

# Data Infrastructure with Amazon Web Services

Le Nguyen The Dat

<https://linkedin.com/in/lenquyenthedat>

<https://github.com/lenquyenthedat>

# Outlines

- Background
- Company profiles & challenges
- Overview & steps: data infrastructure
- Applications & Demo
- Key takeaways, Q&A.

# ZALORA Group

- Biggest online fashion retail in Southeast Asia
- By 12 May 2013:
  - 1 million orders
  - 17.9 millions visitors monthly



# ZALORA Group: Challenges

- Huge amount of data:
  - 20+ of different data sources
  - 10s of TB of data processed daily
  - 1000s of analytical queries daily

# Commercialize TV



▶ LUXURY

▶ FASHION

▶ ACTION & ADVENTURE

- Global digital content distribution, creative, and management company.
- Operates across multi-channel, multi-platform – YouTube, DailyMotion, Baidu Video, TenCent Video, Pandora TV, and so on...

# Commercialize TV: Challenges

- Platform dependent: 10s of different 3<sup>rd</sup> party data sources.
- Cost & Scalability: ability to scale up 100x with minimal effort.

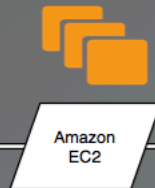
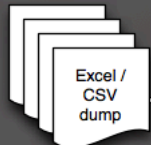
# Team & Technology Stack

- Small team of 1-4 programmers
- Amazon Web Services:
  - No upfront cost
  - Low maintenance
  - Scalability
  - Integrations
- Shell scripts, Python, Haskell, D3.js
- Unix, open-source technologies



