

# **Web-Scale Recommendation Engines**

CS130 Fall 2012

**Andrew Look**  
Senior Software Engineer  
**Shopzilla**

# About Shopzilla

- Comparison Shopping Engine
- Lots of traffic
  - > 200 million products
  - > 100 million impressions / day
  - > 8000 searches / second
  - > 30M unique visitors / month
- Operate web properties in US, Europe
  - Shopzilla.com
  - Bizrate.com
  - Beso.com

# Beso.com

beso THE SITE DEVOTED TO SHOPPING

What can we help you find?

My Beso Sign Up About Beso  

WOMEN MEN HOME KIDS JEWELRY BEAUTY SALE EDITOR PICKS COLLECTIONS BRANDS & Beso Blog

Follow @beso  21k

## THE 10 THINGS THAT MAKE MY LIFE A LITTLE PRETTIER




It's a steadfast morning ritual to click over to EMILY SCHUMAN's blog Cupcakes and Cashmere for a dose of images from Schuman's beautiful world. Now that she's the author of lifestyle book, *A Guide to Defining your Style, Reinventing Your Space, and Entertaining with Ease*, we knew that Emily's Top 10 would make any daily routine more delightful.

[SHOP HER PICKS >](#)


DEAL ALERTS  
Personalize your shopping experience.  
[Start now >](#)

OUR GIFT TO YOU  
Win a \$250 Cash Shopping Splurge!  
[Enter now >](#)







 ELISE LOEHNEN  
editor-in-chief  
See what's on her wish list.

 DIANA RYU  
associate fashion editor  
View all the things she loves right now.

 ANDREEA ABRIA-DEVOE  
contributing editor  
Check out her collection of must-buys.

 CRYSTAL MEERS  
contributing editor  
See what she's all about.

### What We Love Right Now >>

					
J.Crew Cuffed printed wool and silk-blend... New-A-Porter \$228.00	Canvas Lands' End Women's True Slim Chinos Lands' End Canvas \$99.00	Ipad Case By Clare Vivier For Splendid Splendid \$128.00	Canvas Lands' End Women's Rugby Shirt ... Lands' End Canvas \$59.00	Canvas Lands' End Women's Long Sleeve... Lands' End Canvas \$59.00	Wedge Peash By Raggs Tigertree \$16.00

### What Users Are Viewing Right Now

[View All](#)

	NIKE / Men's Shoes Nike Special Field Boot Chukka ID Custom Men's Shoe - Brown Nike Special Field Boot Chukka ID Shoe A CUSTOMIZABLE, HIGH-PERFORMANCE BOOT The Nike Special Field Boot is engineered to ... Nike Official Store \$165.00
	RALPH LAUREN / Men's Shoes Polo Ralph Lauren Vaughn sneaker Men's lace up sneaker Canvas upper Textile lining Lace up closure Rubber outsole Piperlime \$59.00
	IZOD / Men's Pants Izod American Chino Men's Ek Pilot Front Twill Pants

### Most Viewed Brands & Stores

6pm.com	Lacoste
Adidas	Levi's
COTY	Neiman Marcus
Crystorama	Nike
Eddie Bauer	Nordstrom
Hugo Boss	Reebok
Hurley	Target.com

# Beso.com: Personalization

- Focused on personalization for users
- Users can pick favorite..
- Brands
- Stores
- Products

The screenshot shows the Beso.com user profile for Andrew Look. The page features a blue navigation bar with categories like WOMEN, MEN, HOME, KIDS, JEWELRY, BEAUTY, SALE, EDITOR PICKS, COLLECTIONS, and BRANDS & STORES. A search bar is located at the top right. The user's profile section includes a placeholder for a profile picture and links to manage account, mail, and photos. Below this, there are sections for 'FILTER FAVORITES BY:' (ALL, MEN'S, HOME, GADGETS), 'FAVORITE BRANDS' (Ben Sherman, Cole Haan), and 'FAVORITE STORES' (currently empty). The main content area displays a grid of favorite items, including shoes, jeans, and books, each with a price and a '50% off' or '40% off' badge. The items are categorized under 'FAVORITES', 'RECENTLY VIEWED', and 'COLLECTIONS'.

**beso** THE SITE DEVOTED TO SHOPPING

What can we help you find?

