Customers

Discussion: (Karthik Rajagopal —TCC, Isaac Dancy — Warehouse)

TCC Customers
NTR - Knowledge sharing.
IBSS - Recent changes.
VSS
Warehouse Customers

  - ○ Domain knowledge on 3DW and TeklaWH business use cases, statistics etc.,

SketchUp
Tekla

DevOps

Discussion: Roman Laguta

- What are all of the steps required to update the production stack?
- What "binaries" need to be created before a push? How are they created?
- Is anything done differently on a staging/release stack?
- Can we get access to all tracking/statistics/logging tools (Datadog, etc)?
- There's a separate Git repo for DevOps, right? If so, we need access to that.
- How is a hotfix handled? There are two recent examples. What was the process there?
- What's the process of updating/deploying the render server?
- What services does cron manage?
- There is a small pool of elements servers(test servers) still kept for use in addition to spot-stacks (including samarium)
  - what's their use?
  - how do they get transitioned over
- "issart" org in myconnectedsite.com seem to have documentation and other resources. Is it going to be shared?
- Access to production servers. (Need an inventory of servers and access that currently ISSArt has)
- Bamboo setup..
- Merge plugin codebase.
- LucidWorks, AWS and other relations.
- Physical/deployment architecture of the system - From Cloudfront to memchache to Solr to DB.
- Discussions around tools, process documentation/presentation from Operations perspective and support required from warehouse development team.
- What DevOps concerns/considerations are there when it comes to the new Mandrill and Akismet additions?

Exercises

- Dry run release candidate creation
- Dry run hot fix push.
- Dry run production push.
- Dry run staging, load stack update.

Render Server

Discussion: Denis Plakhin

- Where does the .NET / Windows task-runner code live?
- During mass re-renders, what's the process of spinning up multiple render servers?
- Render service infrastructure still in play (ever since VO's gone, what changes have gone in).

Exercises

- Start a mass re-render
- Render server update

Solr

Discussion: Evgeny Chemezov

- How do we make changes to the schema? Diagram of all of the various components/workflow? Is this desired?
- How is the import done?
    - Currently the import takes an hour and a half. It's "required" on every release. How can we make that more efficient?
    - Are the steps different in local development vs. production? How?
- At the development level: How to setup/debug/view data? Is there a way to do it using the embedded solr in a dev. environment?
- When it comes to Irina's research into Lucidworks Fusion, is there anything else not in the search ticket worth discussing?
- What was the thought process in using Solr over, for example, ElasticSearch?
- Below is some assumptions around Solr import. Please correct if wrong
    - There's only two places where the underlying index is modified.
        - Full import - Multiple threads run through all the entries in the database and add documents to index using solrj library
        - Data import - when the entities/collections's updateSolrIndexTime is touched by code. This is run every one hour? and the Cron job calls
- What are the reasons we don't directly modify documents in the index in the java code when there's database changes and imports just left as true imports ?
- For every database change (updateSolrIndexTime column), should we have to wait till the cron job kicks in to update the index ? When are the changes visible to the user ?

- Softcommits and hardcommits - Some more information on the configurations currently in place, corresponding behaviour.
- How are the dynamic fields related to translations in solr schema used ?
- How are the dynamic fields related to stemming in Solr schema used ?
- How and where is parentDate function used. Some explanation on parent* fields in solr schema will be useful and also AddParentFieldUpdateProcessorFactory.
- What are the modifications in solr-dataimporthandler-fork?
- What is the physical architecture around Solr servers ? (Is traffic distributed between the two solr servers?)
- A minor one : Why does WithBinaryDetails class have com.solr as the package name ?
- Physical architecture. How servers are organized and how is traffic routed to two solr servers and how is index shared between the servers.
- Information on the forks and the changes that went into it and the reasons.

