

NEBIUS

Breaking out of the cage: Move complex development to GitHub

Alexander Smirnov
Technology Expert, Nebius



A team of 500+ highly skilled professionals



Nebius is a European public cloud with global expertise



Recognized as an NVIDIA® Preferred Partner



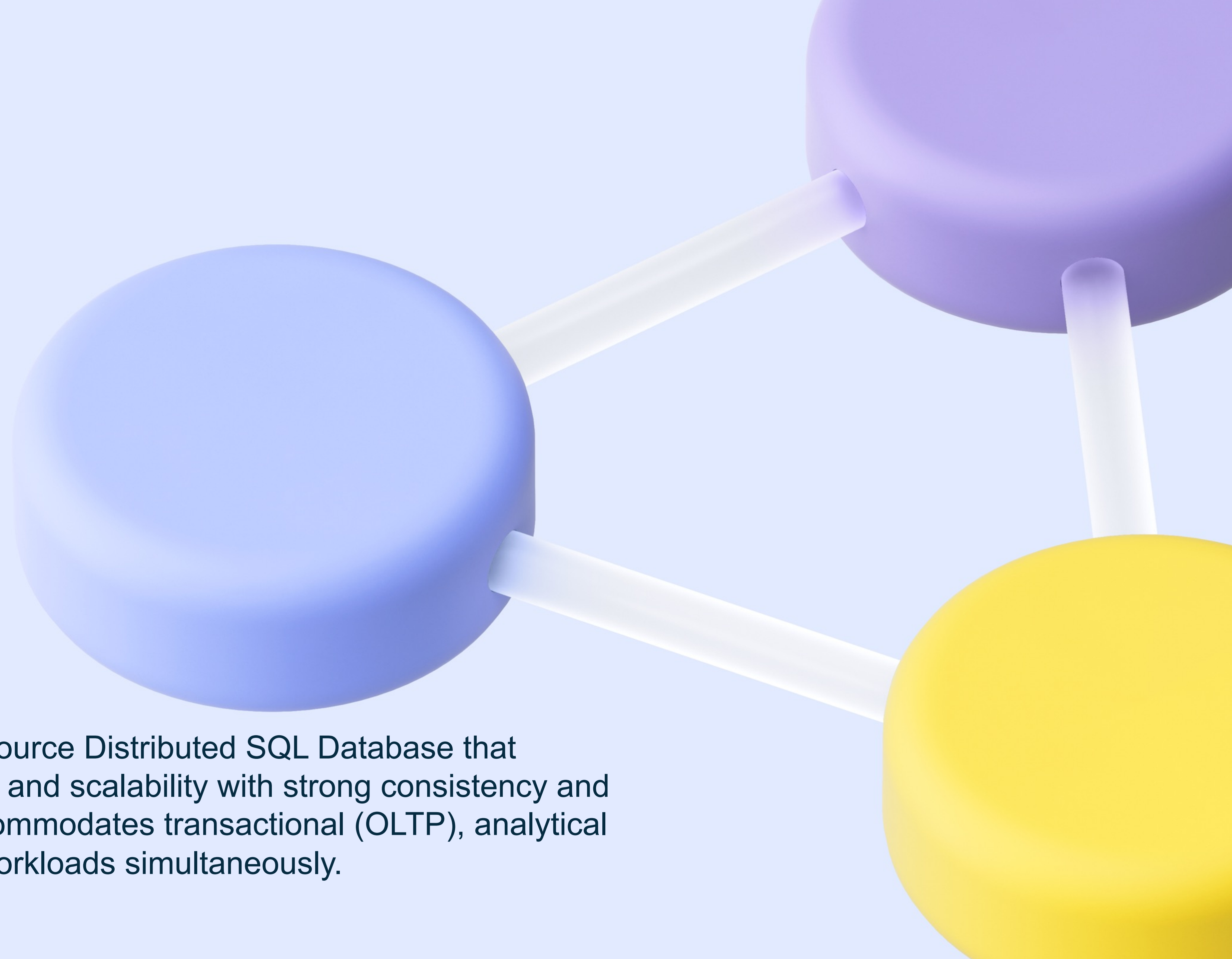
Green data center in Finland featuring modern solutions



Nebius AI Supercomputer ISEG Ranks 16th in the Top 500, exemplifying our cutting-edge HPC capabilities



YDB



YDB is a versatile open-source Distributed SQL Database that combines high availability and scalability with strong consistency and ACID transactions. It accommodates transactional (OLTP), analytical (OLAP), and streaming workloads simultaneously.

YDB Open-Source Timeline

2014

First commit
10.01.2014

2021

Open-Source
project started

2022

- Source code published on GitHub under Apache 2.0
- Website and documentation
- Post-commit sync from the corporate monorepository
- Pull Request import from GitHub to the corporate monorepository

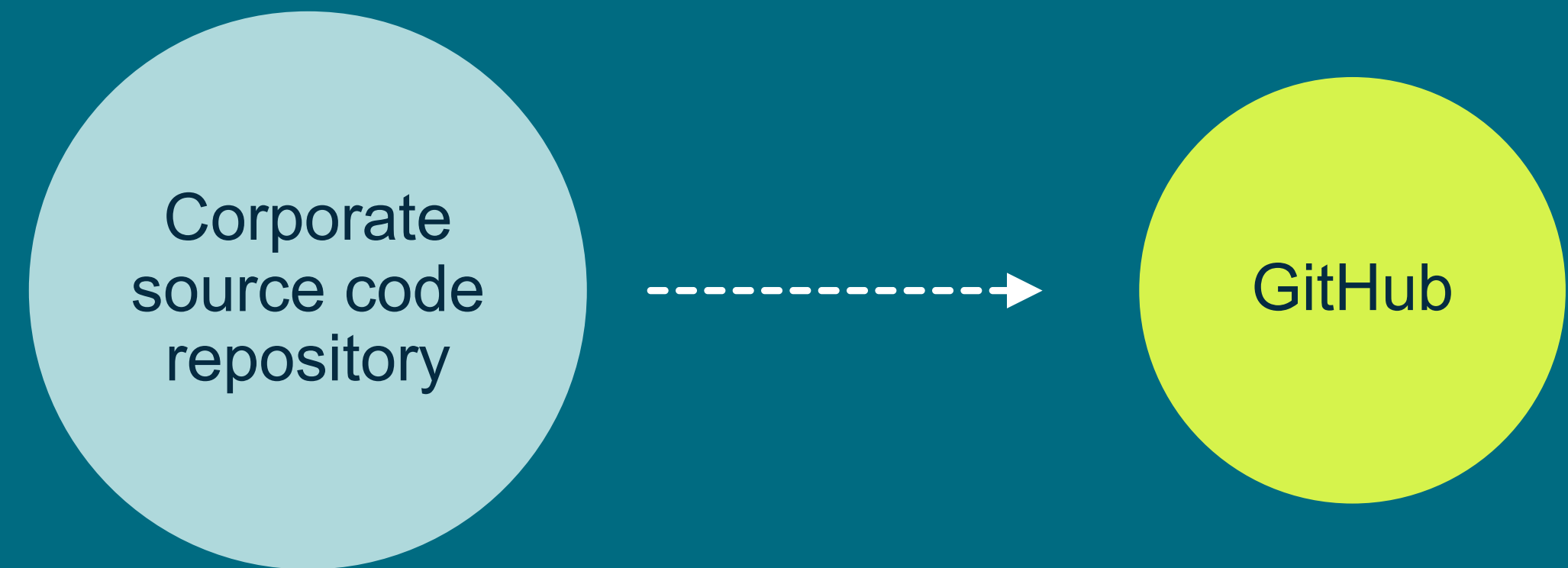
2023

- Development in GitHub
- Cloud infrastructure to run builds and tests
- Snapshot delivery to various corporate contexts
- GitHub Actions CI
- Release management in GitHub

