How Java Powers Large Online Retail Sites

Robert Brazile - ATG
Jason Brazile - Netcetera
218

JAZOON
THE INTERNATIONAL CONFERENCE ON JAVA TECHNOLOGY
1-3 JUNE 2010, ZURICH

atg
netcetera

ORACLE netcetera

Agenda

- Introduction
- The state of e-commerce today
- Major functions of an e-commerce system
- What do we mean by “large scale”?
- Challenges
- Business requirements
- Architecture
- The marketplace
- Trends and the future
- War stories

ATG

- Founded in 1991
- Early Java adopter
  - Dynamo Application Server (1996)
  - Session tracking, page compilation licensed to Sun (1997)
  - Hand in original JSP STL and EL reference app (2002)
- More recently an e-commerce vendor

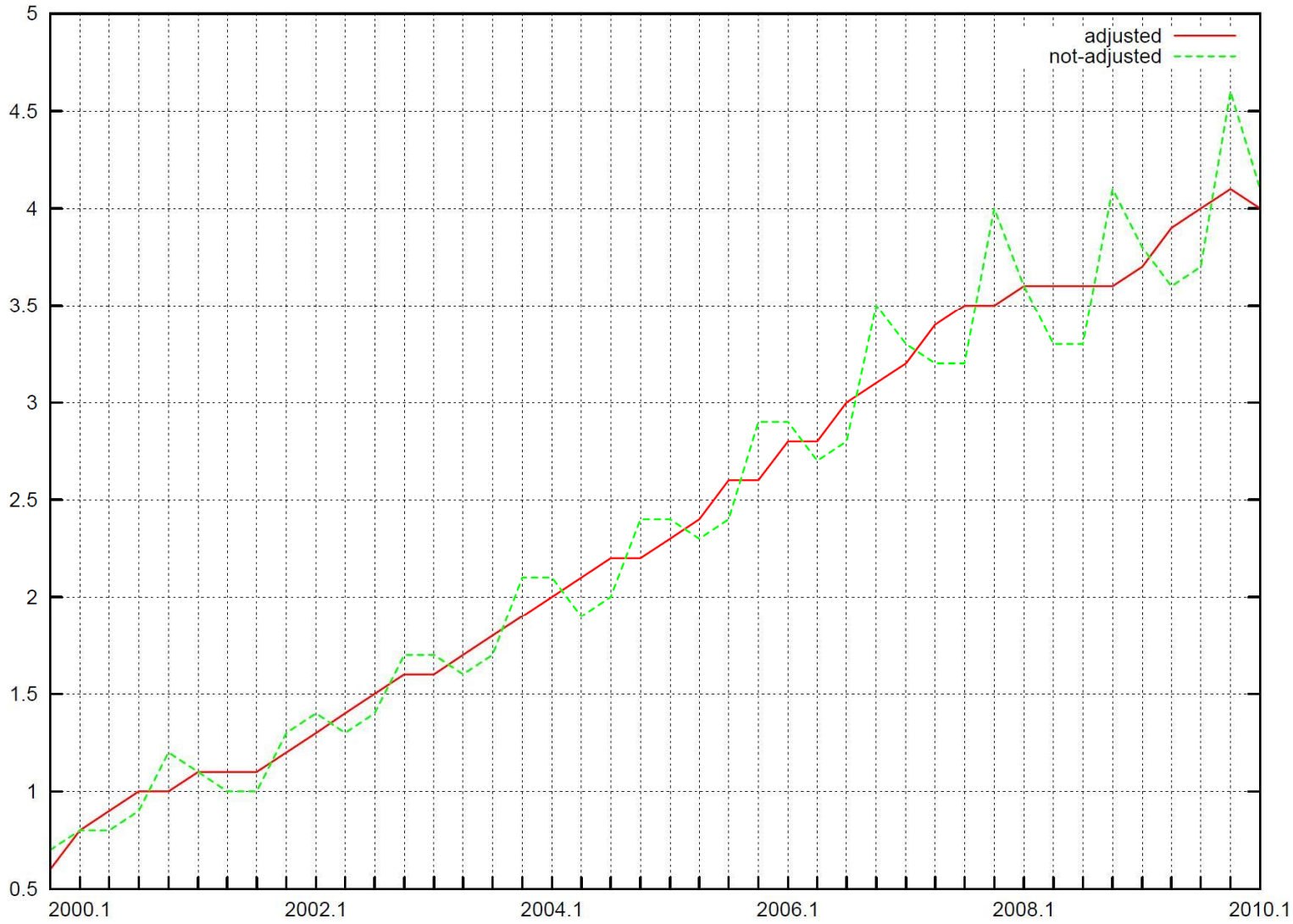
Some ATG customers

Selected ATG Commerce Customers

Selected ATG Commerce Suite Customers

Selected ATG Optimization Customers

US Census: e-Commerce as % of total retail sales



A timeline: some interesting dates

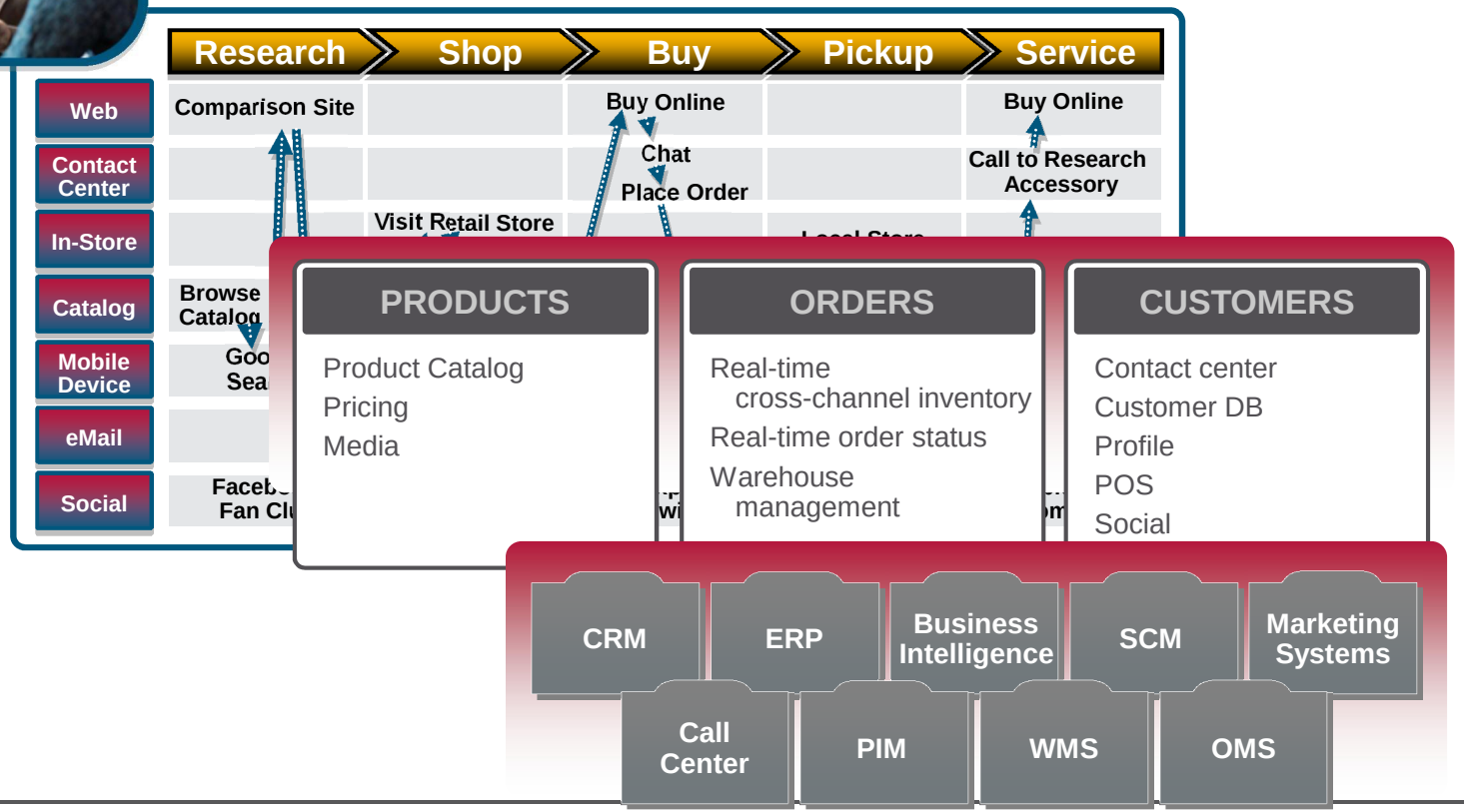
- 1979: Michael Aldrich invents online shopping (videotex with TV and phone line)
- 1982: Online train reservations possible with France's Minitel
- 1984: Jane Snowball, 72, first online home shopper (Gateshead SIS/TESCO)
- 1987: Swreg: First merchant account system supporting online payments
- 1990: Tim Berners-Lee's first web browser
- 1991: Oak (later Java) language invented for Sun's Star7 (PDA)
- 1994: Netscape introduced SSL encryption
- 1995: Amazon and AuctionWeb (later ebay) launched; Gosling invents Servlet
- 1996: JDK 1.0 software is released
- 1997: Java Servlet API 1.0 released
- 1998: PayPal invented; US Census Bureau begins tracking e-commerce
- 2003: Amazon posts first yearly profit
- 2008: Apple's iTunes passes Wal-Mart as #1 music retailer in US
- 2009: China's Alipay passes PayPal as #1 third-party online payment platform

Sources: "Electronic commerce", *Wikipedia*, May 2010 "Servlet History", *Jim Driscoll*, 10 Dec 2005 "iTunes Store Top Music Retailer in the US", *Apple Press Release*, 3 Apr 2008

The evolving shopping journey

A single purchase cycle involves many interactions

Elements required to support the journey



