ATLASSIAN
team '22

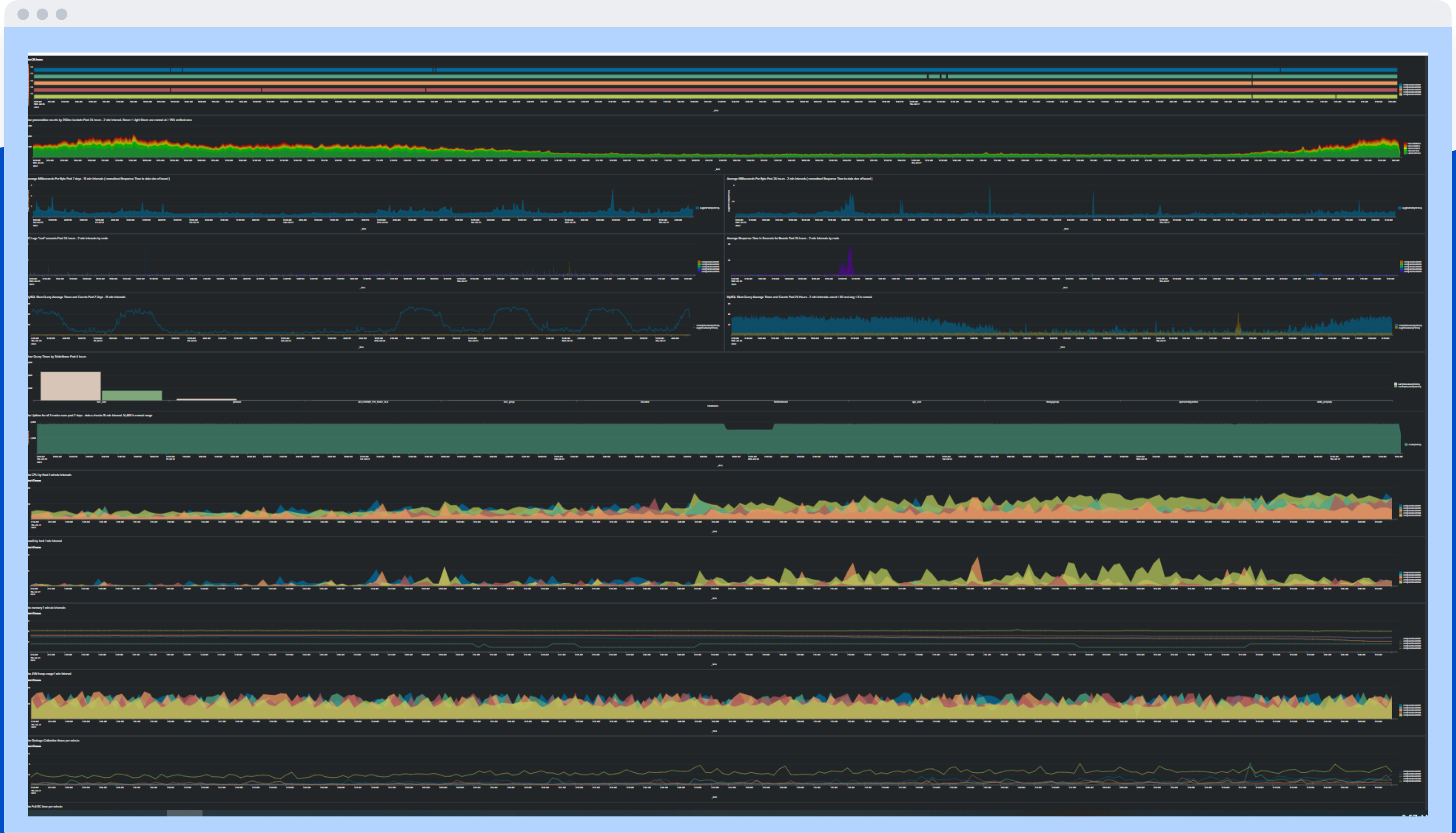
How to Purge Stale Data From Jira Software & Confluence

with governance from legal department, data retention policies, and project owners

Andrew Morin, Director Engineering, Charter Communications
Joe Jadamec, Software Engineer, Charter Communications

JIRA – ALL SYSTEMS RUNNING NORMAL

Charter Splunk Dashboard for Jira



Data Growth - All Systems Running Normal

https://www.wsj.com/articles/SB10001424052702303404704577305520318265602

PROBLEM – BENEFITS - SOLUTIONS

Unlimited Data Growth

Save Time & Money, Improve Performance

Purge Stale Unused Data

DATA GROWTH IS NORMAL

Without regular purging,

Jira issue count > 5 million

Jira customfieldvalues > 300 million

Both shared drives > 5 TB

Key Takeaways

Data Purges

Your Options

Cost Analysis

Know - How to Safely Purge Stale Data

Identify stale data, communicate to users, and save a copy to a test or reporting environment.

