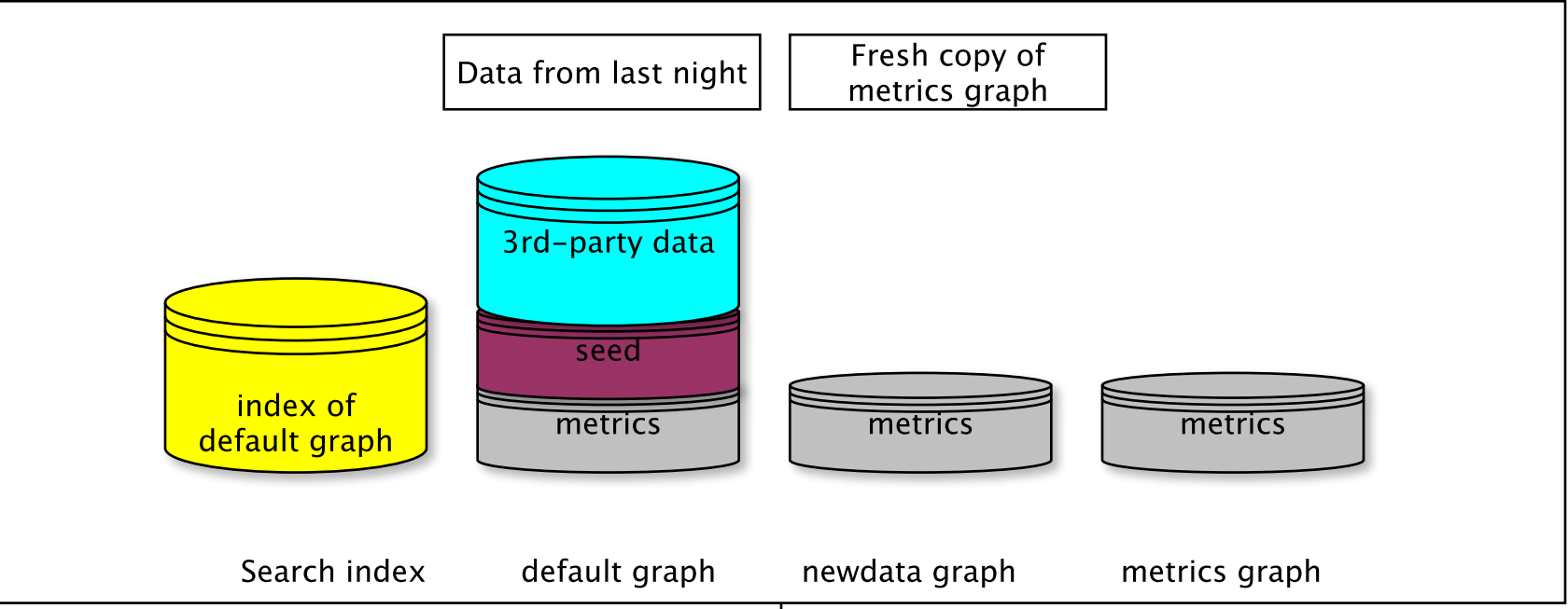


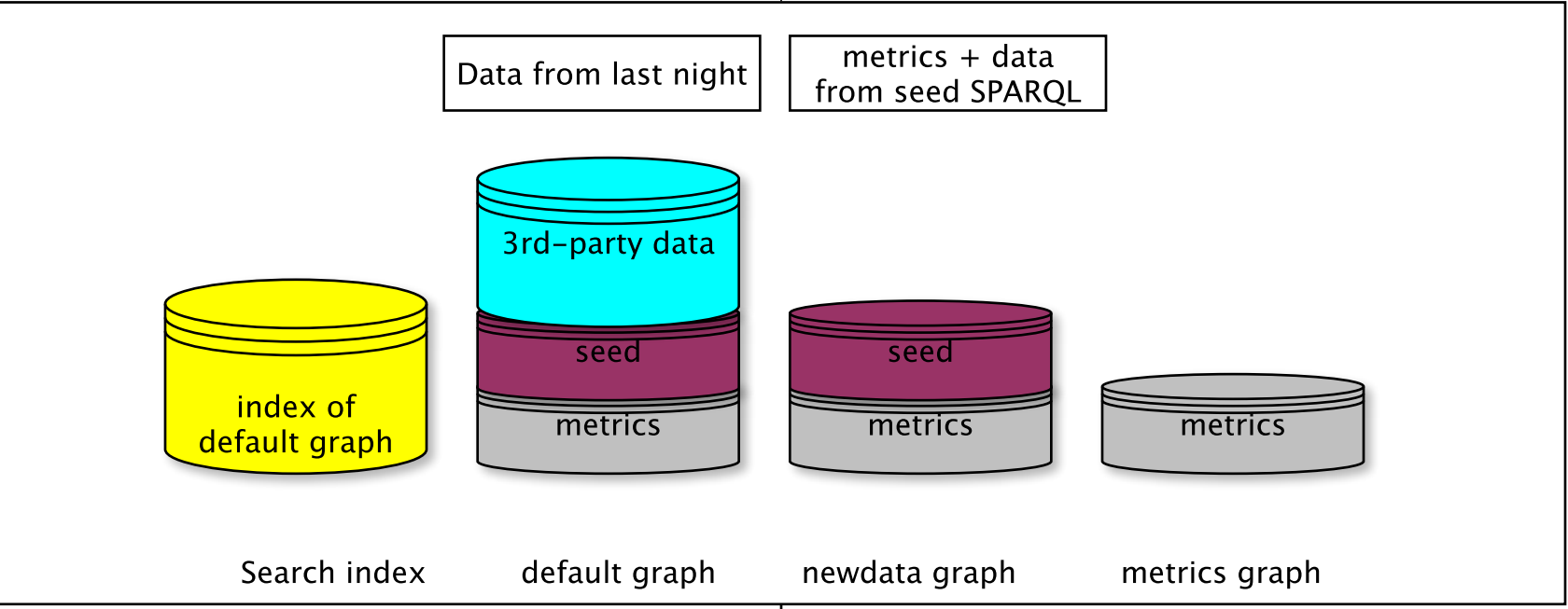
"Search index" is used to return search results
"default graph" is used to show details and diagrams
"newdata graph" and "metrics graph" are not queried by the live system,
but user activity is logged in the "metrics graph"