Major functions of an e-commerce system

- Content management
- Back-office integrations
  - Order management systems
  - Warehouse systems
  - Fulfillment systems
  - Pricing/Promotion systems
  - Combinations of these (ERP, CRM)
- Marketing campaigns
- Payment gateway and tax calculation
- Customer service systems
- Reporting and analytics
- Service integrations
  - Ratings and reviews
  - Product Recommendations
  - “Click to call”

These systems are well-suited to Java implementation

Examples of large scale retail: traffic

Large multinational retailer:

- 10M visitors 4Q09, planned for 1.5M visitors per hour
- 25K orders per hour
- 40 servers x 6 application instances per server
- expected to lose 15% capacity to SEO, scaled up to 57 servers to balance
- mobile and kiosks run from same pile
- actuals: 1.2M visitors per hour, 36K orders per hour
- Thanksgiving-”Cyber Monday” accounted for 1/3 of total
  - 287K orders, >12M visits (3:1 human:bot)
- Holiday peaks are ~10x in general

Examples of large scale: traffic

Large US retailer:

- Registered Users – 16,000,000
- Average Concurrent Users – 8,100
- Peak Concurrent Users – 27,000
- Average Page Views (Hour) – 1,100,000
- Peak Page Views (Hour) – 3,600,000
- Average Orders/Hour – 2,000 – 4,000 (Use 3,000)
- Peak Orders/Hour – 12,300

Examples of large scale: catalog

Sample catalog sizes:

Book retailer:

4 million products, 12 million SKUs, 18-20 million assets

Gen. merchandiser:

5-6 million products, plans to scale to 13.5 million (15M to 40M assets)

Direct merchandiser:

80k products, up to 50 SKUs per product, each SKU has 6 assets (usually translations) = close to 4 million products

Note: different organizations update different amounts and on different schedules, *e.g.,* 30% of the products weekly, say, or all products every night

Key takeaway

- "Large scale" takes on many different aspects
  - Size of catalog in number of products, SKUs, assets
  - Number of customers
  - Average order size
  - Frequency of product update
  - Volume of shopping traffic
  - Volume of transactions completed
  - Number of back-office integrations
  - etc., etc.