Feel – Comfortable That You Have Options

Build new instance every 2 years & import last 1 year?
Warehouse data and load onto reporting servers?
Grow forever or purge data?

Do – A Time & Cost Analysis, Then Choose

Learn your footprint size and growth rate, identify retention policies, and choose what works best for you.

Agenda

Data Growth and Costs

Solution Options for Confluence

How to Clean up Confluence Trash

Solution Options for Jira

How to Purge Issues from Jira

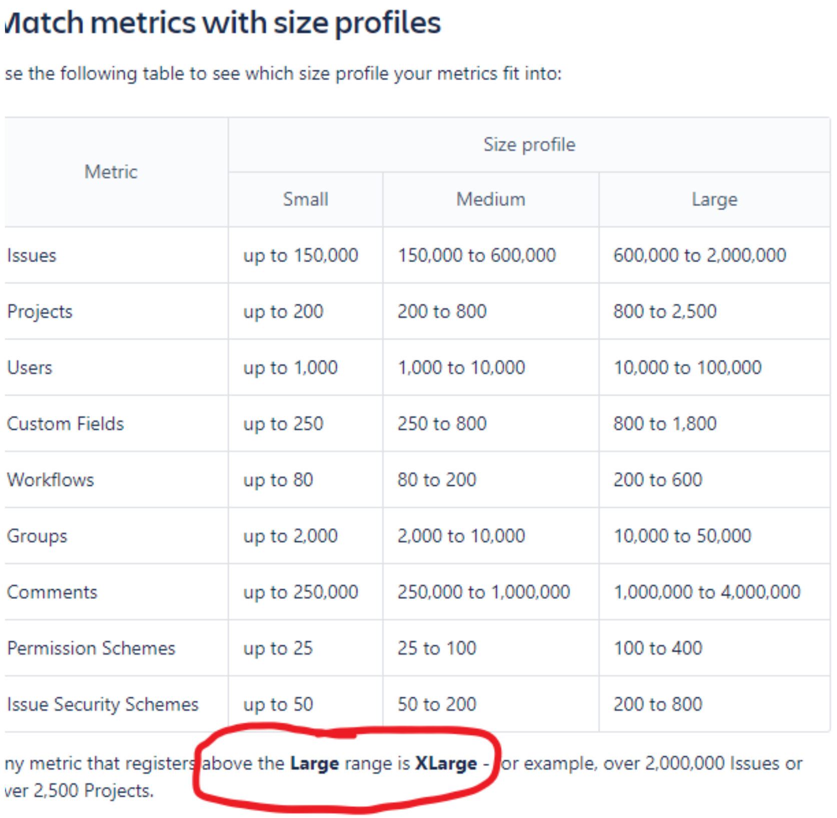
Our Examples and Results

What Size is Your Data Footprint?