My Beso Sign Out About Beso

WOMEN MEN HOME KIDS JEWELRY BEAUTY SALE EDITOR PICKS COLLECTIONS BRANDS & STORES Beso

Home : Profile : Andrew Look

**Andrew Look**

- manage account
- my mail
- manage photos

FILTER FAVORITES BY:

ALL

MEN'S

- Apparel
- Shoes

HOME

- Decor

GADGETS

- Cell Phone Accessories

FAVORITE BRANDS [ADD BRANDS >](#)

- Ben Sherman
- Cole Haan

FAVORITE STORES [ADD STORES >](#)

You currently have no favorite stores.

**FAVORITES** RECENTLY VIEWED COLLECTIONS

- Cole Haan Air Colton Saddle Oxford**  
Cole Haan - Official Store  
\$498.00 \$99.95
- Joe's Jeans Straight Leg The Classic**  
Joe's Jeans Official Site  
\$138.00
- Joe's Jeans 'Brixton' Slim Straight...**  
Nordstrom  
\$189.00
- Joe's Jeans Brixton Straight Narrow...**  
Zappos.com  
\$124.00 \$104.99
- Dunk Low Premium**  
SHOEBACCA.com  
\$57.99
- Banksy Rat Vinyl Decal For Macbook**  
eBay  
\$6.50
- Banksy by Will Ellsworth-Jones**  
BookDepository.com  
\$16.27
- Wall And Piece By Banksy - Multi ...**  
Urban Outfitters  
\$22.95

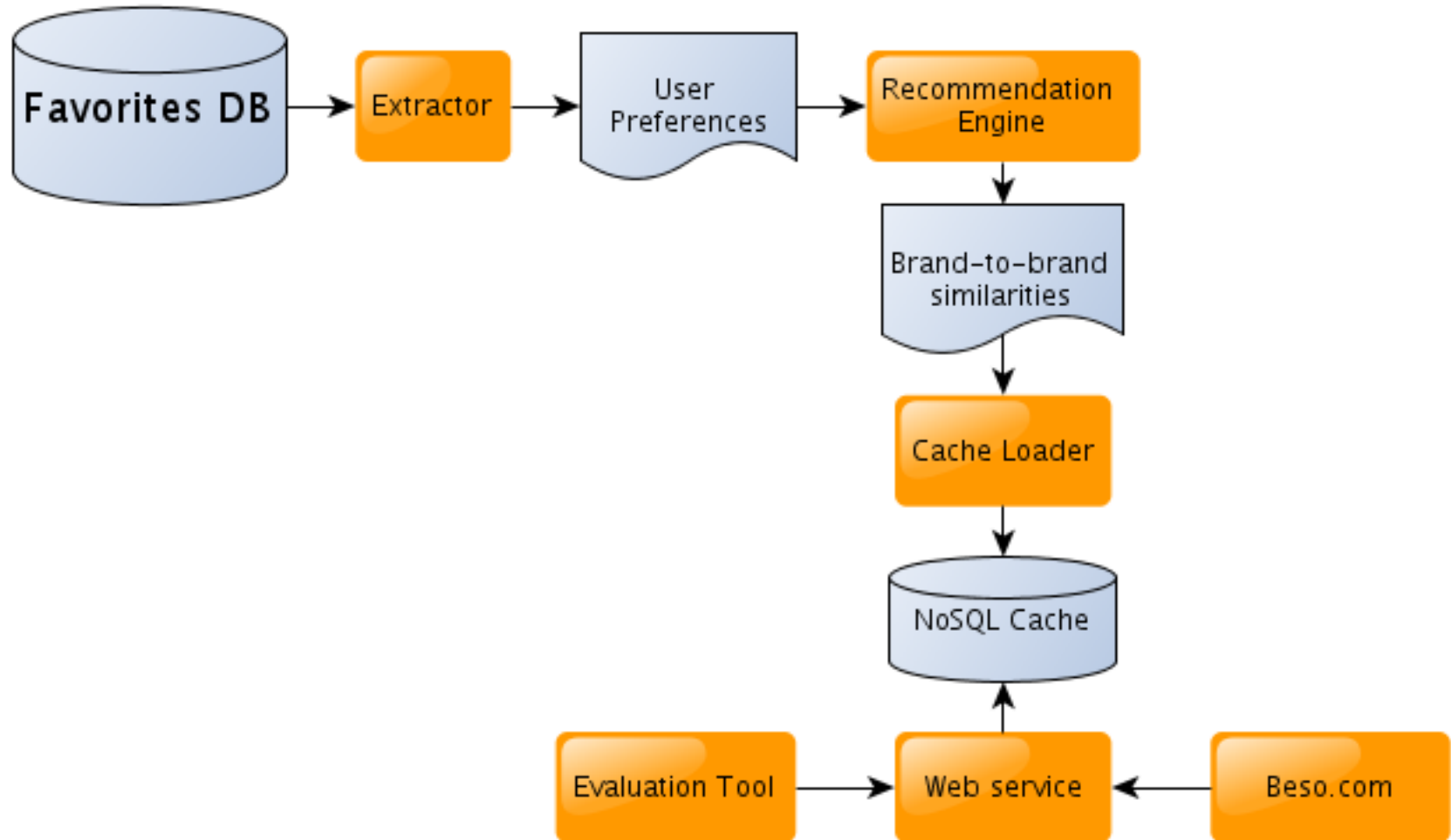
# What's missing?

