Real World Repairs @ Netflix

Vinay Chella

Cassandra MVP, Cloud Data Architect

Agenda

- What is Repair and Why?
- How do we Repair @ Netflix?
- Use cases
- Pain points
- Alternatives to Repair
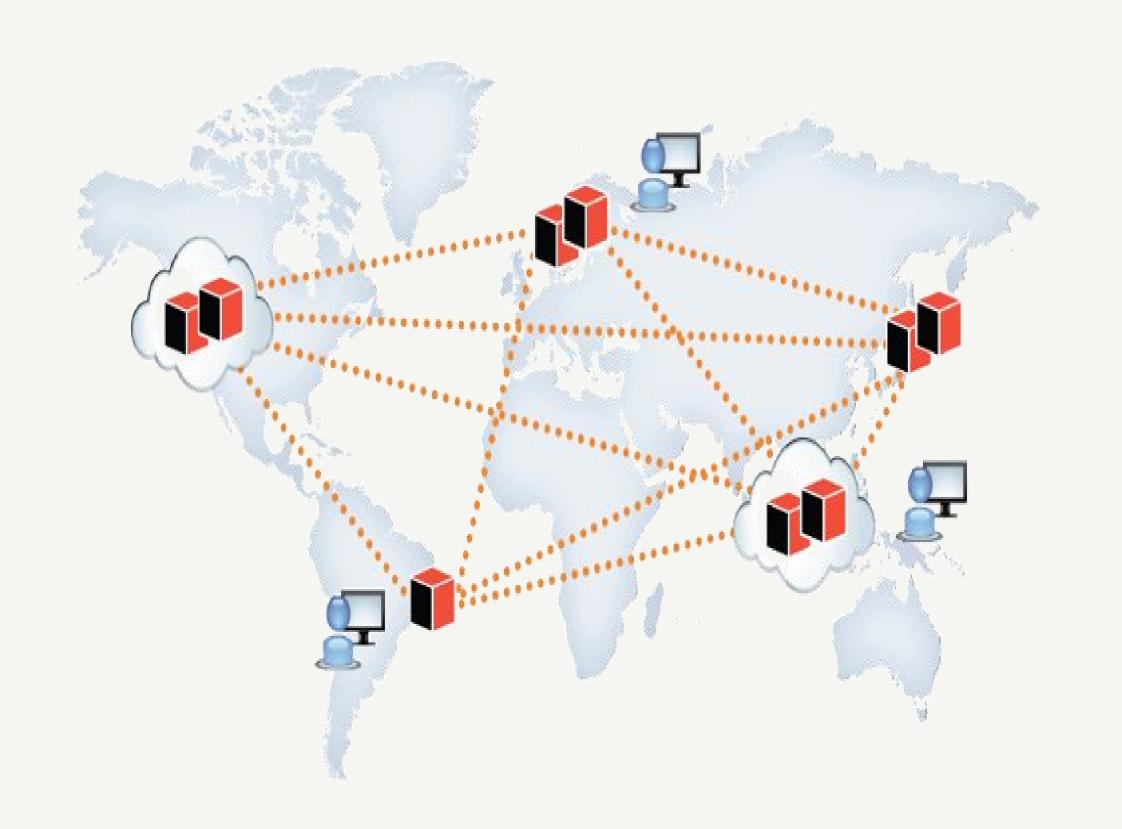- When to/ not to repair
- What is missing in C*

Why is Repair Important?