Phase one

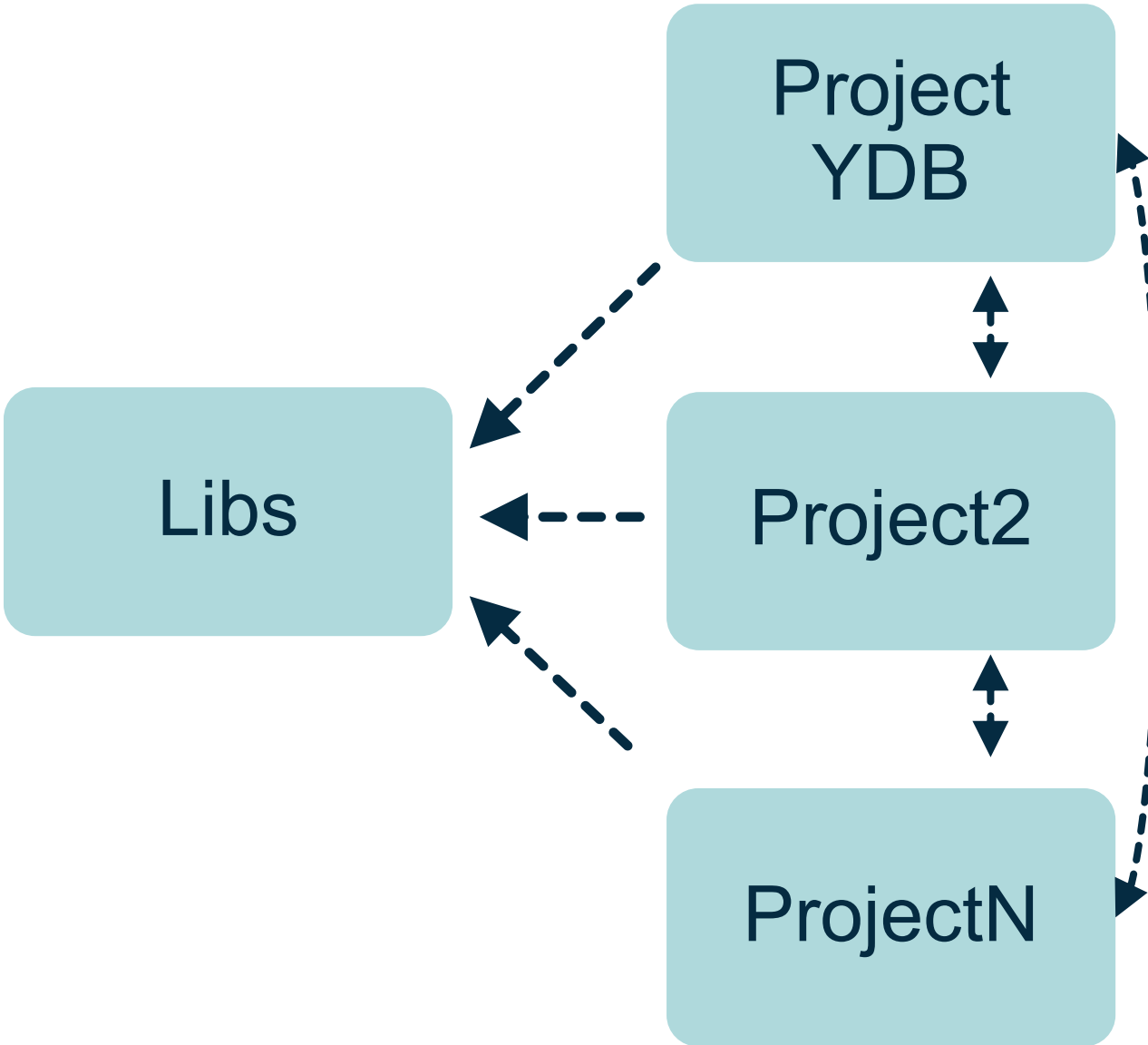
The Mirror

- Increase popularity
- Reduce vendor lock for corporate customers

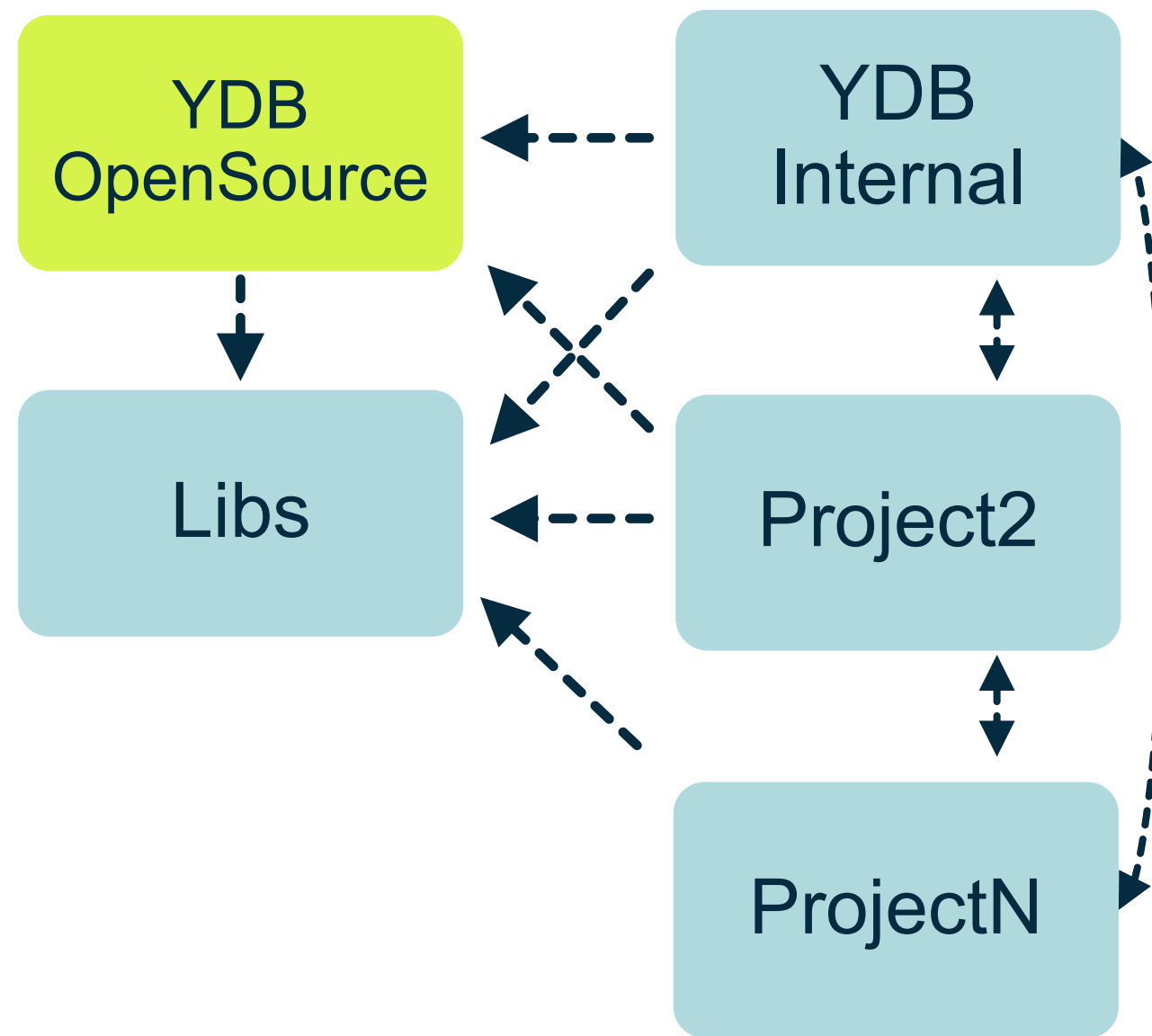


Open-source code isolation

Internal repository structure, arrows indicate dependencies:



BEFORE

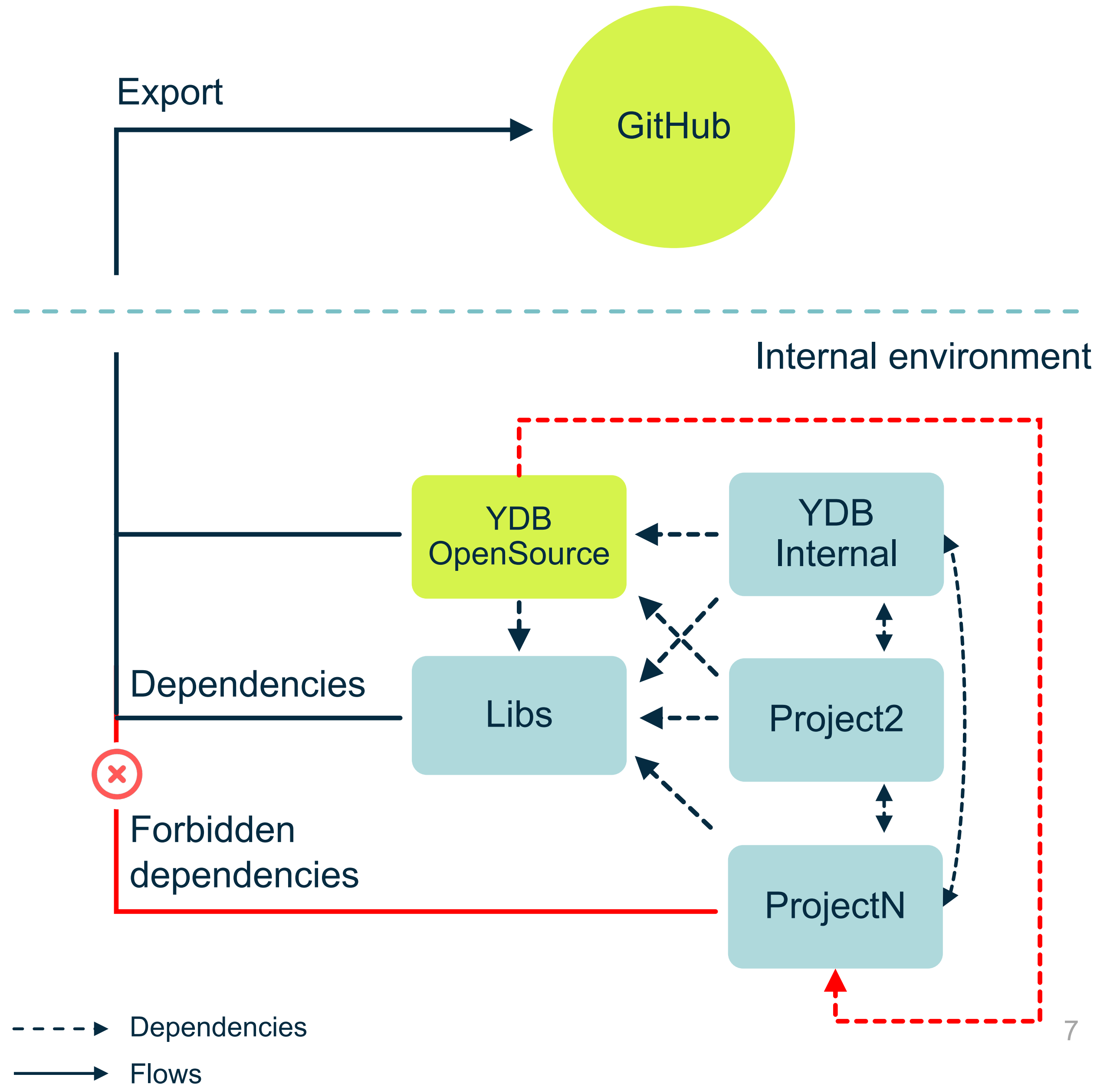


AFTER

Phase one

Export sync

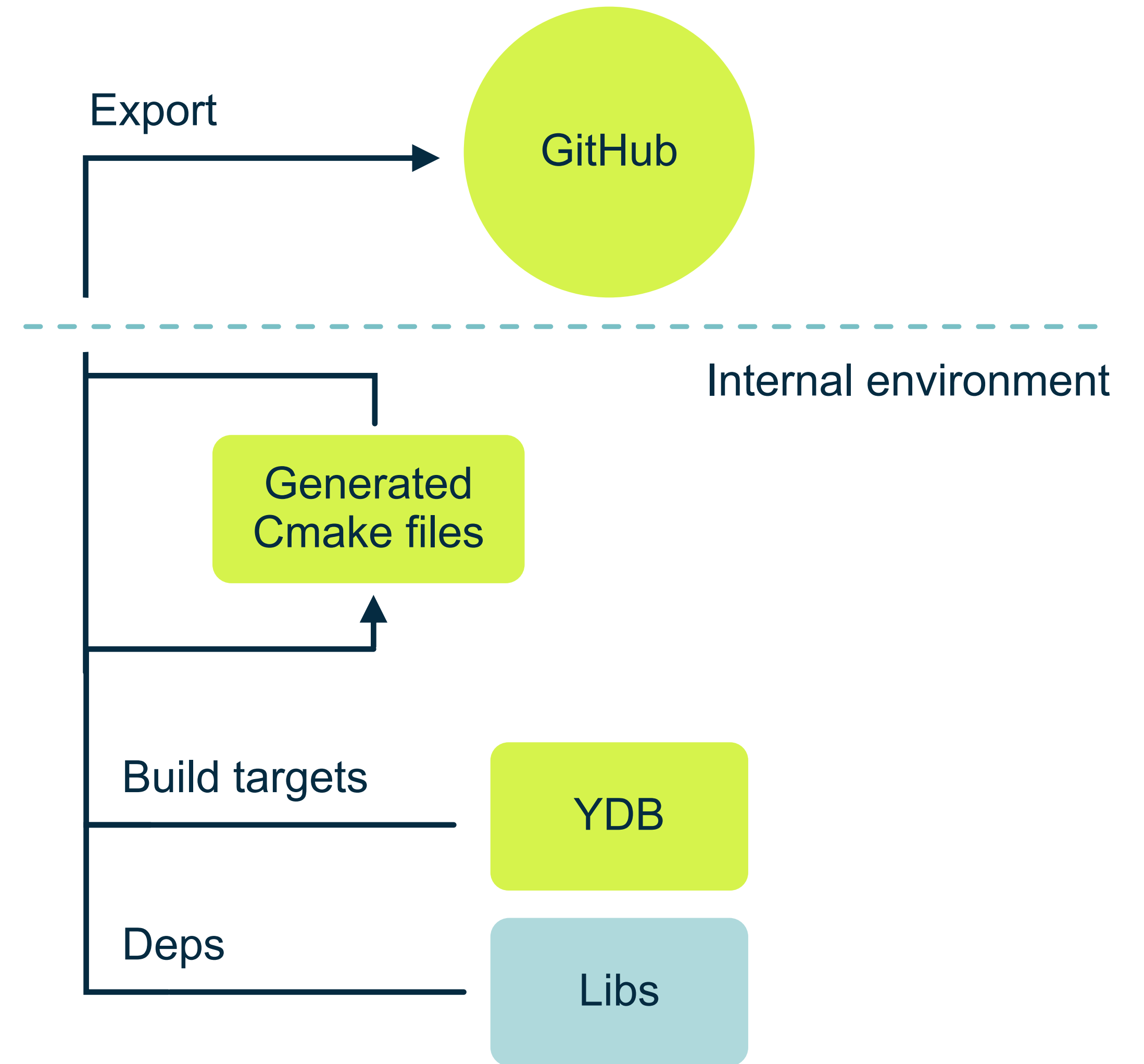
- Capture repository changes
- Build graph dependency discovery
- Filter for allowed directories



Phase one

CMake build

- Internal build system is not available outside
- Generate CMakeLists.txt
- Unit tests supported

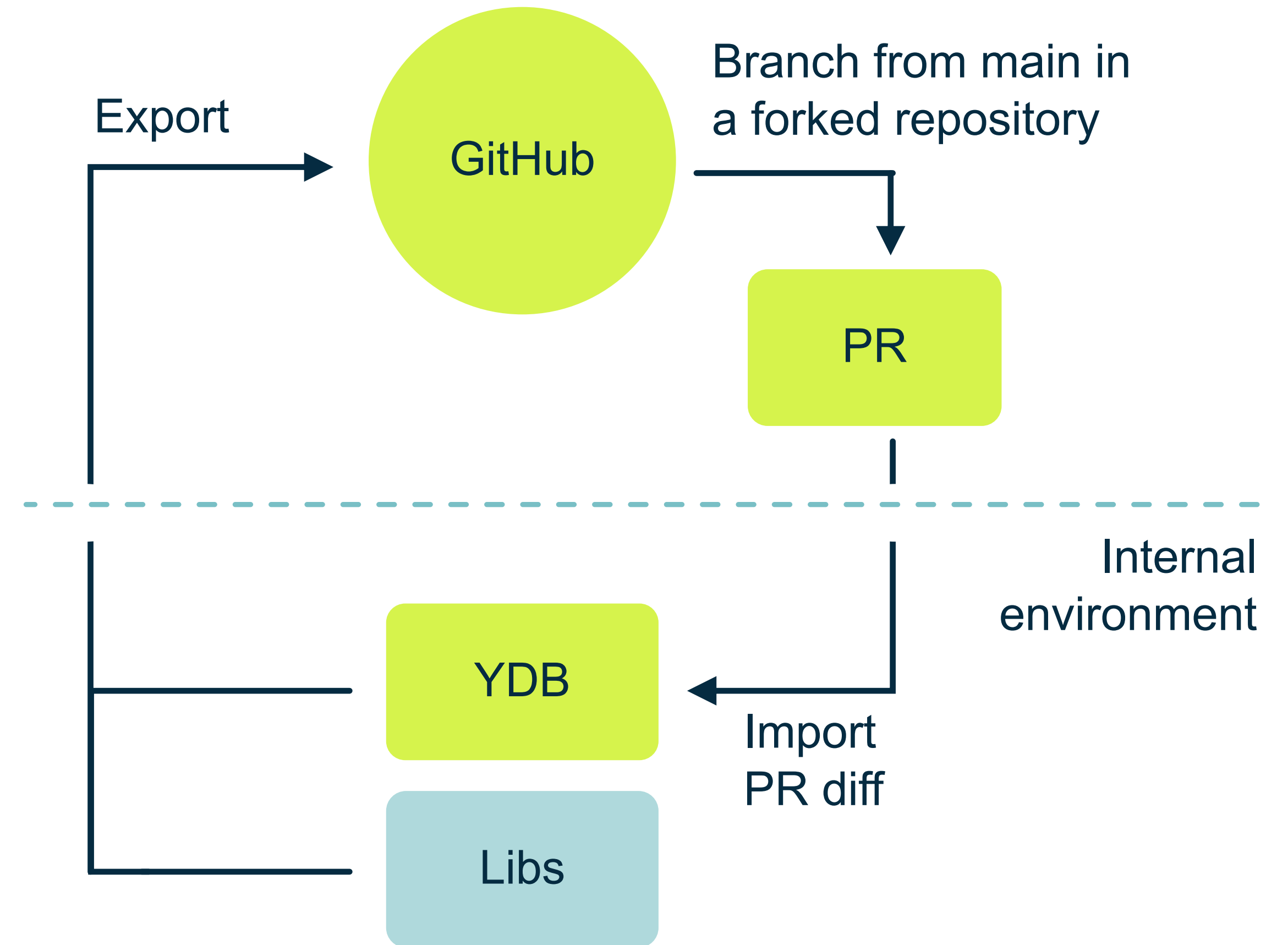


Arrows indicate flows

Phase one

Pull Requests import

- Merge to main in GitHub is blocked
- Import PRs, merge internally
- Commit authorship is preserved



