

# Walmart eCommerce Search System Architecture: intro and evolution

Ning Cao

WalmartLabs Engineering Manager



# CNUTCon 2017

## 全球运维技术大会

上海·光大会展中心大酒店 | 2017.9.10-11

智能时代的新运维

大数据运维  
安全  
SRE  
DevOps  
Kubernetes  
Serverless  
游戏运维  
AI Ops  
智能化运维  
基础架构  
监控  
互联网金融



主办方

Geekbang > InfoQ

极客邦科技



实践驱动的IT教育



<http://www.stuq.org>

**斯达克学院(StuQ)**，极客邦旗下实践驱动的IT教育平台。通过线下和线上多种形式的综合学习解决方案，帮助IT从业者和研发团队提升技能水平。



10大职业技术领域课程

# SPEAKER INTRODUCE

---

曹宁

WalmartLabs Engineering Manager

- Ning Cao is an engineering manager in search runtime team at WalmartLabs. Prior to that, he worked at Google, Huawei.
- Ning received his Ph.D. in Electrical and Computer Engineering at Worcester Polytechnic Institute. His publications have 4000+ citations.

SPEAKER  
ArchSummit 2017 ShenZhen

# TABLE OF CONTENTS 大纲

---

- Walmart eCommerce Search
- Search Architecture Evolution
- Experience & Lessons

# Walmart eCommerce Search

- Search
- Browse
- Category Pages

The screenshot shows the Walmart eCommerce search results page. The header includes the Walmart logo, navigation links (Home, Search, Gift Cards, Registry & Lists, Weekly Ads, Store Finder, Track Orders, Reorder Items, Credit Card, Help, Grocery Pickup), and a user greeting (Hello, Ning My Account). The main content area features several promotional banners and a 'Featured Products' section.

**Rollbacks Low prices in home just got lower.**

**Explore your style with new arrivals**

**Make your patio perfect with new furniture**

**Entertaining essentials that take the party outside**

**Featured Products**

Product	Price	Shipping
Ninja CF081 Coffee Bar Brewer with Glass...	\$149.99	2-Day Shipping
KRUPS, Deluxe Convection Toaster...	\$119.00	2-Day Shipping
KRUPS, 4-Slice Belgian Waffle Maker With...	\$49.88	2-Day Shipping
Dyson DC34 Bagless Cordless Hand Vacuum	\$129.00	2-Day Shipping
Bissell Big Green Deep Cleaning Machine...	\$396.99	
Eureka Quick-UP Bagle Stick Vacuum with...	\$27.83	2-Day Shipping

# Walmart eCommerce Search

- Performance Challenges in Search Backend
  - Increasing index size
    - ~8x in past 3 years
- Real time update