# Let's build it!

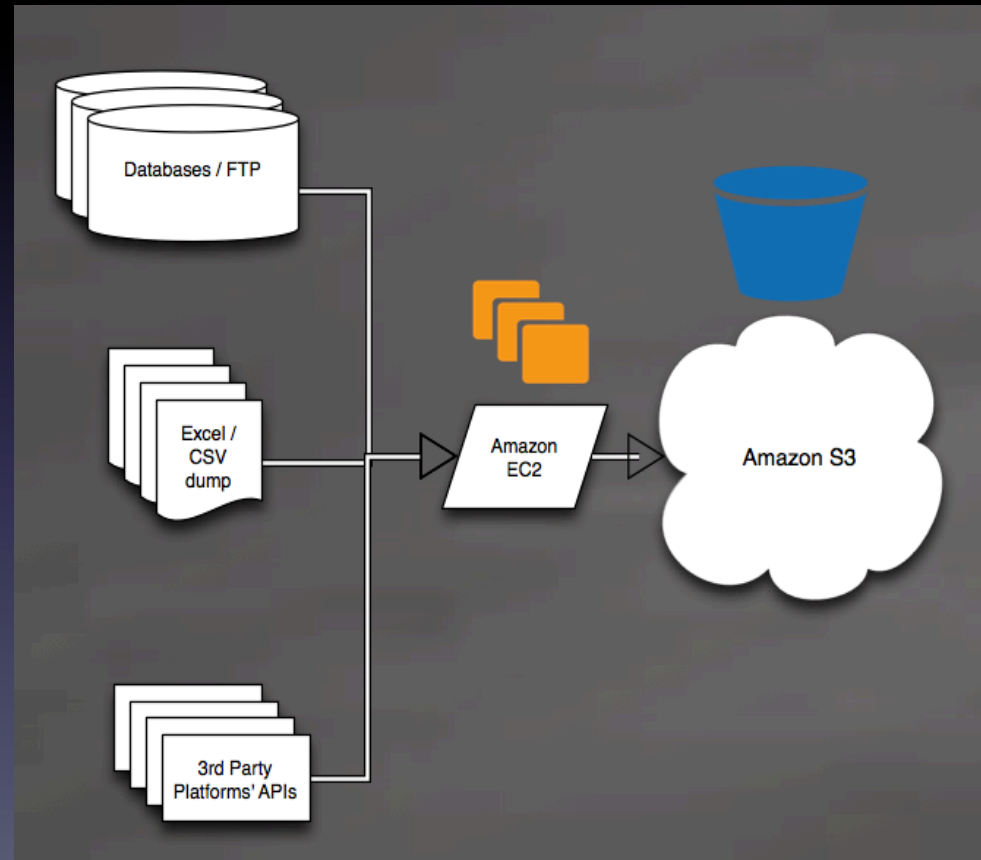
## Infrastructure Overview





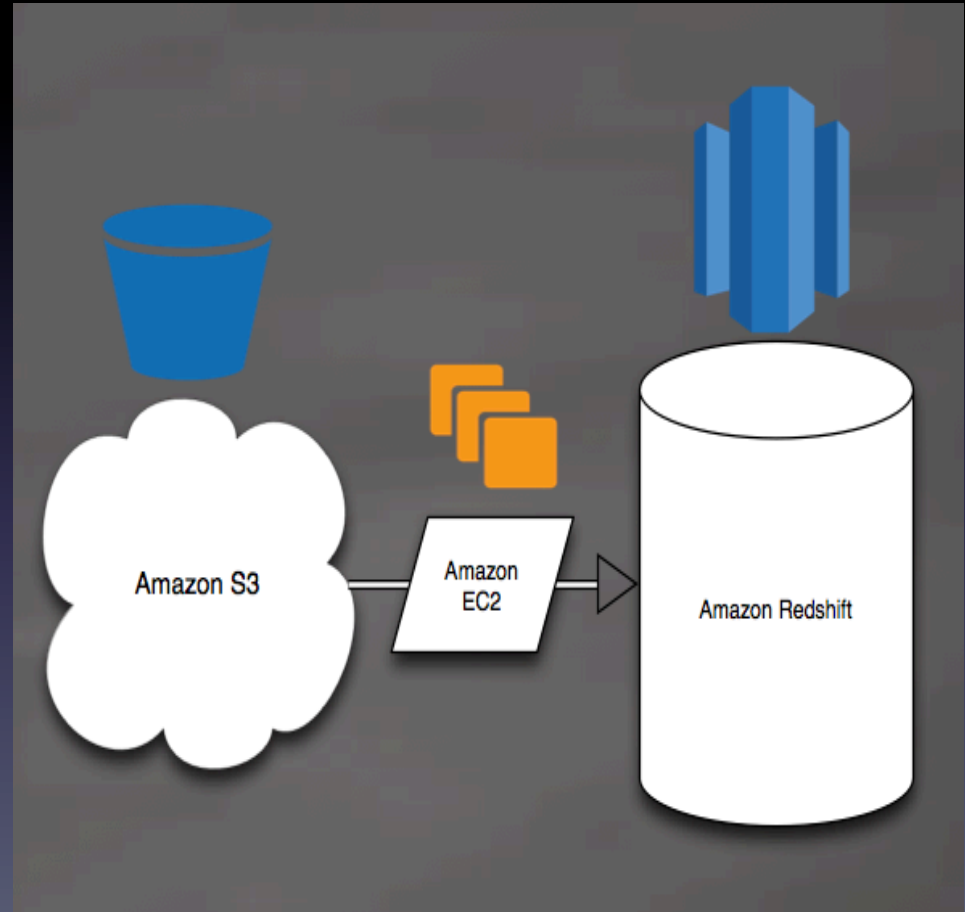
# Step 1: Data Collections

- Amazon S3:
  - Simple to use
  - Scalability & Speed
  - High availability
- Amazon EC2:
  - Programmatic data collectors



