



Solr Power FTW

Alex Pinkin
@apinkin
#solrnosql

Presented by



Produced by



What Will I Cover?

- Who I am
- What Bazaarvoice does
- SOLR and NoSQL
- Can SOLR handle 20K queries per second?
- Lessons learned: large scale multi data center deployment
- Conclusion

Alex Pinkin

- Software Engineering Lead,
Data Infrastructure team,
Bazaarvoice
- Loves to play with SQL and NoSQL



Bazaarvoice

- Bazaarvoice is a software as a service company powering user generated content such as ratings and reviews on thousands of web sites
- 5 billion page views per month
- 230 billion impressions
- 75 million UGC



NoSQL ?

HOW TO WRITE A CV



Leverage the NoSQL boom

SQL vs NoSQL



ORACLE®

PostgreSQL



Cassandra



NoSQL is Not Only SQL

- Departs from relational model
- No fixed schema
- No joins
- Eventual consistency is OK
- Scale horizontally

Types of NoSQL

- Key-value (Redis, Riak, Voldemort)
- Document (MongoDB, CouchDB)
- Graph (Neo4J, FlockDB)
- Column family (Cassandra, HBase)

SOLR as NoSQL

- Non-relational model - Check
- No fixed schema - Check (dynamic fields)
- No joins - Check (denormalization)
- Horizontal scaling - Check (with work)

SOLR stats - Bazaarvoice

SOLR Case Study

RATINGS & REVIEWS ***PRODUCT Q&A**

HOME BROWSE SEARCH

Search: GO

Your search returned 5 results

View All

Q Has anyone experienced an allergic reaction?	8 answers
abmskers (Seattle, WA)	
Q Are Dior mascaras Allergy Tested like Clinique mascaras?	2 answers
Anonymous (Allen, Texas)	
Q Formula Change?	2 answers
CEAB8 (Woodbridge, VA)	
Q How far does the 0.35oz go? (does it dry out fast, mess up the mascara wand). Also, does it contain Phthalates?	2 answers
Anonymous	
Q does this formula stay on well thru water or tears?	47 answers
purp	

1 of 1

ASK A NEW QUESTION

THE HOME DEPOT More saving. More doing.™

TRACK ALL OF YOUR HRTIC ELIGIBLE PURCHASES. REGISTER

PROVINCIAL ECO REBATES & INCENTIVES. LEARN MORE

Building Supplies Indoor Living Outdoor Living Tools & Hardware Know-How

SEARCH **YOU ARE HERE: Home > Camp > Outdoor Living > BBQ > Propane BBQ > Weber BBQ > Propane Barbecue - S620**

PRINT PAGE TELL A FRIEND

WebberSECURE MEMBER SINCE 2007

Back to Propane BBQs

MORE INFO

YOU MAY ALSO LIKE:

Webber Weber Q Grill

20091 Weber Spirit Propane Barbecue - S210

Exo-Mix Exo-Mix Triple Bar Grill

ORDERING ASSISTANCE?

Call us **1-800-628-0525**

Monday - Saturday 8am - 11pm EST

Sunday 10am - 10pm EST

Webber Weber Summit Propane Barbecue - S620

\$2399.00

QUANTITY:

Add to Cart

Check Inventory

Free Returns

Price: \$2499.00 to \$2399.00

AVERAGE CUSTOMER RATING

★★★★☆ 42 OUT OF 5

READ REVIEWS | **WRITE THE PRODUCT**

QUICK RATE ☆☆☆☆☆

Download Instructions

OVERVIEW SPECIFICATIONS WARRANTY REVIEWS REGISTER PRODUCT

PRODUCT REVIEWS **Product Title** **Rating: High to Low**

★★★★★ Overall Rating - 5 users

Spinning a Weber, May 29, 2008

By **klb188** from North Vancouver, BC (read all my reviews)

"The amazing grilling space, the nifty rotisserie, the efficient heat distribution all add up to one mean machine! This barbecue has performed well for us. From grilling shrimp and kabobs to a baron of beef on the rotisserie, it's been easy. We just recently hosted a backyard party for 25 - so that meant a ton of juggling timing at the grill. Quick cooking times and lots of space to grill steaks, chicken, hot dogs and burgers - all at the same time. The side burner is a bonus and is very useful too. We don't expect to buy another barbecue until years to come - the brand, Weber, is synonymous with quality."