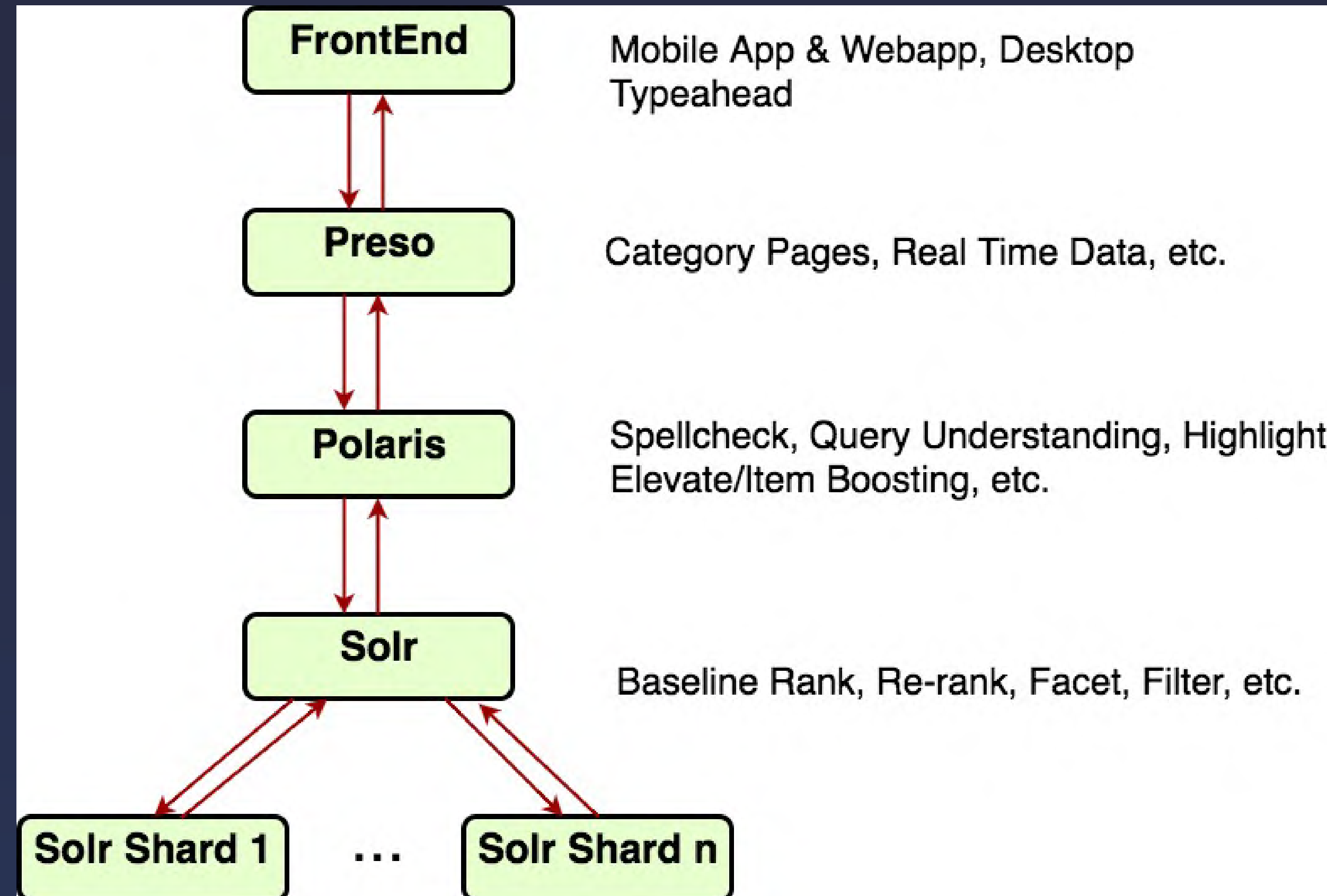
# TABLE OF CONTENTS 大纲

---

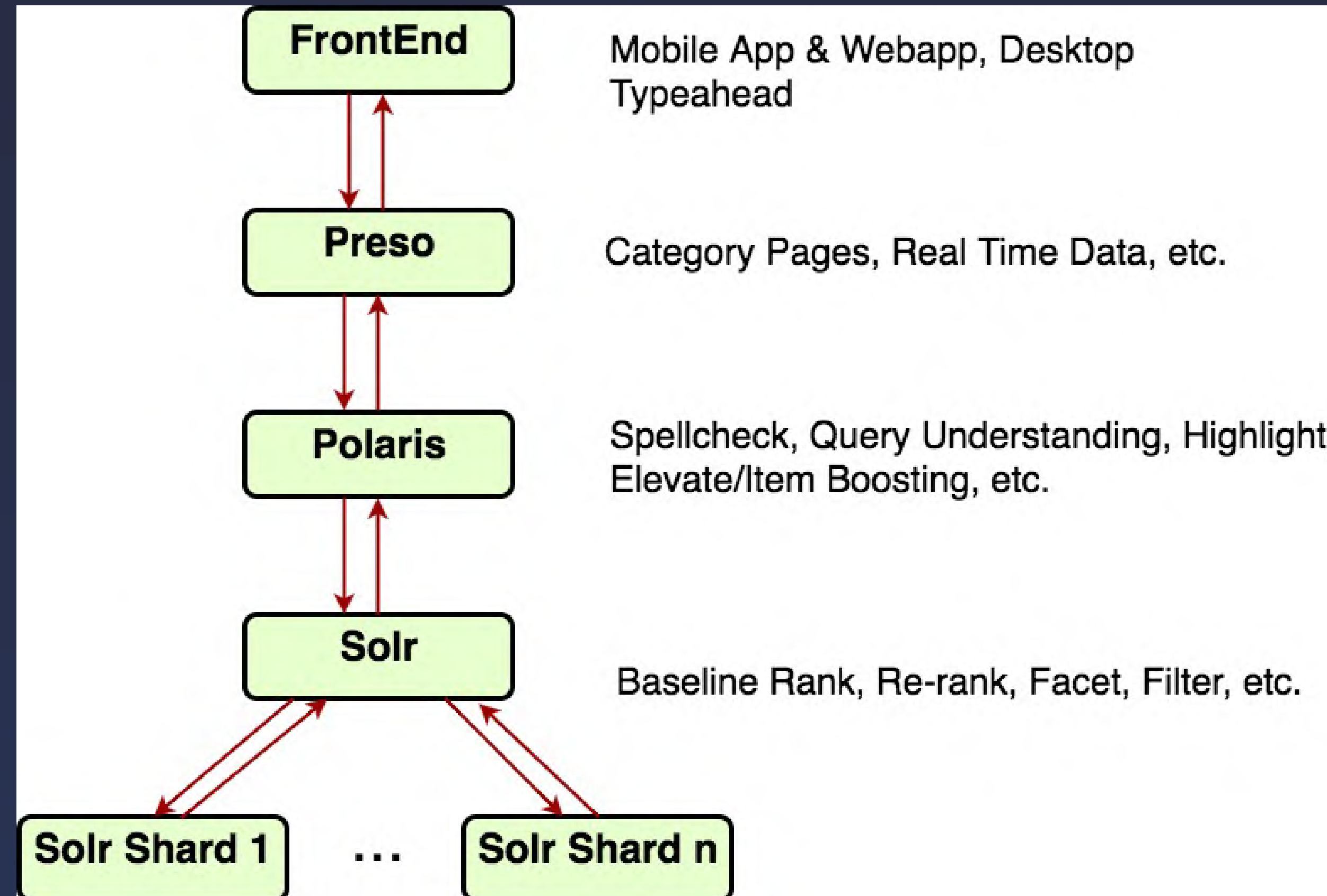
- Walmart eCommerce Search
- Search Architecture Evolution
  - Architecture Overview
  - Distributed Search Cloud
  - Re-rank Migration
  - Metadata Store
- Experience & Lessons