- Entropy is inevitable

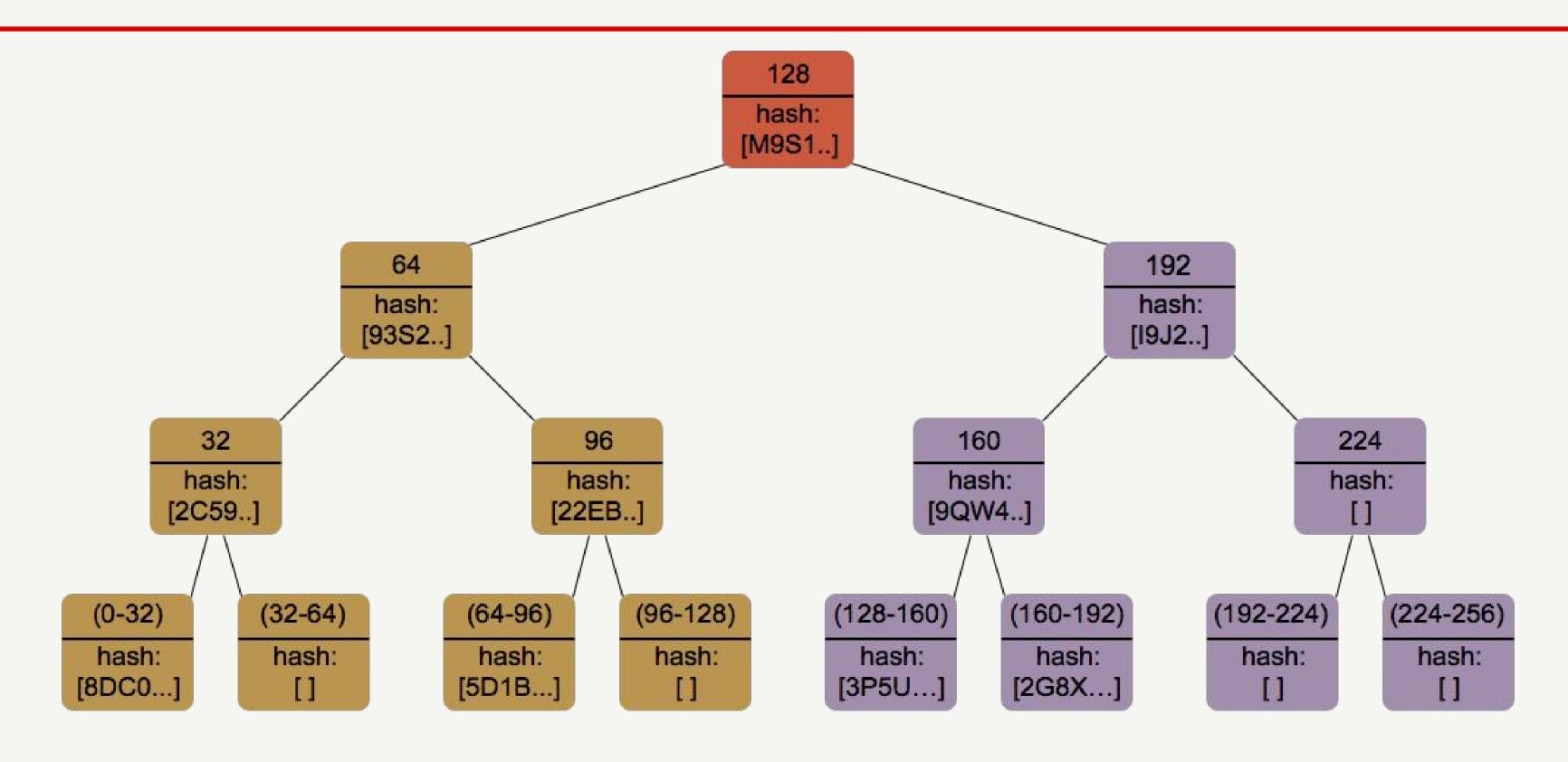- Inconsistent data might impact your business

- Data Resurrection

What is Repair

Repair in Netflix world

Case 1

- ~300 TB cluster size
- ~1 TB per node
- SSDs
- 288 Nodes - Multi Region Cluster
- Critical data, No TTLs

Case 2

- ~180 TB cluster size
- ~3.6 TB per node
- HDDs
- 48 Nodes – Island cluster
- No TTLs, long lived

NETFLIX

How do we Repair @ Netflix?

- Primary Range (**-pr**)

- Subrange (**-st** and **-et**)

- On every **node** in every **dc** sequentially

- Parallel (**-par**) & *SEQUENTIAL*

How do we Repair @ Netflix? Cont..