Was this review helpful to you? Report an inappropriate review

Case #s: Product:

★★★★☆ Overall Rating - 4 users

tsuwant 1-4th, April 4, 2009

By **tsuwant** from GTX, Ontario (read all my reviews)

"Holds heat well even in direct wind. With 6 burners you can control slow cooking or when required searing for steaks. Ignition system works well but in the very cold you have to be deliberate with the pre-ignition process. Found it easy to clean with the stainless catch tray. LED lights do work in the dark. Very pleased and would buy again. Note the rotisserie attachment is specific to the Summit line."

How would you rate your home improvement expertise? Professional

Was this review helpful to you? Report an inappropriate review

Case #s: Product:

SOLR Case Study

▼ RATINGS & REVIEWS

▶ PRODUCT Q&A

Average Rating



▶ [WRITE A REVIEW](#)

- Volumizing** (5,801) **Lengthening** (5,458) **Defining** (4,265)
- Curling (101) Clumpy (60) Amazing (57) Thickening (44) Dramatic (31)
- Long Lasting (26) Flaky (25) Fabulous (24) Sexy (24) Non-Clumping (23)
- Awesome (22) Perfect (22)

Check the boxes above to filter reviews

Filters

Eye Color Skin Tone Age

[See All](#)

Sort by

laa35
beautyinsider

location:
Atlanta, GA

eye color:
green

skin tone:
olive

age:
25-34

***** **Must have!** -11.07.11

Quick Take: lengthening, volumizing

I use this as my base coat and then I add Dior Blackout over the top. People ask me all of the time if my eyelashes are real!

Was this review helpful? [Yes](#) [No](#)

SHARE WITH FRIENDS

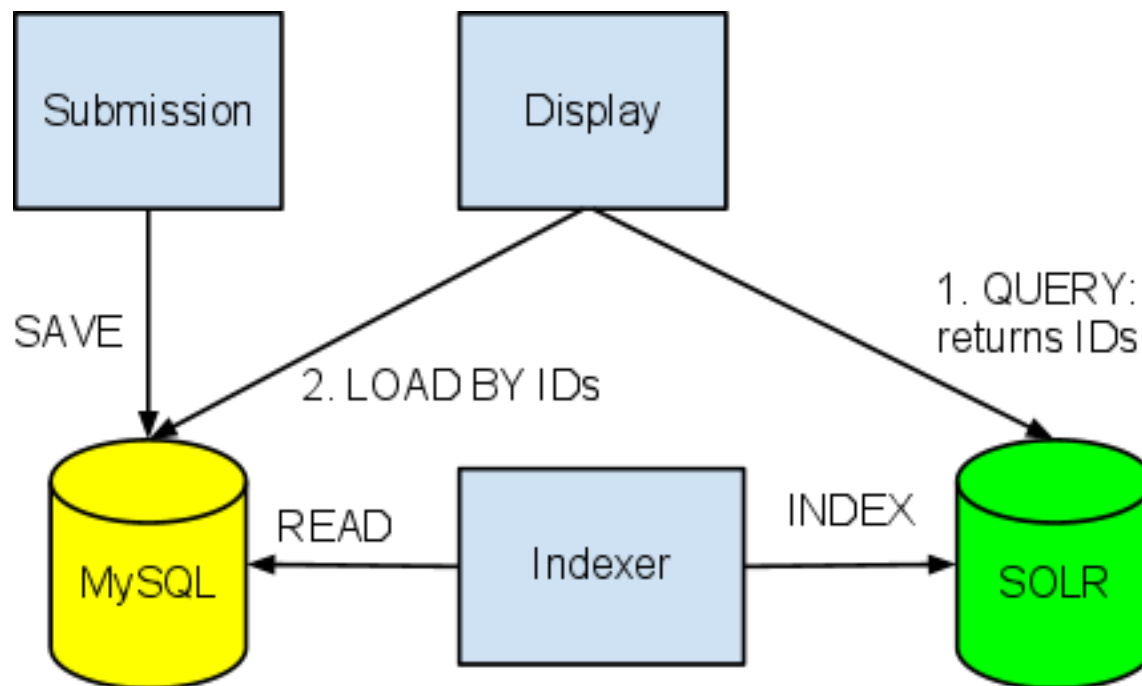
Life Before SOLR

- Indexes for sorting and filtering
- Aggregate tables for stats
- Nightly jobs
- Bugs...