# Search Runtime Architecture Overview



# Search Runtime Architecture Overview



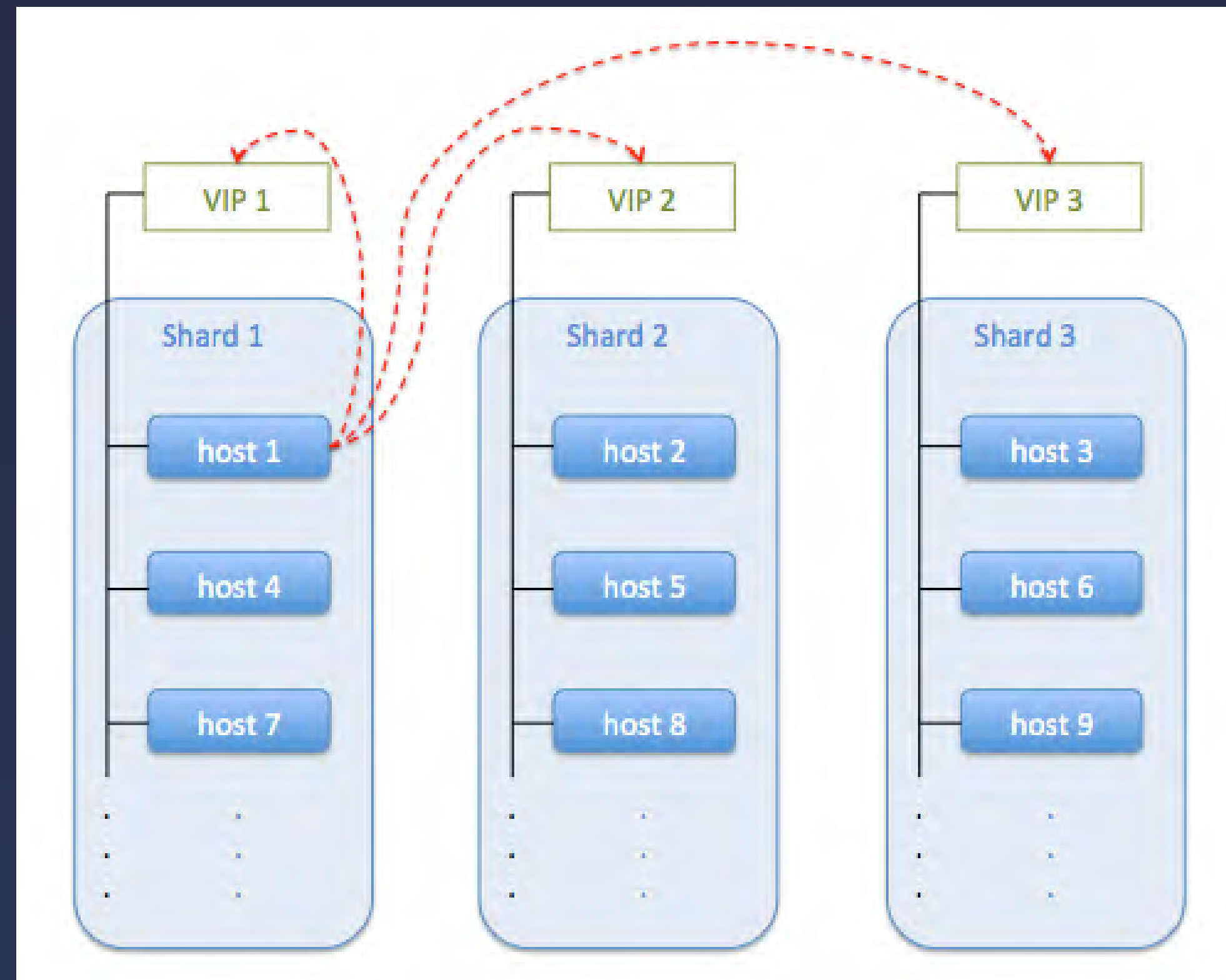
# TABLE OF CONTENTS 大纲

---

- Walmart eCommerce Search
- Search Architecture Evolution
  - Architecture Overview
  - Distributed Search Cloud
  - Re-rank Migration
  - Metadata Store
- Experience & Lessons

# Distributed Search

- Load-balanced Shard VIP



# Distributed Search

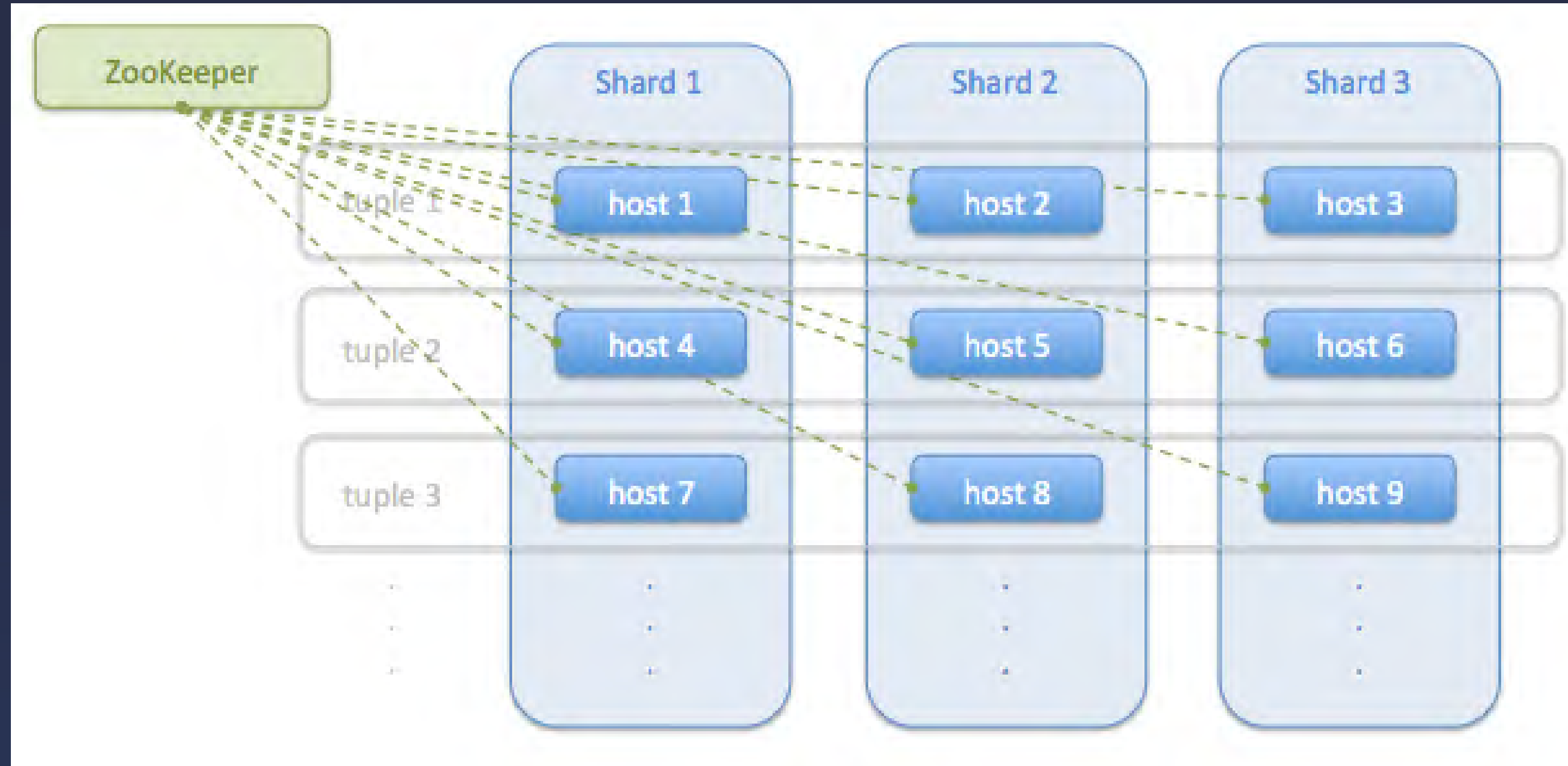
- Problems with Load-balanced Shard VIP
  - Performance bottleneck
  - Hard to troubleshoot
  - Unnecessary re-routing
  - Increasing open connections between VIP and Solr shard