https://www.disneyplus.com/series/gravity-falls-shorts/6TfHWav6TOXv

Atlassian Data Center Load Profiles

Jira



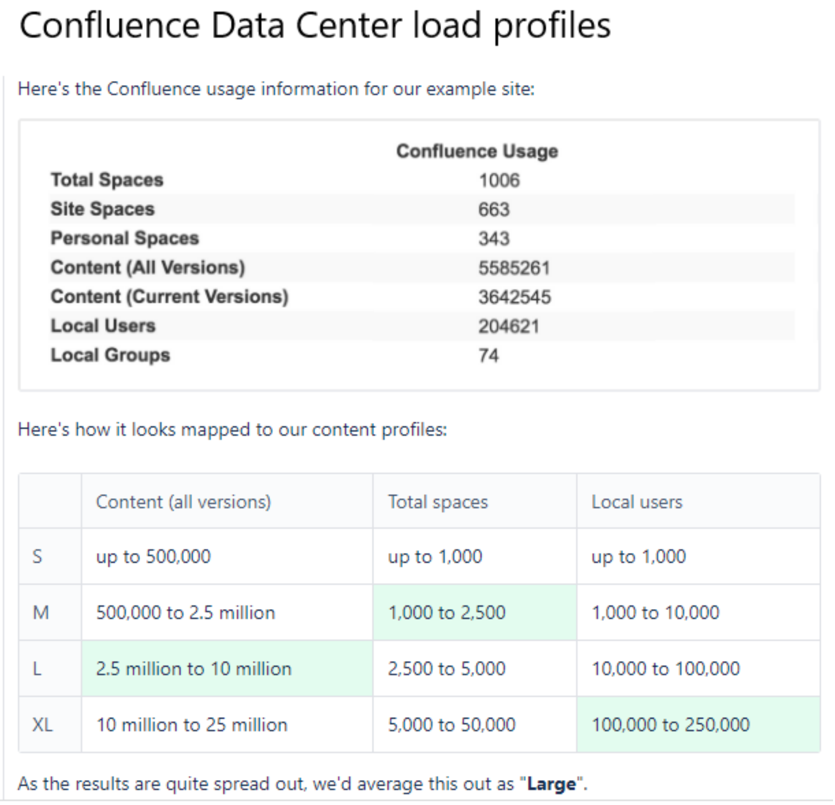
Match metrics with size profiles

se the following table to see which size profile your metrics fit into:

Metric	Size profile		
	Small	Medium	Large
Issues	up to 150,000	150,000 to 600,000	600,000 to 2,000,000
Projects	up to 200	200 to 800	800 to 2,500
Users	up to 1,000	1,000 to 10,000	10,000 to 100,000
Custom Fields	up to 250	250 to 800	800 to 1,800
Workflows	up to 80	80 to 200	200 to 600
Groups	up to 2,000	2,000 to 10,000	10,000 to 50,000
Comments	up to 250,000	250,000 to 1,000,000	1,000,000 to 4,000,000
Permission Schemes	up to 25	25 to 100	100 to 400
Issue Security Schemes	up to 50	50 to 200	200 to 800

ny metric that registers above the **Large** range is **XLarge** - or example, over 2,000,000 Issues or ver 2,500 Projects.

Confluence



Confluence Data Center load profiles

Here's the Confluence usage information for our example site:

	Confluence Usage
Total Spaces	1006
Site Spaces	663
Personal Spaces	343
Content (All Versions)	5585261
Content (Current Versions)	3642545
Local Users	204621
Local Groups	74

Here's how it looks mapped to our content profiles:

	Content (all versions)	Total spaces	Local users
S	up to 500,000	up to 1,000	up to 1,000
M	500,000 to 2.5 million	1,000 to 2,500	1,000 to 10,000
L	2.5 million to 10 million	2,500 to 5,000	10,000 to 100,000
XL	10 million to 25 million	5,000 to 50,000	100,000 to 250,000

As the results are quite spread out, we'd average this out as "**Large**".

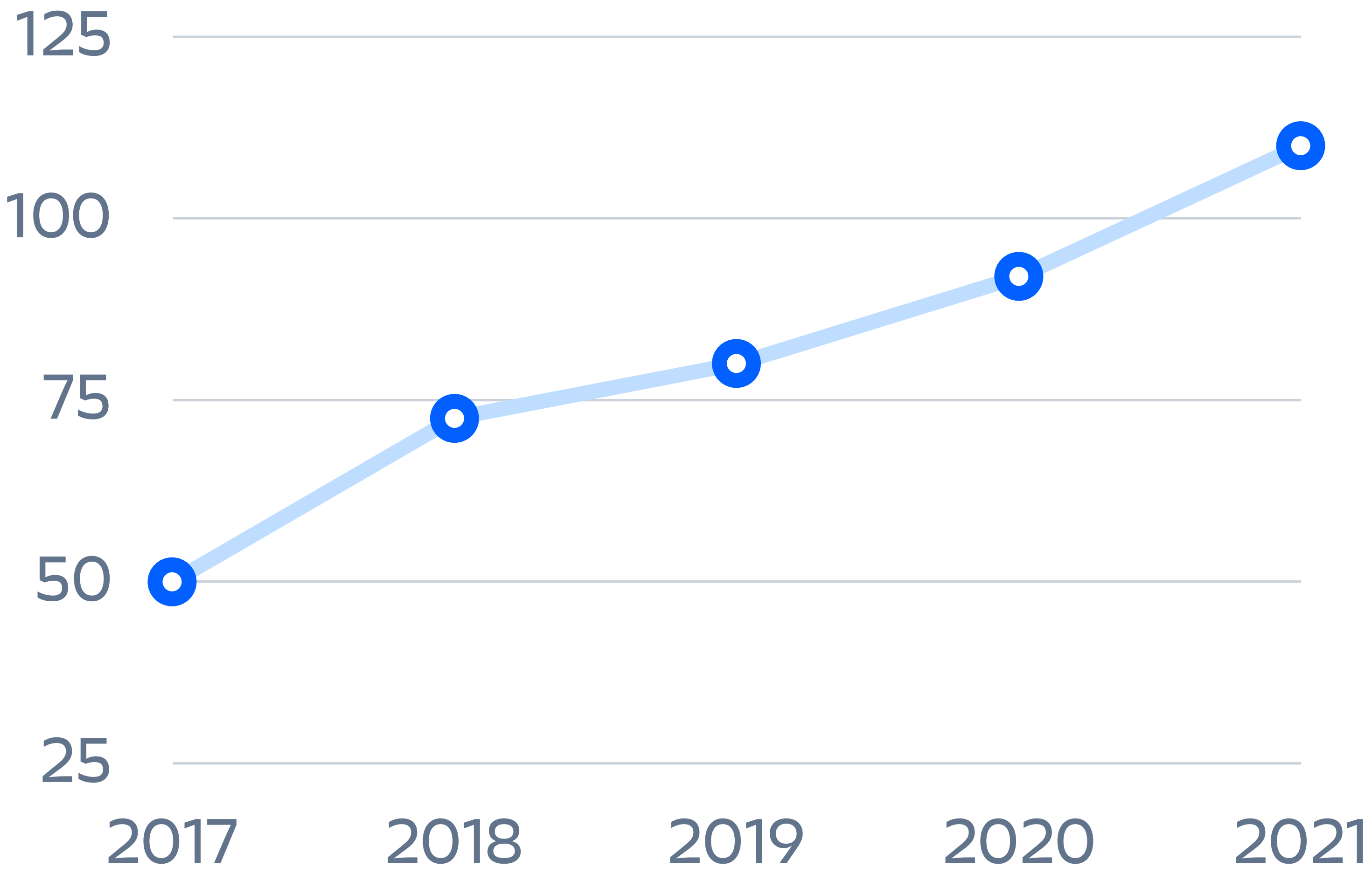
https://confluence.atlassian.com/enterprise/jira-data-center-size-profiles-955171062.html