Challenges

- Business control
  - Reduce business dependency on IT for simple changes
  - Safe changes
  - Quick changes
  - Split testing
  - Continuous results measurement
  - Direct mgmt of business rules
- Operations
  - Monitoring and measurement
  - Deployment
- Speed, speed, speed
  - Responsiveness, refresh, change
  - Speed of interface, speed of change
- UX
  - Clean, usable, reduce clicks!
- Development
  - Thread-safety
  - Tuning and optimization
  - Developers should not be required for trivial changes

Operational challenges

- Scalability/Reliability/High Availability
  - Session and database design are critical
  - Redundancy (component level, device types, app server, DB tier)
  - Scale up vs. scale out
  - Disaster recovery and resiliency (active/passive v. active/active)
  - Capacity for peak demand vs. cost vs. performance
  - Testing: functionality, load and performance
- Integrations are critical
  - Sometimes the master for particular data types
  - Sometimes acts as proxy for other systems
  - What are business rules around availability?
  - Need to be “safe”, not bring the site down
  - Must decouple site performance from that of integrated system

Business requirements

- Managing site content
  - Content management (catalog and marketing content)
  - Personalization (implicit, explicit, manual, automated)
  - Measurement
  - Marketing campaigns
  - Ability to accept and use UGC
- Managing the business
  - Merchandising
  - Split (A/B) and multivariate testing
  - Multichannel (incl affiliate)
  - Different styles of buying and selling (store, auction, bazaar, subscription)
  - Search engine optimization
- Operating the site
  - Site administration, multiple sites
  - Internationalization, localization
  - Delegation of authority, roles
  - PCI DSS/ISO 27001/2

Architecture

- Over-simplified history
  - Largely the history of dynamic, data-driven sites
  - Consider the timeline given earlier
  - Progression of tools favored for this

    CGI, Cold Fusion, ASP, Java, Perl, PHP, Ruby etc.
  - Today quite a mix of scripting languages, Java, and frameworks
- Consider both application architecture and server architecture
- In our case, a subset of Java standard features implements major infrastructure
  - Servlets, Java Beans, JTA, JMS, JDBC, various JAX elements
  - Our own dependency-injection system and dynamically-typed ORM layered on top
- Presentation layer is independent, can be JSP, Struts, Flex/Flash, etc.

Application architecture considerations

- Must be master, or act as proxy for master, for many processes and entities
  - Catalog, prices, customer profiles, orders, etc.
- Reusable components (both backend and site elements), services
  - Often will be used by other applications via web services
- Presentation: reusable/re-targetable components, speed, device- and locale-specificity
- Order processing pipeline
  - Write plug-ins for price, tax, shipping calculations, inventory checks, etc.
- Clean data model for performance, management, and future growth

Server architecture

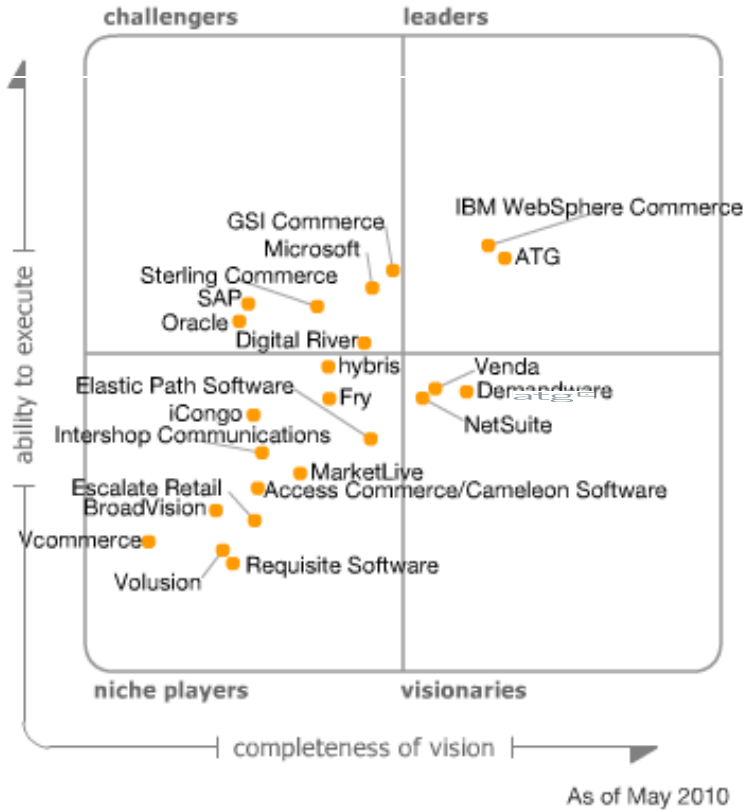
- Cloud computing increasingly a factor
  - In services: analytics, recommendations, ratings and reviews, payment, etc.
  - Cloud hosting: scalability, disaster recovery (DR) benefits
  - Provider perspective: economy of scale through multitenancy
- For a particular site, engineering analysis required
  - n-tier model with session-affinity vs. “shared-nothing”
  - Consider tradeoffs

    Complexity v. scalability

    Potentially massive, distributed relational database installation vs. NoSQL approach
- Truly massive sites may require shared-nothing elements such as external caching and partitioning (e.g., sharding); this is determined by requirements
- Content Distribution Networks (CDN) are heavily used to reduce server load

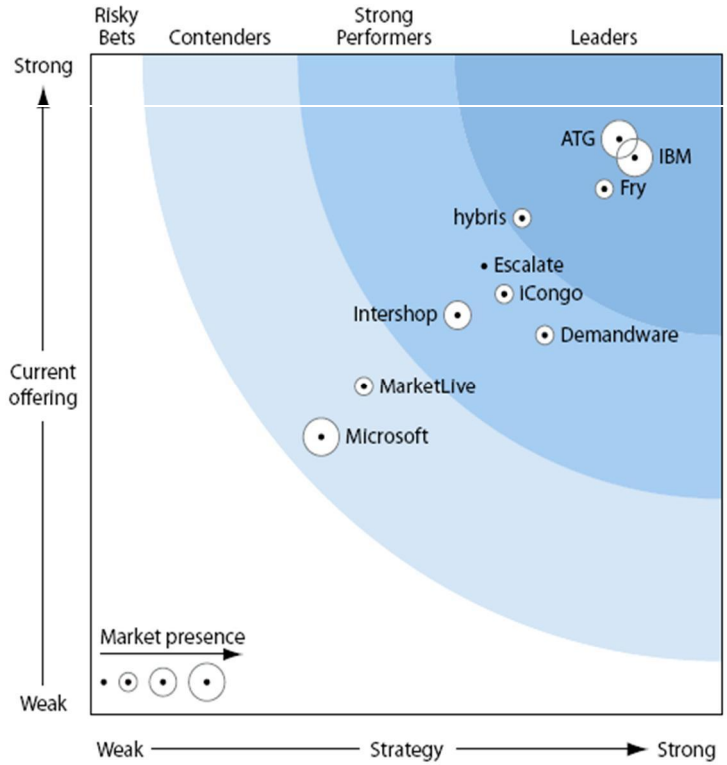
Current industry: analyst view

Gartner Magic Quadrant for e-Commerce



Source: Gartner (2008)