Arrows indicate flows

Achievements

- Source code is publicly available
- Anyone can build and run YDB
- Some tests can be run too
- Increased interest
- Corporate clients attained

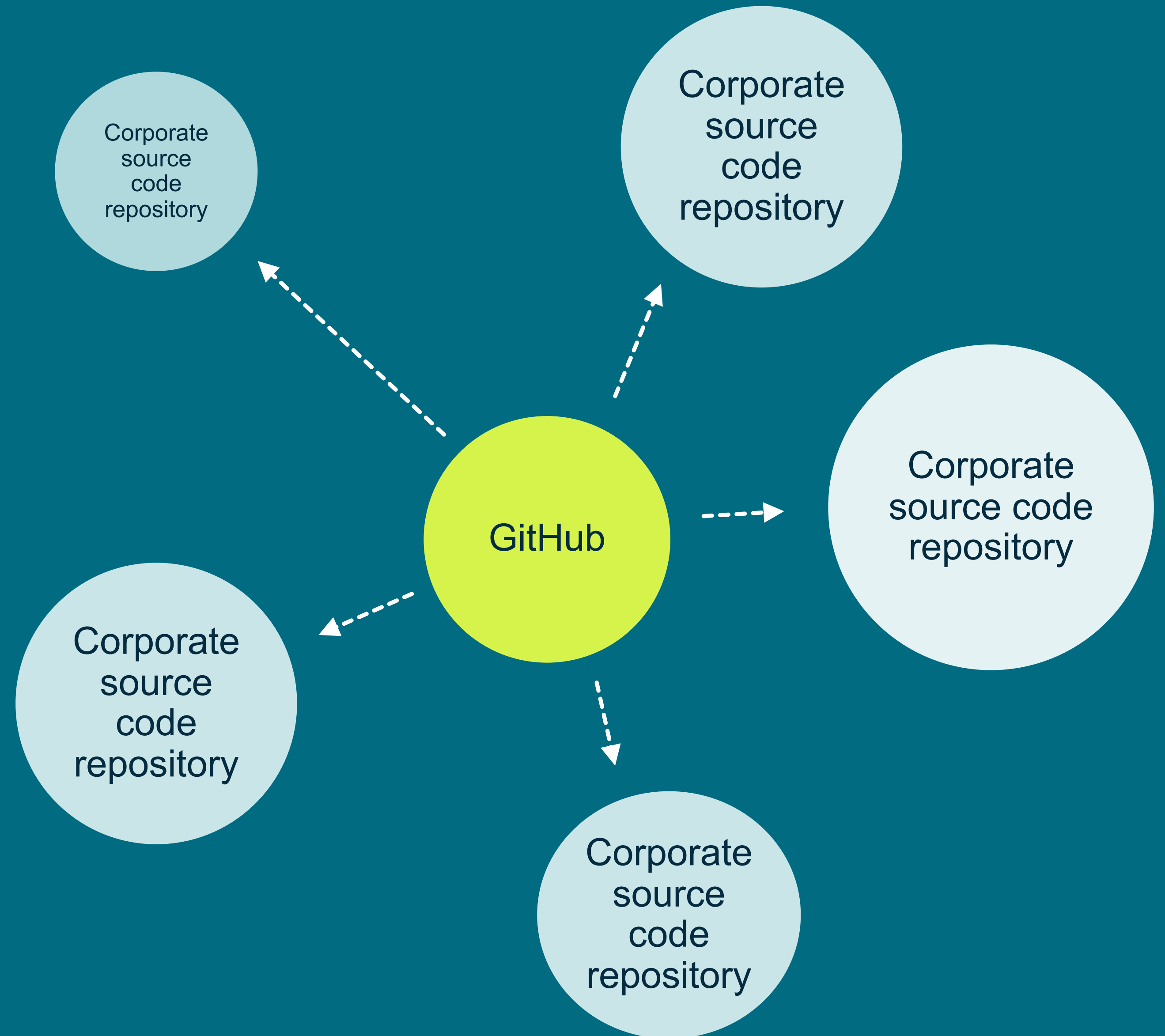
Impediments

- Contributors cannot run tests
- Contributors have no tools and infrastructure to work on a change
- Corporate contributions are almost impossible

Phase two

The Star

- Enable contributions
- No private context for development
- One product for all users
- One process for all contributors



Public cloud

- On-demand virtual machines
- Permanently running VM for on-demand dispatcher
- Permanently running VM for build cache
- S3 object storage

Build cache

- Everyone can read
- Requires authentication to write

Phase two

Continuous Integration

GitHub Actions:

Nightly checks

Pre-commit checks

Post-commit checks

Phase two

Nightly checks, test results

- Run on schedule
- Every supported architecture, operating system, compiler
- Structured test information, JUnitXML
- Logs truncation

Nightly checks

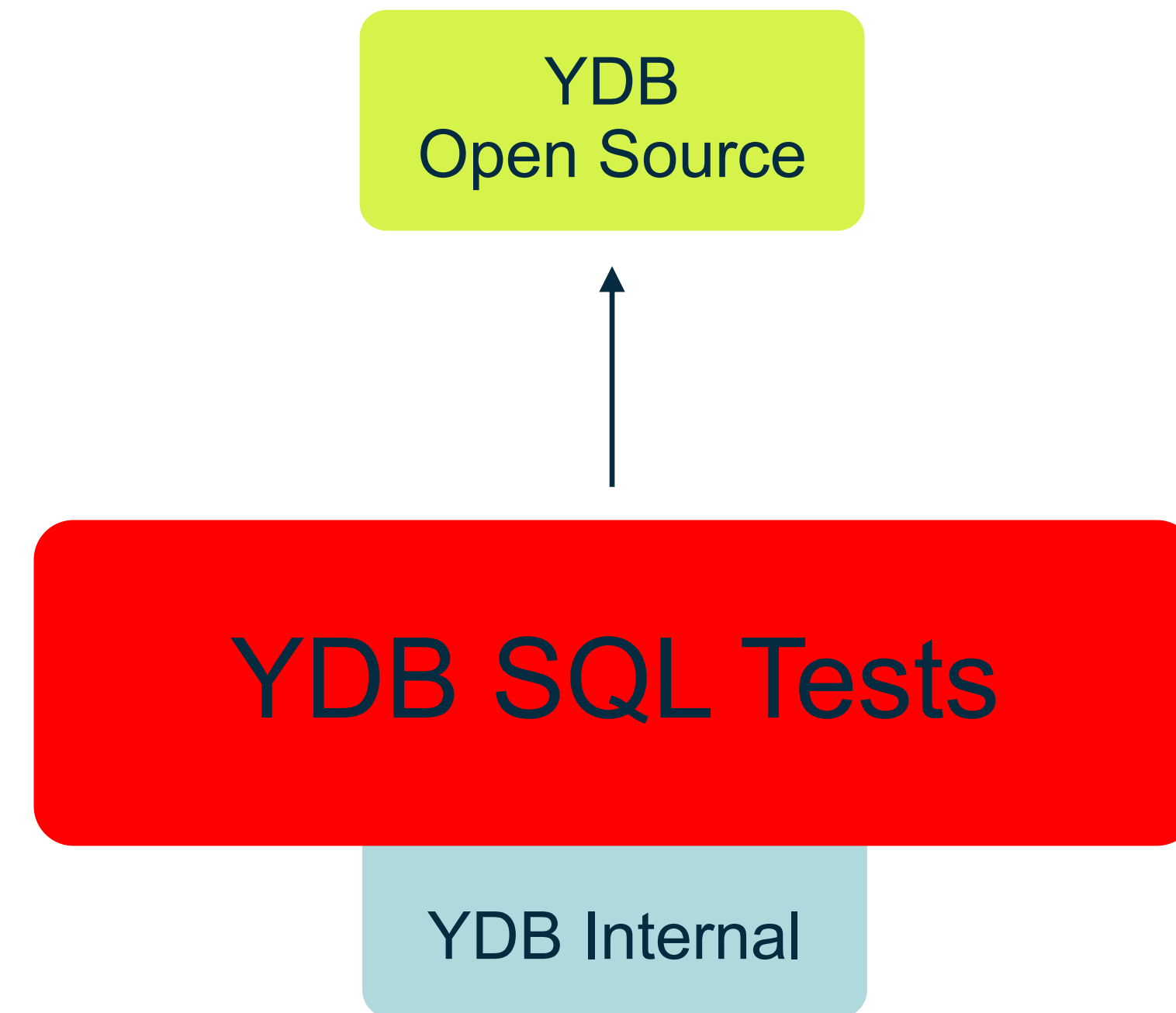
Pre-commit checks

Post-commit checks

Phase two

Add tests to the YDB code

- You need more code to develop an application rather than to build it
- One of the teams brought a very complex test infrastructure
- Their tests could not be run with CTest