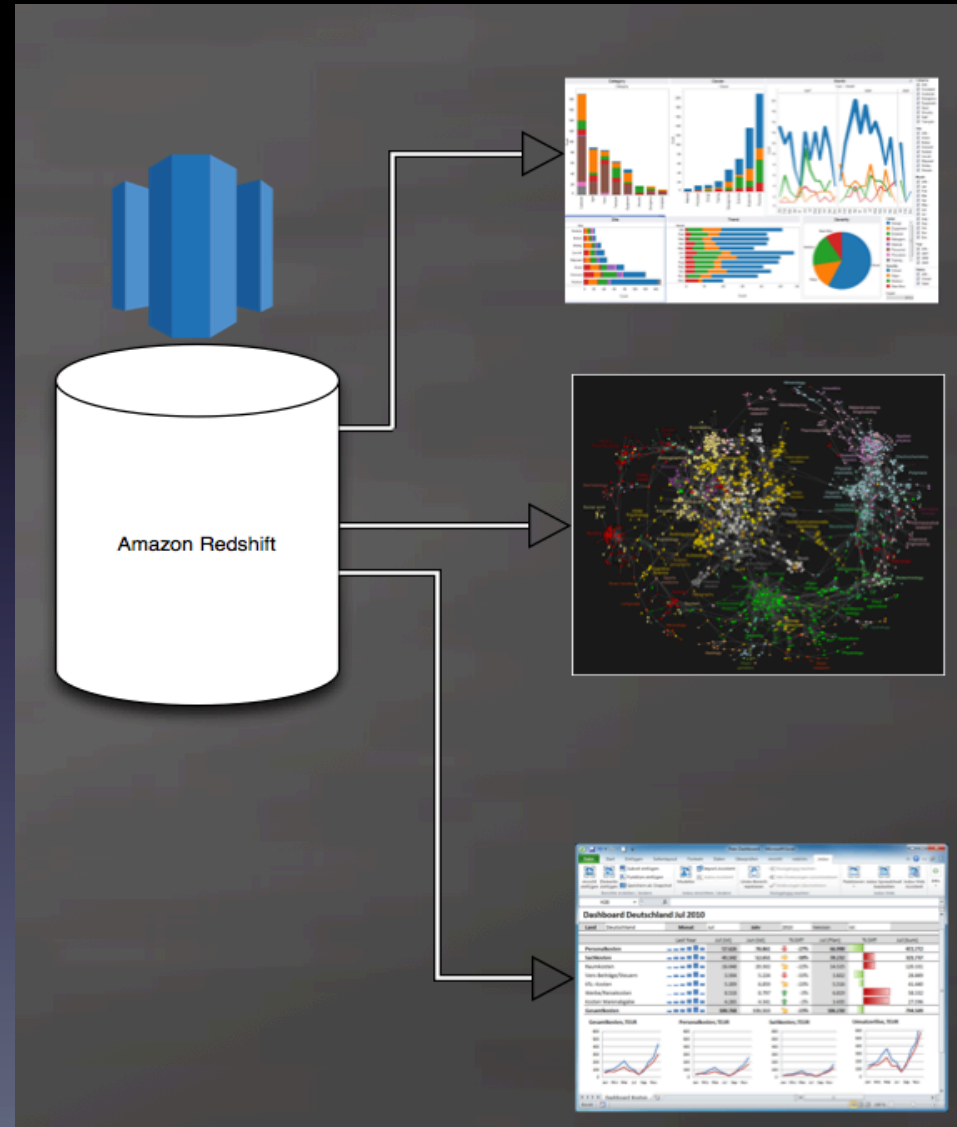
# Step 2: ETL

- Amazon Redshift:
  - Petabyte-scale Data Warehouse
  - Relational data (Pre-transformation needed)
  - [COPY command](#)



