SemanticMediawiki in Wikia

From DevOps perspective
Agenda

● Introduction
● Reasons for using SMW
● Some examples how SMW is used on Wikia
● Issues we had/have with SMW on Wikia
  ○ ... and how we deal with them
● Plans for future
About me

- Krzysztof Krzyżaniak
- eloy@wikia-inc.com
- Almost 6 years in Wikia
- Manager of backend in DevOps team
About Wikia

- About 300000 wikis
  - on 4 MySQL clusters
  - About 10% of them had activity last day
- SemanticMediawiki extensions as option, requested by community
  - We ask them why they need SMW and how they want to use it
  - More than 150 wikis has enabled different SMW extensions
Reasons for using SMW

- "I've heard it's cool!"
- "I've heard it's cool and useful!"
- "I think our wiki could use some features of SMW"
- "We need better categorization tool"
- "We want to have better infoboxes"
- "Our wiki doesn't make sense without SMW"
Family History and Genealogy Wiki

- http://familiypedia.wikia.com/
- http://semantic-mediawiki.org/wiki/Familiypedia
- About 125,000 articles about Genealogy data
- **SemanticForms and SemanticDrillDown**
- Infoboxes
- Forms
- Templates
Biology wiki

- http://lt.biologija.wikia.com/
- Lithuanian wiki about taxonomy of biological organism
- More than 580,000 pages!
- SemanticForms
- SemanticDrilldown
- Templates, templates, templates!
Yu-Gi-Oh Wiki

- Anime/manga franchise wiki
- http://yugioh.wikia.com/
- Almost 60,000 pages
- Infoboxes for characters, game cards etc.
Narutopedia

- Anime/manga franchise wiki
- http://naruto.wikia.com/
- Almost 4800 pages
- Infoboxes
Issues

- Bad concurrency in writing operations
- Heavy database queries
  - search bots traffic!
- Massive number jobs in JobQueue
- Temporary tables in SemanticDrilldown
Master switch for Ops

top of config/CommonExtensions.php

```php
<?php

########################################################################
# Emergency disabling settings
#
# uncomment each line to disable feature(s) sitewide
########################################################################
#
# disables SemanticMediaWiki and related extensions
#$wgEnableSemanticMediaWikiExt = false;
```
Massive number jobs in JobQueue

- Most of writing operations are done using JobQueue engine
  - That's very good!
  - ... but also: delay, problems with managing queues if you have more than one wiki and you use cron for running jobs
    - For example how to check which wiki has queued jobs?
- Custom JobQueue runner
  - additional shared table for all wikis
  - threaded runner
  - runner is aware which wiki has queued jobs
Separated database pool for SMW

- Hack around MediaWiki LoadBalancer
- Separate connection to MediaWiki tables (page, categorylinks etc) and to SMW tables (smw_rels2, semantic_drilldown_values) etc
- Issues:
  - You can't do joins between SMW tables and MW tables
  - You need patched SMW and MW
  - Doubles database connections in application
Separated database pool for SMW (cont.)

For SMW code:
- $dbr = wfGetDB( DB_SLAVE );
+ $dbr = wfGetDB( DB_SLAVE, 'smw' );

For LoadBalancer code:

```php
function getSectionForWiki( $wiki = false ) {
    ...
    if ( isset( $this->sectionsByDB[$dbName] ) ) {
        // this is a db that has a cluster defined in DB.php
        $section = $this->sectionsByDB[$dbName];
    }
    ...
```
Temporary tables in SemanticDrilldown

- **Ops don't like them**
  - because they don't like exceptions
  - painful for database servers
  - additional permission needed
  - replication exceptions

- **We don't have good solutions for this**
  - looks like MariaDB handles them better than Percona and Oracle MySQL
RDF store

- We tried and we resigned
- Problematic from ops perspective
  - lack of experience
  - additional service (exception!)
- No business value for Product Department
- Didn't resolve our problems with Semantic Drilldown
Plans for future

- Second attempt for RDF storage
- 2nd screen applications for gamers
  - We don't decided yet what's will be used as backend