https://confluence.atlassian.com/enterprise/confluence-data-center-load-profiles-946603546.html

GROWTH OF DATA FOOTPRINT FOR JIRA AND CONFLUENCE

Data Footprint

In Terra Bytes over past 4 years.

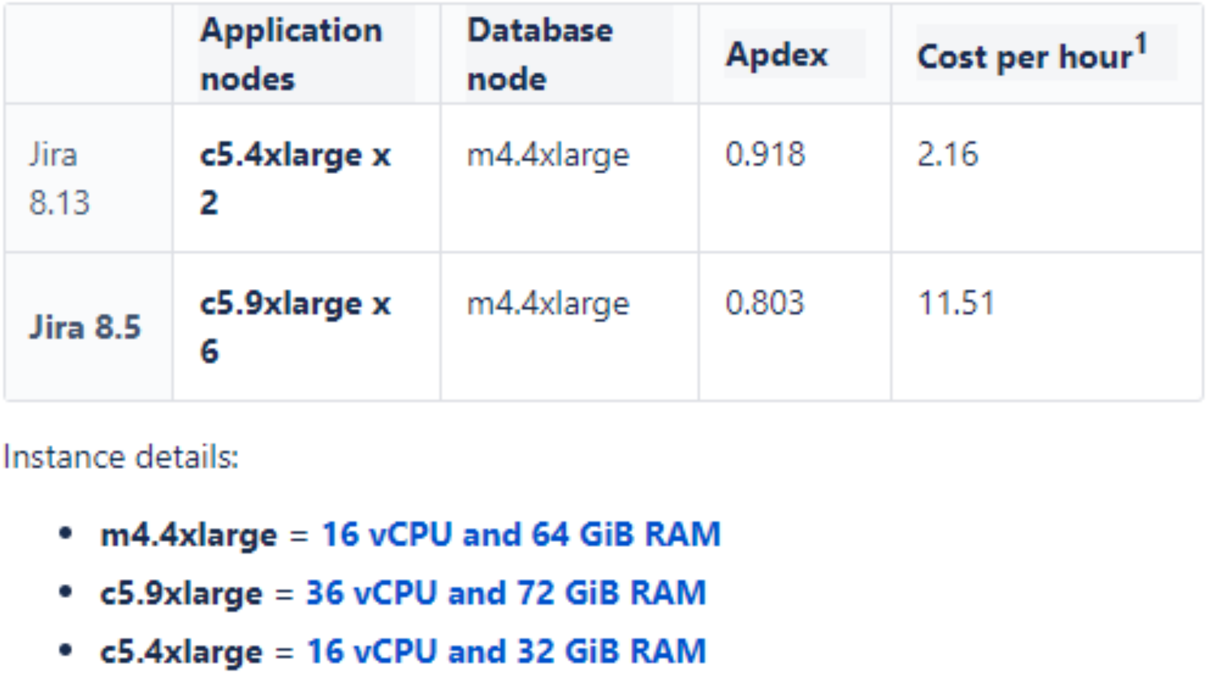


THAT'$ A BIG-Foot Print

https://www.etsy.com/listing/795814200/bigfoot-sasquatch-footprint-casting

What does it Cost Me?

