Unit 13: Merges

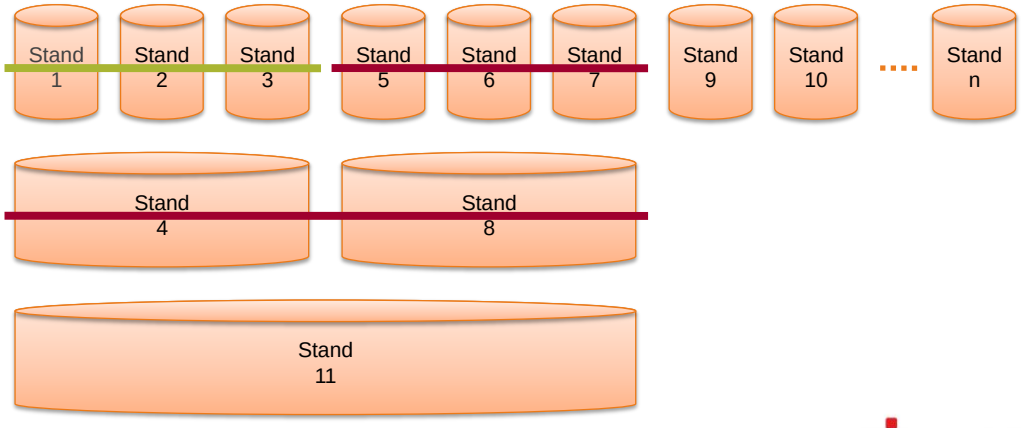
Learning Objectives

- Describe merge processes
- Define requirements for merge optimization
- Manage merge policy

MarkLogic UNIVERSITY

Merges: What?

- Stands merge in background to drive performance
- Query performance impacted by number of stands
- Small merges provide a significant performance boost and are less expensive than large merges in terms of CPU and disk space



MarkLogic UNIVERSITY

Merges: Why?

- Merges are a way for the system to self-tune
  - The server continually assesses the state of the database and decides if a merge is necessary
  - This is a Good Thing™
    - Reclaims disk space and improves performance
- I/O vs. CPU intensive operation?
  - Depends on the architecture of the system
  - Don't use your laptop to decide your merge policy

MarkLogic UNIVERSITY

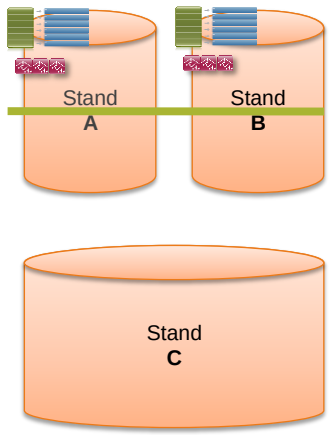
Merges: Why?

- Combining stands increases performance
  - Indexes are held per stand
  - More stands, more indexes to intersect
- Any deleted fragments are cleaned off the disk (subject to merge timestamp restrictions)
  - Reclaims disk space
- Indexes and lexicons are combined and optimized based on their new size
  - Increases performance

MarkLogic UNIVERSITY

Merges: How?

- Merging stand **A** and **B** into **C**
- Disk space is allocated for the new stand [2 × (A_{size}+ B_{size})]
- All of the non-obsolete fragments in **A** and **B** are copied into **C**
- Any index and lexicon data in **A** and **B** is merged together, optimized, and copied into stand **C**
- Stand **C** takes over for **A** and **B**. Any new index is now available
- Stands **A** and **B** are deleted



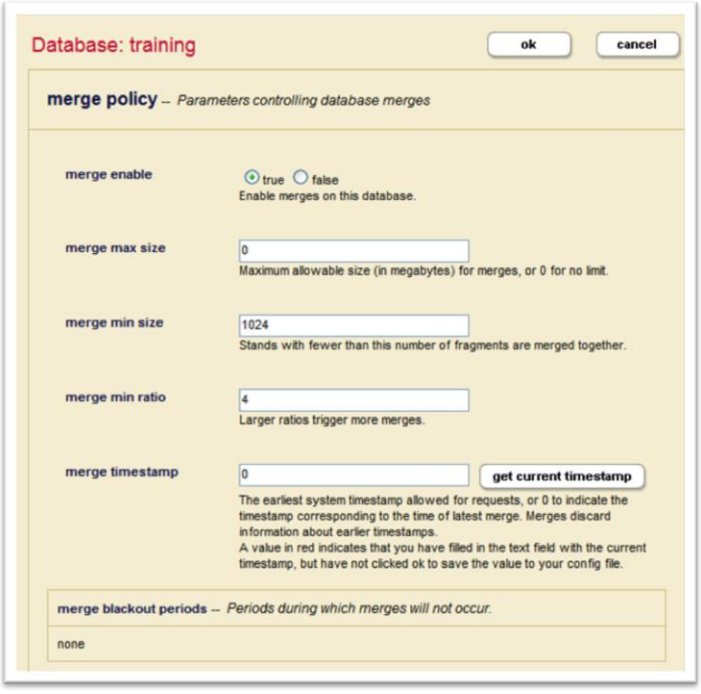
MarkLogic UNIVERSITY

Merges: When?