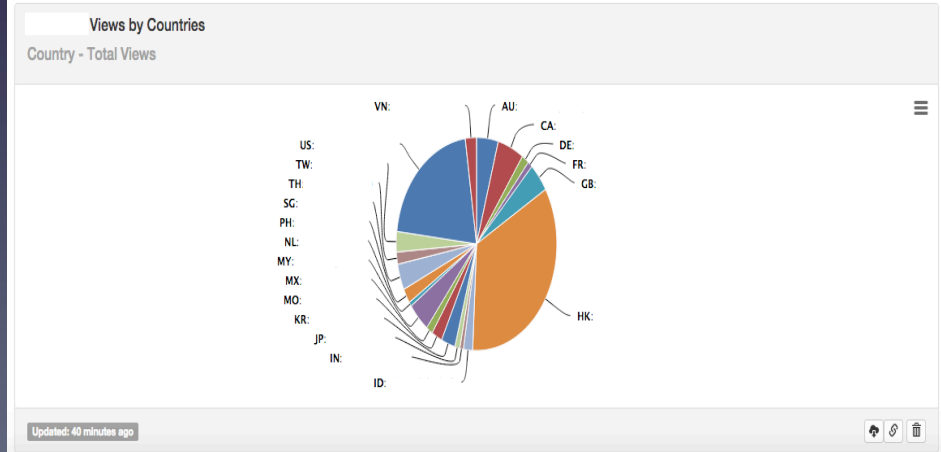
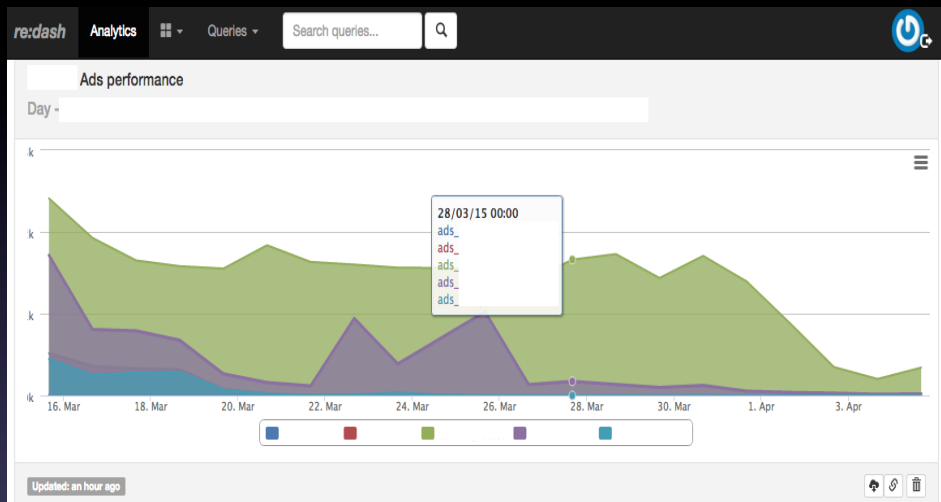
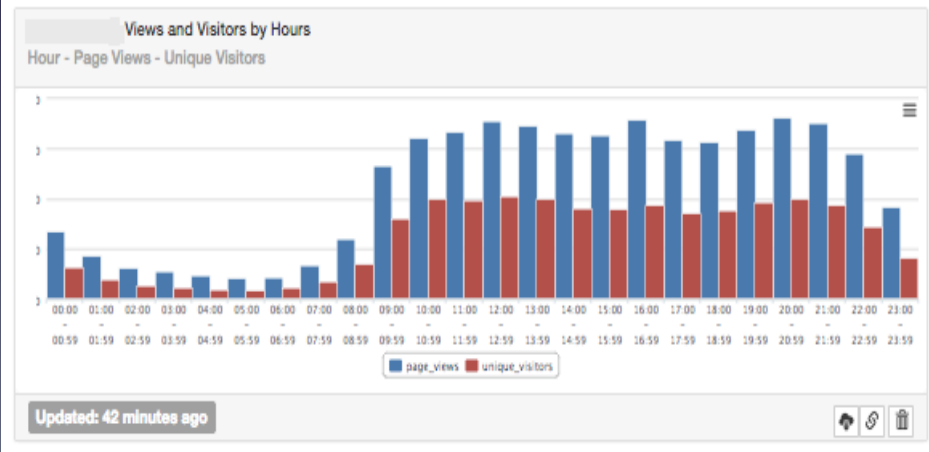
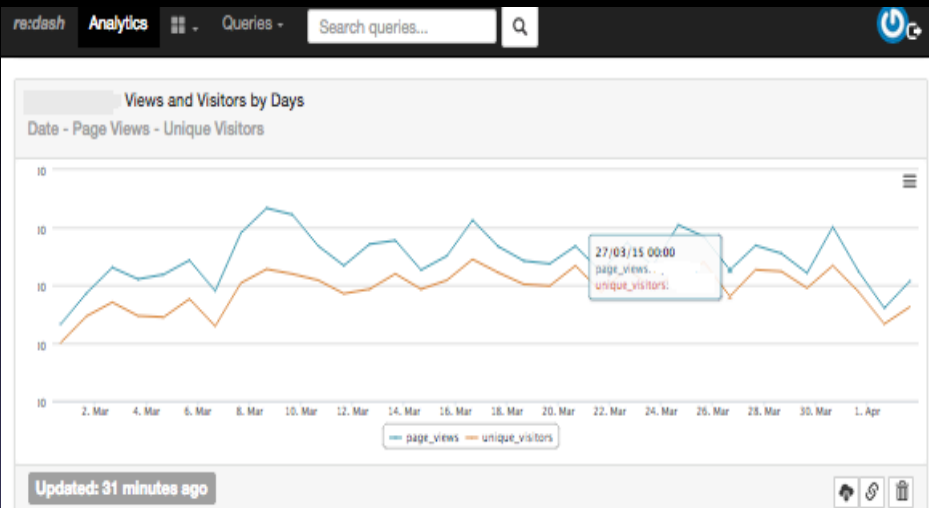
# Step 3: Visualization

