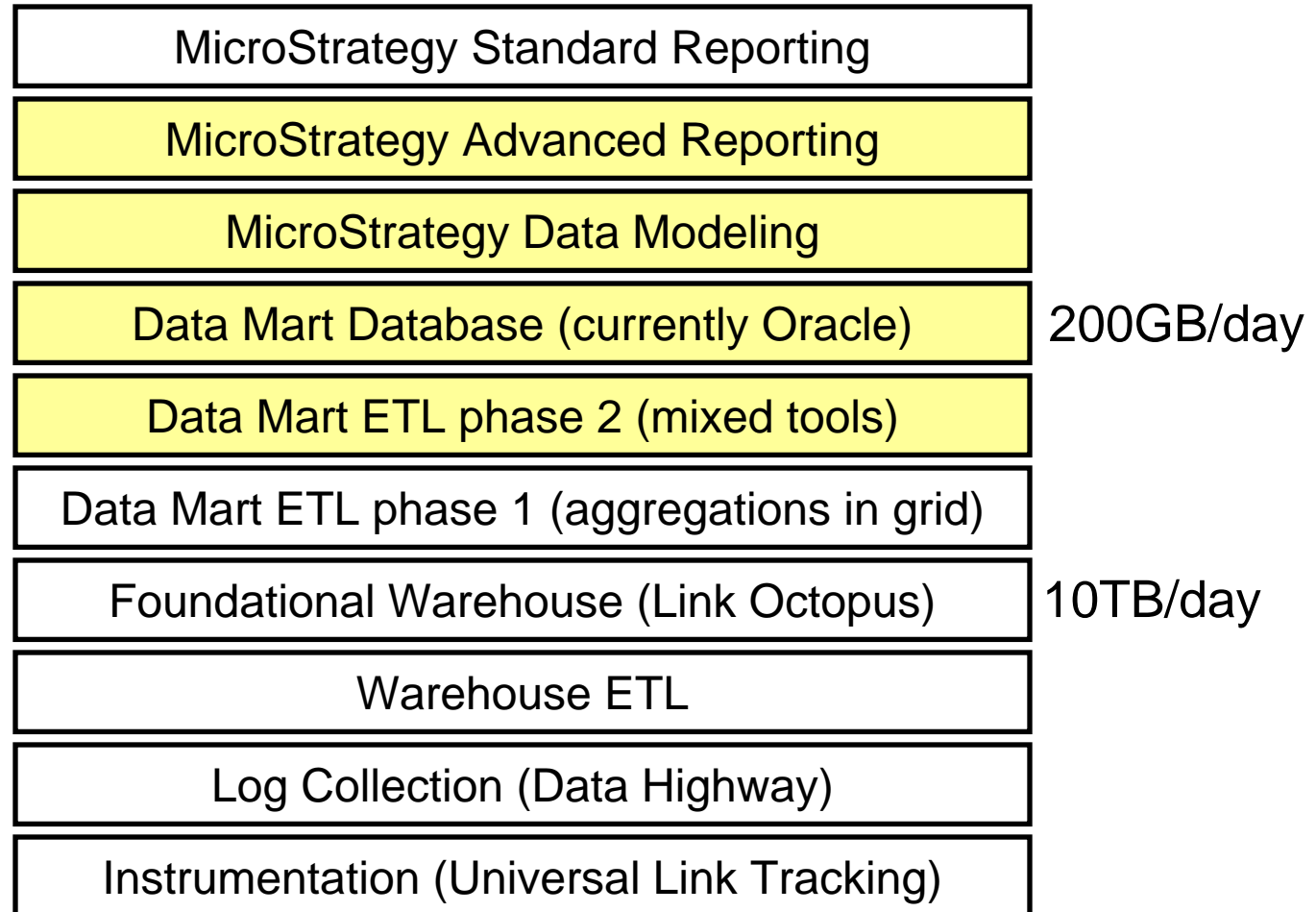
- “If you can’t measure it, you can’t fix it”
- “If you can’t measure it, you can’t grow it”
- “If you can’t measure it, you can’t build it”



We Support More Than a Dozen Datamarts

-  /Frontpage
-  Search
-  MAIL
-  TOOLBAR
-  MESSENGER
-  LOCAL
Maps
-  VIDEO
-  SPORTS
- 
-  GROUPS
-  ANSWERS
-  NEWS
-  FINANCE
-  TRAVEL
-  SHOPPING
-  ENTERTAINMENT