- Recommendations
- Relevancy
- ...but we have data!

The screenshot shows the beso website's brand directory page. The header includes the beso logo with the tagline 'THE SITE DEVOTED TO SHOPPING', a search bar with the placeholder text 'What can we help you find?', and user account options 'My Beso' and 'Sign Out'. A navigation bar below the header lists categories: WOMEN, MEN, HOME, KIDS, JEWELRY, BEAUTY, SALE, EDITOR PICKS, COLLECTIONS, and BRANDS & STORES. The main content area is titled 'Home : Store and Brand Directory - Brands' and features a 'Brands' button and a 'Stores' button. A letter navigation bar shows 'A' through 'Y'. The 'A - Popular Brands' section displays a grid of brand names, each with a heart icon to its left, indicating a favorite or watchlist feature. The brands listed are:

A Peace Treaty	Alfred Dunhill	Anna Sul
abc home	Ali Ro	Annette
Abercrombie & Fitch	Alice + Olivia	Annick Goutal
Acca Kappa	All-Clad	Anthropologie
Acne	Almay	Apple
Acorn	Alno	Aquaphor
Acqua di Parma	Alora Ambiance	Aramis
Adesso	Alterna	Archipelago Botanicals
Adia Kibur	Alternative	Arcona
Adidas	Alternative Apparel	Arcosteel
Adirondack	Altesse	Ardell
Adrienne Landau	American crew	Arienne
Aero Bed	American Eagle Outfitters	Aroma
Aesop	American Essentials	Ash
Africa's Best	American Range	ASICS
African Pride	American Retro	ASOS
AG Adriano Goldschmied	American Standard	Astara
Agatha Ruiz de la Prada	Amrita Singh	Athleta
Agent Provocateur	Anastasia	Atlas
Ahava	Andis	Aubrey Organics
Akris	Andrea	Australian Gold
Alessi	Andrew Marc	Avalon Organics
Alexander McQueen	Angela Devine	Aveda
Alexander Wang	Ann Taylor	Aveeno

# Recommender: System Overview



# Extracting User Preferences

- Shopzilla Engineers built this component
- Important in tuning recommender algorithm



# Extracting User Preferences

--which brands did each user favorite?

```
SELECT user_id, brand_id  
FROM favorites
```

--which brands did they click on?

```
SELECT C.user_id, P.brand_id  
FROM clicks C, products P  
WHERE C.product_id = P.id
```



# Recommendations: Overview

- I like a brand. Which other brands do i like?
- Use latest available user data to answer this



# Recommender: What will you learn?

- Leverage **Machine Learning** at scale
- Use **Hadoop** on **BigData**
- How to evaluate solutions to data problems

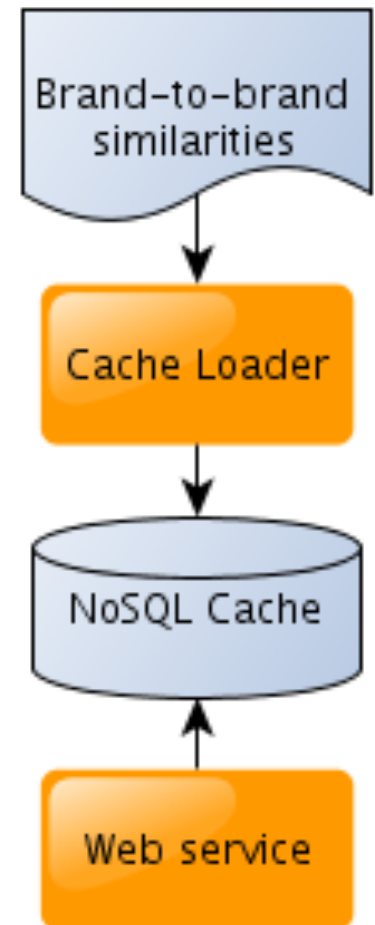
Ex. [Recommendations With Hadoop Streaming](#)