- More granular Repair
  - Selective tables only
- On schedule basis
  - Jenkins
- Repair progress tracking

Impacts of UnRepaired C*/ Use-cases

- Incorrect bookmark index

- Losing the ratings

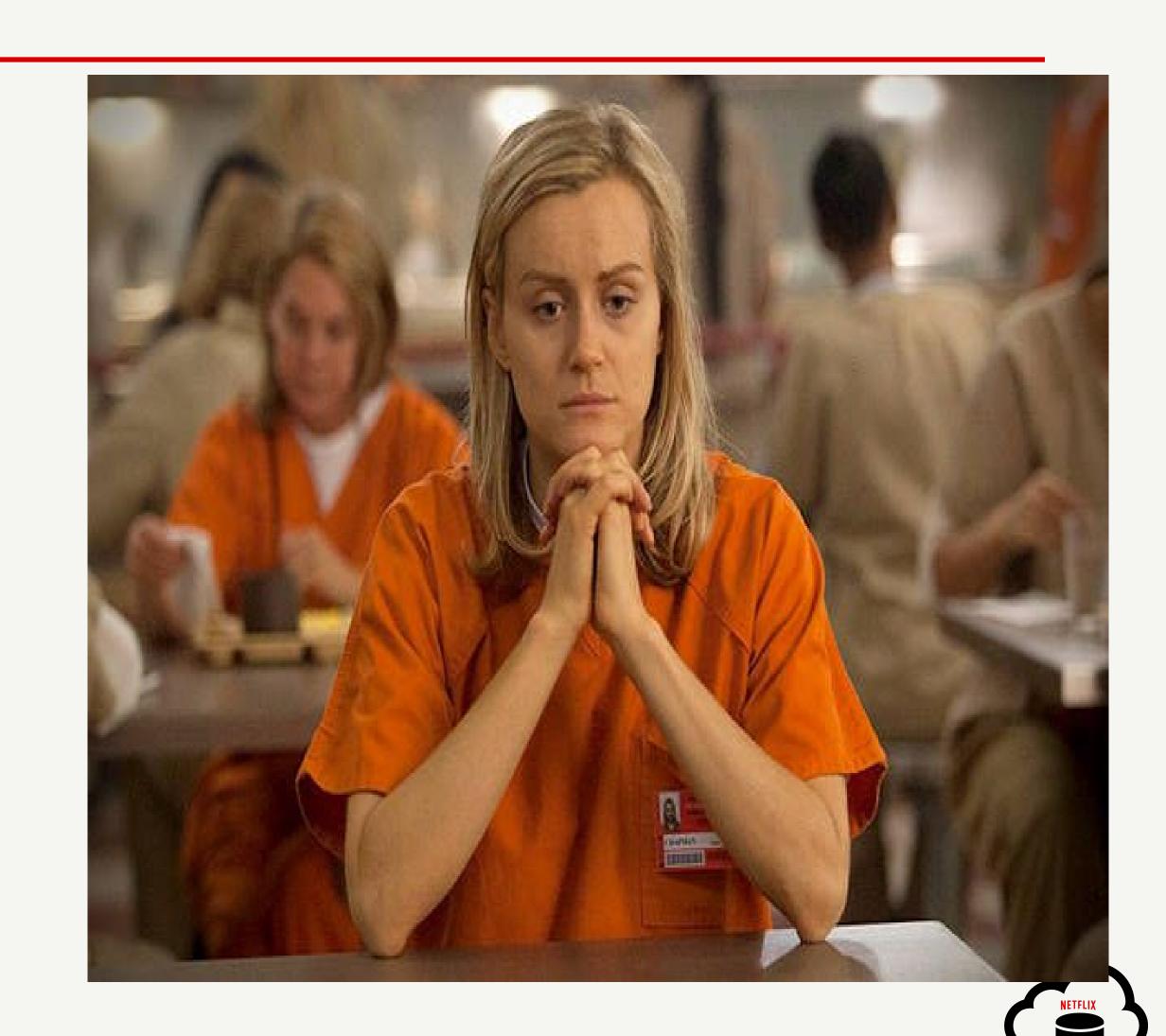- Recommendations impact

- Broken recently watched

Pain Points

- Over streaming

- Stream timeouts

- Stuck repairs

- Tracking/ resuming repair

- Compactions bottlenecks

- Disk I/O

- CPU usage

- Latencies/ SLA

- Wide partitions

NETFLIX

How do we live with it?

Well, we don't, we fix it.

