#ASLI ENGINEERING

Dissecting Atlassian's Mega Outage

SWIPE

BY

ARPIT BHAYANI

Dissecting Atlassian's Outage

April 2022

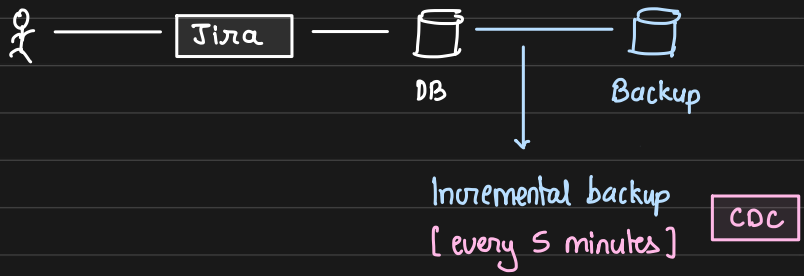
What happened? 400+ Cloud Customers experienced full outage across Atlassian products

Data of these customers was "permanently deleted" and it will take a few weeks to restore the data

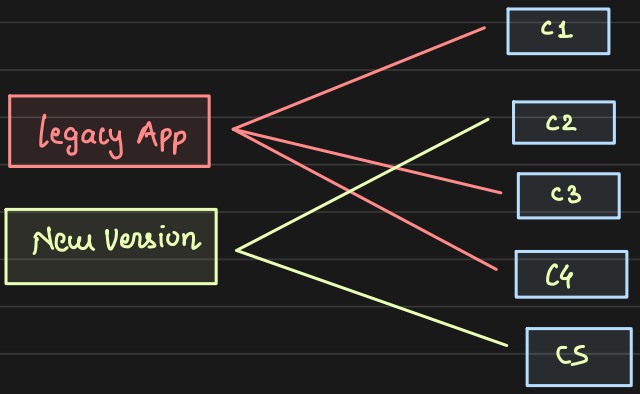
This outage teaches us so much about

- Multi-tenant architecture
- Importance of having a Disaster Recovery Strategy
- provides us the insights about their architecture and engineering practices

Insight 1: reported data loss for up to 5 minutes prior to incident



Insight 2: Product upgrade: Deactivating legacy app



Deactivate legacy app infavour of a shiny new version!

* They used their existing scripts to deactivate the legacy version

Insight 3: Mark for deletion v/s Permanently Delete

Mark for deletion : - soft delete — is_deleted = True
- recovery is possible

Permanent deletion : - hard delete — DELETE FROM t ...
- recovery not possible

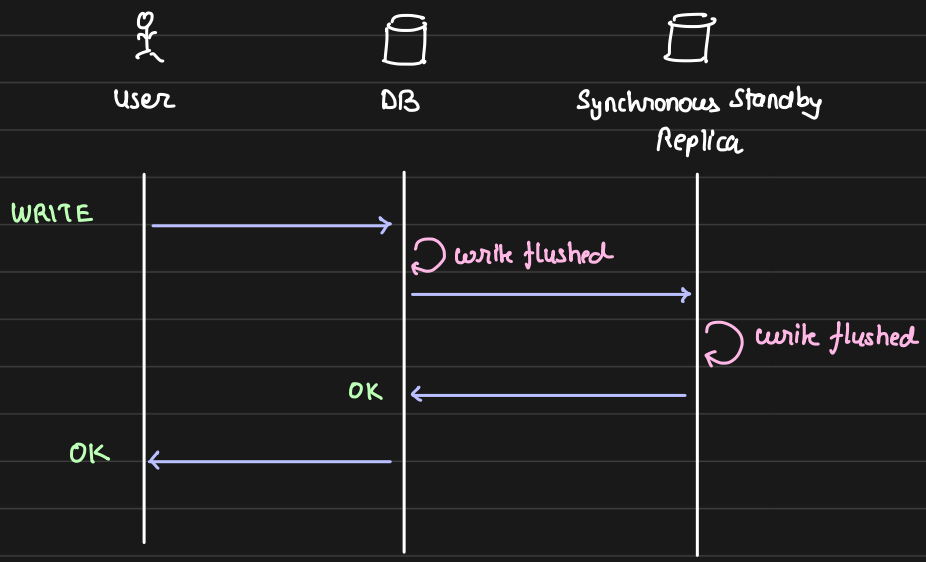
Why we ever permanently delete? Compliance

GDPR gives users the

'Right to be forgotten'

Insight 4: High Availability using Synchronous replication

To maintain high availability : Synchronous Standby Replicas



Insight 5: Immutable Backups

Standby Replica

Serialized into some format

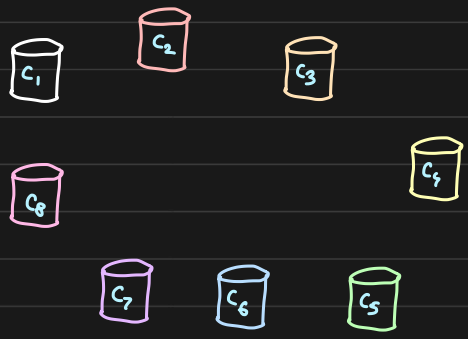
Protection against data corruption

Error correction codes

Immutable backup

Insight 6: Not truly a multi-tenant architecture [data]

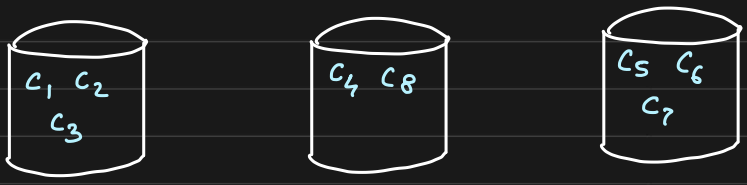
Each customer should have an isolated & independent setup in a multi-tenant arch



Making it easy to backup and restore data of individual customer

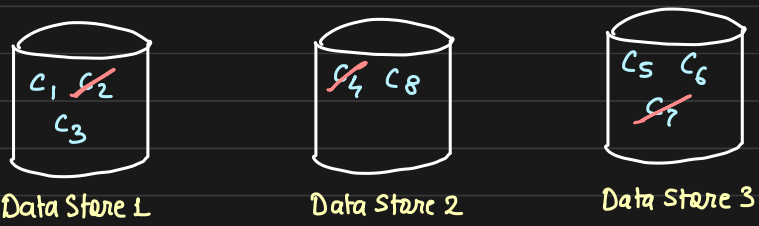
But if we have individual DB for each customer, the cost will be MASSIVE

So, atlassian, like many other companies does this



- Data of multiple customers are present in single data store
- Because the incremental backups are customer-agnostic, they are backed up together on the backup site. (S3)

Why it taking time to restore the data ?



Say they deleted data of customers c_2, c_4 and c_7

But because in the archive the data of the entire data store is kept, it is extremely time consuming to restore the fragment of information

id	c_id	.....
100	c_1	
125	c_2	
722	c_1	
967	c_3	
1096	c_2	
2709	c_2	

eg: Restoring the entire data of Data Store 1 will also restore it for unaffected c_1 and c_3

Restoring the entire DB is easy

Snapshot → restore

But it will also restore the data of undeleted customers

↓

leading to data loss !!