# Why Not SolrCloud

- Problems with SolrCloud
  - Unable to utilize offline SolrCloud for index update
  - Inefficient indexing: # of shards
  - Must build two sets of SolrCloud to index all shards at same time

# Distributed Search Cloud

- Tuple-based Polaris Cloud



# Distributed Search Cloud

- Tuple-based Polaris Cloud
  - Fallback
  - Last-known-state use
  - Cached sharding data from ZooKeeper
  - Sharding data from live update



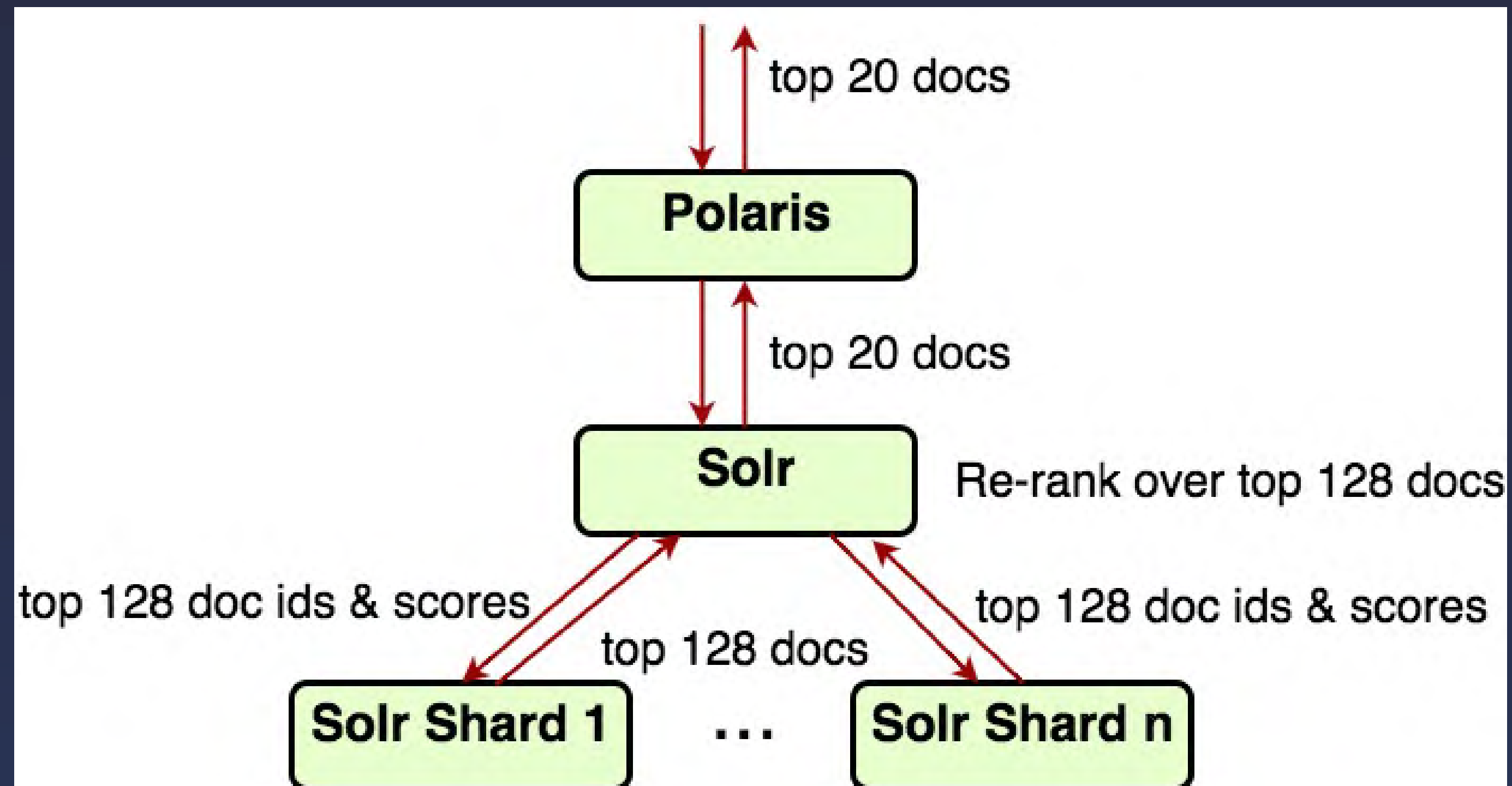
# TABLE OF CONTENTS 大纲

---

- Walmart eCommerce Search
- Search Architecture Evolution
  - Architecture Overview