Over streaming

- Subrange repair
  - Techniques
  - Algorithm

Subrange Repair

- Get subranges https://github.com/pauloricardomg/cassandra-list-subranges
  - Thrift: *describe_splits_ex*
- Repair those ranges using https://github.com/BrianGallew/cassandra_range_repair

Subrange Repair

Regular Repair

- Size per Node – **950 GB**
  - Time to repair – **~11 Hours**
    - Primary Range
    - *SERIAL*

- Size per Node – **550 GB**
  - Time to repair – **~24 Hours**
  - Primary Range
  - *SERIAL*
  - *Wide partitions*

Subrange Repair

- Size per Node – **950 GB**
  - Time to repair - **~2.5 Hours**
  - Primary Range
  - Threads #: 5
  - Split size 64K
- Size per Node – **550 GB**
  - Time to repair - **~5.5 Hours**
  - Primary Range
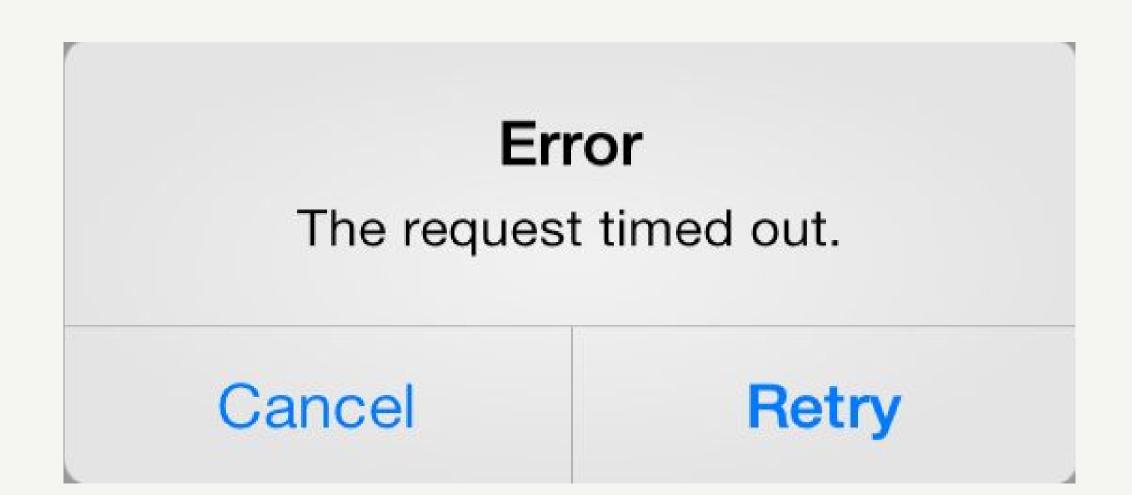 - Threads # - 10
  - Split size - 32K
  - Wide partitions