00:00
Clear newdata

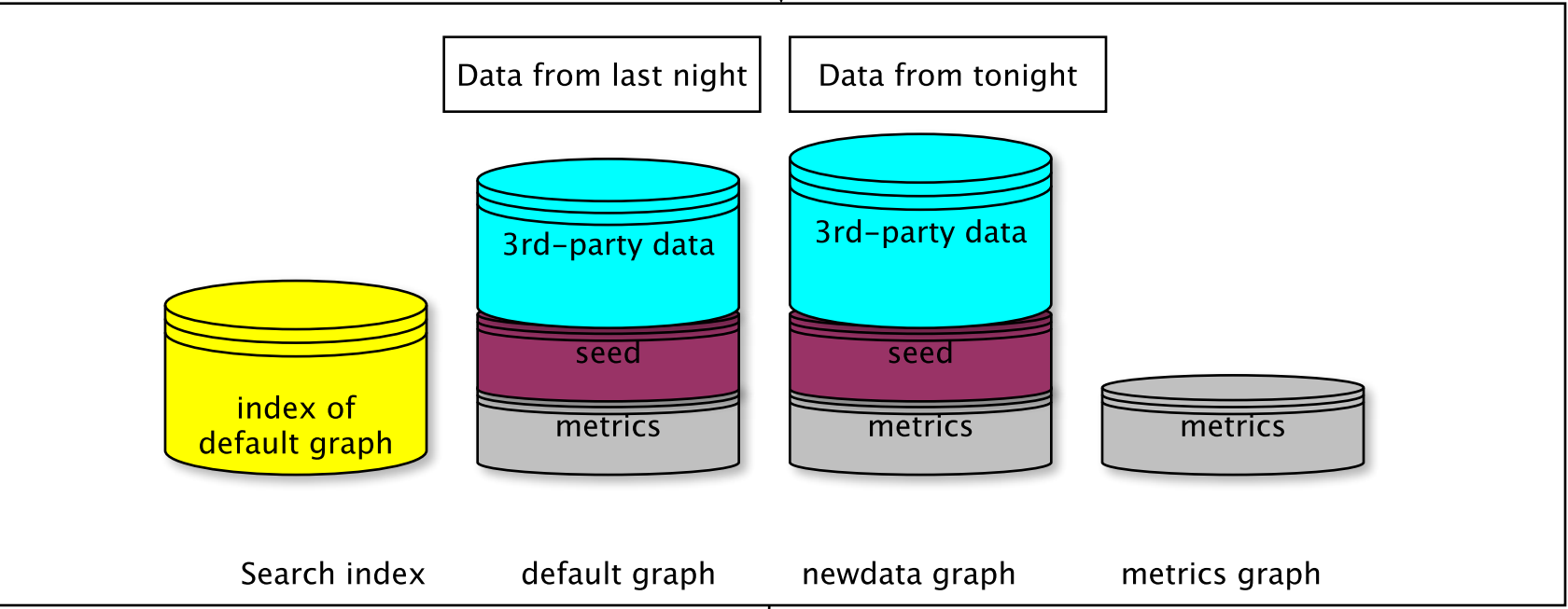


00:02
Add seed graph to newdata

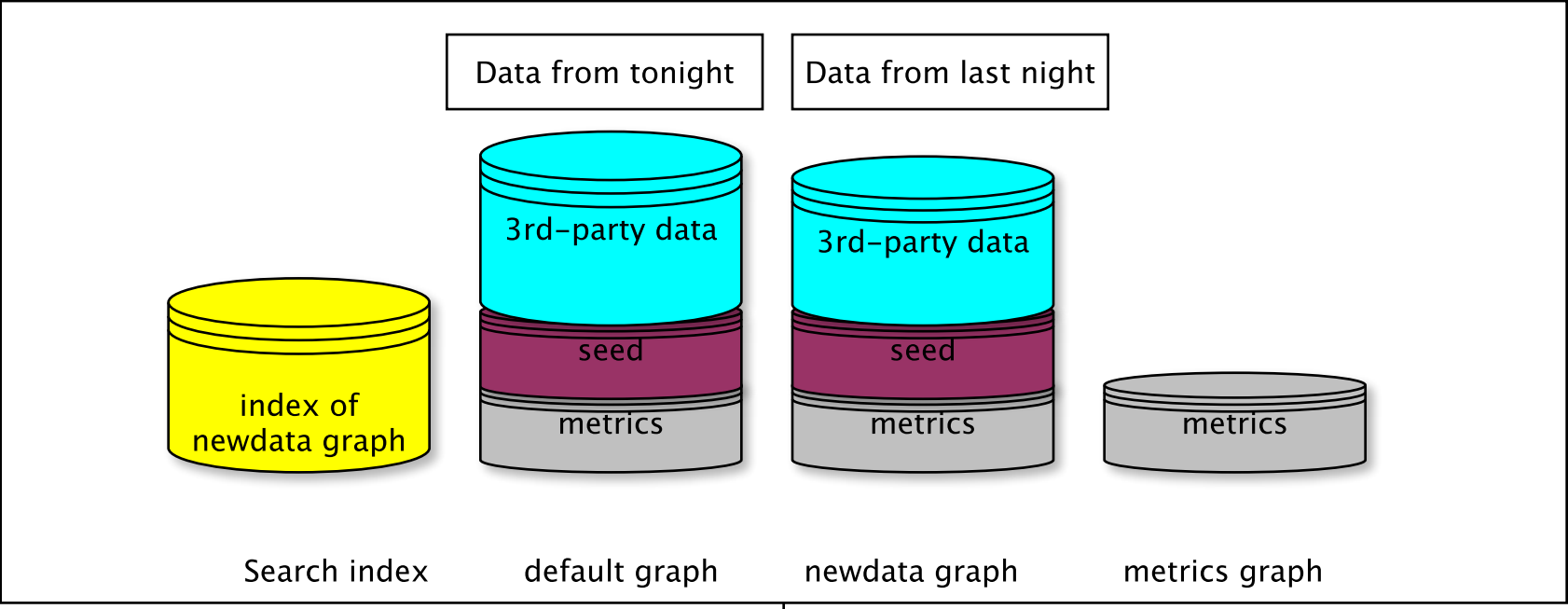
The seed SPARQL file resides in src\main\frontend\sparql-backend\seed.sparql.txt
It defines the 3rd-party sources to harvest in the next cron job



