

# Architectures for massive data management

MapReduce Session Lab

Albert Bifet

[albert.bifet@telecom-paristech.fr](mailto:albert.bifet@telecom-paristech.fr)

université  
PARIS-SACLAY



September 22, 2015

# Recommendation Engines

- Websites as Amazon help customers to find items that could interest them
- For example: “Customers who bought this item also bought”

Customers Who Bought This Item Also Bought



A screenshot of the Amazon website showing a row of five book recommendations under the heading "Customers Who Bought This Item Also Bought". Each recommendation includes a book cover, title, author, star rating, and price. A left-pointing arrow is visible on the far left.

Book Title	Author	Rating	Price
Predictive Analytics For Dummies	Anasse Bari	29	\$17.72 ✓Prime
Predictive Analytics: The Power to Predict Who...	Eric Siegel	229	#1 Best Seller in Econometrics \$16.88 ✓Prime
Quantifying the User Experience: Practical...	Jeff Sauro	8	\$40.63 ✓Prime
Marketing Analytics: Strategic Models and...	Stephan Sorger	29	\$50.52 ✓Prime
Data Driven Marketing For Dummies	David Semmelroth		\$20.49 ✓Prime

# “Customers who bought this item also bought”

## Customers Who Bought This Item Also Bought



The screenshot shows five recommended books:

- Predictive Analytics For Dummies** by Anasse Bari. 4.5 stars, 29 reviews. Paperback: \$17.72. Prime.
- Predictive Analytics: The Power to Predict Who...** by Eric Siegel. #1 Best Seller in Econometrics. Hardcover: \$16.88. Prime.
- Quantifying the User Experience: Practical...** by Jeff Sauro. 4.5 stars, 8 reviews. Paperback: \$40.63. Prime.
- Marketing Analytics: Strategic Models and...** by Stephan Sorger. 4.5 stars, 29 reviews. Paperback: \$50.52. Prime.
- Data Driven Marketing For Dummies** by David Semmelroth. Paperback: \$20.49. Prime.

- Input
  - Set of items in transactions:  
book12, book34, cd12, cd42, dvd32  
book32, book34, dvd32
- Output
  - key-value pairs in which the key is the item and the value is the list of the items most commonly bought by customers who also bought this item

# MapReduce Session Lab

- Install Hadoop
  - <http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/SingleCluster.html>
- Follow the MapReduce Tutorial
  - <http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>
- Write a Hadoop MapReduce application for “Customers who bought this item also bought”