Our Data Stack



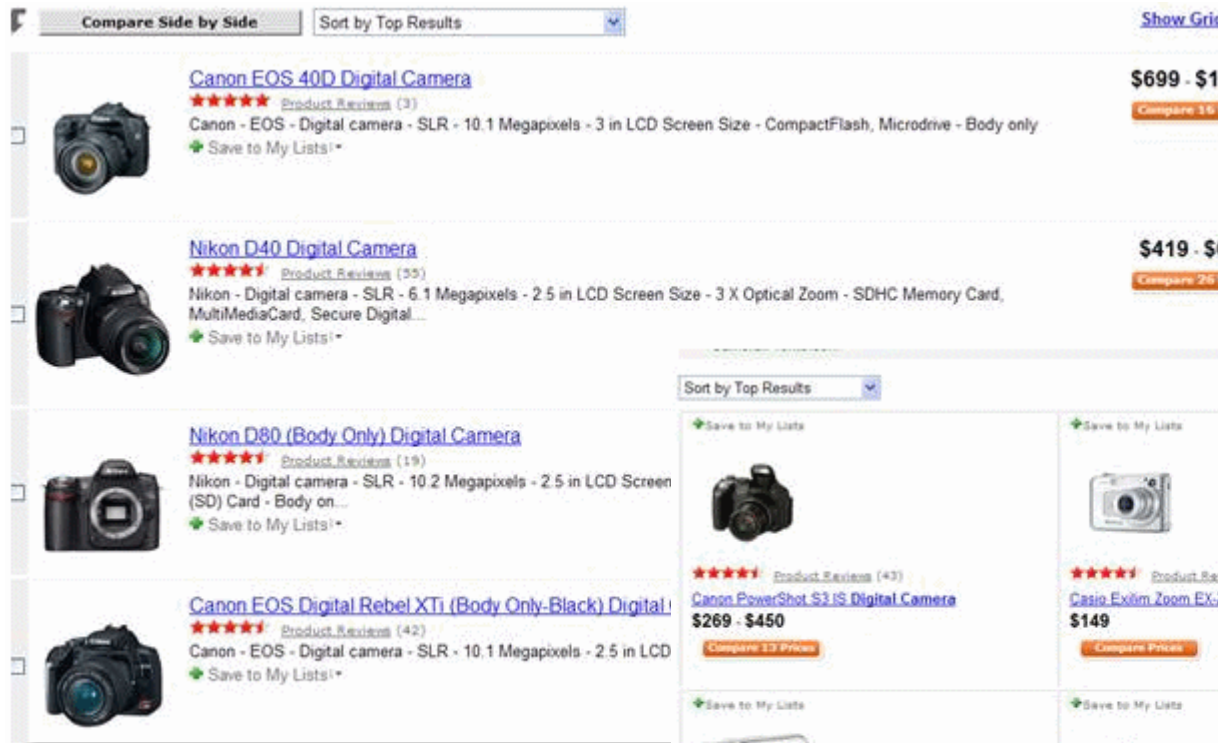
Groups That Use Our Systems

- Business Operations/Finance
- Product Management
- Research and Development
- User Experience and Design
- Product Marketing
- Advertising Sales



