.conf18

splunk>

Indexes – a Splunk Admin´s best friend

Victor Rosberg
Operational Intelligence Manager
Betsson Group

victor.rosberg@betssongroup.com

October 2018

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

splunk> .conf18

Indexes – an Splunk Admin´s best friend

- what stands between you and that next cup of coffee

- **Introduction** > to Betsson Group and yours truly
- **Indexes basics** > the core concepts and key variables for administration purposes
- **Principles** > the “immutable laws” behind successful Splunk governance
- **Design** > Best practices and considerations for an immaculate index & role structure
- **Execution** > how to leverage the principles and design in practice
- **Q&A** > ask me anything

Introduction

to Betsson Group and yours truly

Betsson Group – Online Gaming Company

“The best customer experience in the industry”

1
Platform

17
Brands

1700+
Employees representing 56 nationalities in 12 locations

692k+
Active Customers Q2 2018

4.72b
Revenue 2017 (SEK)

splunk> .conf18

VICTOR ROSBERG

- Leads Operational Intelligence team
- IT Ops: 12+ years, Tech & Leadership
- Splunk: 5+ years, Architect
- Co-leads Splunk User Group Sweden

splunk> .conf18

OPERATIONAL INTELLIGENCE

- Manager, Architect and 3x Engineers
- Splunk Platform capacity: 2.5TB/day, typically 365days retention including Enterprise Security & ITSI
- 700+ users spanning from operations, development, QA, NOC, SOC and more