Forrester Wave: B2C e-Commerce Platforms



Source: Forrester (2009)

Plus open source providers, such as Magento and osCommerce

JAZOON
THE INTERNATIONAL CONFERENCE ON JAVA TECHNOLOGY
1-3 JUNE 2010, ZURICH
atg
ORACLE
netcetera

Trends and the future

- Business
  - Mobile is growing rapidly, *is* e-commerce in developing countries, and changing business processes as well
  - Social networking
  - Convergence of these and other channels
  - Growing use in the developing world
  - Ease of use by the business user
  - Spawning of many smaller sites rather than changing big one
- Technology
  - Virtualization/Cloud Computing
  - NoSQL
  - Scripting (PHP, Ruby, Groovy, Scala, Clojure, Erlang, etc.)
  - Frameworks (Rails, Grails, Lift, etc.)
  - Multi-core, more concurrency: STM?

Social and mobile convergence

Social example: Tor.com

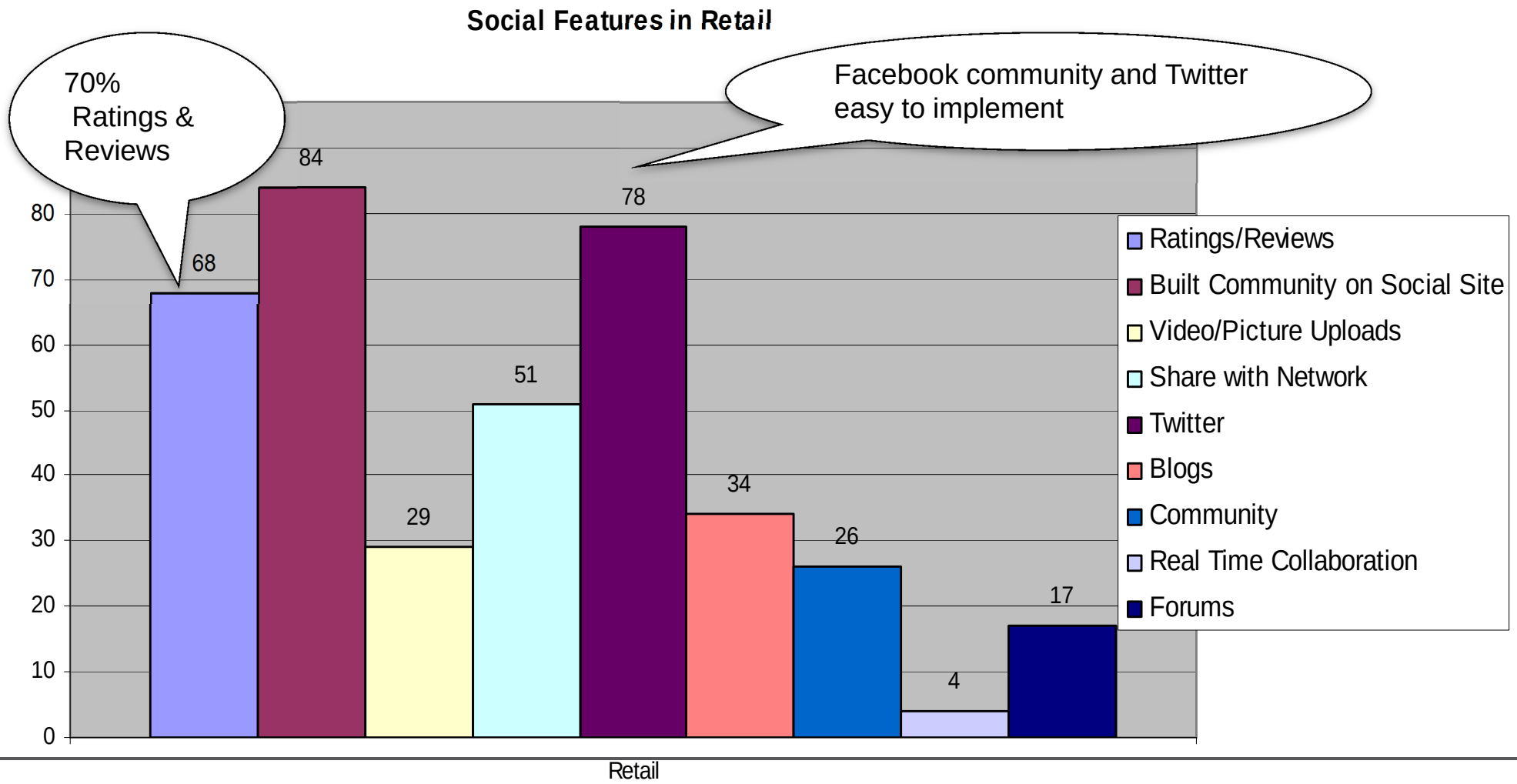
Science-fiction and fantasy publisher, owned by Macmillan

