Migrating An Application Database to Splunk

Baltimore Splunk Usergroup
Feb 2024

Agenda

- What this talk is
- Project background
- Analyze your database activity and Splunk Deployment
- Implementation Options
- Wrap up

About me

Phil Meyerson

Systems Developer
a.i. Solutions

AppDev SplunkDev CICD

https://conf.splunk.com/watch/conf-online.html?search.event=conf21&search=metaprogramming#/ (conf `21 on Oversight asset inventory app high leve talkl)

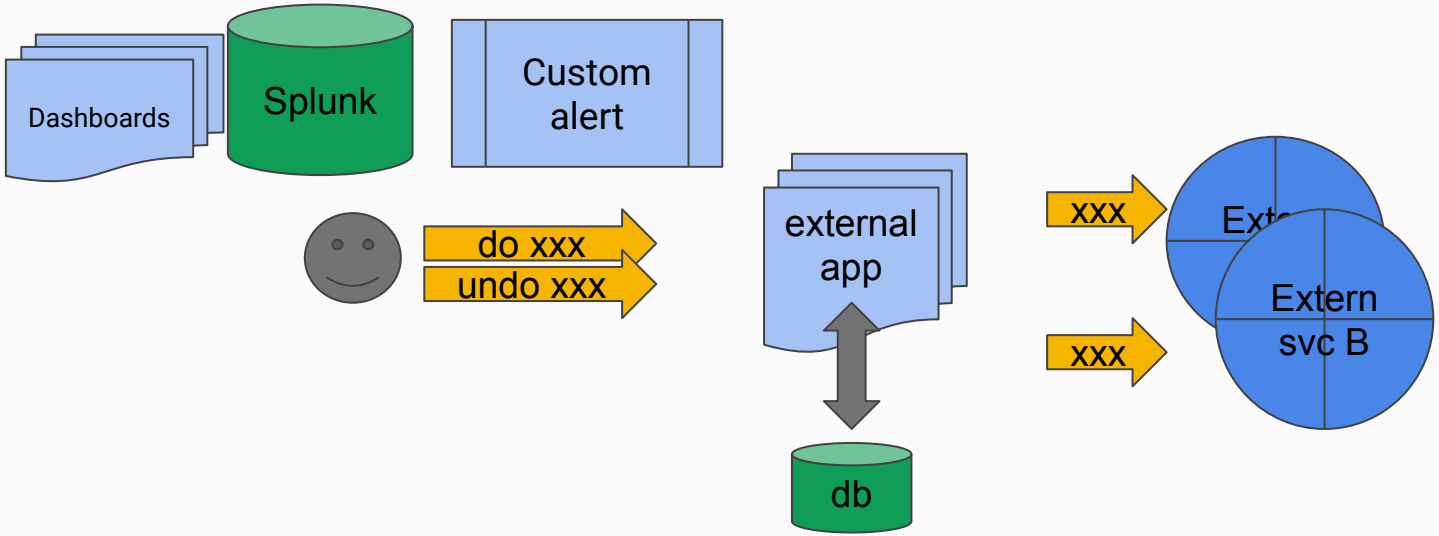
Bsides Splunk `21 OVERSIGHT: Building an Asset Inventory Data Pipeline

Conf `19 - Public Data Exploration with Splunk

What this talk is

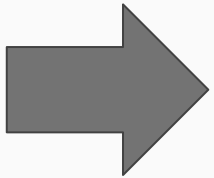
- For “application developers”
- How to decide if this is might be a good solution for you
- How to maintain application state in Splunk

Existing Architecture



Cost factors to “use Splunk” DB

- Ops and Maintenance of DB/Platform
- Security of data



- Install UF
- Developer costs
- Splunk resource utilization costs?

Splunk resource utilization

- Available splunk instance resources -- room to grow?
- Available splunk license -- room to grow?
- Splunk instance uptime dependency
- Splunk Admin Time

Splunk resource utilization

- Ingest $$$
  - Are you already logging from this app?
- Splunk Uptime dependency
  - Buffer requests in app or*
- Splunk Resource utilization by app
  - Ingest is typically very performant, searching is heavy
  - Review Distributed Monitoring Console
- Splunk Admin Time
  - Provide documentation; especially in-app or in-dashboard
  - Follow better and best practices for application logging

Distributed Monitoring Console

Existing CPU resource contention could be made worse by high frequency of remote search jobs