Indexes basics

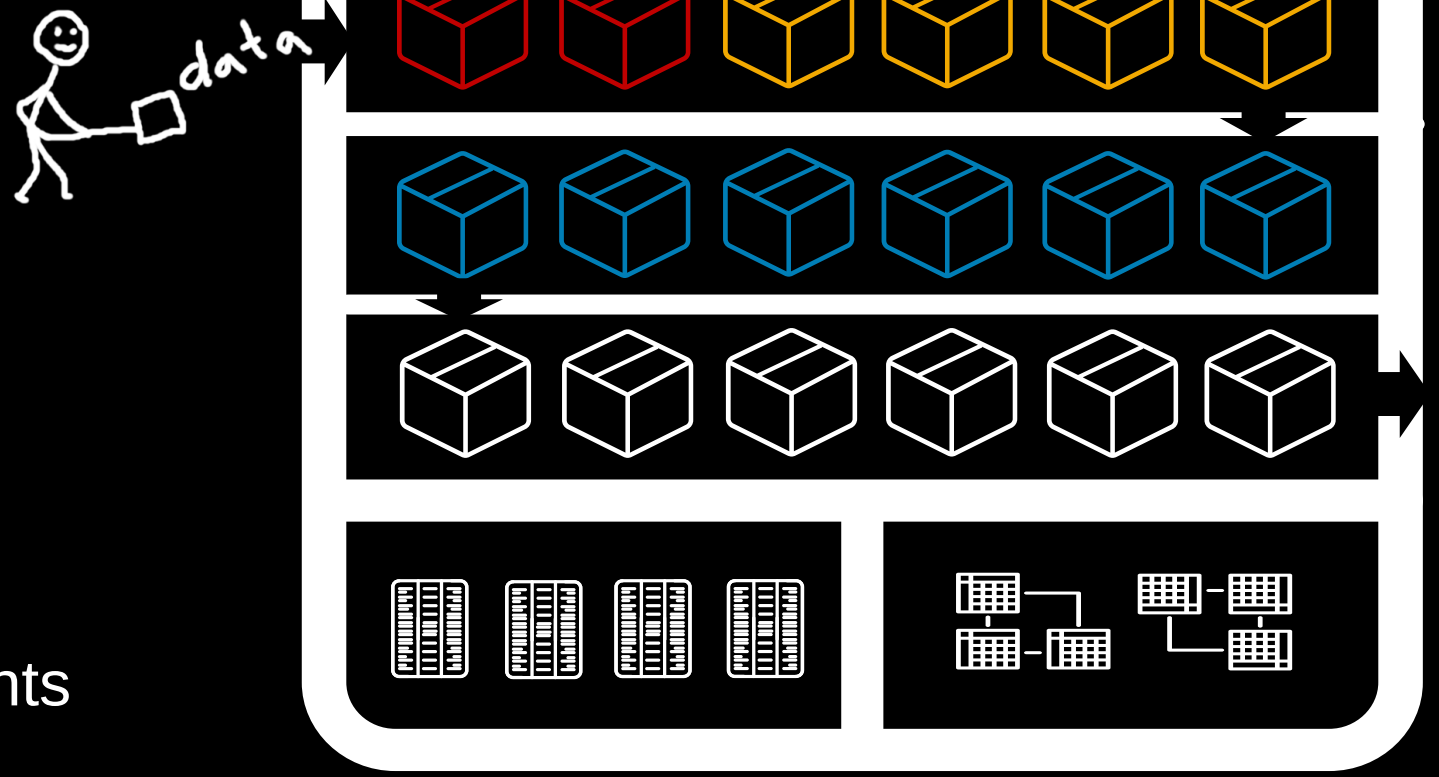
the core concepts and

key variables for administration purposes

splunk> .conf18

Indexes Basics > General Concept

Think of an index as a **container of data**

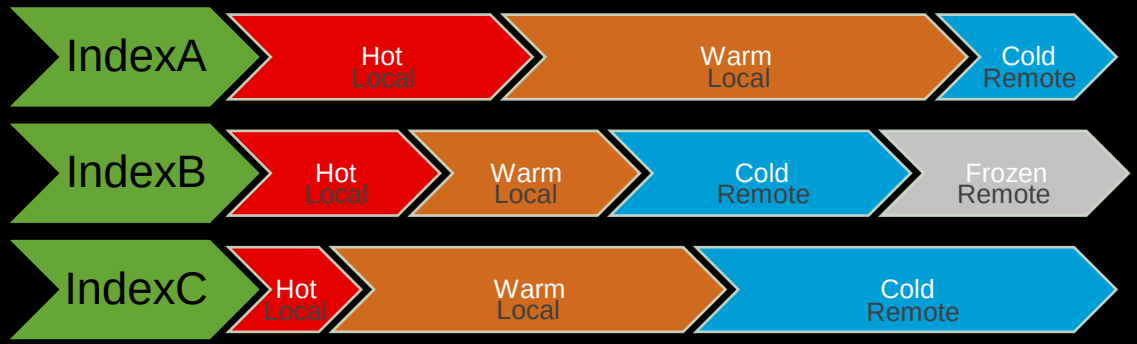


An index contains:

- ***Raw data*** stored as ***events,*** bundled into ***buckets***
- ***Meta-data*** about the events
- ***Catalogue of pointers*** to events

Indexes Basics > how buckets flow

- Buckets are **bundles of events**, received close in time **within an index**
- **Buckets** move with **age,** split **over different volumes**
  - **Hot: read/write / Warm: read-only**
    - Typically on fast local disk
  - **Cold: read-only**
    - Typically on slower storage like a SAN
  - **Frozen: archive/delete**
    - Archived data typically on persistent storage
    - Archived data not-searchable until **thawed**
    - If not Archived, data will be **deleted**



Indexes Basics > key variables

These are the key variables we´ll work with:

- **Name** > Each index have a unique name we can use
- **Roles** > Each index´s access can be controlled via Splunk roles
- **Volumes** > Each index´s buckets move through volumes, with different storage options, as they age

Principles

the “immutable laws” behind successful Splunk governance

Principle #1: clearly defined responsibilities

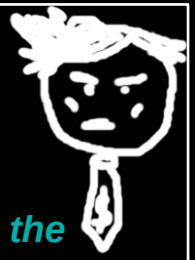
Identify these three roles within your organization