Atlassian Recommendations for best performance and reliability - Jira XLarge



	Application nodes	Database node	Apdex	Cost per hour[1]
Jira 8.13	**c5.4xlarge x 2**	m4.4xlarge	0.918	2.16
Jira 8.5	**c5.9xlarge x 6**	m4.4xlarge	0.803	11.51

Instance details:

- **m4.4xlarge** = **16 vCPU and 64 GiB RAM**
- **c5.9xlarge** = **36 vCPU and 72 GiB RAM**
- **c5.4xlarge** = **16 vCPU and 32 GiB RAM**

https://confluence.atlassian.com/enterprise/infrastructure-recommendations-for-enterprise-jira-instances-on-aws-969532459.html

Agenda

Data Growth and Costs

Solution Options for Confluence

How to Clean up Confluence Trash

Solution Options for Jira

How to Purge Issues from Jira

Our Examples and Results

Confluence Options

Empty Trash

Delete Unused Spaces

Warehouse

Take Out the Trash

When a user deletes a page or attachment, it goes to the trash folder for the space – and stays there.

/pages/viewtrash.action?key=SPACEKEY

Identify and Delete Unused Spaces

Use analytics or SQL queries to identify unused Spaces, then delete with automation.

Warehouse Unused Spaces

Use snapshots to store and restore spaces from year *n*. Restore to warehouse app cluster as requested.

ONE SIDE

Trash: no downtime, reduces shared drive costs. Custom script or add-on.

Delete: no downtime, reduces shared drive costs, reduces DB size, improves performance

Warehouse: no data lost

THE OTHER SIDE

Trash: no performance increase, does not reduce content.

Delete: data is lost

Warehouse: costs more to store snapshots, have to maintain a warehouse cluster

Agenda

Data Growth and Costs

Solution Options for Confluence

How to Clean up Confluence Trash

Solution Options for Jira

How to Purge Issues from Jira

Our Examples and Results

How To Take Out the Trash

Notify users

Refresh Test Environment

Space Keys

Parallel Scripts

Notify Users in Advance

At least 2 weeks - Explain only trash items are being removed, no content will be deleted.

Refresh Your Test Environment

In case trash items were being used as 'draft' storage.

Assign Space Keys

Get list of and assign Space keys to scripts. We assign about 1500 spaces per script.

Run Scripts in Parallel

Two scripts per node. Each script works on different Spaces. One minute wait after each Space.

REST Endpoints

Delete Trash by Item

```
/admin/permissions/pagepermsadmin.action
```

Delete Trash by Space

https://confluence.atlassian.com/confkb/how-to-purge-all-remove-all-trash-in-a-space-using-rest-api-1063559677.html

```
rest/api/content?spaceKey={KEY}&status=trashed
```

```
curl -u user:password -X POST -H "Content-type: application/json" http://localhost:8090/rpc/json-rpc/confluenceservice-v2/emptyTrash -d '["SPACE_KEY"]'
```

Delete Spaces

https://docs.atlassian.com/atlassian-confluence/REST/6.6.0/

```
DELETE /rest/space/{spaceKey}
```

DO NOT

Run Scripts over 24 hours.

Run more than 2 scripts per node.

DO

Rebuild Ancestor tables.

Notify Users.

Run on Test Server.

Restore Test server.

Run in DEBUG mode.

Agenda

Data Growth and Costs

Solution Options for Confluence

How to Clean up Confluence Trash

Solution Options for Jira

How to Purge Issues from Jira

Our Examples and Results

Jira Options

Archive

Warehouse

Cloud

Purge

Archive

Unused projects and issues

Warehouse

Take snapshots every year, and restore to warehouse cluster as needed.

Cloud

Migrate to Atlassian Cloud, or store snapshots in lower cost Cloud storage.

Purge

Delete Jira issues, attachments, database history and relationships.   Custom script or bulk change.

PROS

ARCHIVE: no downtime, improves performance, reduces index size

WAREHOUSE: No data lost, may improve performance

CLOUD: lower cost for long-term storage. And see Atlassian booth

CONS

ARCHIVE : does not reduce shared drive space or database size

WAREHOUSE: cost of storage and app cluster. Do you purge production data?