Phase two

Back to internal build system

- Yatool goes Open-Source
- Supports existing development process
- Zero conversion for code import from GitHub
- Developers do not want need to relearn



The Return of the Prodigal Son, Rembrandt

Phase two

Pre-commit checks

- Green light to merge changes
- On a merge commit
- Dependency graph
- Summary in the Pull Request comments

Nightly checks

Pre-commit checks

Post-commit checks



github-actions bot commented 1 hour ago · edited

- 2024-02-23 16:02:45 UTC Pre-commit [check](#) for [de64eac](#) has started.
- 2024-02-23 16:02:48 UTC Build `linux-x86_64-relwithdebinfo` is running...
- 2024-02-23 16:04:07 UTC Build successful.
- 2024-02-23 16:04:24 UTC Tests are running...
- 2024-02-23 16:53:31 UTC Some tests failed, follow the links below.

[Test history](#)

TESTS	PASSED	ERRORS	FAILED	SKIPPED	MUTED?
67841	56832	0	3	10953	53



Phase two

Test reporting

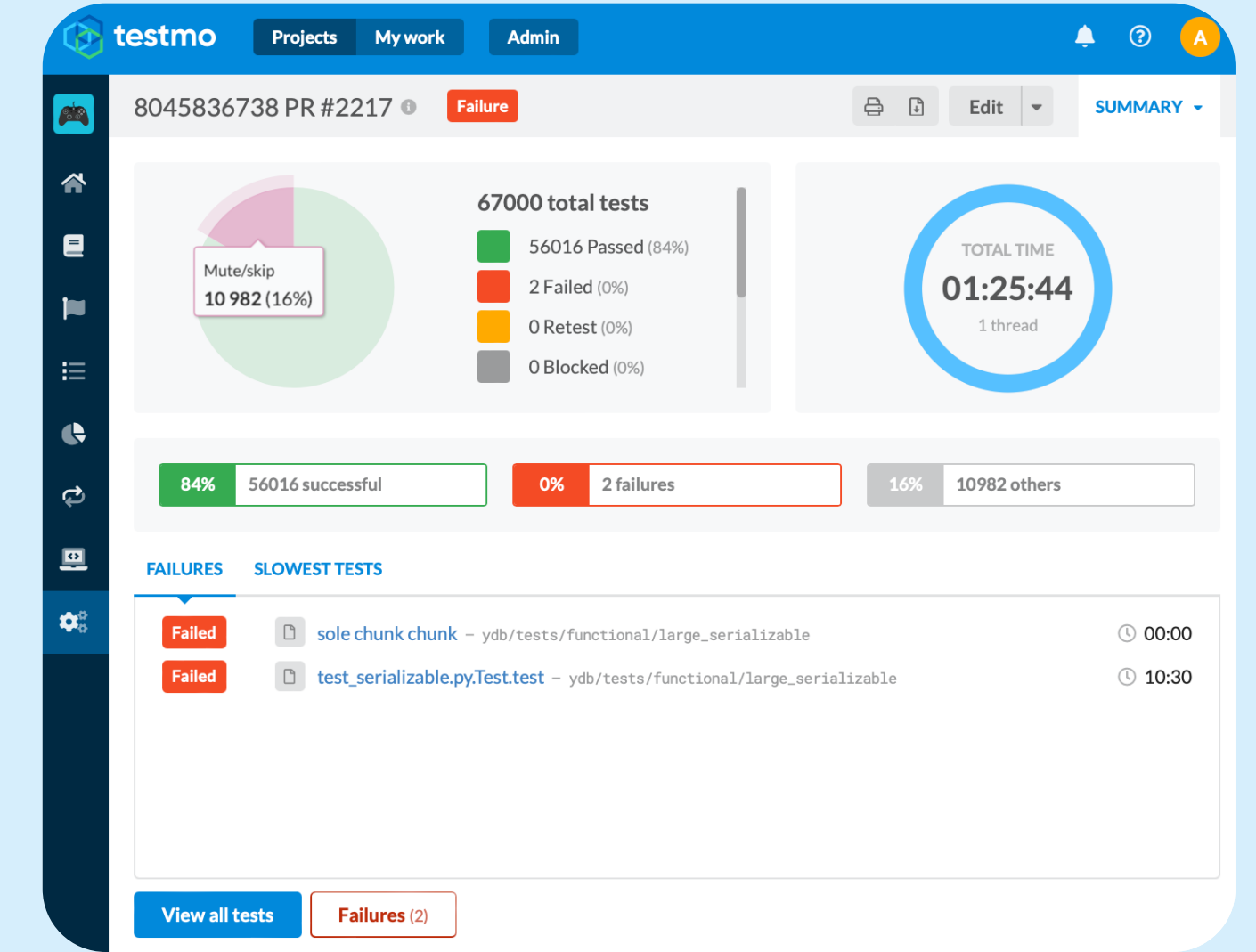
- Muted tests list
- Report to Testmo app
- Simple HTML report

```
ydb / .github / config / muted_ya.txt
```

GrigoriyPA YQ-2786 partly unmute tests fq s3 (#2143) ✓

Code Blame 104 lines (104 loc) · 6.08 KB

```
1 ydb/core/blobstorage/dsproxy/ut TBlobStorageProxySequence
2 ydb/core/blobstorage/dsproxy/ut_fat TBlobStorageProxyTest
3 ydb/core/client/ut TClientTest.PromoteFollower
4 ydb/core/client/ut TClientTest.ReadFromFollower
5 ydb/core/client/ut TFlatTest.AutoSplitMergeQueue
6 ydb/core/cms/ut_sentinel TSentinelTests.BSControllerCantC
7 ydb/core/quoter/ut QuoterWithKesusTest.PrefetchCoefficien
8 ydb/core/kafka_proxy/ut KafkaProtocol.CreatePartitionsSce
9 ydb/core/kafka_proxy/ut KafkaProtocol.ProduceScenario
```



0046615553 POST Failure

Test	Elapsed	Status
test.py.test[insert-multiappend_sorted-default.txt-Analyze]	3.921s	Passed
test.py.test[insert-multiappend_sorted-default.txt-Debug]	7.165s	Failed
test.py.test[insert-multiappend_sorted-default.txt-ForceBlocks]	12.87s	Passed
test.py.test[insert-multiappend_sorted-default.txt-Plan]	0.003s	Passed
test.py.test[insert-multiappend_sorted-default.txt-Results]	1.076s	Passed
test.py.test[insert_monotonic-truncate_fail--Analyze]	0.005s	Mute/skip
test.py.test[insert_monotonic-truncate_fail--Debug]	0.003s	Mute/skip
test.py.test[insert_monotonic-truncate_fail--ForceBlocks]	0.003s	Mute/skip
test.py.test[insert_monotonic-truncate_fail--Plan]	0.003s	Mute/skip
test.py.test[insert_monotonic-truncate_fail--Results]	0.73s	Passed
test.py.test[join-inner_with_order-off-Analyze]	0.806s	Passed
test.py.test[join-inner_with_order-off-Debug]	1.046s	Passed
test.py.test[join-inner_with_order-off-ForceBlocks]	0.004s	Mute/skip
test.py.test[join-inner_with_order-off-Plan]	0.009s	Passed
test.py.test[join-inner_with_order-off-Results]	0.004s	Passed
test.py.test[join-mapjoin_on_complex_type_optional_left_only_...]	0.867s	Passed
test.py.test[join-mapjoin_on_complex_type_optional_left_only_...]	1.242s	Passed
test.py.test[join-mapjoin_on_complex_type_optional_left_only_...]	2.9s	Passed
test.py.test[join-mapjoin_on_complex_type_optional_left_only_...]	0.021s	Passed
test.py.test[join-mapjoin_on_complex_type_optional_left_only_...]	4.003s	Passed
test.py.test[join-mapjoin_on_tablerecord-off-Analyze]	0.965s	Passed
test.py.test[join-mapjoin_on_tablerecord-off-Debug]	3.748s	Passed
test.py.test[join-mapjoin_on_tablerecord-off-ForceBlocks]	0.004s	Mute/skip
test.py.test[join-mapjoin_on_tablerecord-off-Plan]	0.012s	Passed
test.py.test[join-mapjoin_on_tablerecord-off-Results]	0.013s	Passed