Service Provider

- *"that Splunk guy"*
- *Splunk subject matter experts*
- *Responsible for the Splunk platform; governs, supports, monitors and develops the platform*
- *Develops the processes, policies and strategy for the platform*

In our case:
Operational Intelligence

Business Owner



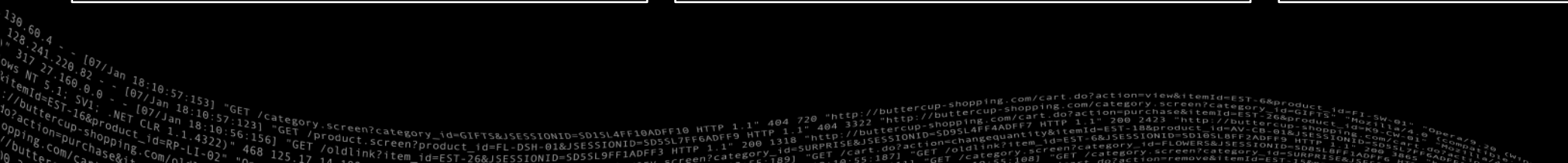
- *"Mr Moneybag"*
- *Stakeholder representing the Business*
- *Has a budget – justifies the cost*
- *Owns the data - responsible for it´s content and who gets access to it*

In our case:
Department Managers

Technical Owner

- *"Team OnCall"*
- *Splunk Super Users*
- *Service subject matter experts*
- *Interacts with Service Provider and Business Owner*
- *Delegated by Business Owner for technical execution*

In our case:
Department DevOps



splunk> .conf18

Principle #2: full data control

Answer these questions before adding any data to Splunk

"Who owns the data?"

- ***Every event should have an owner***

Useful follow up questions:

- ***Retention / archive time requirements?***
- ***Estimated volume now and future prognosis?***
- ***Expected business value for total footprint?***

"Who can access it?"

- ***Every event should be controlled securely***

Useful follow up questions:

- ***Public or restricted access?***
- ***Confidential or sensitive content?***
- ***Can parts of the data be useful without full access to the full events?***

"How will it be used?"

- ***Every event should be stored for optimal performance***

Useful follow up questions:

- ***What´s the primary use cases?***
- ***What´s the typical timeframe this data will be searched for?***
- ***Can we correlate with other data?***
- ***Do we need all or just parts of the data?***

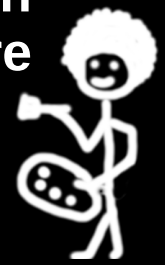
splunk> .conf18

Principles - Summary

- These principles provides you with:
  - Business justification
  - Responsibility distribution
  - Expected use cases
  - User expectations
  - Footprint estimation
  - Security template

Design

Best practices and considerations for an immaculate index & role structure



splunk> .conf18

Design - #1: Design a solid index naming convention

Each index name should convey:

- **Owner** > Who owns the data?
  - Owners domain
- **Function** > What´s the source and/or purpose?
  - Sub-division within owners domain
- **Access** > Who can access it?
  - Optional for public indexes
  - Security levels like sec1/sec2/sec3

Index name example: Owner ***platform***_Function ***app***_Access ***sec1***

Design - #2: Define data configuration

Each index should be defined to reflect:

- **Retention time** > how long does the data need to exist?
  - Consider storage type, sizing and volume setup
- **Searchability** > how long does it need to be searchable / performant?
  - Consider bucket transition between volumes for optimal performance vs cost
  - Consider storing only events that's typically searched together
- **Total volume** >

  - Consider capacity and archiving options to avoid data loss

Design - #3: Integrity and admin

Each index should be tweaked to reflect:

- **Data Integrity** > Slows down performance but may be necessary for audits
  - Consider requirements around data ***At rest*** and ***In transit*** and possibly split indexes
- **Complexity** > **ended up with loads of indexes?**
  - Consider specifying ***General***, ***Volume*** and ***Index*** specific settings
  - Consolidate indexes where requirements are similar and performance a non-issue