Enter SOLR

- Index content and product catalog
- De-normalization
- Filtering and sorting
- Index every 15 minutes (20 seconds NRT)

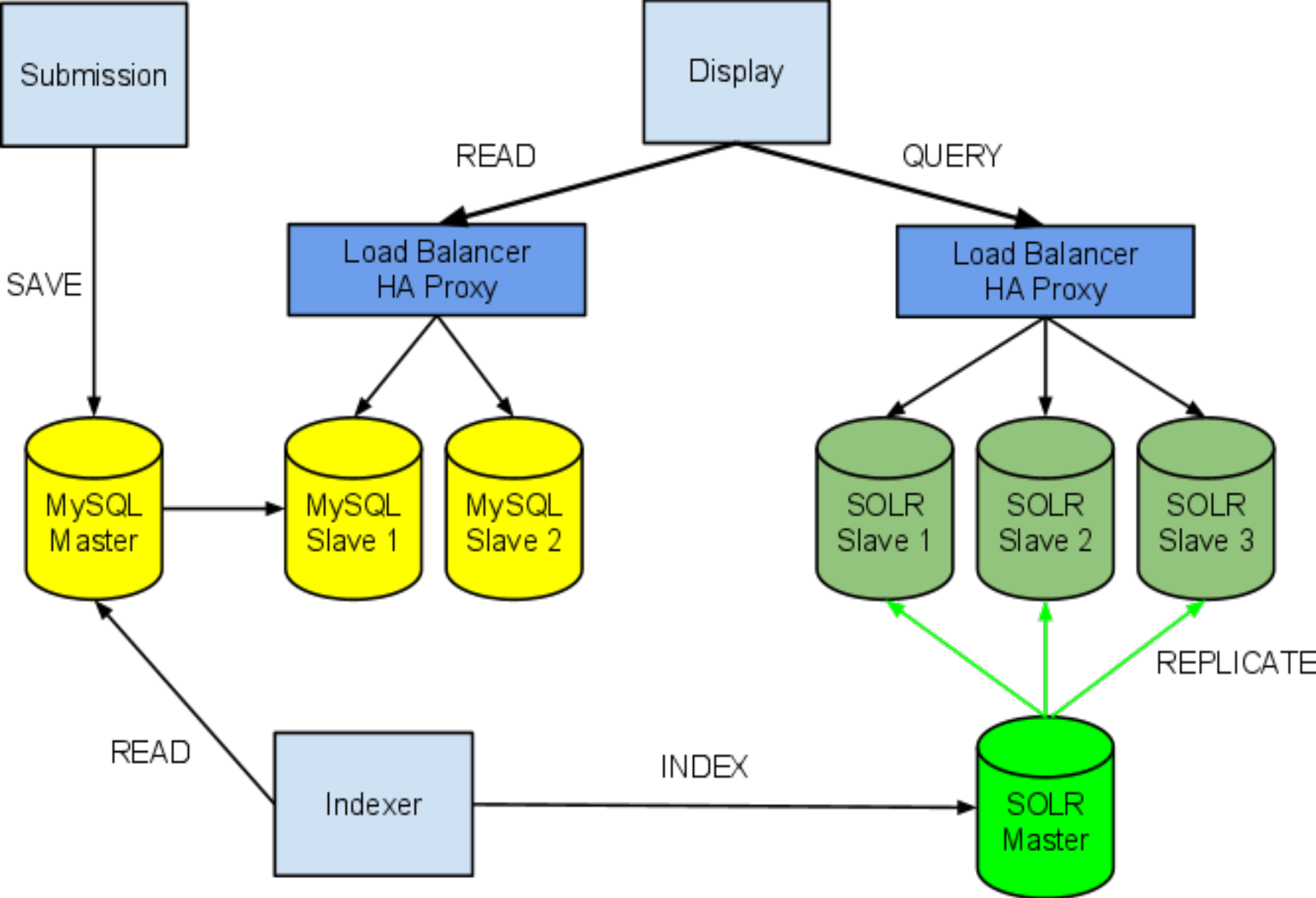