CLOUD: security concerns, less customizations. See Atlassian booth

PURGING PROS

Run after hours, no downtime

Reduces shared drive and DB size

Lower row counts on X-Large tables in DB

Works well for > 2K custom fields, and > 3TB of attachments

Can be used in combination with warehousing data

PURGING CONS

Can take over 2 days to complete. Bulk changes can lockup node

Issues and attachments are removed permanently

Legal and corporate exclusions can limit the number of issues to delete

Can delete issues still in use

Agenda

Data Growth and Costs

Solution Options for Confluence

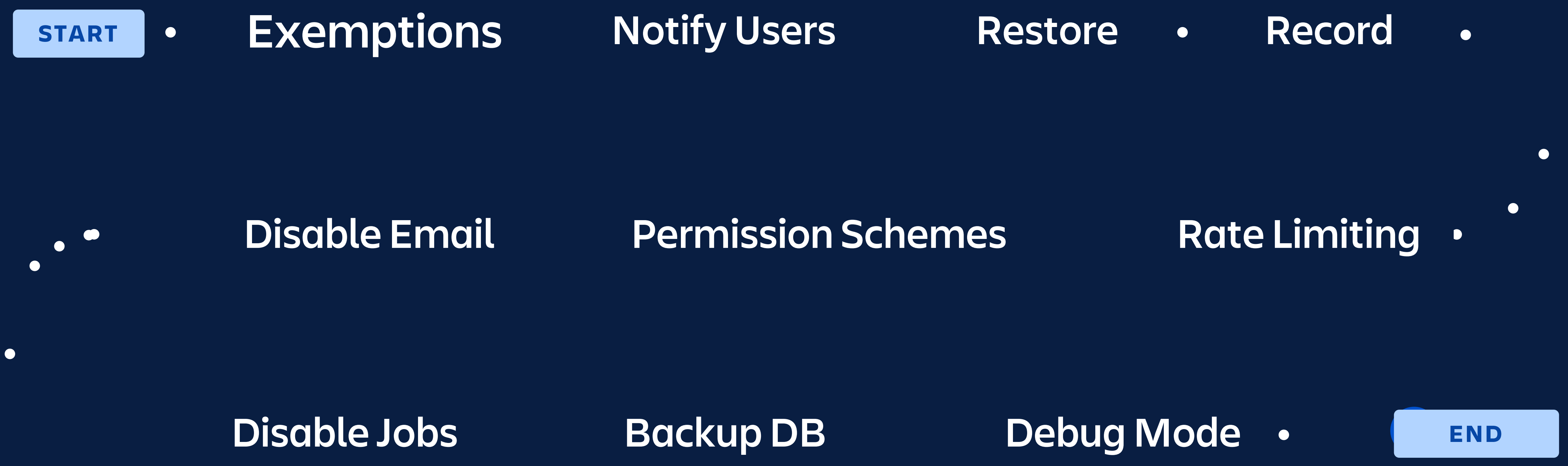
How to Clean up Confluence Trash

Solution Options for Jira

How to Purge Issues from Jira

Our Examples and Results

STEPS BEFORE DELETING JIRA ISSUES



START

Exemptions

Notify Users

Restore

Record

Rate Limiting

Permission Schemes

Disable Email

Disable Jobs

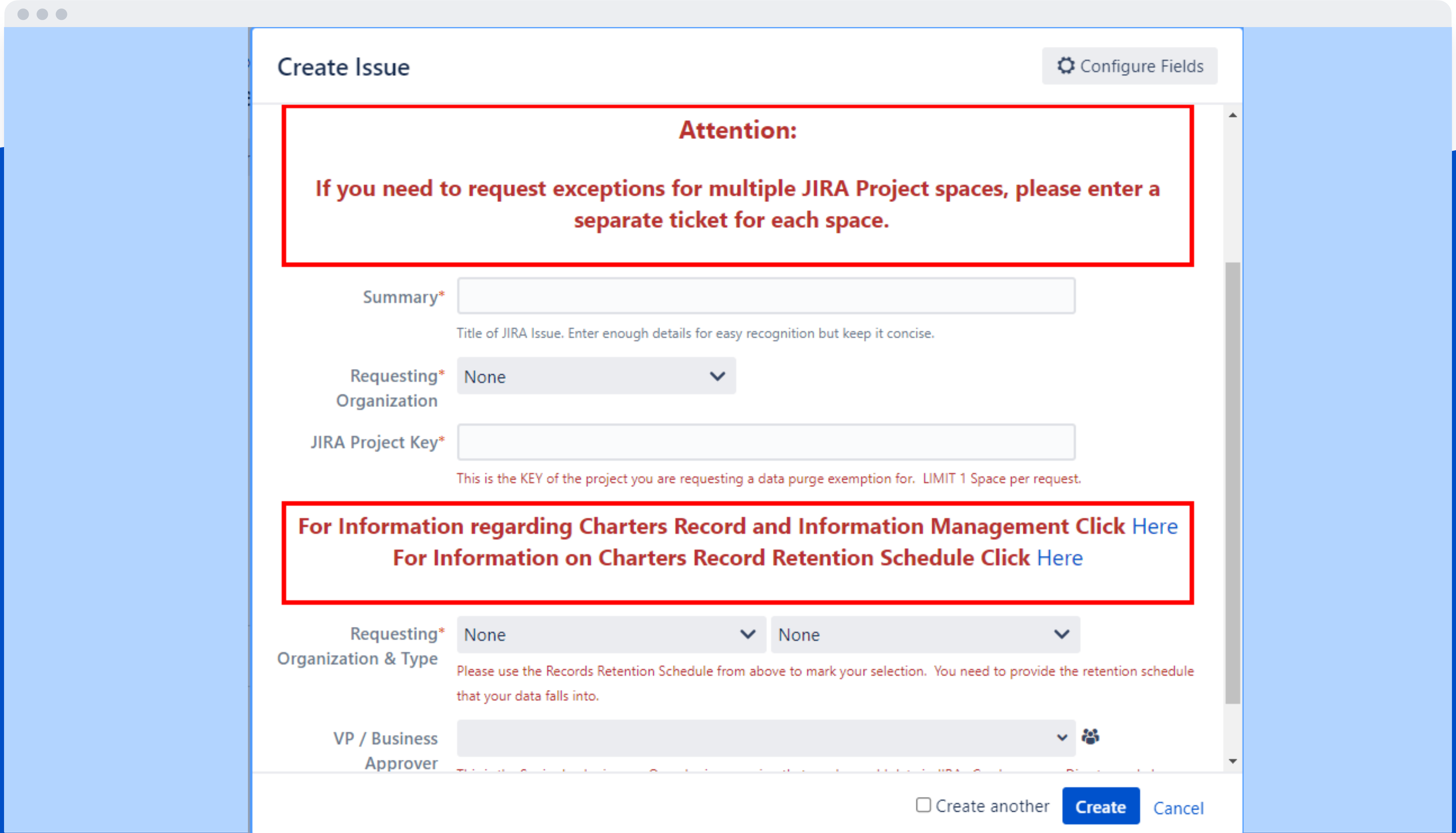
Backup DB

Debug Mode

END

JIRA – EXEMPTION REQUEST EXAMPLE

Request » Review » Confirm » Approve » Encode » Close



Attention:

If you need to request exceptions for multiple JIRA Project spaces, please enter a separate ticket for each space.

Summary*

Title of JIRA Issue. Enter enough details for easy recognition but keep it concise.

Requesting* Organization: None

JIRA Project Key*

This is the KEY of the project you are requesting a data purge exemption for. LIMIT 1 Space per request.

For Information regarding Charters Record and Information Management Click Here
For Information on Charters Record Retention Schedule Click Here

Requesting* Organization & Type: None | None

Please use the Records Retention Schedule from above to mark your selection. You need to provide the retention schedule that your data falls into.