- PostgreSQL interface
- Amazon Redshift's partners:
  - Re:dash (Open Source)
  - Tableau (14days Free Trial)



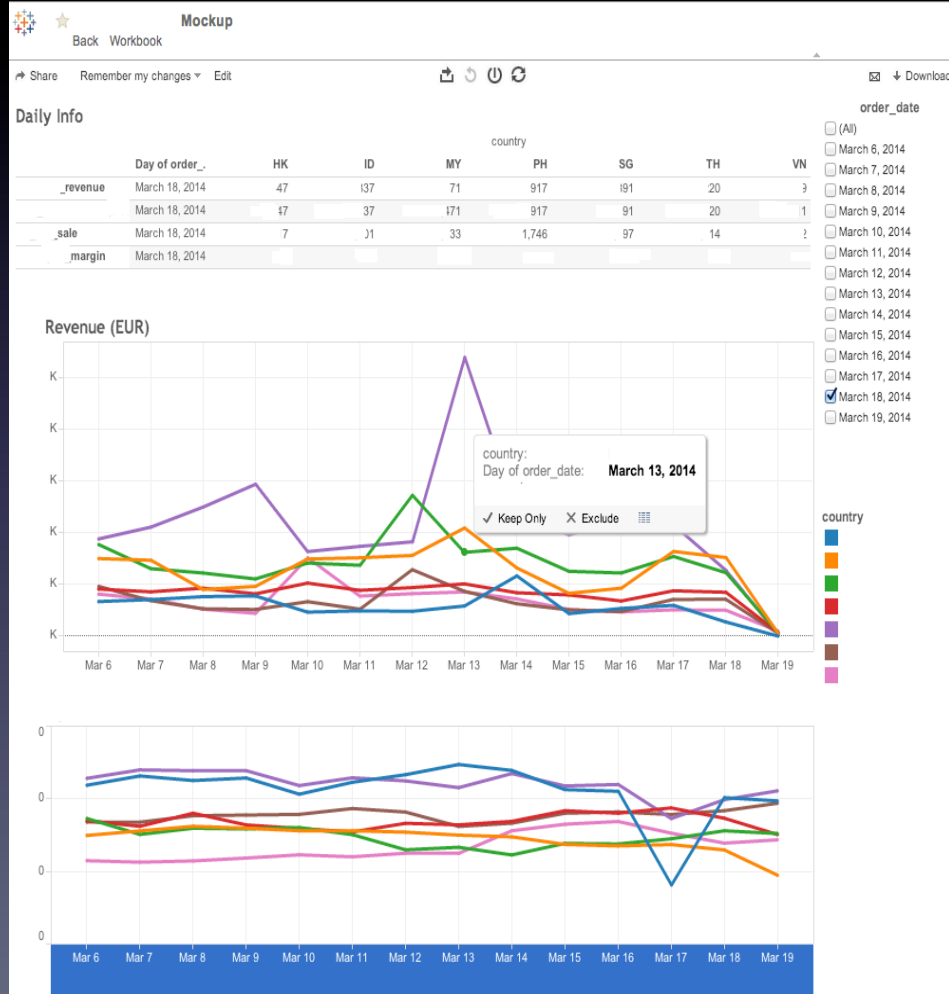
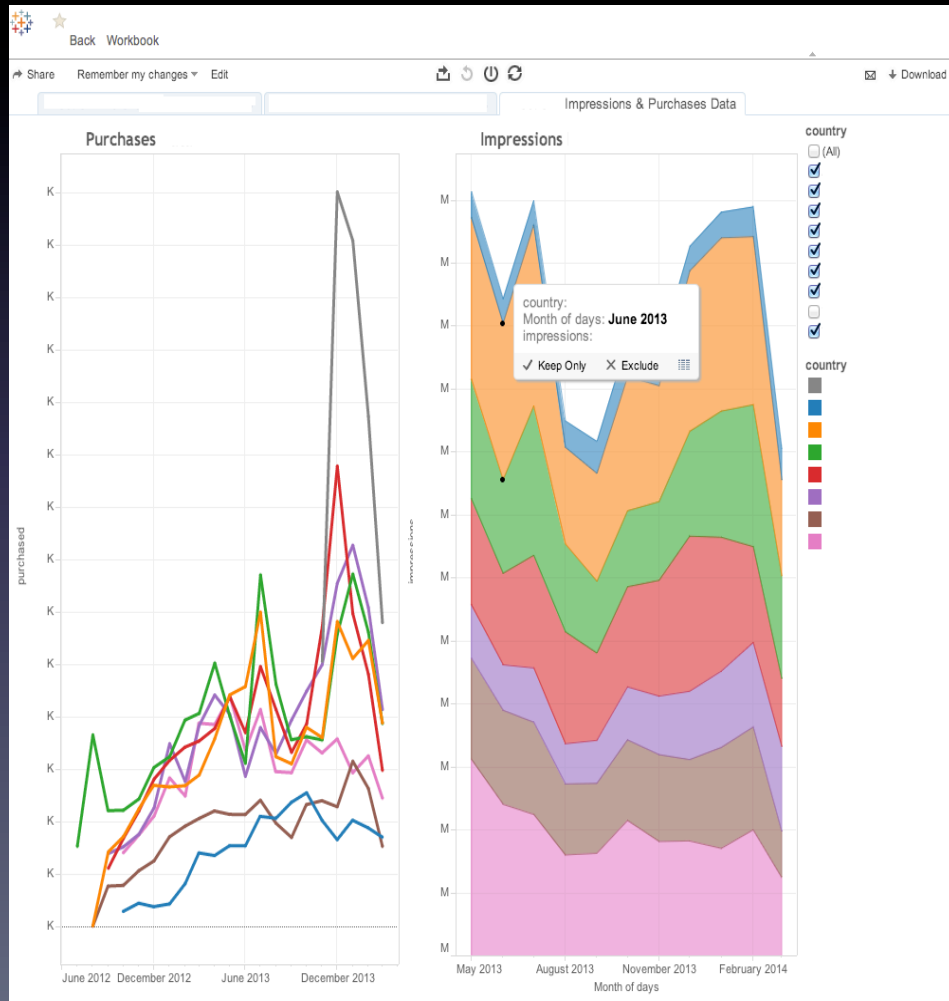
# Application & Demo