SOLR - Statistics

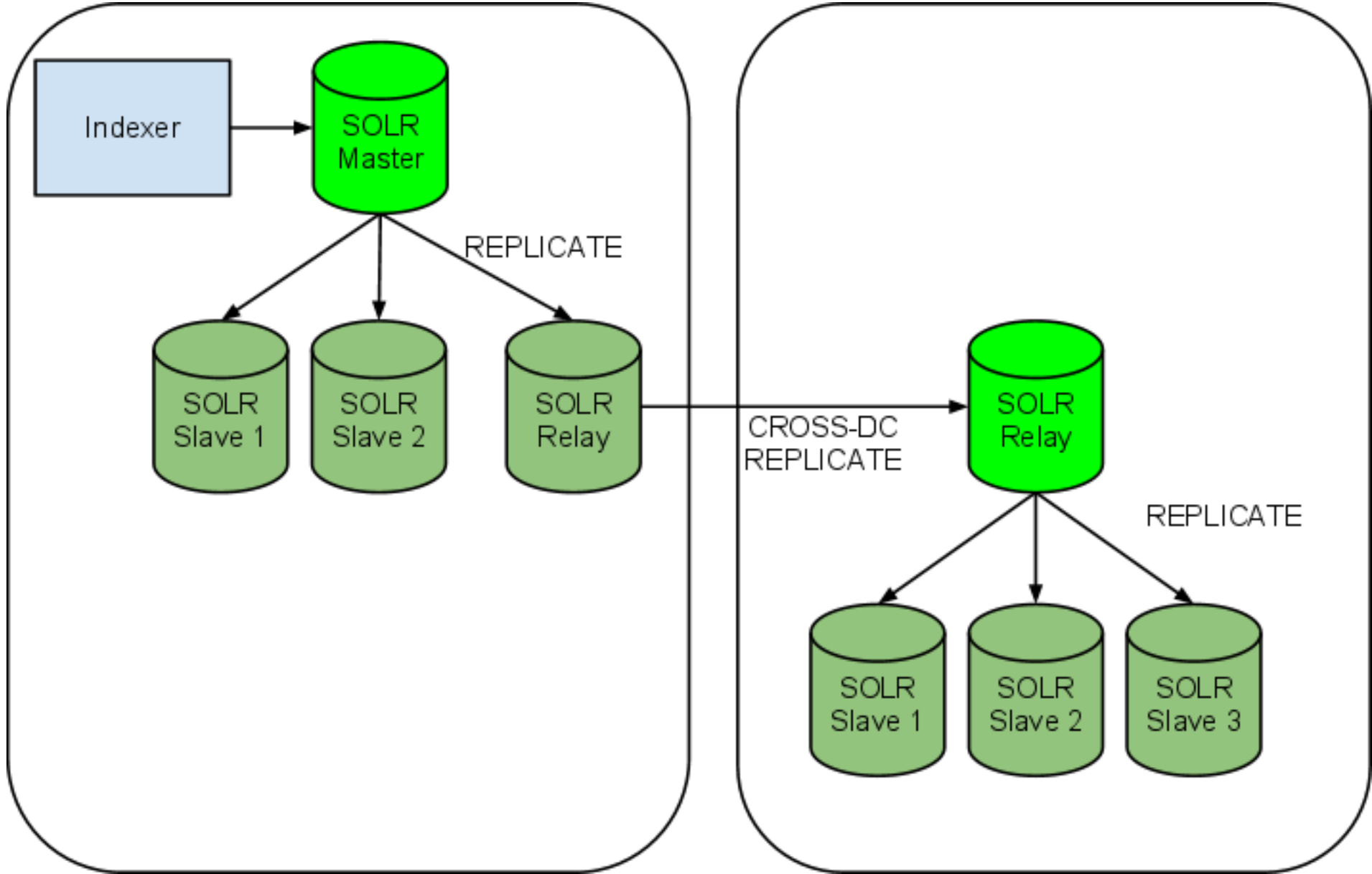
- COUNT, SUM, AVG, MIN, MAX (StatsComponent)
- Stored fields
- Whenever content changes, re-calc stats for all affected subjects