New site is pure social commerce

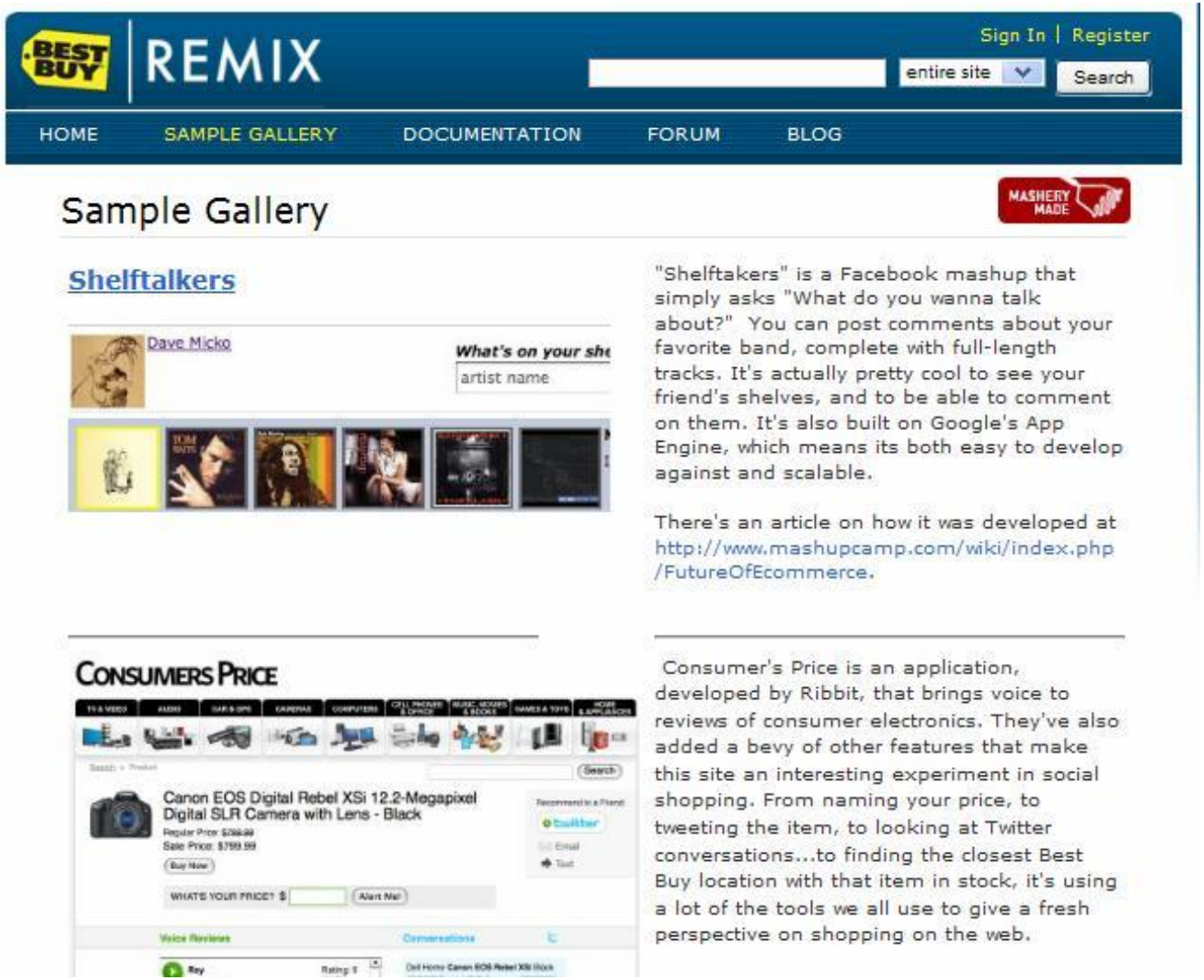
Content, Community, Commerce

- Short stories, art, podcasts, reviews
- Moderated forums (“conversations”)
- All content tagged and accessible via a tag cloud
- Specific entries promoted via “bookmarks”
- Event calendar
- Store for purchasing books and m’dise, as well as link to Main Macmillan store

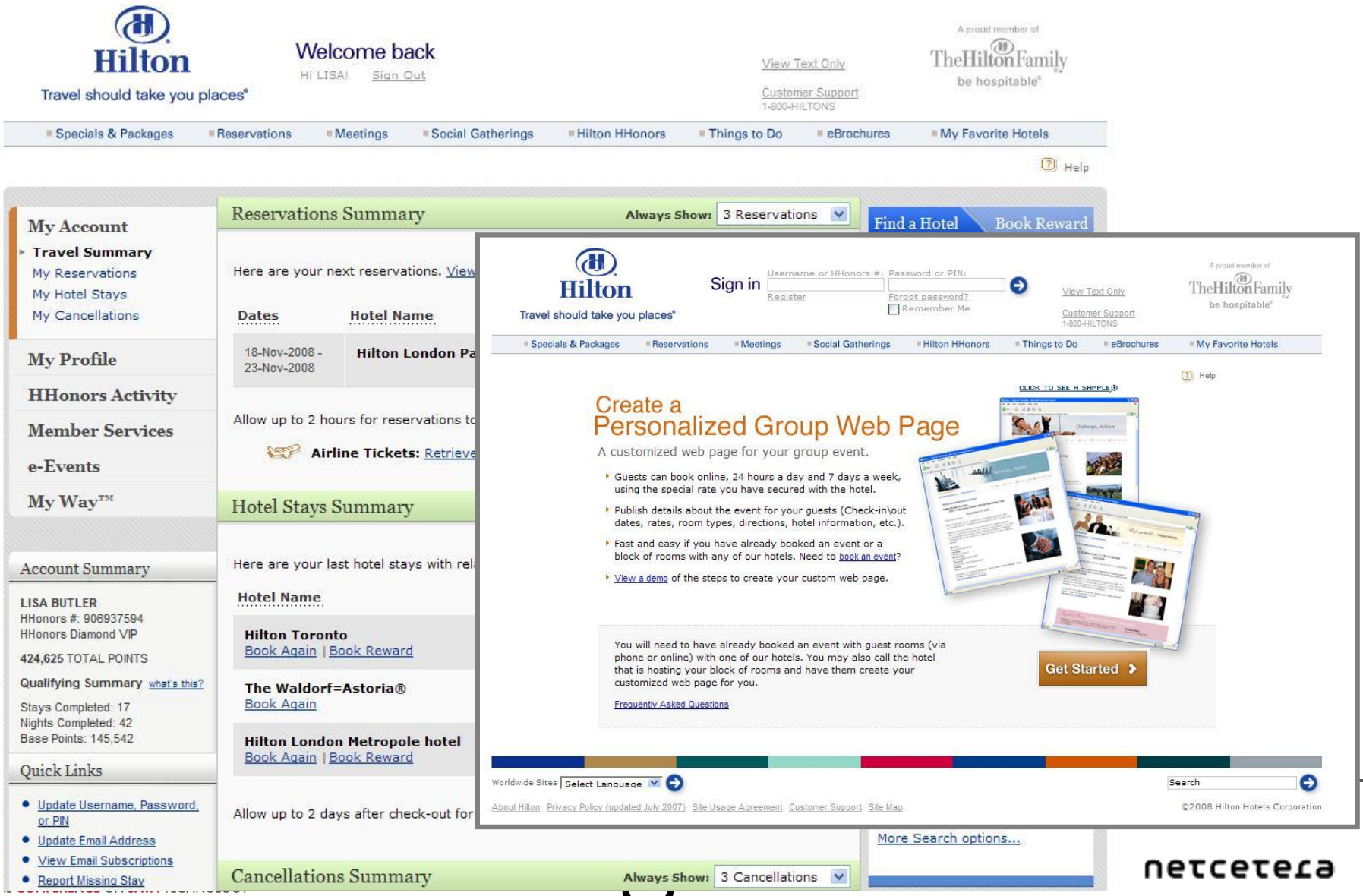
Retail


Social Features in Retail
70% Ratings & Reviews
Facebook community and Twitter easy to implement
0
10
20
30
40
50
60
70
80
68
84
29
51
78
34
26
4
17
Ratings/Reviews
Built Community on Social Site
Video/Picture Uploads
Share with Network
Twitter
Blogs
Community
Real Time Collaboration
Forums
Retail