NETFLIX

To overcome stream timeouts

- YAML Tunings

    - *streaming_socket_timeout_in_ms*

    - *stream_throughput_outbound_megabits_*
      per_sec

    - *inter_dc_stream_throughput_outbound_*
      megabits_per_sec

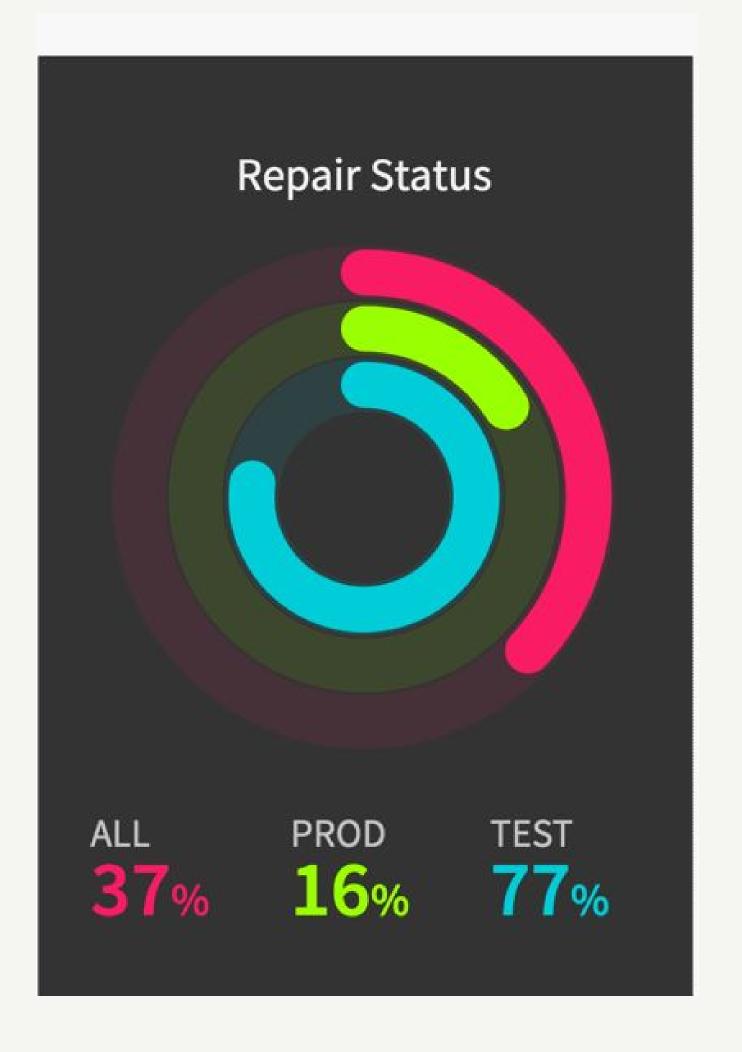**Error**

The request timed out.

Cancel   Retry

To overcome resume issue

- Track it
  - Record the progress
  - Record the status
- Resume-ability
  - Resume from failed or paused range

Repair Status

ALL	PROD	TEST
37%	16%	77%

NETFLIX

To fix stuck repairs

- Use Repair Tracking data
- Timeouts for long running repair
  - JMX notifications
- Self healable repair (Automate it !!)
  - Grep for repair logs
  - *nodetool compactionstats*
  - *nodetool netstats*
  - *JMX: forceTerminateAllRepairSessions()*

NETFLIX

Compaction bottlenecks

- Why compactions?
  - Validation phase

- Tune compaction settings
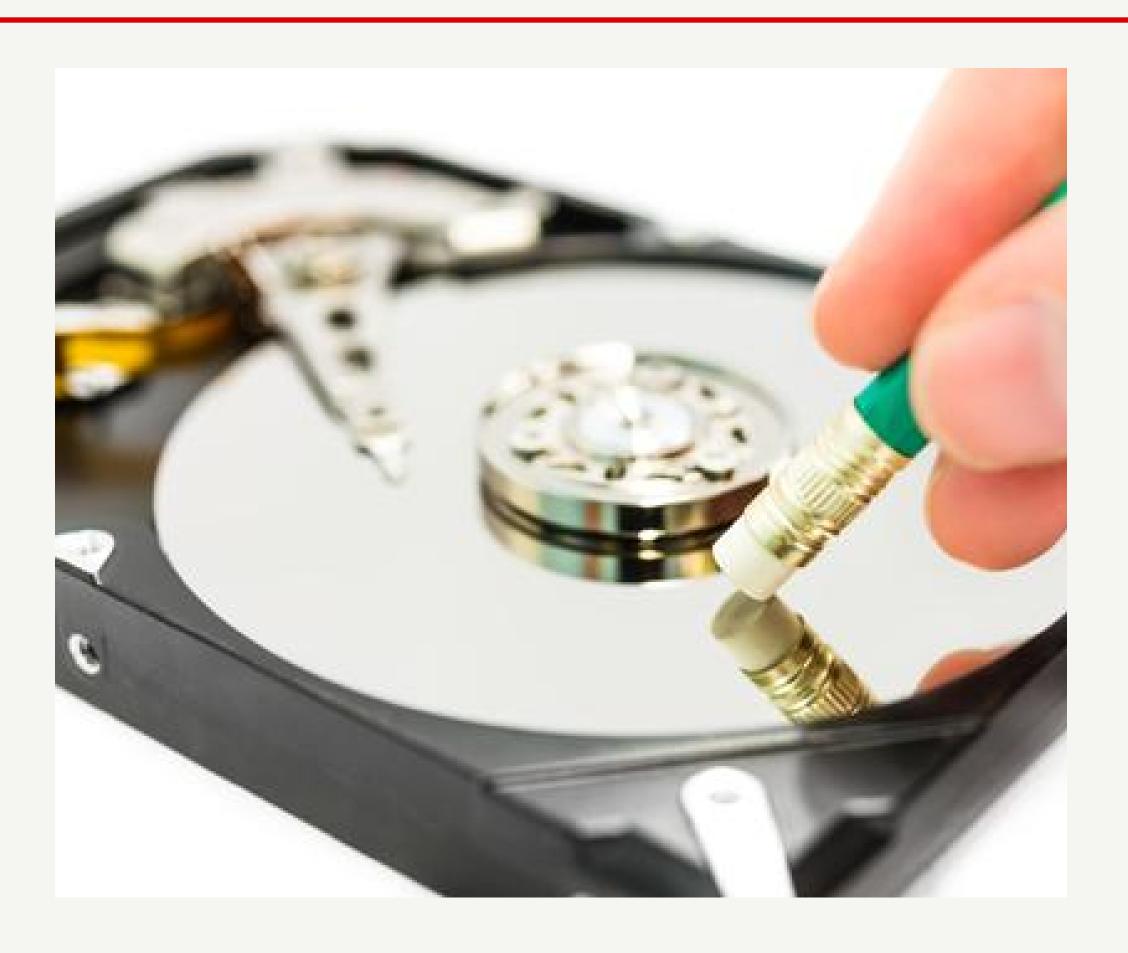 - *compaction_throughput_mb_per_sec*