Every 10 minutes from 00:05 to 03:55
Harvest 3rd-party data

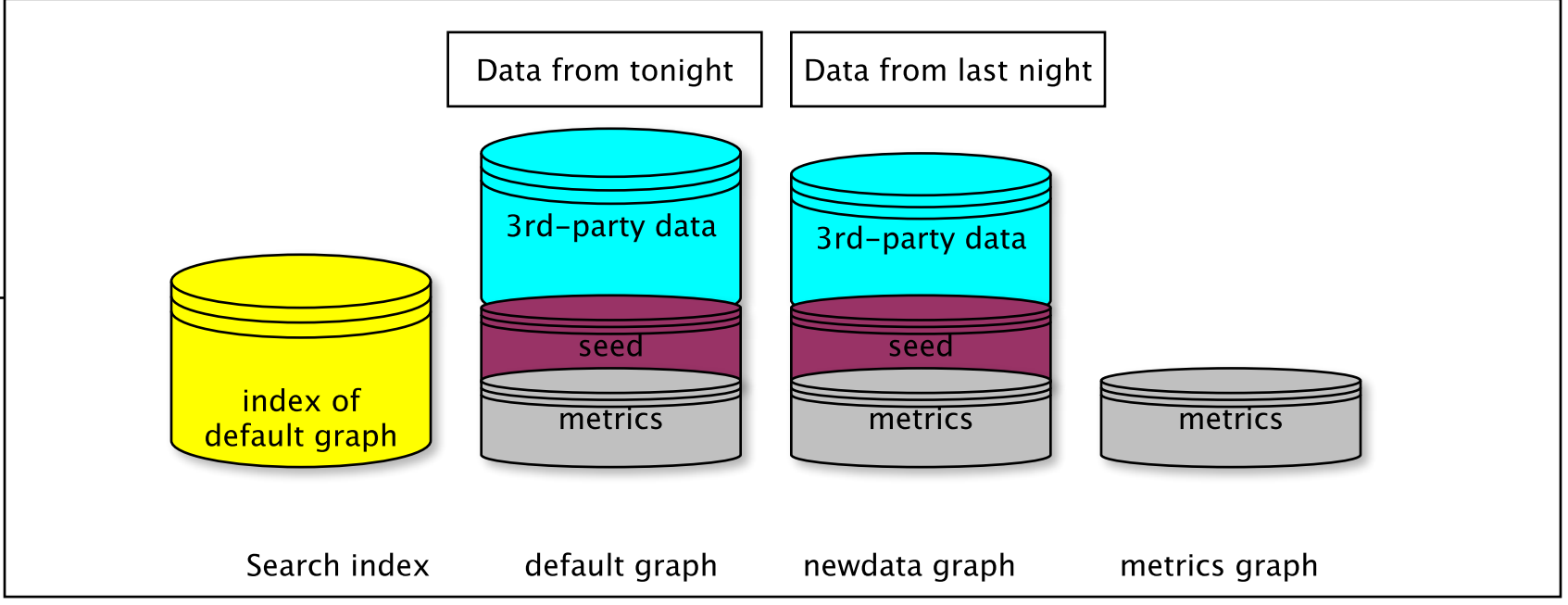


04:00
Switch default and newdata



04:03
Delete search index and rebuild from default

04:00-04:05 the search index is out of sync with the default graph,
causing search results to be inconsistent with the actual data.



No more cron jobs
until the following night

Each invocation of this job takes at most 10 minutes and performs one task.
The task will be on one of the four below types.

Harvest datasets from NERC's RSS feed.

Source URL in the seed SPARQL file (abbreviated):
<http://data-search.nerc.ac.uk/search/full/catalogue.rss?...options>

Harvest all projects from a single PURE repository.

Example source URL in the seed SPARQL file:
http://pure.abdn.ac.uk:8080/ws/rest/getprojectrequest?rendering=xml_long

Harvest one year's papers from a
single PURE repository.
Only IDs are collected.

Example source URL in the seed SPARQL file:
http://pure.abdn.ac.uk:8080/ws/oai?verb=ListRecords&set=publications:year2014&metadataPrefix=oai_dc

Harvest paper details for 30 papers.

URLs are created by the above task. Example:
<http://pure.abdn.ac.uk:8080/ws/rest/publication?uuids.uuid=227d2cef-e9e7-45db-b80e-d9ffb509f42a>