  - Distributed Search Cloud
  - Re-rank Migration
  - Metadata Store
- Experience & Lessons

# Re-rank Migration

- Re-rank plugin in Solr

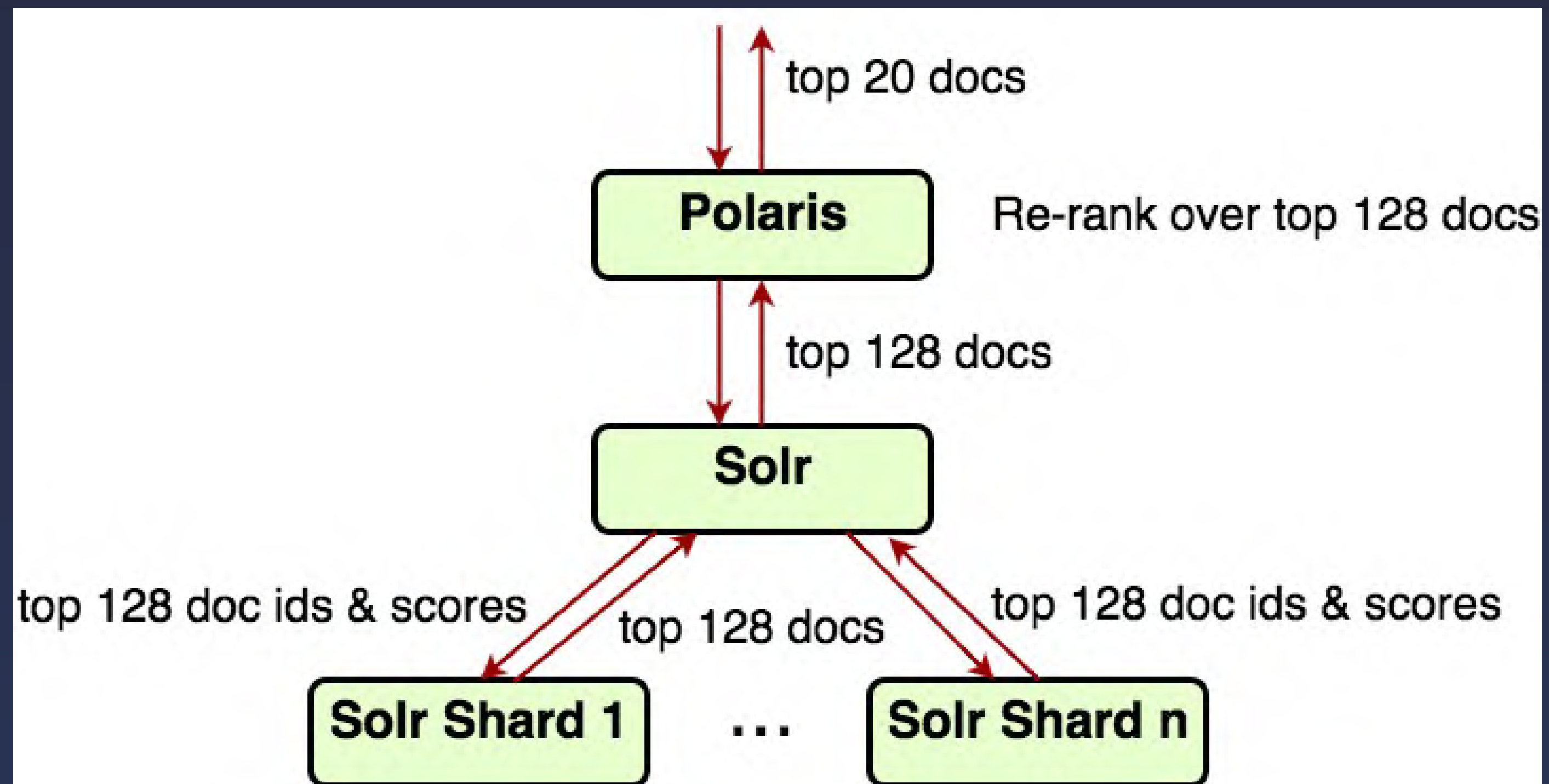


# Re-rank Migration

- Re-rank plugin in Solr
  - Implemented before Solr Re-rank
  - Need code change during Solr update
  - Unable to evaluate/migrate to other search engines

# Re-rank Migration

- Re-rank in Polaris



# Re-rank Migration

- Re-rank plugin in Polaris
  - Pros:
    - Solr update
    - Search engine migration
  - Cons:
    - Network overload between Polaris and Solr
    - Serialization & Deserialization