VP / Business Approver

How To Delete Issues

Common Files

Unique Projects

REST API or Java

Re-Index

Parallel Scripts Use Common Files

Script 1 runs all SQL, reads all files, assigns projects, and writes common files for all scripts.

Unique Set of Projects Per Script

Un-Archiving, Re-Archiving, and weighted delete counts per script are more efficient with unique project lists.

Use the REST API or Java to Delete

Exemption rules for each issue. Atlassian API's to delete. Includes sub-tasks, attachments, and DB.

Java Component

Scriptrunner

deleteIssue()

bulk edit

```
#Scriptrunner

https://community.atlassian.com/t5/Jira-questions/Ho-to-quickly-delete-40-50k-issues-from-JIRA/qaq-p/461530

import com.Atlassian.jira.component.ComponentAccessor

def issueManager = ComponentAccessor.getIssueManager();

issueManager.deleteIssueNoEvent(issue);


#Bulk Change

https://confluence.atlassian.com/jirakb/bulk-editing-more-than-10000-issues-will-result-in-xsrf-security-token-missing-961791994.html

#jira-application.properties
jira.bulk.edit.limit.issue.count = 1000;


#tomcat server.xml
maxParameterCount="10100"
```

REST Endpoints

Delete Issue

Unarchive Project

```
my_url = jira_node_url + '/rest/Api/2/issue/'
+ str(issue_id) + '?deleteSubtasks=true'

response = requests.delete(my_url,
timeout=120, headers=my_headers,
data=post_data, verify=my_verify)

my_url = jira_node_url +
'/rest/api/2/project/' + str(project_id)

response = requests.put(my_url + '/restore',
timeout=240, headers=my_headers, data='{}',
verify=my_verify)
```

While In Progress

Check Logs

Monitor Nodes

Send Updates

Estimate Completion

Measure Delete and Error Rates

The delete rate should be about 1 issue per second. The error rate should be < 2 %.