Failure

```
Test results differ from canonical:
test_result[0]: files content differs:
'ydb/library/yql/tools/astdiff/astdiff'
has finished unexpectedly with rc = 5
stdout:

stderr:
Programs are not equal!
Node in
/home/runner/actions_runner/.work/ydb/ydb/
multiappend_sorted-default.txt-
Debug/_opt_yql_patched_at_186-10:
#717 [Atom] &lt;tmp/19787512-e36f837b-
a1188e13-13&gt;:
Node in
/home/runner/.ya/build/build_root/92sa/081
... (186-10):
```

COMMENTS CONTEXT ISSUES

Statuses 1/243

- Passed 8046571434 PR #2064
- Passed 8047926907 POST
- Failed 8046615553 POST
- Passed 8045996293 PR #2233
- Passed 8045836738 PR #2217
- Passed 8045217545 PR #2185
- Passed 8045598188 PR #2232
- Passed 8045457194 POST
- Passed 8045123659 PR #2228
- Passed 8042483634 schedule
- Passed 8040834336 PR #2228

0046615553 POST Failure

Folder

- library/yql/providers/s3/object_listers/ut
- library/yql/providers/s3/path_generator/ut
- library/yql/providers/s3/provider/ut
- library/yql/providers/s3/range_helpers/ut
- library/yql/providers/yt/codec/codegen/ut
- library/yql/providers/yt/codec/ut
- library/yql/providers/yt/gateway/native/ut
- library/yql/providers/yt/provider/ut
- library/yql/public/decimal/ut
- library/yql/public/embedded/ut
- library/yql/public/embedded/ut/no_llvm
- library/yql/public/fastcheck/ut
- library/yql/public/issue/ut
- library/yql/public/purecalc/io_specs/mkql/ut
- library/yql/public/purecalc/io_specs/mkql/ut/no_llvm

ydb/library/yql/tests/ (625) 11/25

ql/dq_file/part17

Test	Status
test.py.test[insert-multiappend_s...	Passed
test.py.test[insert-multiappend_s...	Failed
test.py.test[insert-multiappend_s...	Passed
test.py.test[insert-multiappend_s...	Passed
test.py.test[insert-multiappend_s...	Passed
test.py.test[insert_monotonic-tru...	Mute/skip
test.py.test[insert_monotonic-tru...	Mute/skip
test.py.test[insert_monotonic-tru...	Mute/skip
test.py.test[insert_monotonic-tru...	Mute/skip
test.py.test[insert_monotonic-tru...	Mute/skip
test.py.test[join-inner_with_order...	Passed
test.py.test[join-inner_with_order...	Passed
test.py.test[join-inner_with_order...	Passed

Phase two

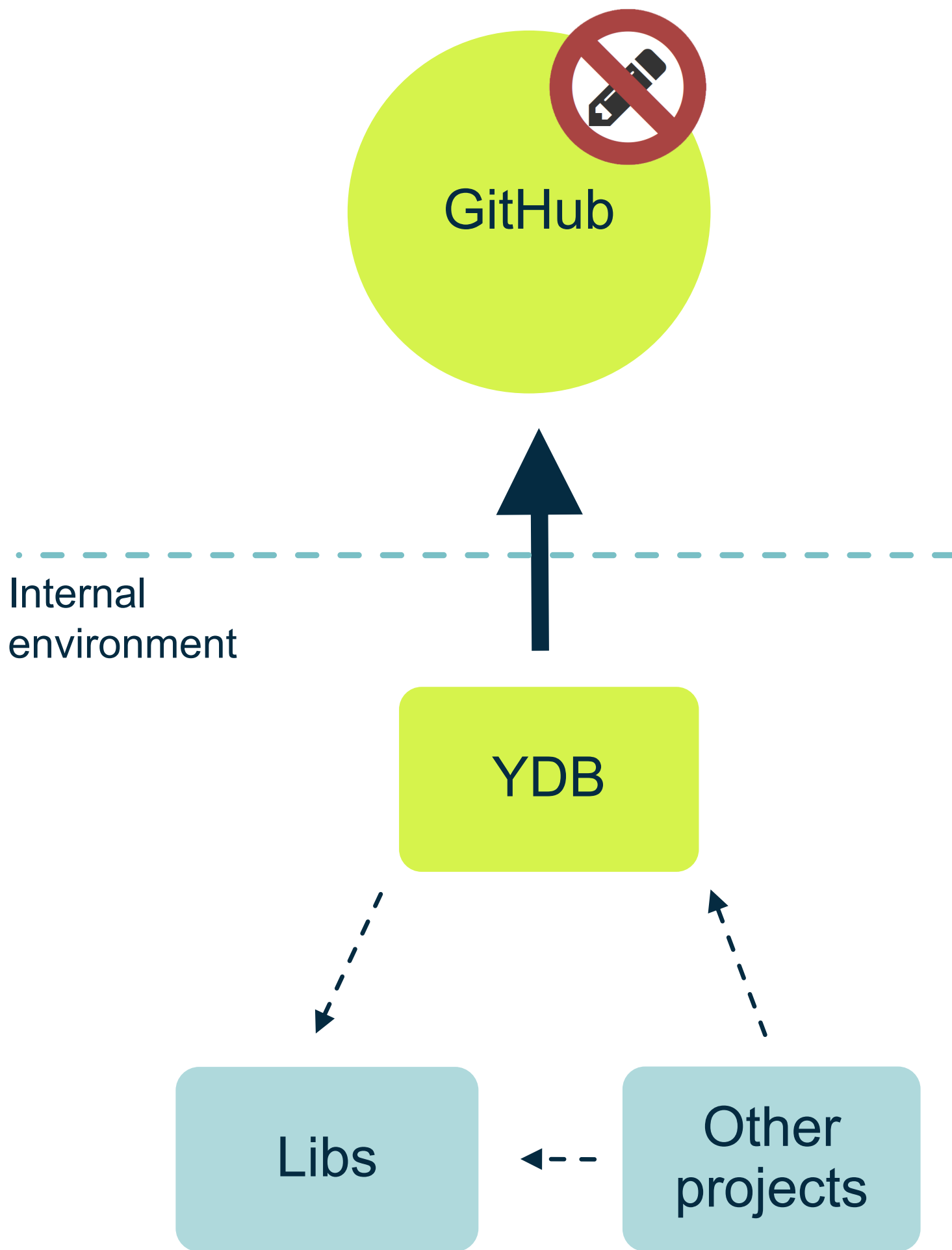
Post-commit checks

- To identify revision when a problem had appeared.
- Helped to estimate the compute resources consumption.