# TABLE OF CONTENTS 大纲

---

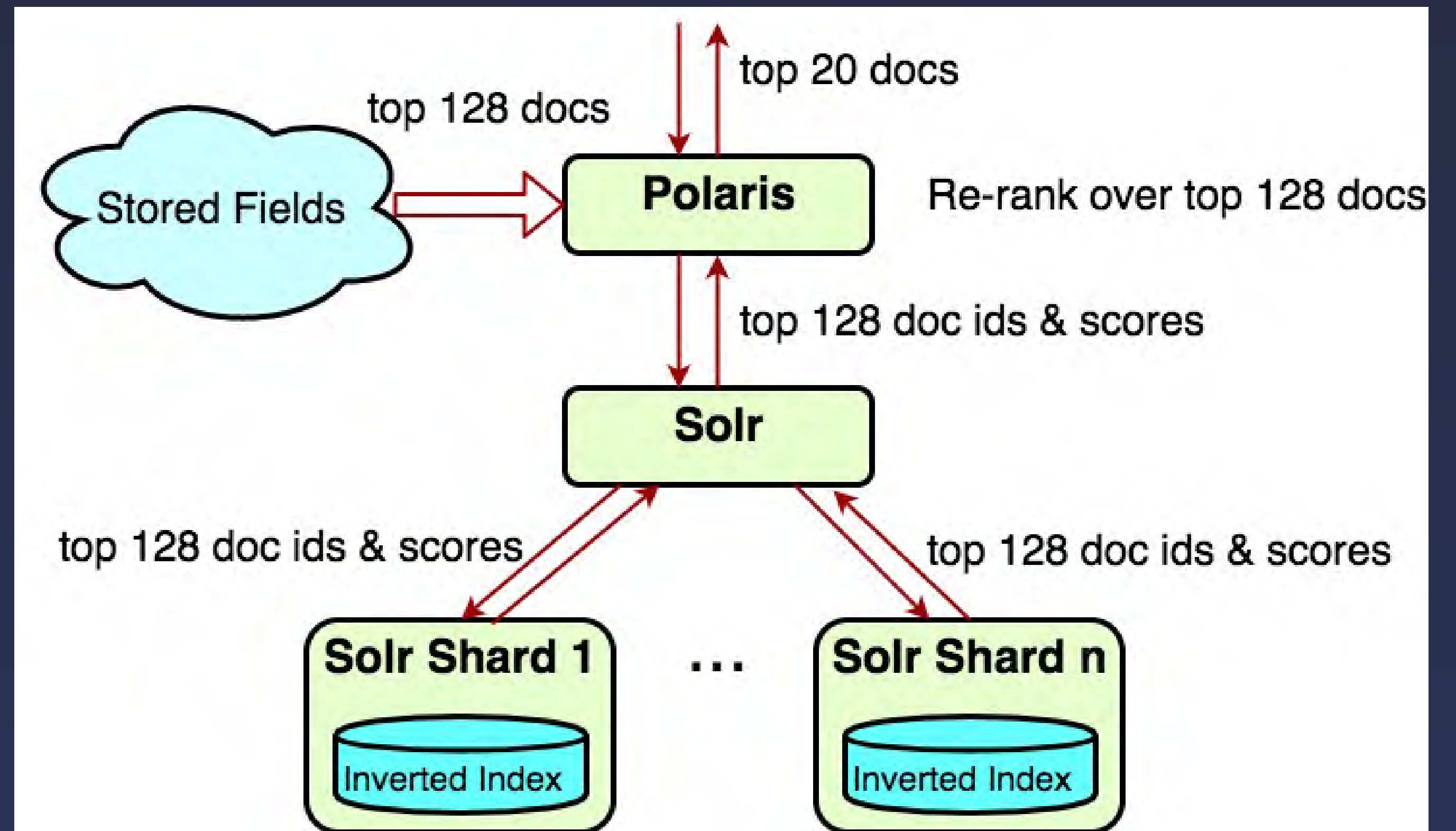
- Walmart eCommerce Search
- Search Architecture Evolution
  - Architecture Overview
  - Distributed Search Cloud
  - Re-rank Migration
  - Metadata Store
- Experience & Lessons

# Metadata Store

- Backend challenges caused by large Solr index
  - Search performance
  - Full index generation time
  - Index replication overhead
  - Real time update throughput