Best Buy Remix - Syndicates Product Content

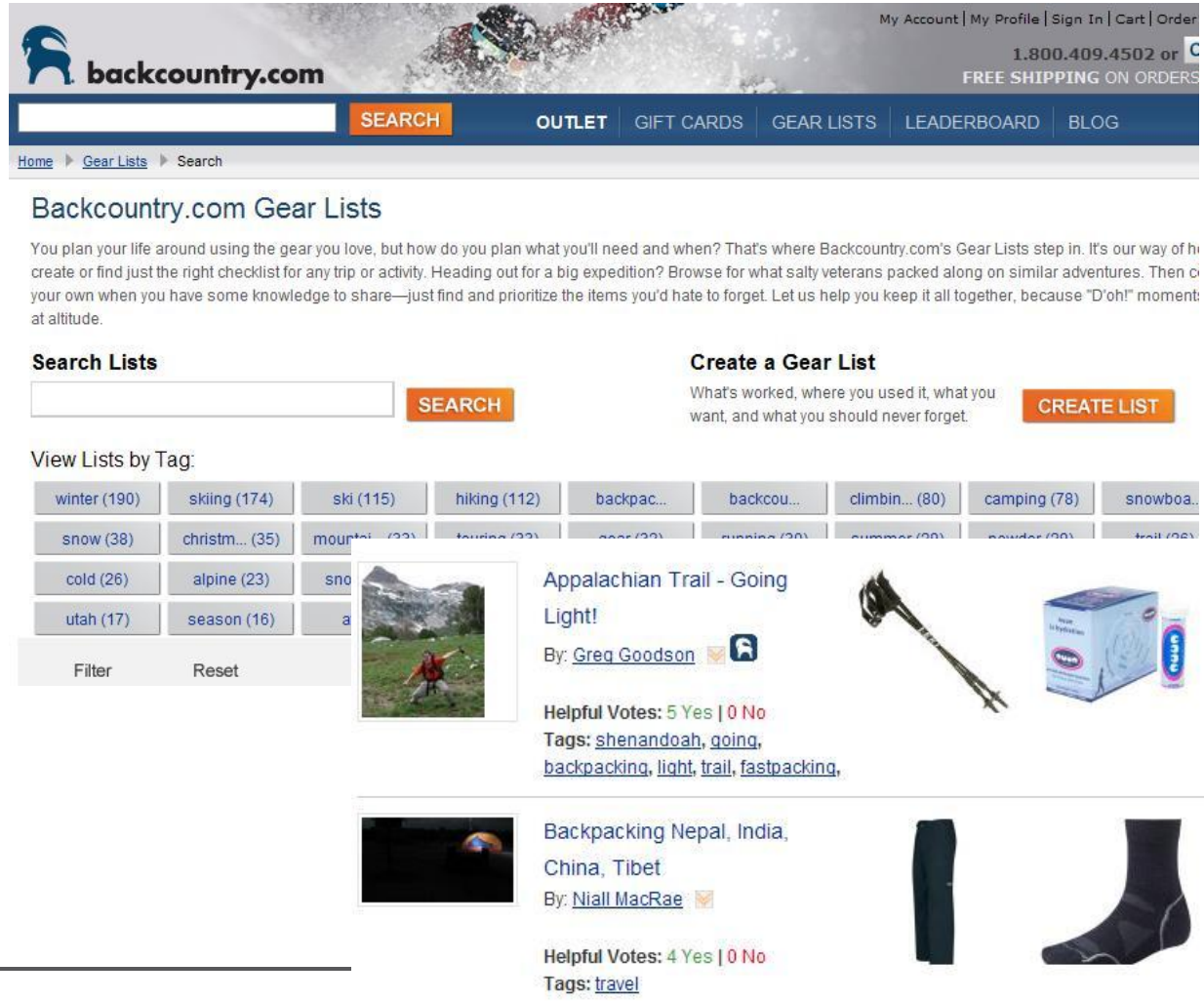


Personalized Pages at Hilton



Backcountry.com- Gear Guru

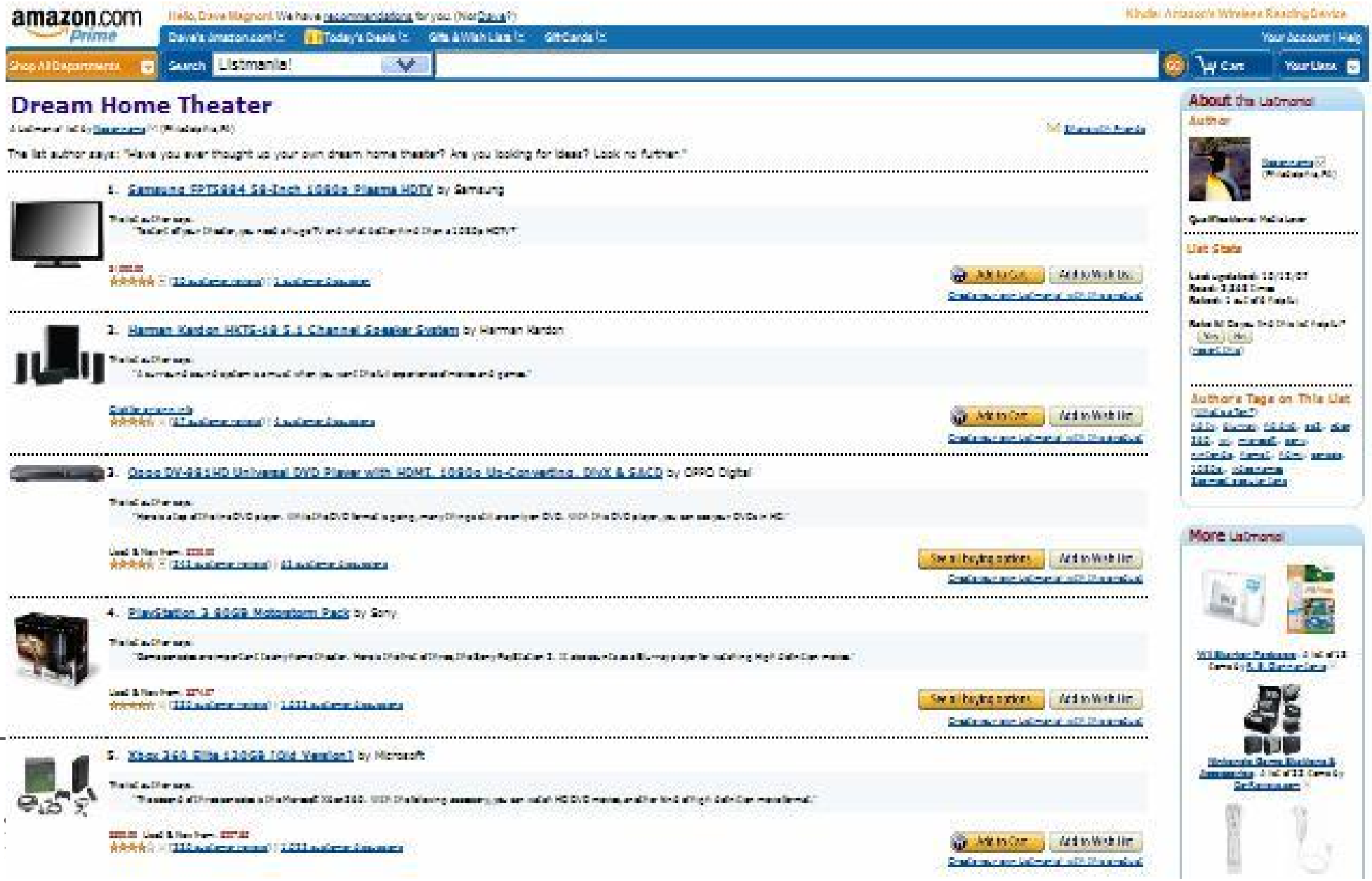
backcountry.com



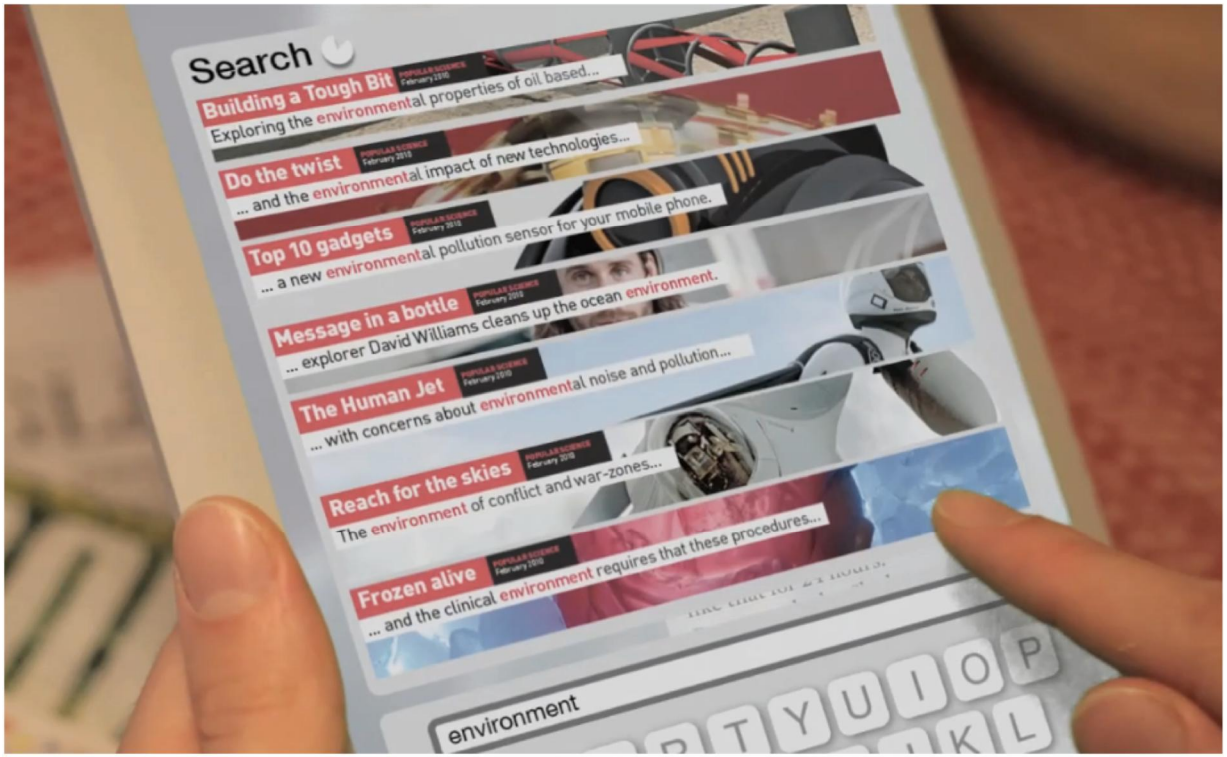
A cool section where you can create or find the perfect checklist for any activity or trip. You can browse lists other veterans created and modify or make a gear list of your own and post it to the site.

They also have a Leaderboard section, which ranks the top “Gear Guru’s” who contribute the most to the blog, reviews, and question and answers section.