Nightly checks

Pre-commit checks

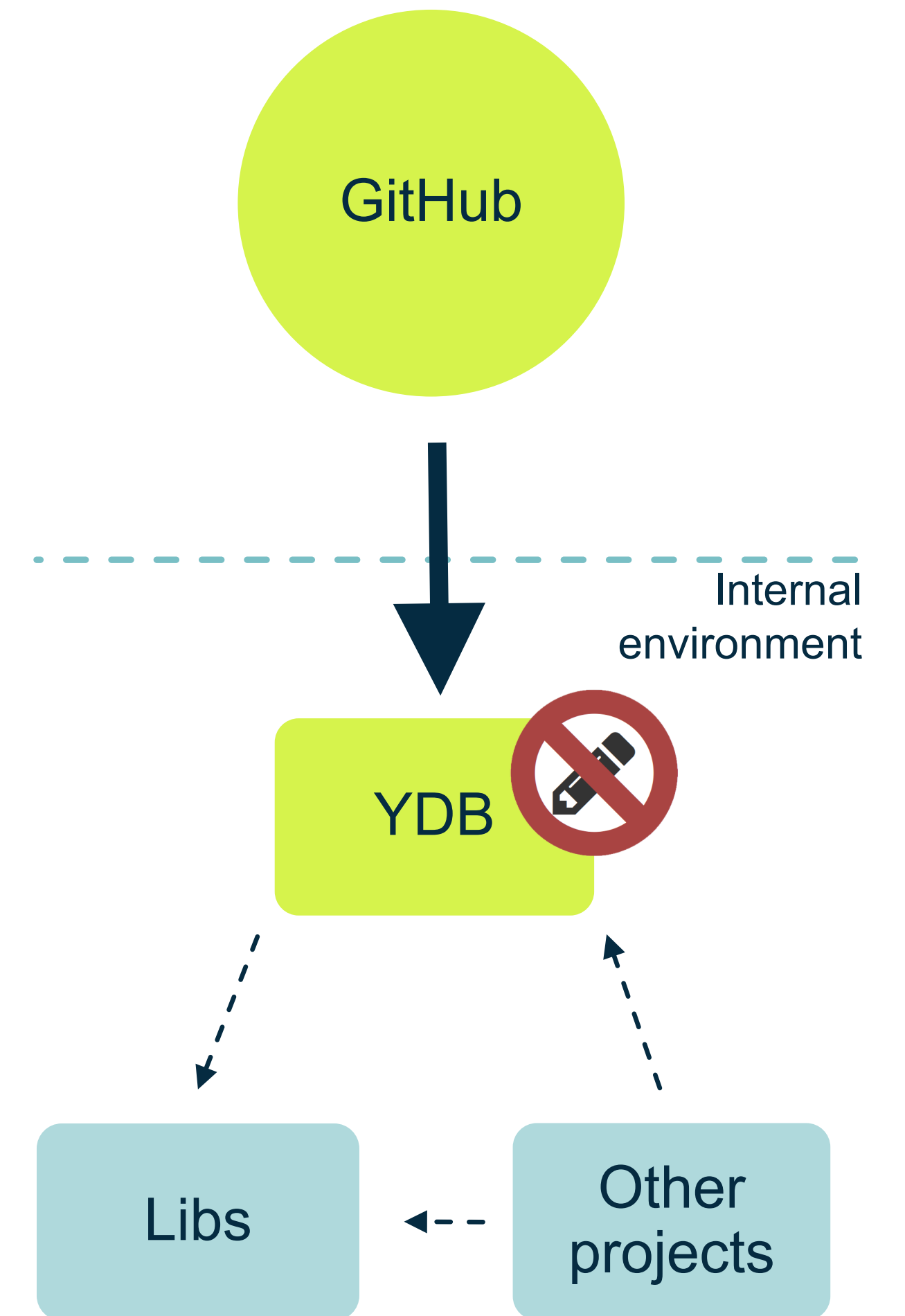
Post-commit checks



BEFORE
Export to GitHub

Switch now?

- Import flow may fail as it is not tested
- Developers may fail to develop outside the monorepository
- We need to mitigate the risks above before actual switching



AFTER
Import from GitHub

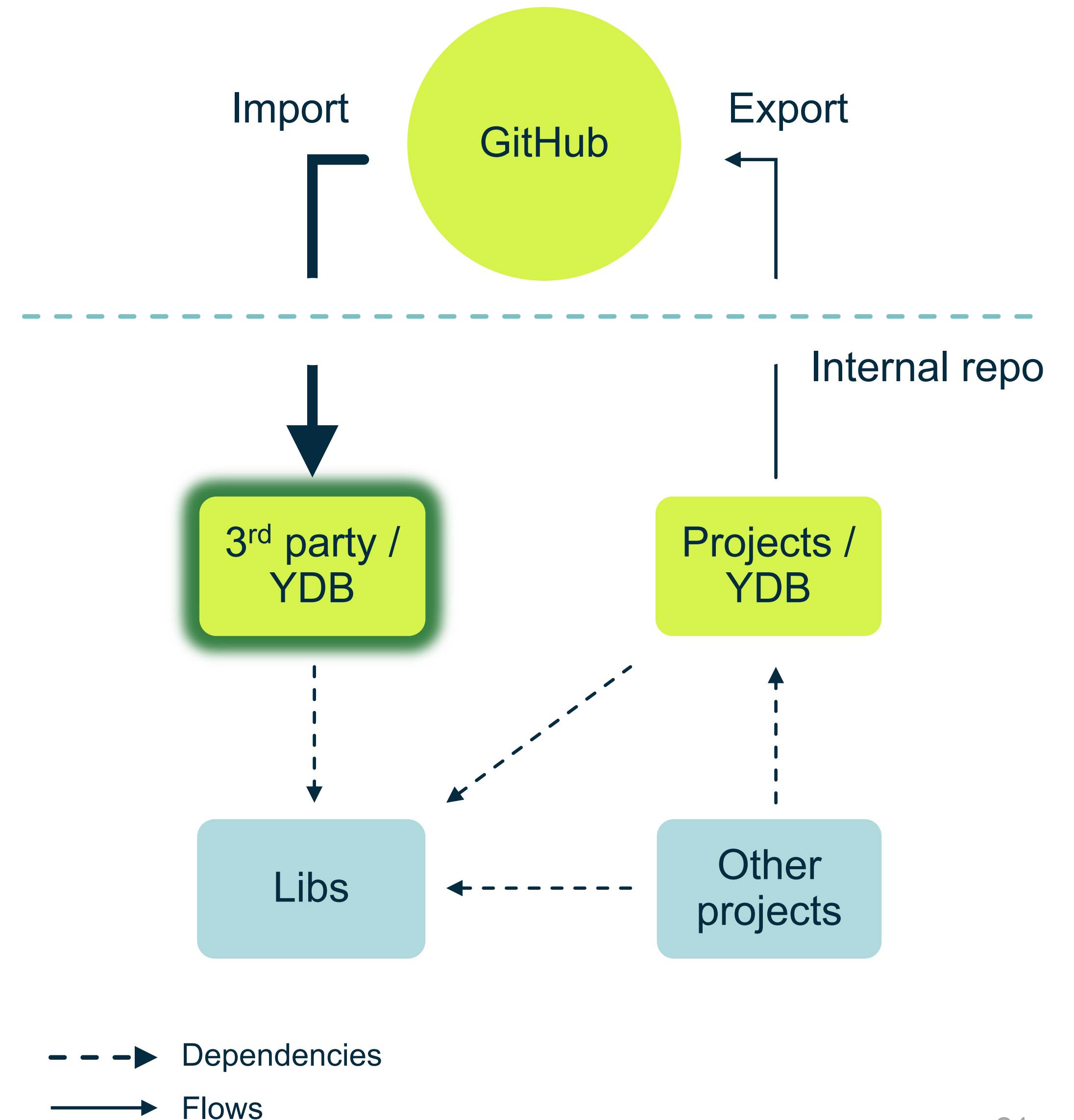
---> Dependencies
 —> Flows

Phase two

Prepare the switch: Step 1

Enable import to the 3rd-party/ydb

- Revision-based, 4 times a day
- Pull Request, required tests run
- Patching code to replace references
- Only YDB code, without libraries
- Stabilize import automation