# Metadata Store

- How to reduce Solr index size?
- Move stored fields out of Solr



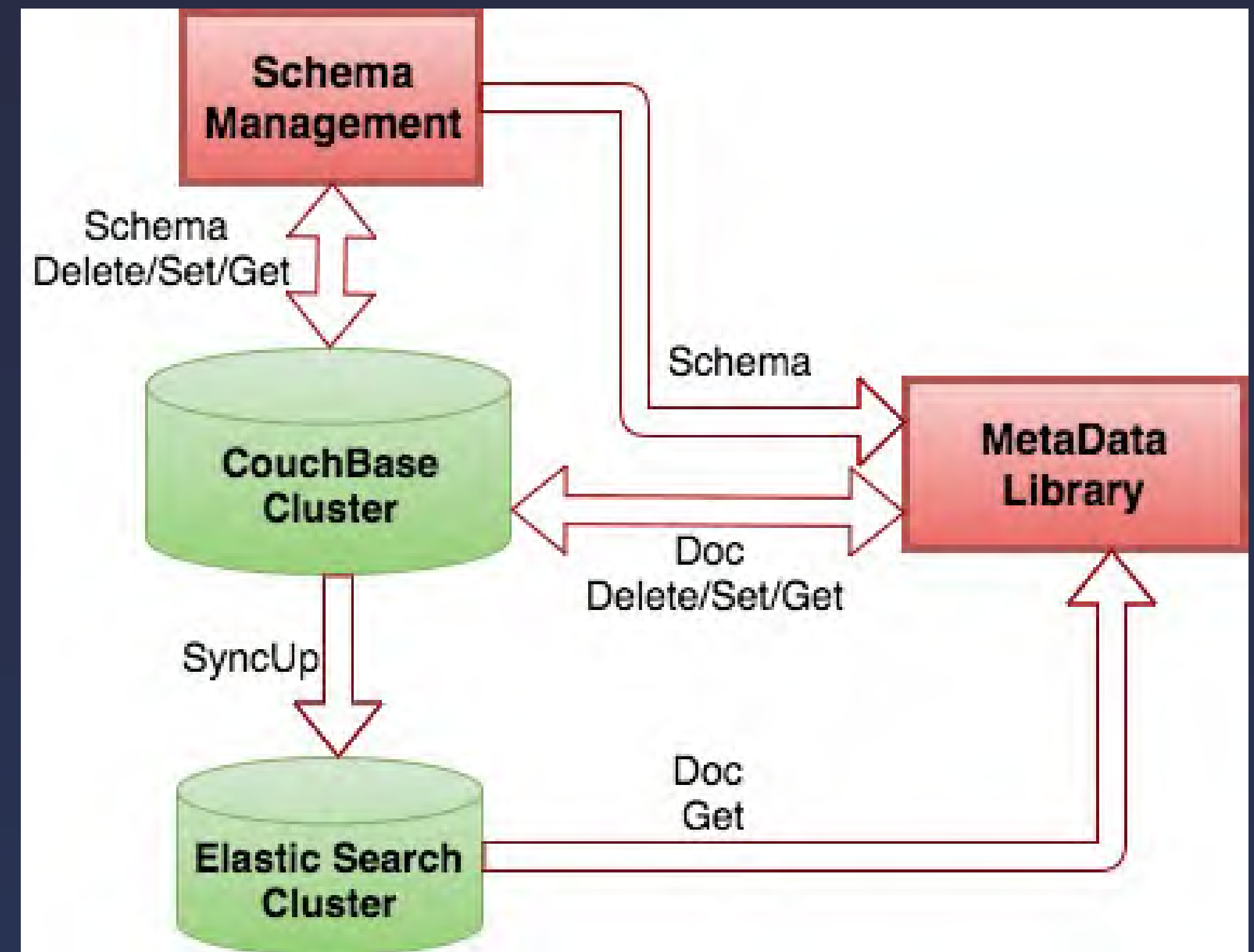


# Metadata Store

- Design Goal: store fields for re-rank and response
  - Scalability: easy to scale
  - Performance: fast data retrieval, high read and write throughput
  - Functionality: support structured queries

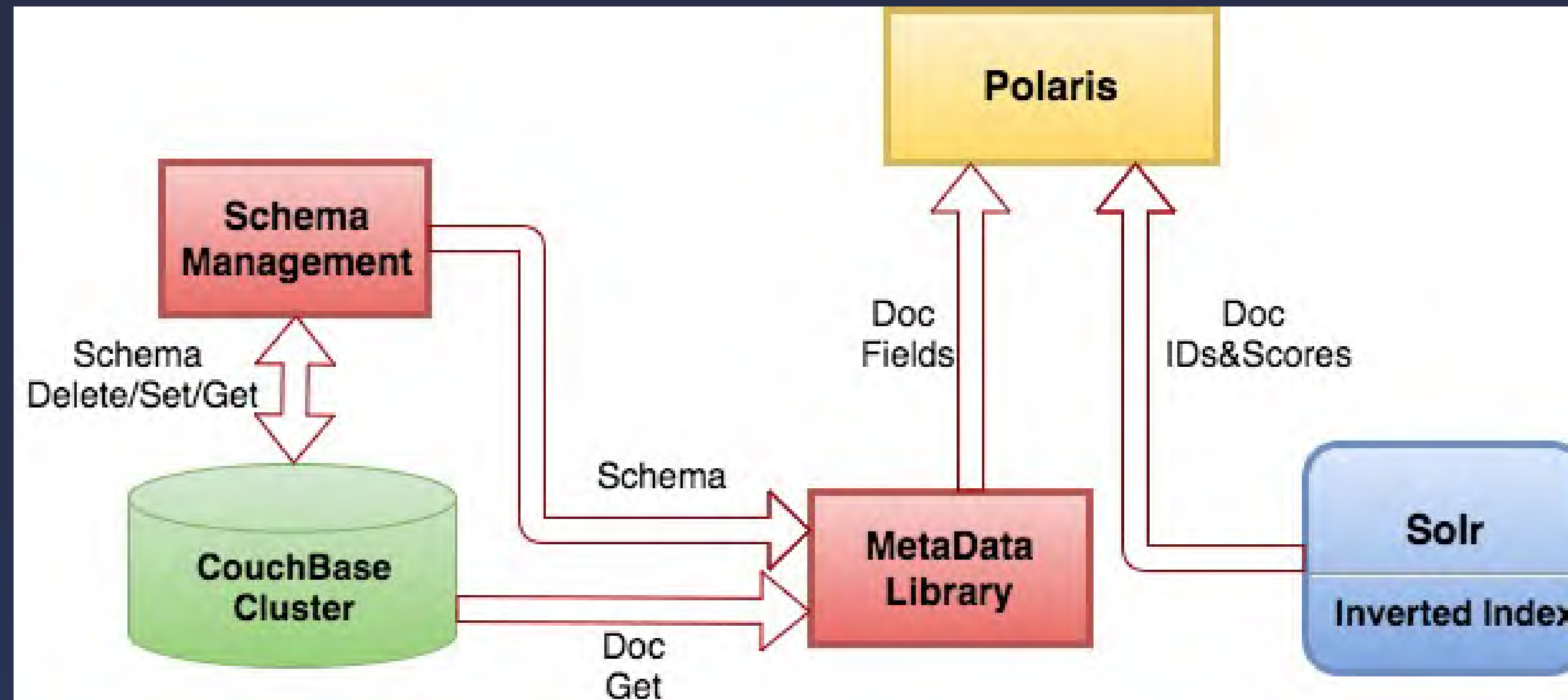
# Metadata Store

- Metadata Store Design
  - Primary data store: Couchbase
  - Secondary data store: Elastic Search
  - Data format: Avro