Re-evaluate & Reiterate

splunk> .conf18

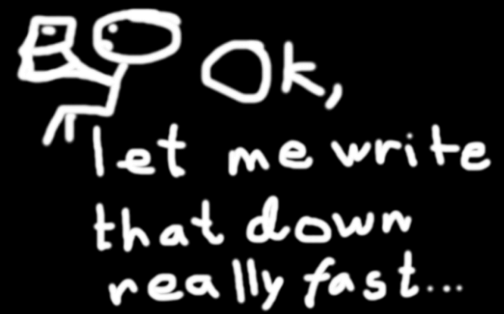
Design – example of an index definition

```
## GENERAL CONFIG FOR ALL INDEXES ##
memPoolMB = 2048
indexThreads = 8
maxTotalDataSizeMB = 4294967295
maxRunningProcessGroups = 16
maxRunningProcessGroupsLowPriority = 2
serviceInactiveIndexesPeriod = 120
frozenTimePeriodInSecs = 31556926
repFactor = auto


## VOLUMES CONFIG FOR ALL VOLUMES ##
[volume:internal]
path = /data/volumes/internal
maxVolumeDataSizeMB = 1100000
[volume:_splunk_summaries]
path = /data/_splunk_summaries
maxVolumeDataSizeMB = 1500000
[volume:cold]
path = /data/cold
maxVolumeDataSizeMB = 18000000
[volume:hot_warm]
path = /data/volumes/hot_warm
maxVolumeDataSizeMB = 14500000
```

```
## INDEX SPECIFIC CONFIG ##

[sportsbook_web_sec]
maxDataSize = 5000
maxMemMB = 10
maxHotBuckets = 10
maxWarmDBCount = 250
maxConcurrentOptimizes = 6
enableDataIntegrityControl=true
tstatsHomePath =
volume:_splunk_summaries/sportsbook/datamodel_summary
coldPath = volume:cold/sportsbook/colddb
homePath = volume:hot_warm/sportsbook/db
thawedPath = /data/thawed/sportsbook/thaweddb
```



Design #4 - Splunk Roles

Make use of three types of Splunk roles to control capabilities and access

Capabilities

- *Limits users resources usage*
- *Create or reuse default roles*
- *Set default index access for public indexes on default user role, inherit upwards hierarchy*
- *Custom roles as necessary*

In our case:
user, power, admin, dashboard

(the ninja club)

Data Access

- *Limits users event access*
- *Map to searchable indexes*
- *1 per non-public index for granularity*

In our case:
<index name>_users

App Access

- *Limits users read/write access*
- *1 per department app, assign default app*
- *1 per non-public app*