Amazon's Listmania! Drives AOV



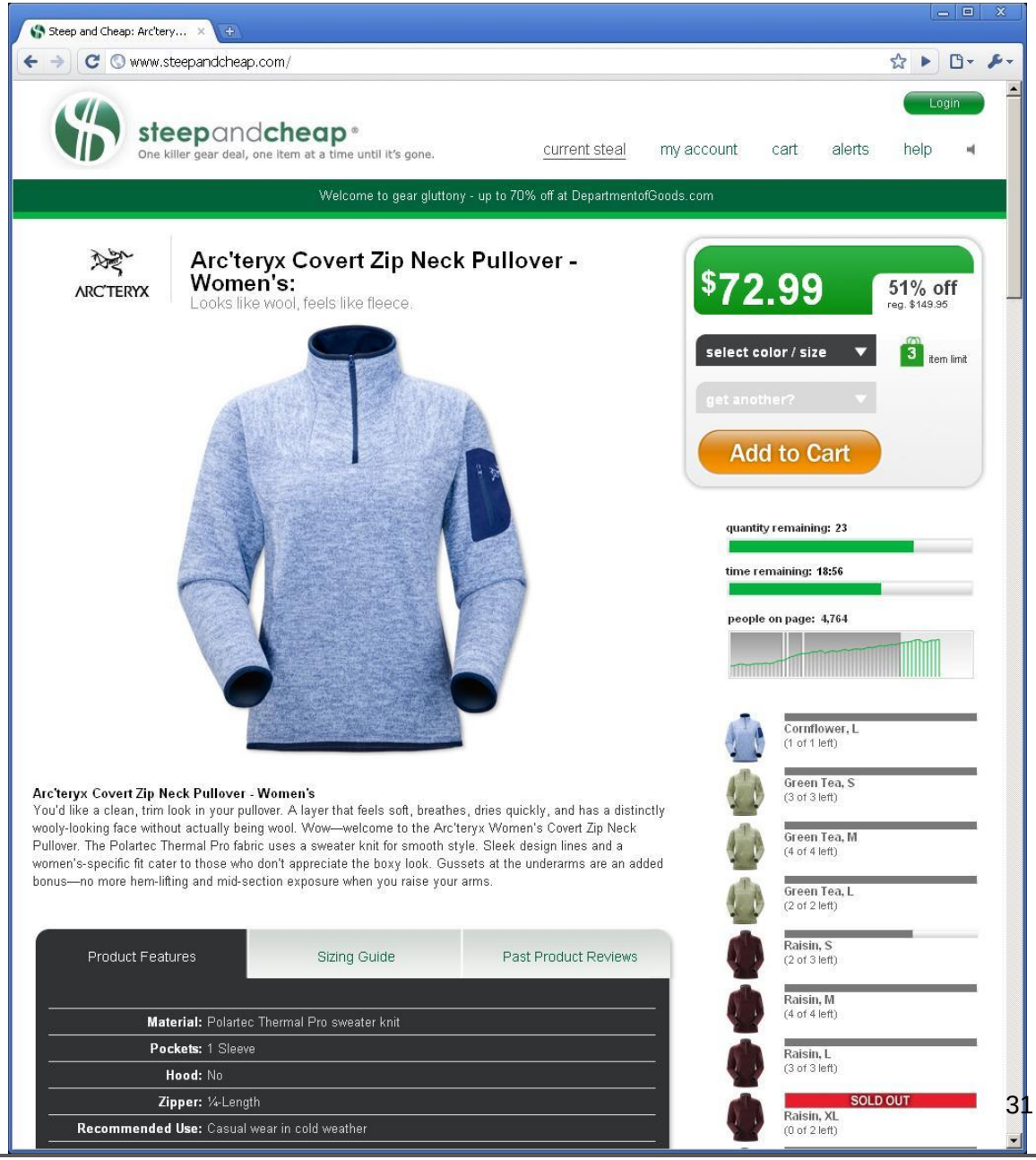
Social and mobile convergence



E-magazine

http://bit.ly/8w6Anf

steepandcheap.com

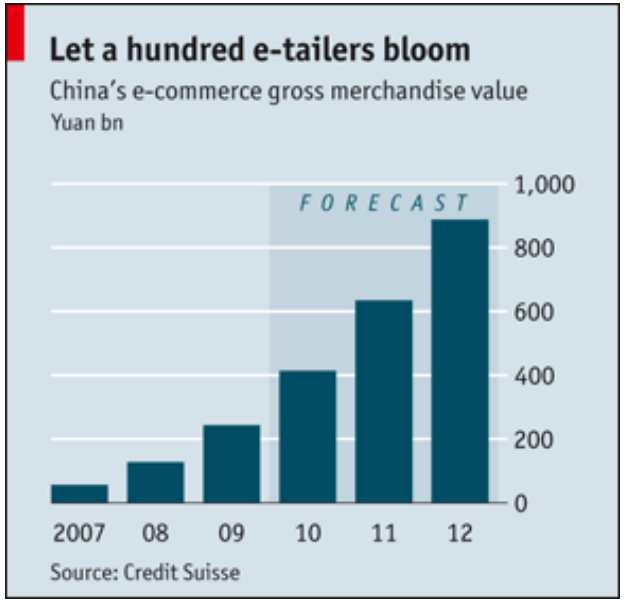


JAZOON
THE INTERNATIONAL CONFERENCE ON JAVA TECHNOLOGY
1-3 JUNE 2010, ZURICH
atg
ORACLE
netcetera

Multichannel?

Economist: “Internet shopping explodes in China”

- Computers, internet cafes, smart phones are now ubiquitous
- Local alternative payment technology established: **Alipay** (China’s PayPal)
- Scooter delivery: <1 hour, 5 yuan ($0.73) in major cities
- Clothing and electronics led early
- 66% online bought within last 6 months
- 50% online with children bought diapers, formula
- Average online discount 21%



From: “Clicks trump bricks”, *The Economist*, 22 Apr 2010

B2B players: Alibaba.com, HC360.com, Myekoo.com

C2C players: Taobao.com, Paipai.com, Eachnet.com

B2C Online Retailers: 360buy.com, Joyo.com, Dangdang.com

From: “Chinese E-Commerce Tops $38.5 Billion; What Comes Next?”, *ReadWriteWeb*, 19 Apr 2010

War Stories

- Unanticipated consequences of integrations
  - PS3 holiday promotion gone awry
- Problems with automation:
  - Amazon: Searched for abortion, got “Did you mean *adoption*?”
  - MLK/Black History Month/Planet of the Apes fiasco
- Effect of design choices
  - “Show all shoes”
- Struggles with outsourcing and education
  - Outsourcer builds entire Shopping Cart, purchase pipeline when in product
  - Outsourcer builds mail sender, scheduler, SQL messaging when in product

War stories (cont'd)

- Perils of testing (or not testing, which is far more common)
  - Testing production site ended up allowing orders using CC test number
  - Testing production site resulted in case of whisky arriving at the test lab
  - Campaign testing: gibberish email sent to 200K people
  - Same coupon promo worked over and over again

Robert Brazile
ATG

http://www.atg.com/
brazile@atg.com

Jason Brazile
Netcetera

http://netcetera.ch/
jbrazile@netcetera.ch