Disk I/O issues

- SSD

High CPU Usage?

- Throttled streaming

  - *nodetool setstreamthroughput*

- Throttling subrange repair

  - Keysplit
  - #threads

To minimize impacts on latencies

- Running Repairs during off-peak hours

- Switch to SERIAL repair
  - Instead of PARALLEL

- If Subrange
  - Reduce #threads
  - Smaller subrange

Wide partition issues

- Reduce the partition size
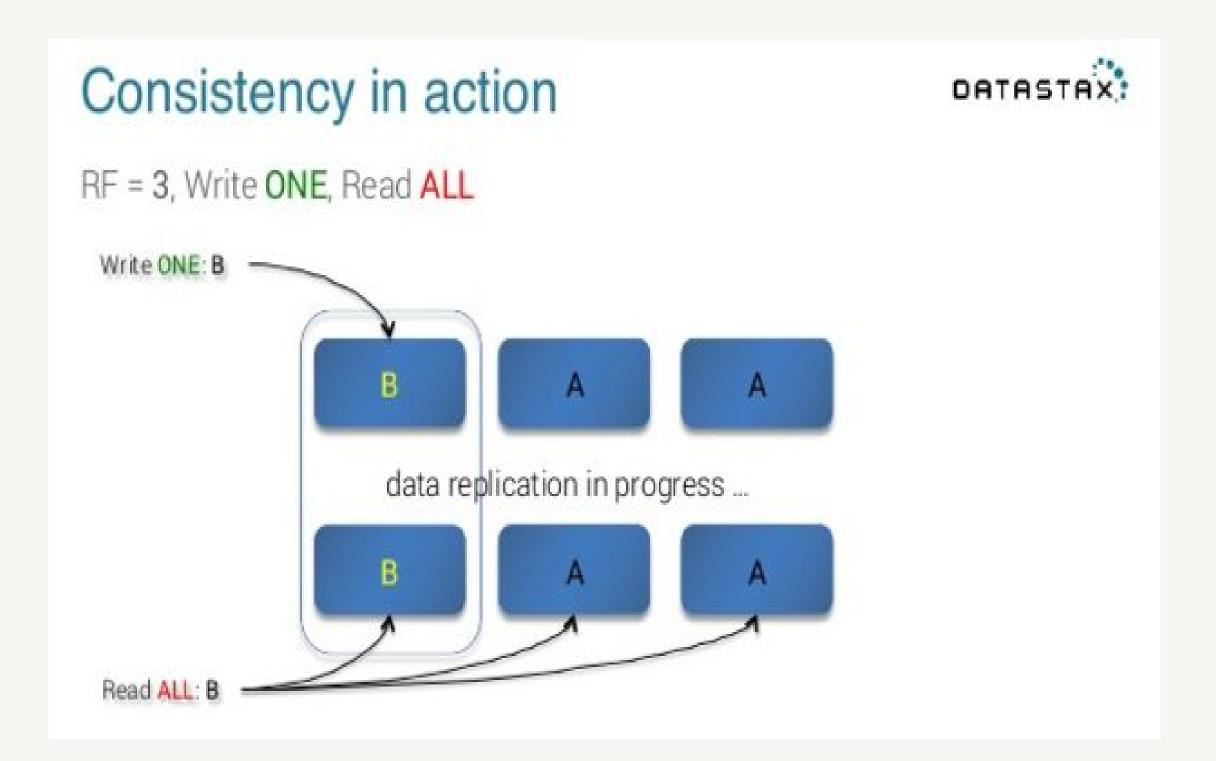
Other ways to Repair