In our case:
<department name>_users

splunk> .conf18

Design - Summary

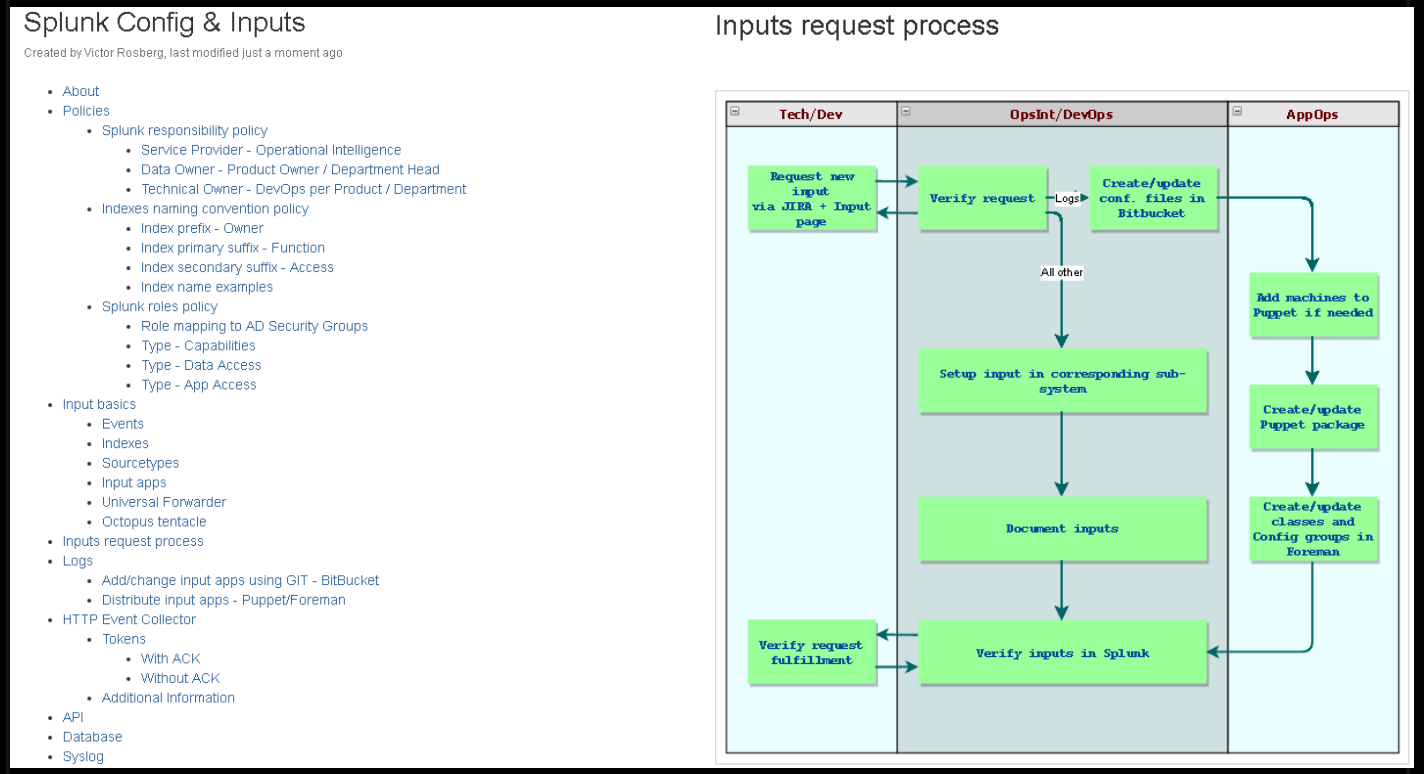
- The index and role structure should now provide:
  - Clear ownership
  - Granular data storage
  - Easy access management
  - Full resource control
  - Footprint estimation
  - Optimized search performance
  - A predictable standard

Execution

how to leverage the principles and design in practice

Execution - document

Make the structure, processes and policies transparent via documentation



- Document the structures, policies and processes
- Make it openly available
- Keep it updated