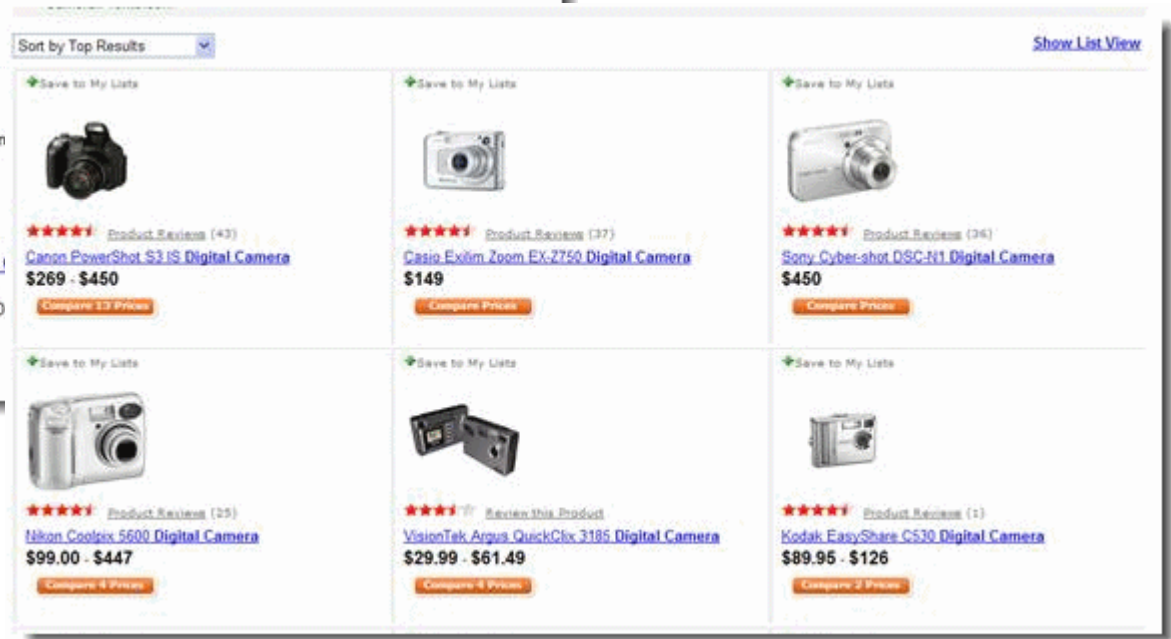
Case Study: A/B testing for Shopping.yahoo.com

Is it better to show items in a list or in a grid?



← Test Bucket A: List

Test Bucket B: Grid →



Dashboard Advice

- Its very hard (**impossible**) to reach a **single universal metric** that summarizes how the product is doing
- Proper **visualization tools are very important** since there are a lot of numbers to examine
- Averages are nice, but **histograms tell the real story**
- Trends are your friend, **keep history**, lots of it
- **Sampling** is very dangerous if not used properly



Example Metrics & Dimensions

Example Metrics:

- Click Probability (Cookies Clicked/Cookies Visited)
- Click Yield (Clicks Per Thousand Cookies)
- RPC (Revenue Per Cookie)
- Sessions/Cookie/Week (or Month)
- Time/Cookie/Week (or Month)
- Retention Rate (percent of Cookies that returned)

Example Dimensions:

- Demographics (Gender, Age, Income, Tenure)
- Geographics (Country, State, Zip, DMA)
- Content ID
- Access Modality (PC, PDA, Cell Phone, Net Speed, Browser, OS)
- Traffic Source (Organic, SEM, Affiliate, Marketing Campaign)
- Bucket Test ID



Evolution of Our BI systems

- Started with **grep** ☺ and generated **static html** dashboards
- Evolved to load a few aggregates into **MySQL** with a dynamic **Perl** dashboard
- Today we load a lot of aggregates for many metrics/dimensions into **Oracle** and use **MicroStrategy** for reporting.
- Next?



Next: A Unified Datamart (aka EDW)

- We currently have **18+ separate datamart** silos across all of Yahoo!'s products.
- We will bring all of these datamarts **under the same umbrella** so that we can easily do **cross-Yahoo! analytics** in the datamart.
- The data-model will need to be a **hybrid data-model** that supports **horizontal uniformity** but allows for **vertically deep** metrics and dimensions based on the product (e.g. mail sends is unique to mail)



Next: Click and View-stream Analytics

- Our current datamarts have **aggregate data only** which **limits the number of questions** that can be asked (we can still answer these questions from the LinkOctopus warehouse, but this requires an engineer to develop SQL due to complex schema).
- We will expand our datamarts to include **event-level data** (both click and view-stream events), this will cause a large **explosion in size** and number of rows (from 200GB/day to 10TB/day).
- The data-model will need to be a **hybrid data-model** that supports **event level data but also aggregates** (for performance and longevity)



Challenges we have

- Load data before new business day starts
- Operational stability
- Data quality: Bot filtering, Cookie churn
- Instrumentation Automation
- Columnar access control
- Scalable dimension, segmentation, and event-level processing



MicroStrategy: Things we like, Things we want

Things We Like:

- It **writes SQL** for us 😊
- It **creates Web GUIs** 😊
- We love the **new flash functionality**
- We like **Personalized Page Execution** in NCS

Things We Want:

- **Cross-mart** dashboarding
- URL functionality to **send out report links** 😞
- Better **Portal SDK**
- **Clickstream visualization**
- Better **NCS Debugging**
- **Intelligent Prompts**
- **Better search** on support.microstrategy.com