Exercises

- Overview of browser-based Solr query/document-view mechanism.
- Can we go through adding a sample value to each Solr document?
- Solr import
- Dev environment Solr

Functional Testing

Discussion: Konstantine Firsanov

- Can we get more information on the general process? Anything that isn't obvious in JIRA?
    - There are QA-specific scripts in the repo right now. After review, is there any questions about those?
- Is the current process dependent on "stacks on demand"? Will that require a licensing cost after the contract ends?
- Admin rights to extend the spot-stack duration should lie with Clay/Allen instead of Alexander Malyshkin?
- There's a whole other project on the TCC JIRA with test steps and documentation. Can we get an export of everything related to the warehouse and its TCC dependencies?

Exercises

- Run through a functional test of a ticket.

Load Testing

Discussion: Konstantine Firsanov

- How is load testing done? What tools are used to generate the load testing scripts?

- Where are the scripts located? In the same Git repo as the code? If not, we really need these available.
- What are all of the steps needed to launch a production stack and perform testing?
- Is there a way to track what functions/processes are contributing to the load?
  - For example, maybe a change to SetCollectionContains is drastically changing CPU load. Can we trace it to a specific function in the Java code?
- Discussion around the current coverage through jmeter and testing methods and documentation surround it.
- Demo/Documentation from ISSArt on current load testing methods/techniques/tools.
- Access to Load stack and performance testing
- Discussion on release candidate testing ticket and steps for Chennai team to take ownership.
- Render server related testing. (Anything more than blackbox testing - performance, load etc.,).
- Load testing demo (The scripts and the techniques)
- Gaps
  - Admin access to Stack on Demand (and demo on what needs to be done when testing period needs to be extended beyond 16 hours.
  - Access to JMeter scripts.
-

Exercises

- Start a load test run
- Add a new API call (CachedSearch, GetUsers) to load testing

---

Memcached

Discussion: Evgeny Chemezov

- From the Omsk team's perspective, what are the pros/cons of current setup? Are there any pain points?
- What other caching tools were considered in the past? For example, what about ElastiCache or Redis?
- What's the ratio of hits/misses you should shoot for in effective caching?

Log access/storage

Discussion

- Where do the logs go? Short term and long term?
- How do we query them?
- Is there anything related to this ticket worth discussing?
- Honestly the collection/storage/retrieval of logs is a big unknown right now.

Exercises

- Crack open a log. Find something interesting.
- Statistics of [insert API endpoint] over [insert timeframe]?

MySQL

Discussion

- The TCC team uses Percona's distribution of MySQL as well as their consulting services. Is that something we want to continue?
- What's the general workflow of applying/organizing database patches with every release? For example, there's a folder in the repo for each RC with patches. How is that folder setup?
- Why was liquibase forked? If it is updated to the newest version, is it still necessary to use a forked version?

Exercises

-

Splitting TCC and Warehouse

Discussion

- Major dependencies
  - Login workflow
  - File Storage
  - Rewrite rules
  - Build information
  - Auto-generated API documentation
- DevOps considerations
  - Any unknowns here?

Exercises

Grab Bag

- What are the modifications in solr-dataimporthandler-fork?
- SqlStat - What’s the logic behind SqlTimer.
- doBuild.sh walkthrough.
- Explain hibernate hooks in TCC (Transaction management etc.,)
- Explain liquibase-fork and resin-quercus fork and how is it used ?
- GDataSearch - Index updation workflow.
- User roles - How is it done today ? How does one normal user get a special role.

Development Process / Documentation

- Can we get an export of Warehouse-related JIRA tickets?
- Can we export of review-board data? This would be a great resource on why things were changed, but I'm not sure if it would be worth the effort.
- Can we export of JIRA wiki? We probably don't need the custom plugin they're using, but being able to import that data into a hosted MediaWiki, for example, would be great.
- Where is the legacy Google data housed right now? Is it in an accessible place for quick lookup?
- What is worth knowing when it comes to login/authentication?
- Exporting the Git repo
    - At what point in the transition will this be easiest to do?
    - Add to current frontend Git? Or keep separate? Pros/cons?
    - Every JIRA ticket that required a code change has its own origin branch. Do we want to keep all of those branches?
- Gathering Statistics on downtime
    - What caused it? TCC or Warehouse?
    - Any recurring themes?
- Other documentation sources
    - What is available? What are we lacking documentation in?
    - Is there anything really worthwhile that's not translated to English?
-