Execution – visualize

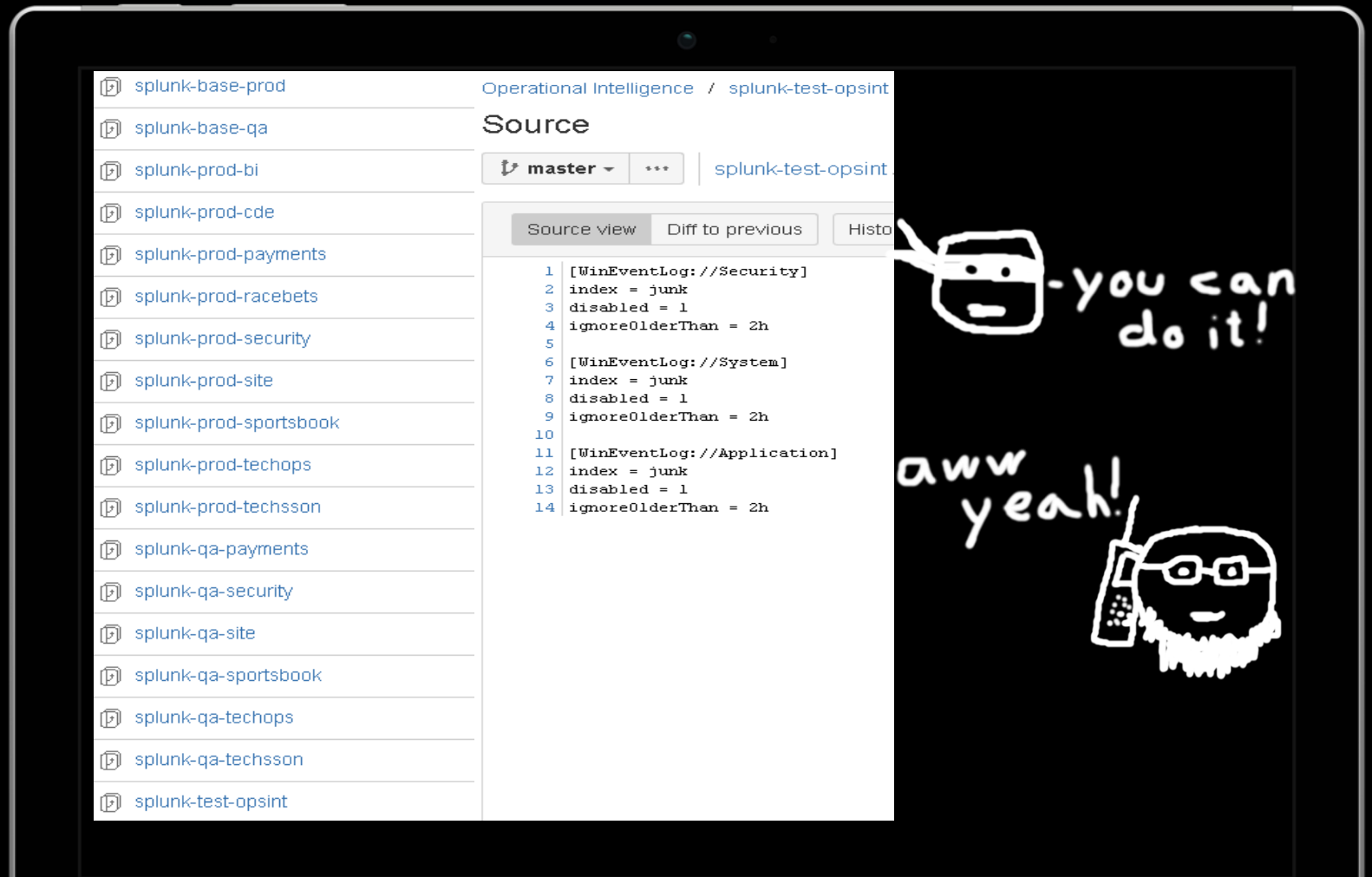
Make it easy and transparent for everyone

- Publish usage footprint
- Keep Business Owners accountable
- Optimize user resource utilization

