Spidey ¢ents

Insight Data Engineering, Boston
Dylan Russell

Online Retail Insight
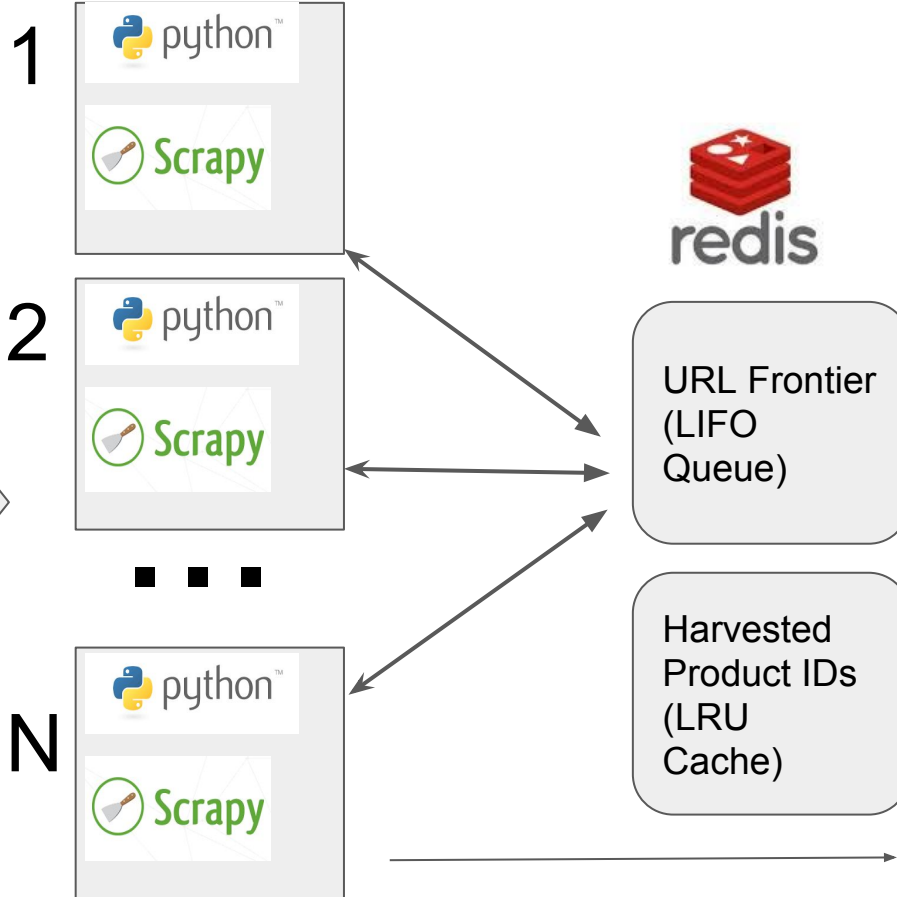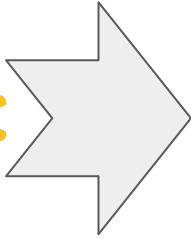
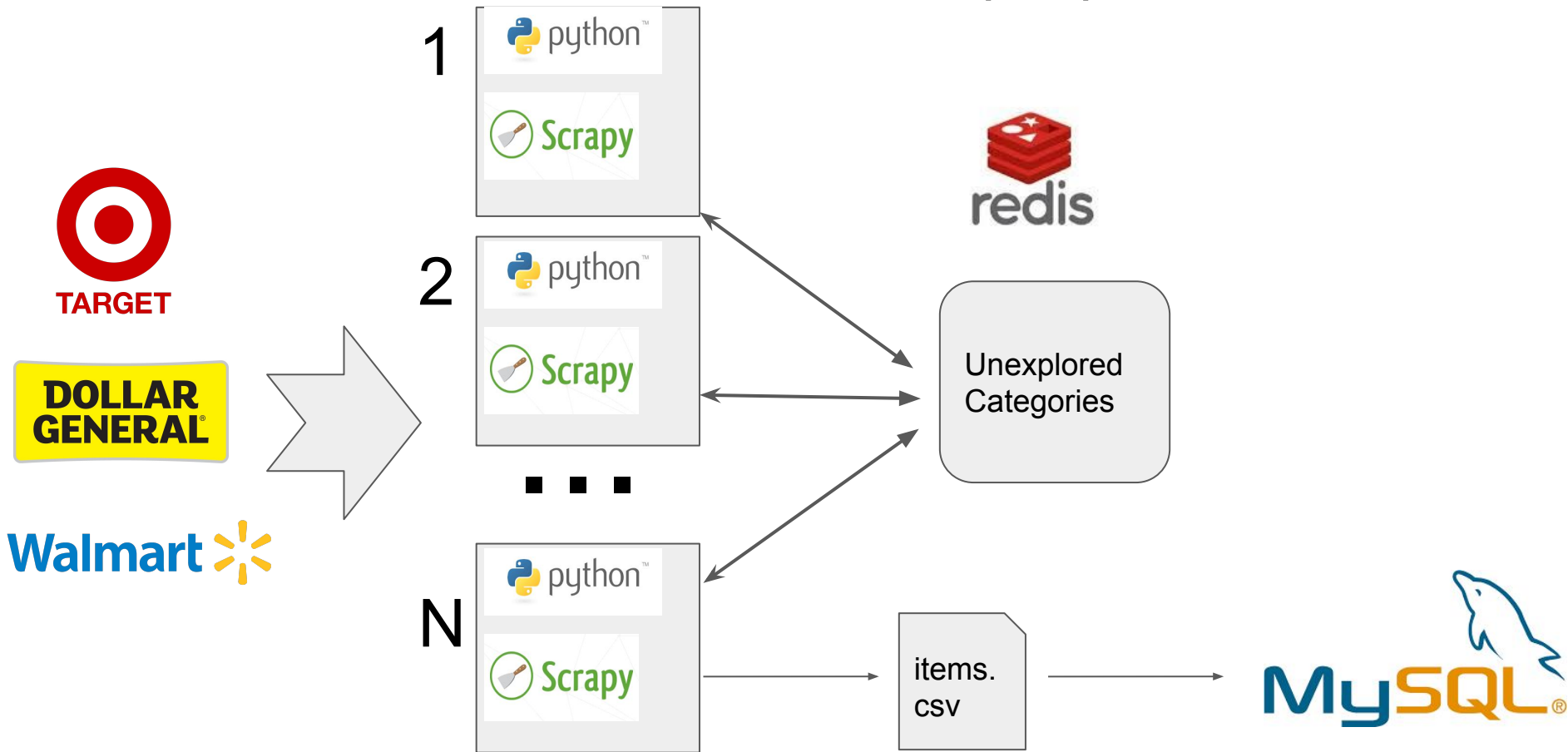# Product Demo

https://www.youtube.com/watch?v=yj1G3NrjB4A&feature=youtu.be

Web Crawler

Web Crawler (v2)

Crawling Strategy

Crawling Strategy

Crawl Statistics

	Dollar General	**Target**	**Walmart**
Product Catalogue Size	~12,000	~1,000,000	~34,000,000
# Items per Crawl	~12,000	~280,000	~16,000,000
# Requests per Cawl	~800	~7,000	~1,500,000
Length of Crawl	~5 minutes	~20 minutes	~6 hours * 6 nodes

Dylan Russell

Soccer

Basketball

Programming

Memorization