- By design, merges happen naturally
- We don't merge (completely) if:
  - Disk space requirements are not met
  - Merges have been disabled
- **Unless**
  - xdmp:merge() is issued
  - Admin interface merge button is clicked
    - Both of these override all merge settings
      - Still need adequate disk space

MarkLogic UNIVERSITY

Merges: Management



MarkLogic UNIVERSITY

Merges: Management

- You can prevent some or all merges from occurring
  - Menu in admin interface
  - This is not considered normal operation
  - The admin interface will warn you that merges are disabled
- Three ways to prevent merges
  - Completely disable for all forests in a database
  - Block merges above a certain size from occurring (using max merge size)
  - Set merge blackout periods
- While merges are disabled your stands will multiply quickly leading to performance degradation.
- Eventually a stand limit (64) will be hit and further **updates** will not be possible.

MarkLogic UNIVERSITY

Merges: Additional Details

- Whenever merges are disabled (completely or by a blackout) a message to that effect is logged once per hour
- Disabling merges, setting a max merge size, or entering a blackout period does not cause currently running merges to stop or abort
- If a merge is reduced in size due to the limit or disk space, a message to that effect is logged.
- Merge settings are at the database level
  - Merges always occur at the forest level
  - Unattached forests do not merge

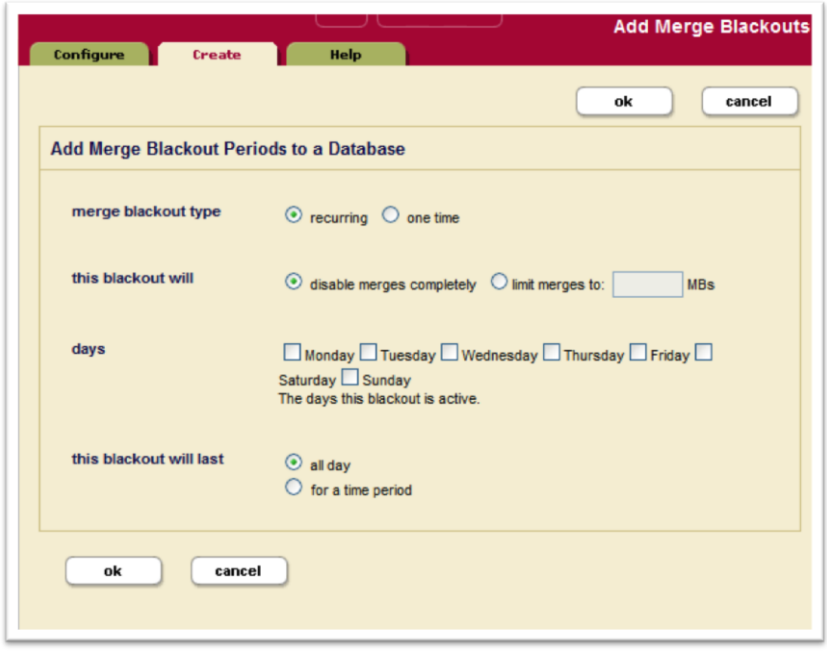
MarkLogic UNIVERSITY

Merges: Blackout Periods

- While it is recommended to let MarkLogic naturally handle merges, you may consider blackout periods for certain scenarios
- During known periods of high activity
  - Not wanting any non-essential CPU or I/O effort
- Going to run a lot of updates without many queries
- Sometimes good to keep small merges enabled and not do big ones
  - You get more benefit out of a small merge
  - Assess the tradeoff between processing time and query performance
- Any time you want to disable merges temporarily

MarkLogic UNIVERSITY

Merges: Blackout Periods



MarkLogic UNIVERSITY

Merges: Unable to Merge

- If a merge won't fit
  - Greater than max merge size
  - Not enough disk space (pessimistic)
- Do the best merge we can
  - Throw out big stands until merge fits
  - Log a warning
- If no two or more stands fit in the disk space limit forest is placed in the error state

MarkLogic UNIVERSITY

Merges: Cancelling

- Merges can be canceled from the database and forest status pages
- Each merge runs on its own forest
- If you don't set some merge control settings before cancelling, it is likely that another merge will begin immediately

MarkLogic UNIVERSITY

Unit 13: Applying the Learning Objectives

- Describe merge processes
- Define requirements for merge optimization
- Manage merge policy
  - Exercise 1

MarkLogic UNIVERSITY