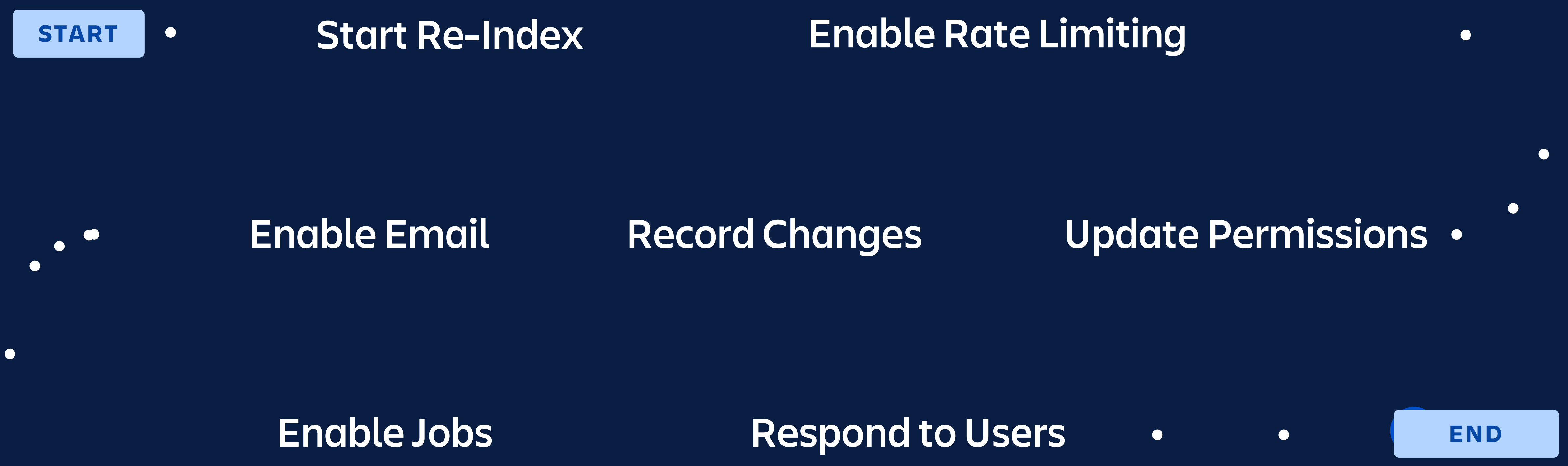
Check Instance Health and Node Status

Use the troubleshooting page in Jira, Cluster status, and any custom alerts to verify all nodes online.

Send Updates, Schedule Next Run

It may 2 – 3 days. Each run is 12 – 20 hours. Re-index and send updates after each run. Estimate completion.

STEPS AFTER DELETING JIRA ISSUES



START

Start Re-Index

Enable Rate Limiting

Update Permissions

Record Changes

Enable Email

Enable Jobs

Respond to Users

END

DO NOT

Run Scripts over 24 hours.

Run more than 2 scripts per node.

Use bulk changes.

DO

Notify Users.

Run on Test Server.

Restore Test server.

Run in DEBUG mode.

Re-Index.

Agenda

Data Growth and Costs

Solution Options for Confluence

How to Clean up Confluence Trash

Solution Options for Jira

How to Purge Issues from Jira

Our Examples and Results

Charter Results

Jira

Shared Drive and DB

1.5 million issues deleted - 50%

80 million rows removed from customfieldvalue - 58%

600GB removed on shared drive - 20%

10GB removed from Index - 42%

Confluence

Shared Drive

750GB removed on shared drive - 15%

DATA GROWTH IS NORMAL

Without regular purging, our

Jira issue count > 5 million

Each shared drive > 5 TB

Follow Up Items to Research

Jira

Purge Avatars

research how to identify and remove unused Jira avatars

Confluence

Delete Spaces

Use analytics and SQL queries to identify unused pages and spaces

Questions?

ATLASSIAN
team '22

Charter
COMMUNICATIONS

Thank you for attending!