PRODUCT
id
name
description
categoryID
categoryHierarchyID
categoryHierarchyID_0
categoryHierarchyID_1
categoryHierarchyID_2
reviewAverageRating
reviewCount
reviewLastSubmission
questionCount

Scaling reads - Replication



Replication - Multiple Data Centers



Replication - Multiple Data Centers

Chatty if using multiple cores

Relay

- Core auto-warming disabled
- Connection wait and read timeouts increased
- Replication poll interval increased (15 min)
- Compression enabled

```
...  
<str name="httpConnTimeout">20000</str>  
<str name="httpReadTimeout">65000</str>  
<str name="pollInterval">00:15:00</str>  
<str name="compression">internal</str>  
...
```

SOLR Cloud - Bazaarvoice version

- Multiple cores (100+ per server)
- Re-balance indexes across cores and servers
 - Automatic
 - Manual
- Deployment map stored in MySQL
 - Host - Core - Partition
 - Statistics
- Partition lifecycle

Schema Changes

Re-indexing is time consuming for large indexes

Process

1. Full re-index off-line prior to the release
 - Incremental indexing after the release

Bottleneck: reading from MySQL

Goal: Transparent re-indexing

Performance Tuning

- Heap size
- Cache sizing
- Auto-warming
- Stored fields
- Merge factor
- Commit frequency
- Optimize frequency

Process: Simulate and measure

- Replay logs
- Analyze metrics
- Monitor GC

Performance Tuning - GC

```
# Java memory usage settings
# Force the NewSize to be larger than the JVM typically allocates.
# In practice, the JVM has been allocating an extremely small Young generation which
objects to be prematurely promoted to the Tenured generation
```

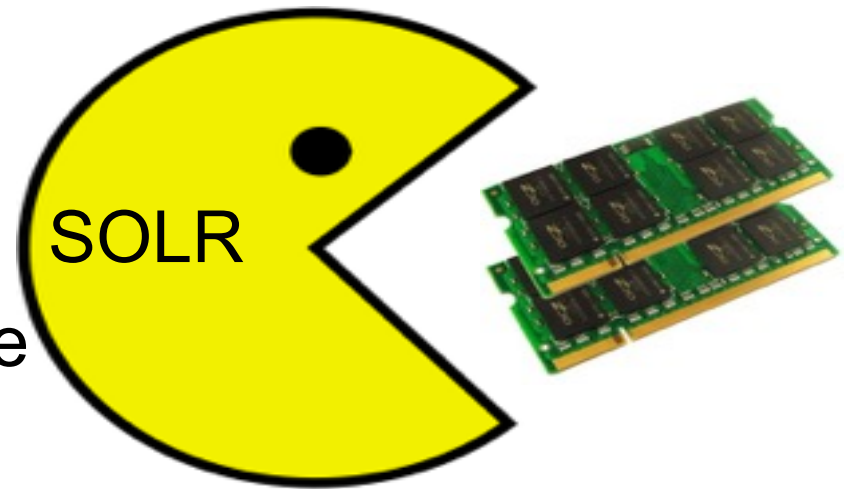
```
JAVA_MEM_OPTS="-Xms27g -Xmx27g -XX:NewRatio=8"
```

```
# -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCDateStamps --> Turn on GC Logging
# -XX:+UseConcMarkSweepGC --> Use the concurrent collector
# -XX:+CMSIncrementalMode --> Incremental mode for the concurrent collector
# -XX:+CMSIncrementalPacing --> Let the JVM adjust the amount of incremental
collection
```

```
JAVA_GC_OPTS="-verbose:gc -XX:+PrintGCDetails -XX:+PrintGCDateStamps -XX:
+UseConcMarkSweepGC -XX:+UseParNewGC -XX:CMSInitiatingOccupancyFraction=55 -
XX:ParallelGCThreads=8 -XX:SurvivorRatio=4"
```

SOLR Performance - Summary

- SOLR loves RAM!
- Log replay
- Same config, same hardware
- Get the most out of one instance



Conclusion - SOLR Strengths

- Lightning fast given enough RAM
- Good scale out support including multi-data center
- Great community

Conclusion - SOLR's Gaps

- Not fully elastic
- Real time takes work
- Secondary data store = sync overhead
- Schema changes

Questions



@apinkin