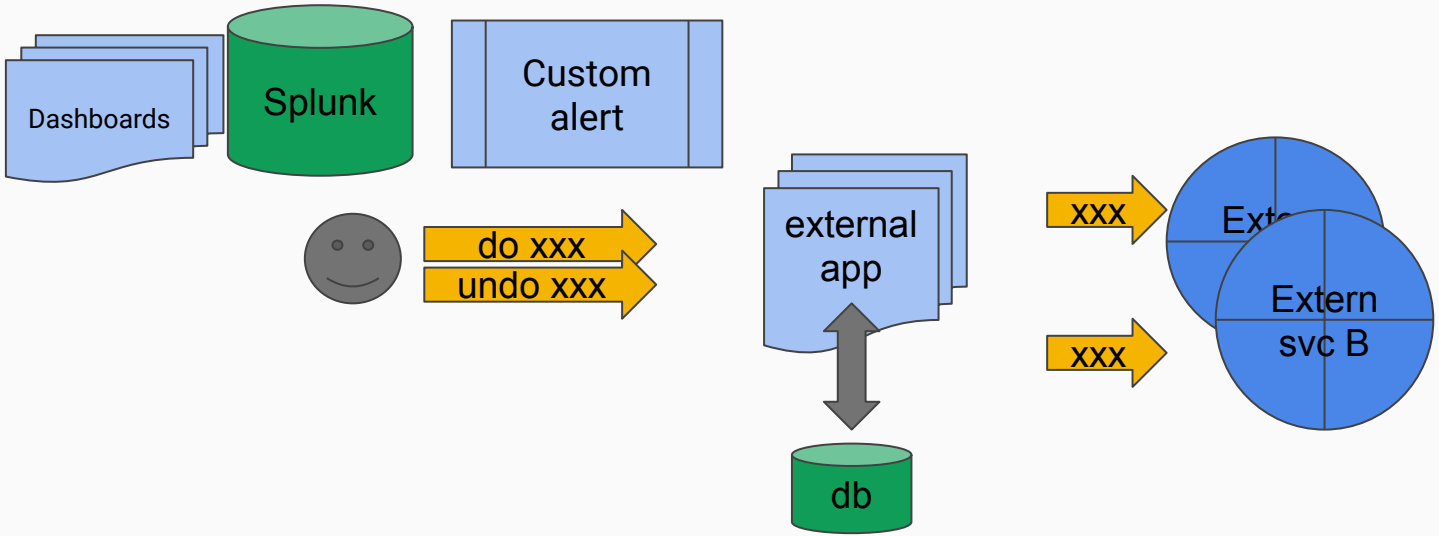
1 search consumes 1 cpu core until completion

- Scheduler Activity:
  - Instance for failed or skipped scheduled searches
  - Execution Latency
- Resource Usage:
  - Excessive CPU

Analyze your database usage

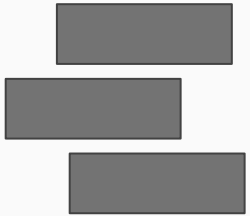
- What functions of the app need db access
  - Read?
  - Write?
  - Modify*?
- Typical write frequency/day? 10x? 100x?
- Type read frequency/day? 10x? 100x?
- Typical amount of data written/day? 10x? 100x?