Other ways to Repair: Tickler

- Row Tickler
- All Row Tickler

When to Repair

- Multi-Region clusters
- Low consistent read and writes
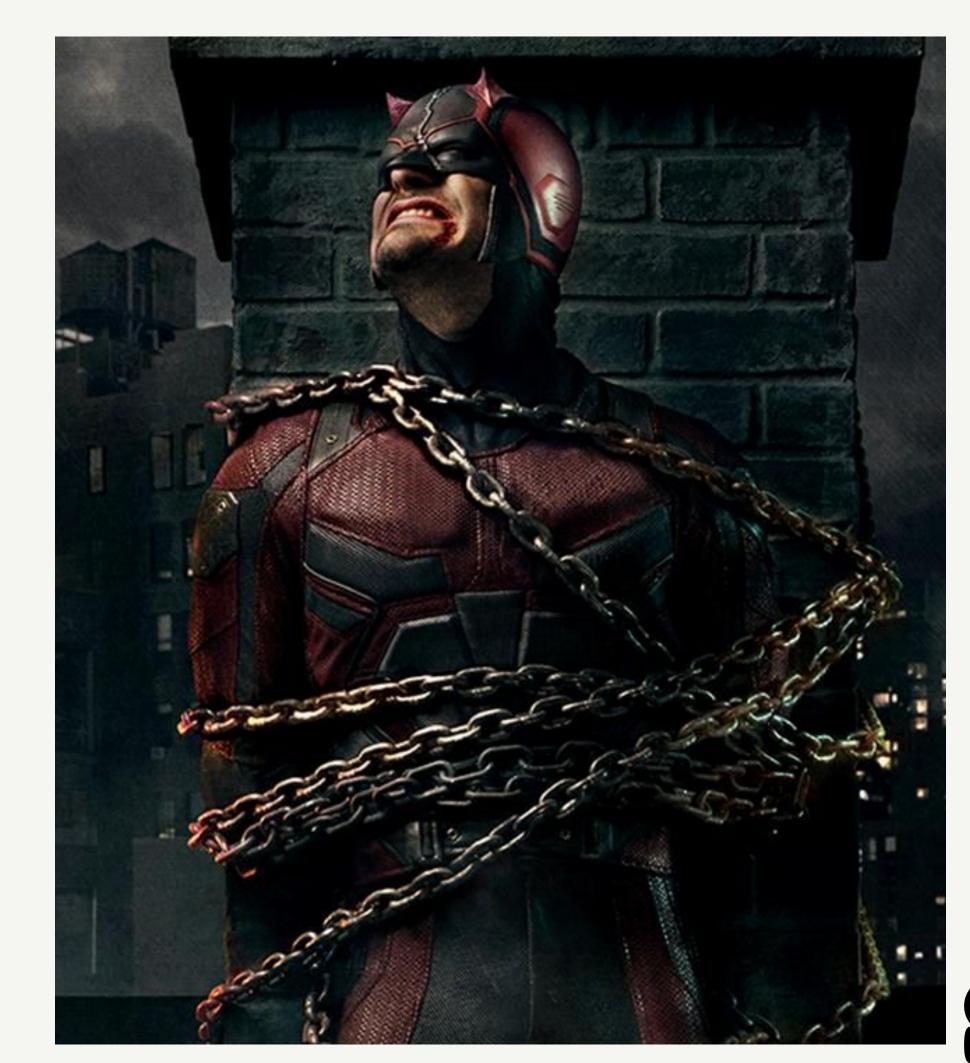- Frequent node outages
- Low read-repair chance
- Flaky networks

NETFLIX

When not to Repair

- Short TTls
- Highly consistent read and writes
- High read repair chance settings in island clusters

What is missing in C*

- Metrics and insights

- Incremental Repair bug fixes

- CASSANDRA-10070 - Automatic repair scheduling

- Mutation-based Repairs?

Photo Not Available

Q & A

NETFLIX