- Re:dash



# Application & Demo

- Tableau



# Application & Demo

- Machine Learning & Data Science

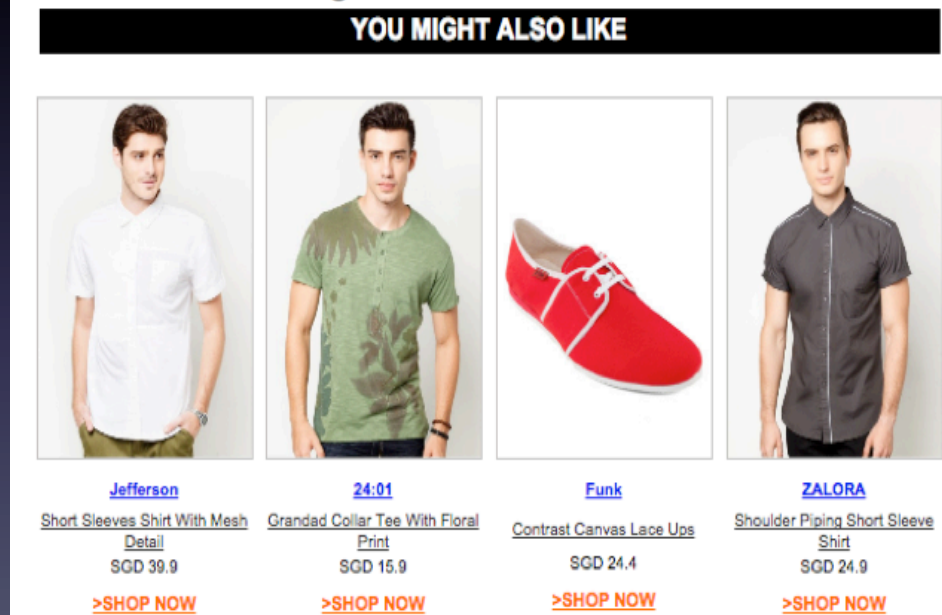
## Recommendation Engine - Onsite



This screenshot shows a product page for a black cardigan. Below the main product image, five recommended items are displayed in a row. Each item includes a model wearing the product, the brand name, the item name, and the price in VND.

Brand	Item Name	Price (VND)
Casways	Áo Len Nữ	379,000 VND
Mango	Áo Khoác Cardigan...	999,000 VND
Casways	Áo Len Nữ	399,000 VND
Mango	Cotton-Blend Swea...	449,000 VND

## Recommendation Engine - CRM



This screenshot shows a CRM recommendation engine interface. A black banner at the top reads "YOU MIGHT ALSO LIKE". Below this, four product recommendations are shown in a row. Each recommendation includes a product image, a brand name, a product name, a price in SGD, and a ">SHOP NOW" link.

Brand	Item Name	Price (SGD)	Action
Jefferson	Short Sleeves Shirt With Mesh Detail	SGD 39.9	>SHOP NOW
24:01	Grandad Collar Tee With Floral Print	SGD 15.9	>SHOP NOW
Funk	Contrast Canvas Lace Ups	SGD 24.4	>SHOP NOW
ZALORA	Shoulder Piping Short Sleeve Shirt	SGD 24.9	>SHOP NOW

# Key Takeaways

- Invest in your programmers
- Understand data technology in-depth
- If you have not: try Amazon Web Services - it's cheap and easy to do so!