Existing Solution



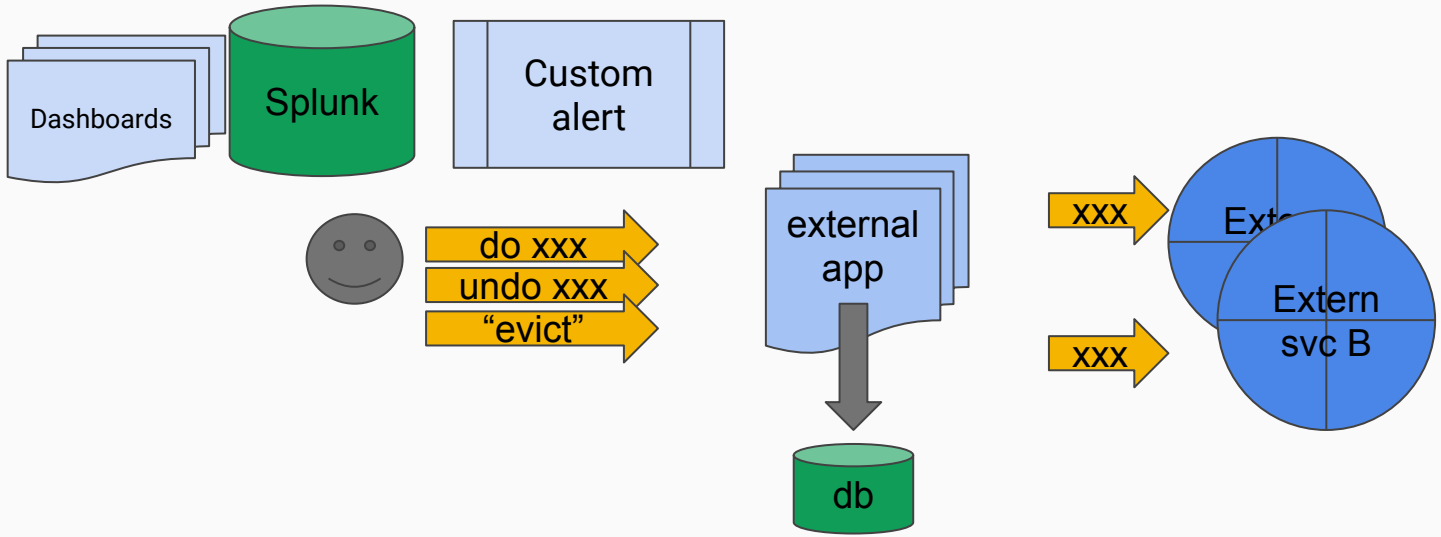
Our database usage

- Average requests per day
  Peak requests per day
  10x Peak requests per day
  Read frequency: very low

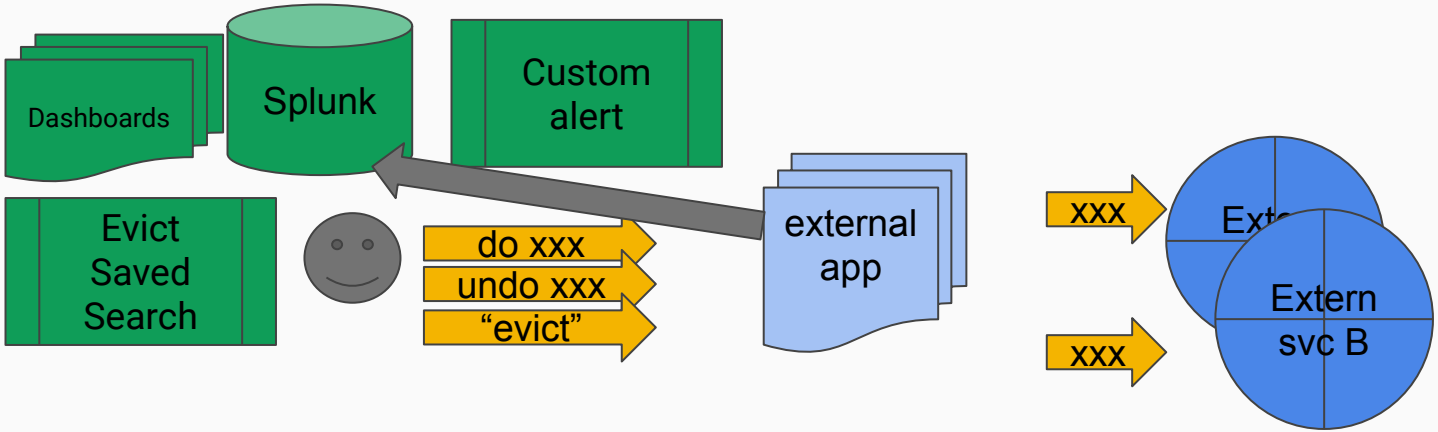


- DB functionality used:
  - Read: determine records for eviction and “undo” requests
  - Write: successful requests
  - Modify*: repeated requests update eviction conditions
  - Write*: record incomplete or unsuccessful requests to application log

New Solution



New Solution



Implementation: Splunk

Database “reads”:

- Saved alert = “eviction”
  - Search summary index for records matching eviction condition
  - Use custom alert action to submit to external app

Database “writes”:

- Write to application log file
- Scheduled updates to a summary index

Implementation: Splunk

Database “updates”:

- Factor out - just write a new event and construct search to filter by _time

Implementation: Code Refactor

Separate library/module to write log messages

- Can update logging statements without touching service logic
- Easier to follow Better Application Logging practices

Implementation: Better Application logging

Reduce the cognitive load to explore your data

- Consistent field order for regex
- Consistent field names
- Consistent case and tense

- Consider structured logging libraries/approach

Alternative Solutions

1. If app needs direct read access:
    a. Splunk REST API to search

2. If app needs to modify records:
    a. Could use kvstore collection -> more traditional db interface
    b. Could just log new event, newest timestamp wins

Kvstore gotcha

“Don’t use splunk as a database”

kvstore is replicated across SHC per oplog max Size

If oplog can’t hold all transactions , you’ll lose data

Default 1GB “contact support before increasing”

Final Takeaway

- Analyze the way your app uses the database today… can this be simplified?
- Consider resource utilization impact of your app before deployment
- Follow Better and Best application logging practices
- Feasible to use Splunk as a database in many use cases!

Notes

https://docs.splunk.com/Documentation/Splunk/9.2.0/Admin/ResyncKVstore

https://dev.splunk.com/enterprise/docs/developapps/manageknowledge/kvstore/usetherestapitomanagekv/

https://dev.splunk.com/enterprise/docs/developapps/addsupport/logging/loggingbestpractices/

Server.conf: oplogSize -- kvstore collection replication file