splunk> .conf18

Execution – control

Take full control of all the configuration

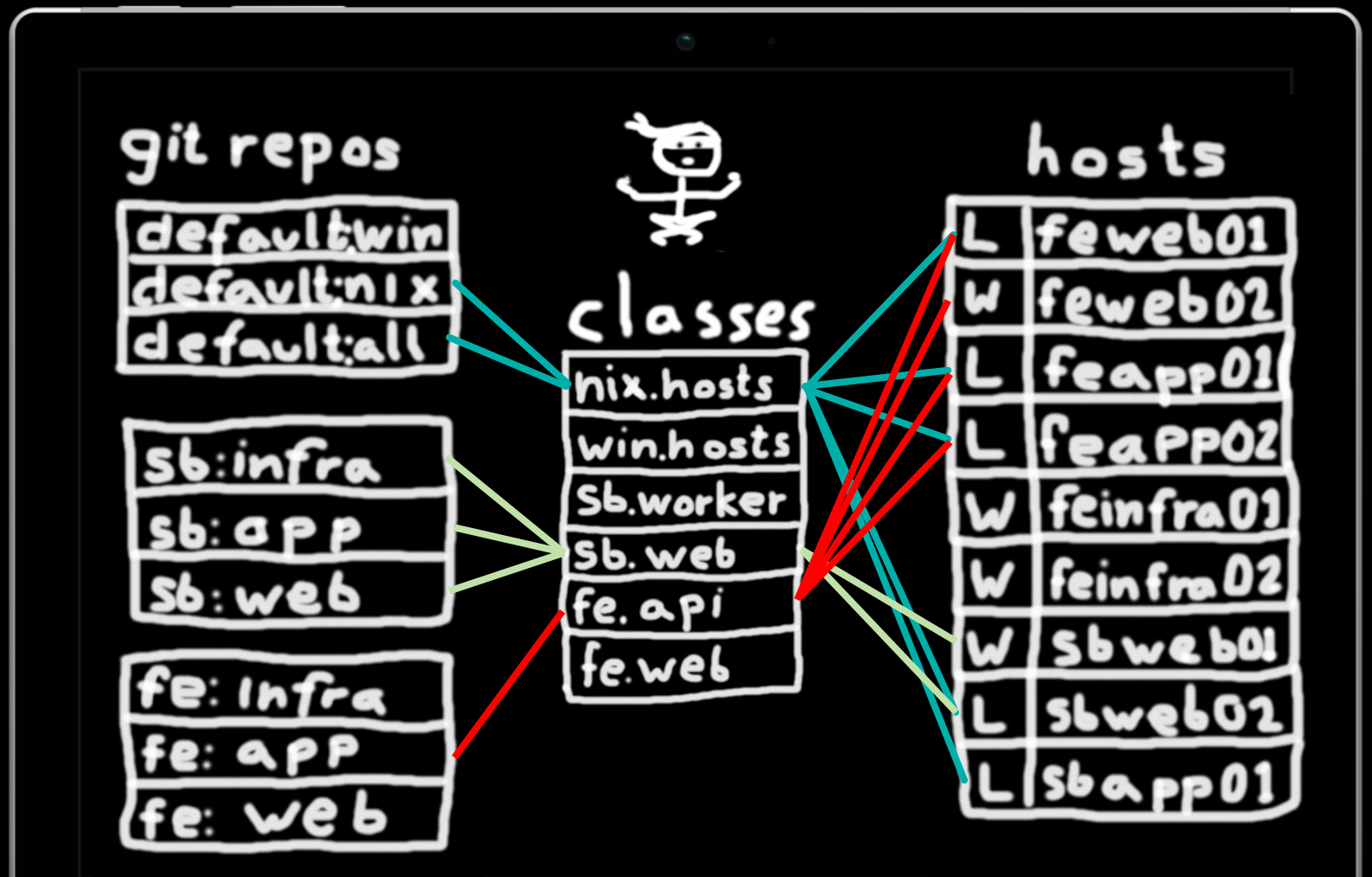


- Central, version controlled and secure – use Git!
- Separate configuration by Owner
- Give Technical Owners access