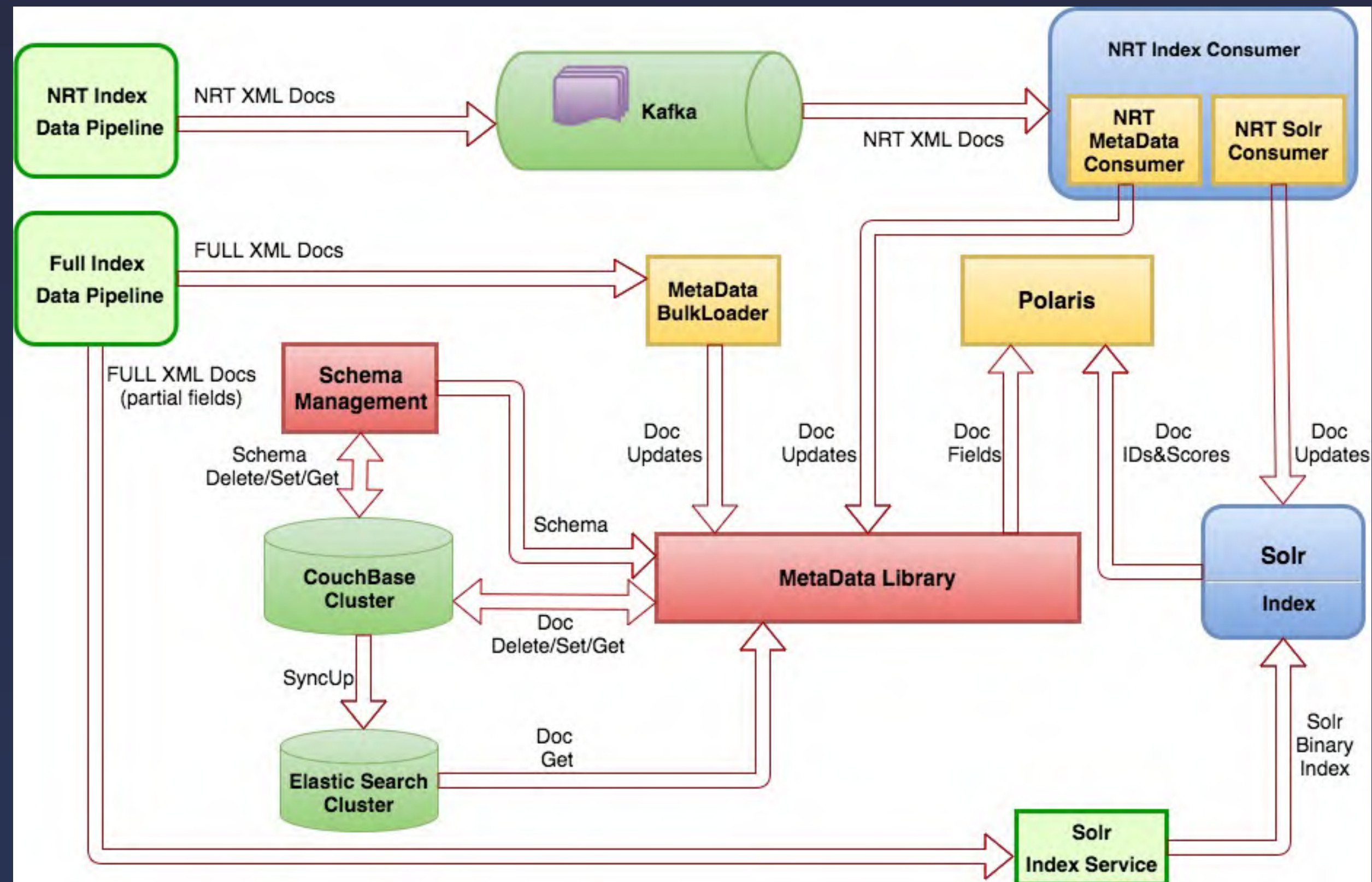
# Metadata Store

- Search with metadata store



# Metadata Store

- Update with metadata service



# Metadata Store

- Gains:
  - Search latency: 10%
  - Index size: 25%
  - Real time update
  - Larger re-rank size

# TABLE OF CONTENTS 大纲

---

- Walmart eCommerce Search
- Search Architecture Evolution
  - Architecture Overview
  - Distributed Search Cloud
  - Re-rank Migration
  - Metadata Store
- Experience & Lessons

# Experience & Lessons

- How to utilize open source software
  - Adopt
  - Customize
  - Replace/Self-dev

# Experience & Lessons

- Microservices
  - Different tech stacks
    - Operation overhead, Performance tuning
- Microservices framework
  - Rate control, perf monitor, config management, authentication, logging, etc.



# THANKS

让创新技术推动社会进步

HELP TO BUILD A BETTER SOCIETY WITH  
INNOVATIVE TECHNOLOGIES

# Geekbang >

## 极客邦科技

**InfoQ**  
ueue

专注中高端技术人员的技术媒体



**EGO** EXTRA GEEKS' ORGANIZATION  
NETWORKS

高端技术人员学习型社交平台



**StuQ**  
ueue

斯达克学院

实践驱动的 IT 教育平台