# Recommender: Why is it cool?

- Run recommender on cluster of 332 CPU's
- Patterns emerge from behavioral data
  - Nike <---> Adidas, Reebok
  - Prada <---> Gucci, Louis Vuitton
  - Levi <---> Wrangler, Lee'
- Data from 100's of millions of users
- Creative ways to represent user preferences

# Web Service: Overview

- Load recs into NoSQL cache
- Build web service for recs
- Build client for Beso.com



# Web Service: What will you learn?

- Build a high-performance web service
  - 100's of requests per second
- Learn enterprise Java best practices
- Understand Service-Oriented Architecture

# Web Service: Why is it cool?

- **Scaling**
  - how many servers do we need?
  - how fast does cache need to be?
  - how much storage do we need?
- **NoSQL Technology selection**
  - read/write performance
  - reliability
  - availability
  - consistency
- **See evolving data affect web layer**


# Evaluation Tool: Overview


- Build UI to compare algorithms side-by-side




Go!  WATMS max  Bid Max  PID/OID  Description  Atom  Match Type  Match Field  Match Score  Noun Weight  Context Weight  Offer Count  Price  Redirect Score


**Query Product Nouns: ipod**  
**Query Brands:**  
**Total Results:** 1098  
WATMS: 0.864  
WANMS: 0.617  
AB: 92.175  
WAB: 90.777  
NWAB: 92.266  
AMR: 8.725  
MD: 0.556  
ARDS: 0.672  
SQ-UDA: 0.533


  
Apple 32 GB iPod touch (5th Generation) - Yellow  
**Product Noun:** ipod touch  
**Relevancy:** 12496816159326208.00000  
**Leaf:** MP3 & Media Players

  
Apple iPod touch 32GB (4th Gen), White  
**Product Noun:** ipod touch  
**Relevancy:** 7471025281302528.00000  
**Leaf:** MP3 & Media Players

  
Apple 32 GB iPod touch (5th Generation) - White/Silver  
**Product Noun:** ipod touch

**Query Product Nouns: canon**  
**Query Brands:**  
**Total Results:** 7647  
WATMS: 1.000  
WANMS: 0.010  
AB: 128.950  
WAB: 97.583  
NWAB: 128.950  
AMR: 7.284  
MD: 0.475  
ARDS: 0.647  
SQ-UDA: 0.010  
SQ-MDA: 0.602  
SQ-DDA: 113.266  
SQ-SDWA: 0.366  
SQ-SDMA: 0.663

  
Canon Powershot SX40 HS 12.1 Megapixel Digital Camera  
**Product Noun:** digital camera  
**Relevancy:** 32882164760576.00000  
**Leaf:** Digital Cameras  
**Bid:** \$1.35  
**Merchant:** 42nd Street Photo **Rating:** 8.50

  
Canon EOS Rebel T3i (600D) Digital Camera with 18-55mm IS II Kit  
**Product Noun:** digital slr camera  
**Relevancy:** 17695430934528.00000  
**Leaf:** Digital Cameras  
**Bid:** \$1.35  
**Merchant:** Primotronic

# Evaluation Tool: What will you learn?

- Build enterprise Java webapp from scratch
- Use frontend technologies (CSS/JS/AJAX)
- Consuming web services using REST client



# Evaluation Tool: Why is it cool?

- Like CS144, with modern Java webapp tools
- Visualizing data problems
- Inform recommendation tuning process

# Logistics

- Pre-configured VM to get you started
- Knowledgeable project mentors
  - available for questions / help daily
- Agile Tracking system
- Github / Continuous Integration build

# Schedule

- Week 1: bootcamp
- Week 2: technology selection
- Week 3: interface definition
- Week 4: mock implementation
- Week 5: component buildout
- Week 6: component integration
- Week 7: hardening, performance
- Week 8: deployment

# Big Picture

- Technology selection & platform rollout
- Mirrors process used by Shopzilla's Senior Architects and Engineers
- Program designed to teach enterprise engineering techniques
- Equips you with skills for your first job

# Questions?

alook [at] shopzilla [dot] com  
@andrewlook