splunk> .conf18

Execution – distribution & management

Make configuration distribution seamless

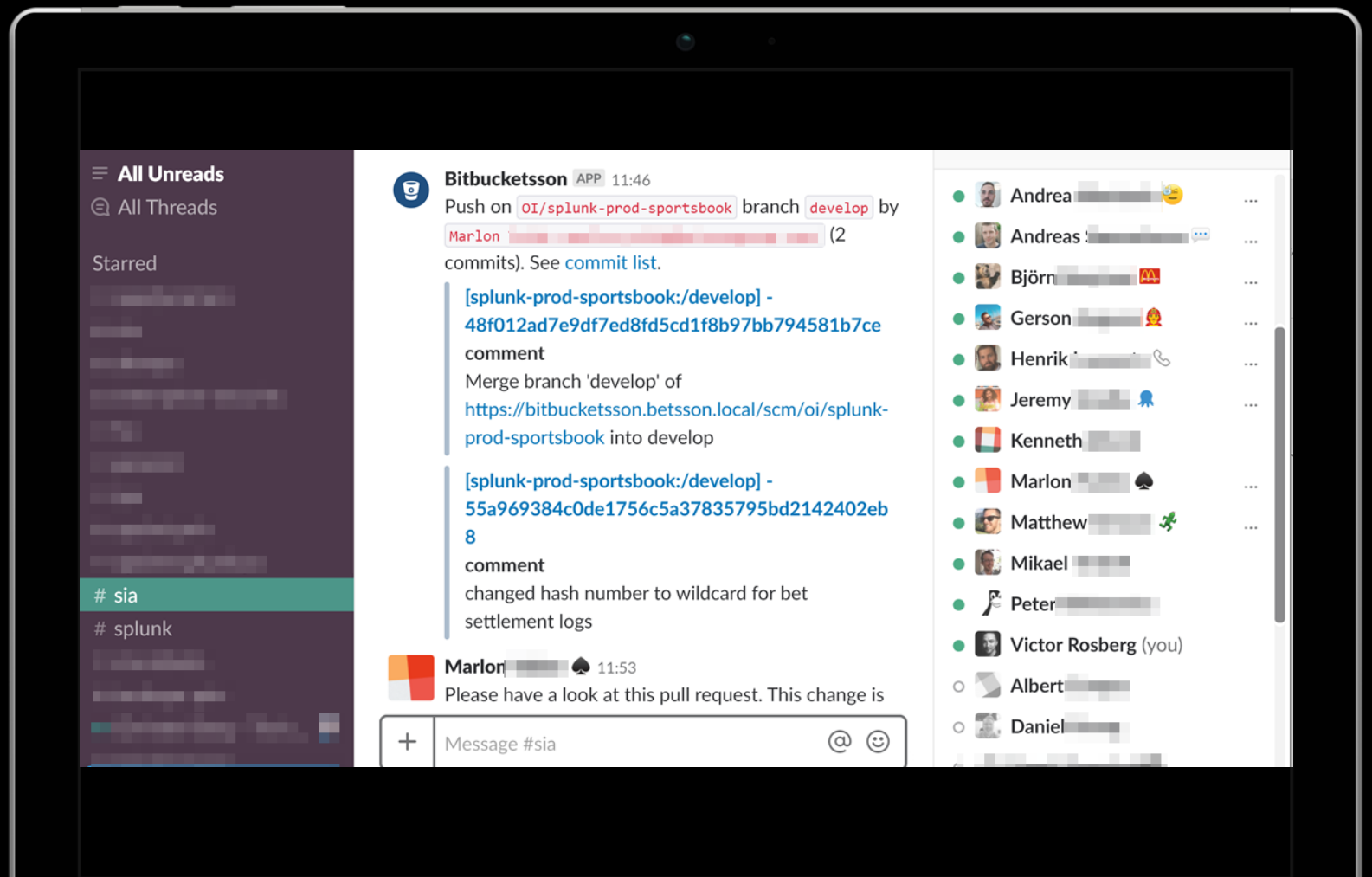


- Class-based system
- Automatic distribution
- Automatic remediation
- Remote management

splunk> .conf18

Execution – self-service data onboarding

Move towards governance rather than administration, step by step



- Standardize
- Automate
- Integrate
- Orchestrate
- Delegate
- Monitor

splunk> .conf18

Key Takeaways

- for a better Splunk admin experience

1. **Build a solid foundation –** early shortcuts will “probably” not scale
2. **Stand by your principles** – deviations will inevitable break integrity
3. **Be transparent** – play with open cards and provide constant feedback

Q&A

Ask me anything

Questions?

Psst! You can also just reach out to me during the event

…or drop me a mail:
victor.rosberg@betssongroup.com

…or connect on LinkedIn:
https://www.linkedin.com/in/victor-rosberg-7a373bb/

Thank You

Don't forget to rate this session in the .conf18 mobile app

.conf18

splunk>