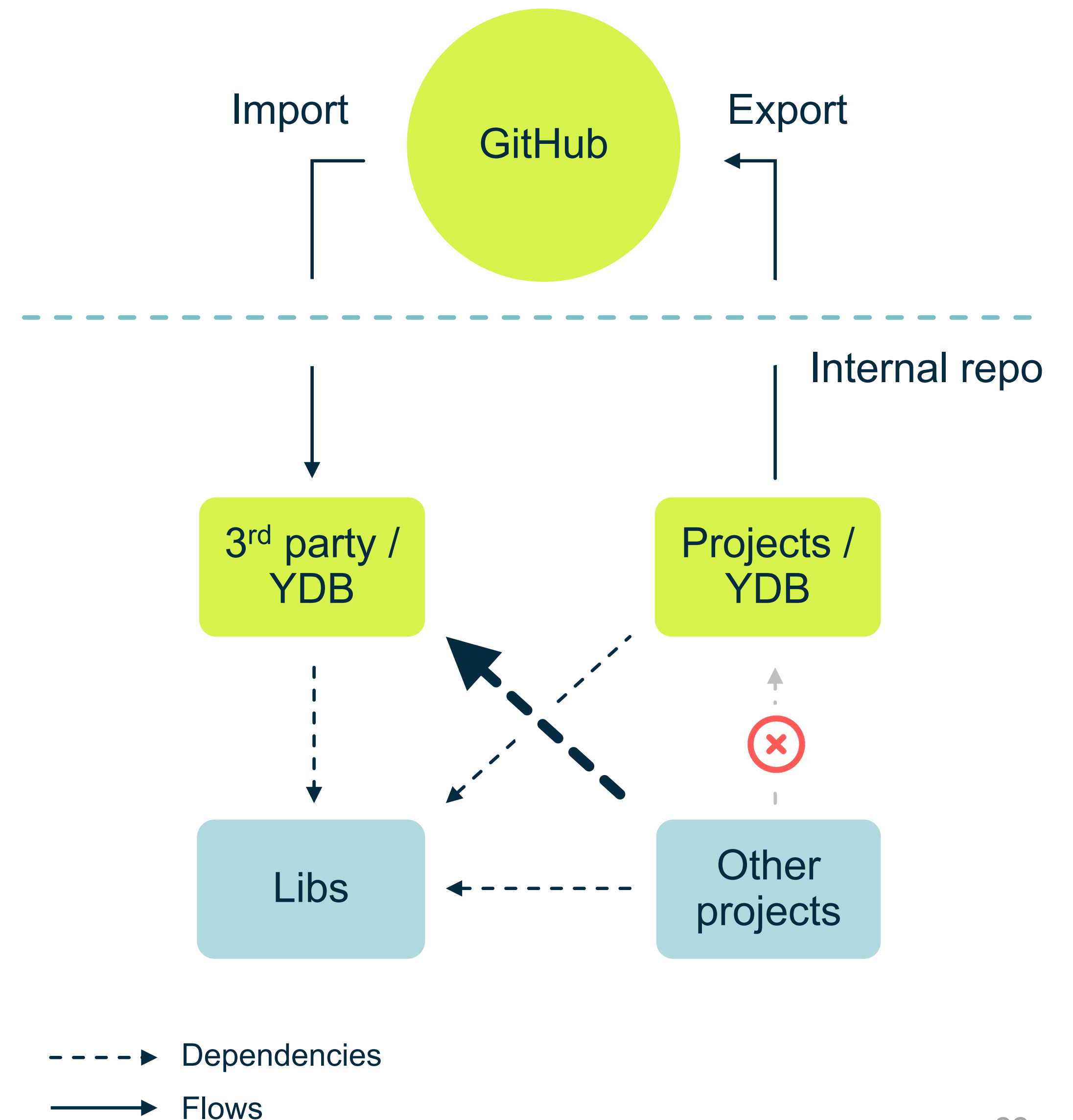


Phase two

Prepare the switch: Step 2

Move all projects dependencies to YDB code as part of the 3rd party components

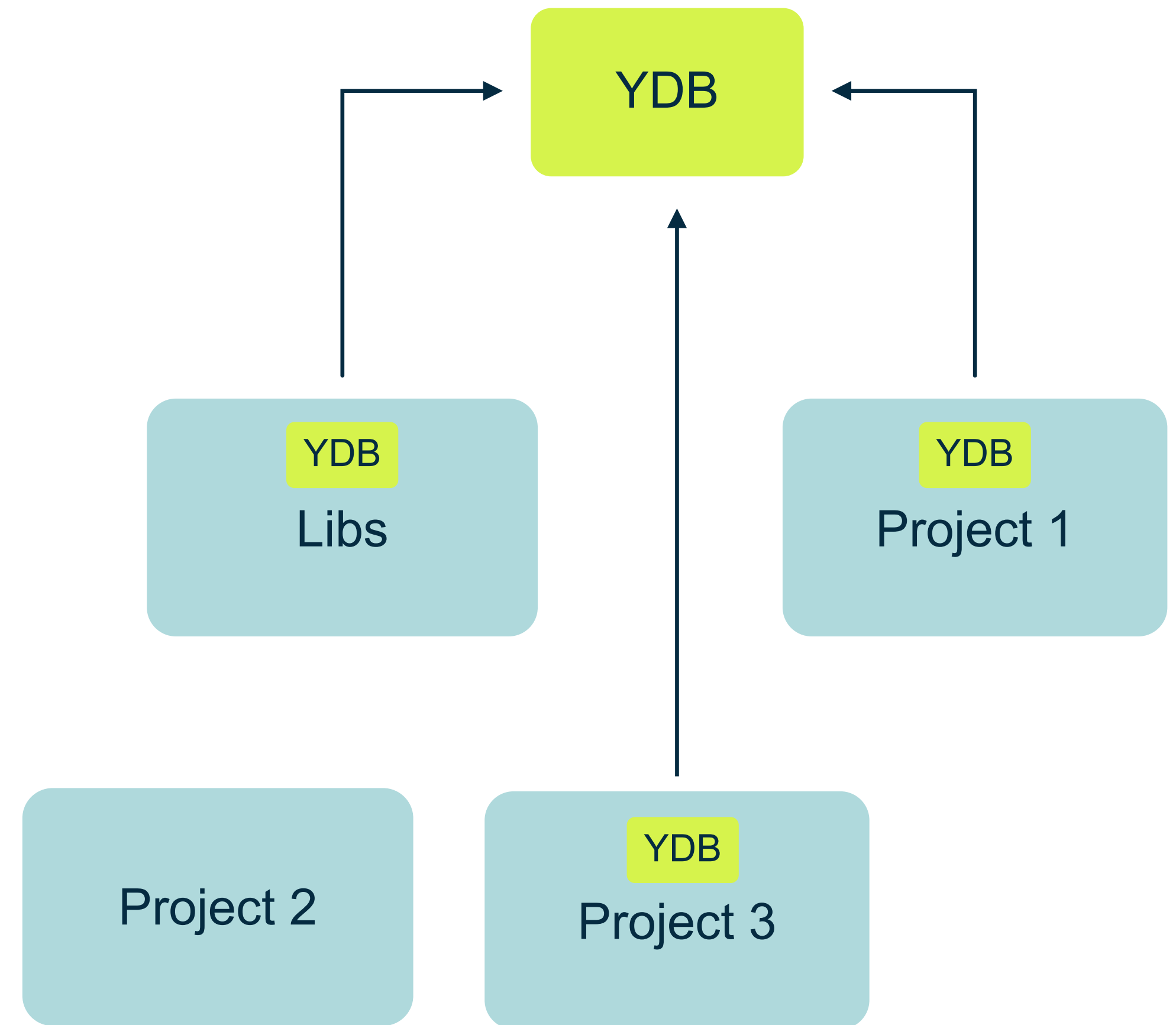
- Iterative, project by project
- Now running all tests of dependent projects on import iterations
- Cease running tests of dependent projects on YDB code changes



Phase two

The real development perimeter unveiled

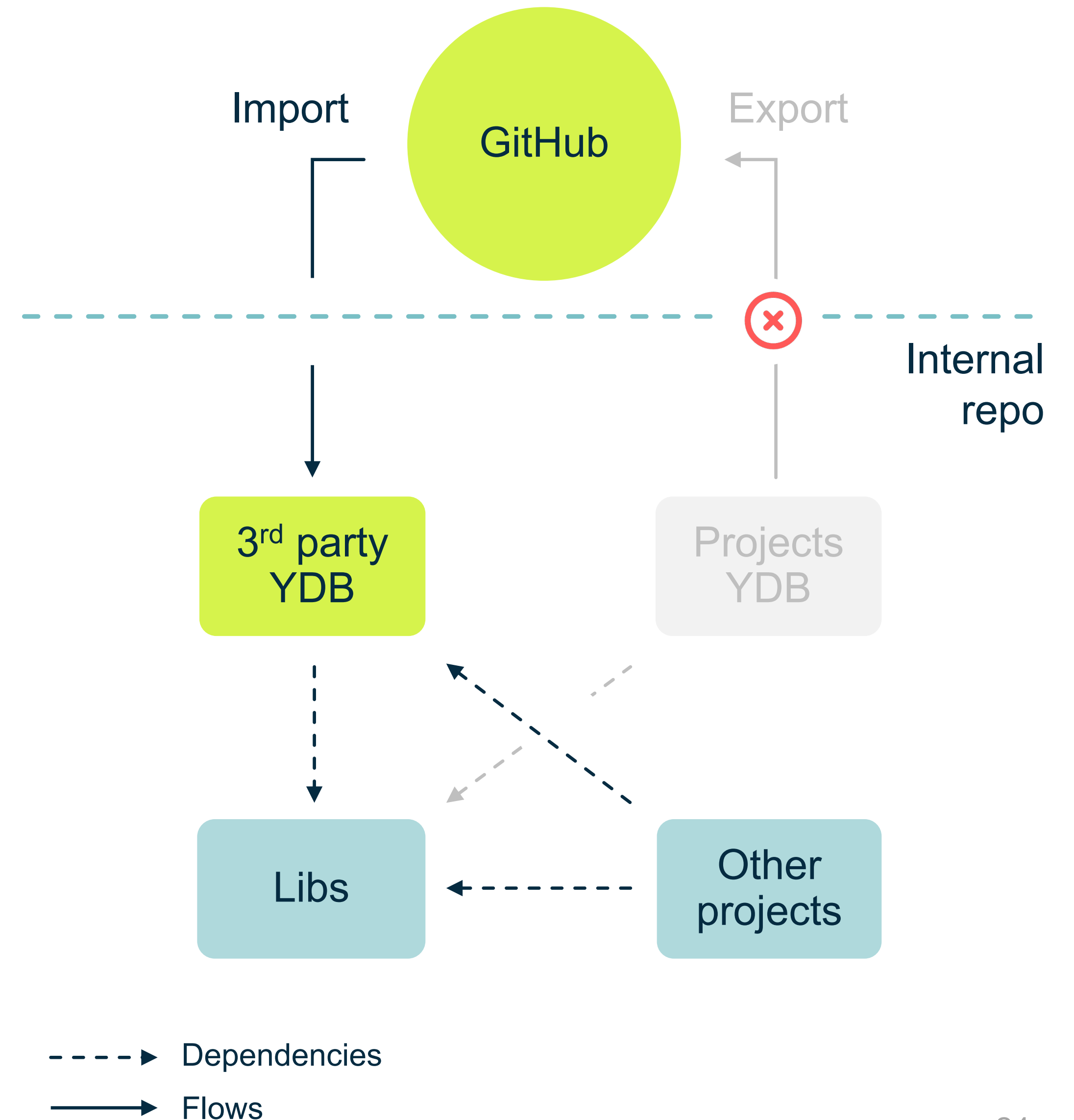
- Highly-cohesive code from common libraries
- More and more tests
- Auxiliary utilities



Phase two

The switch

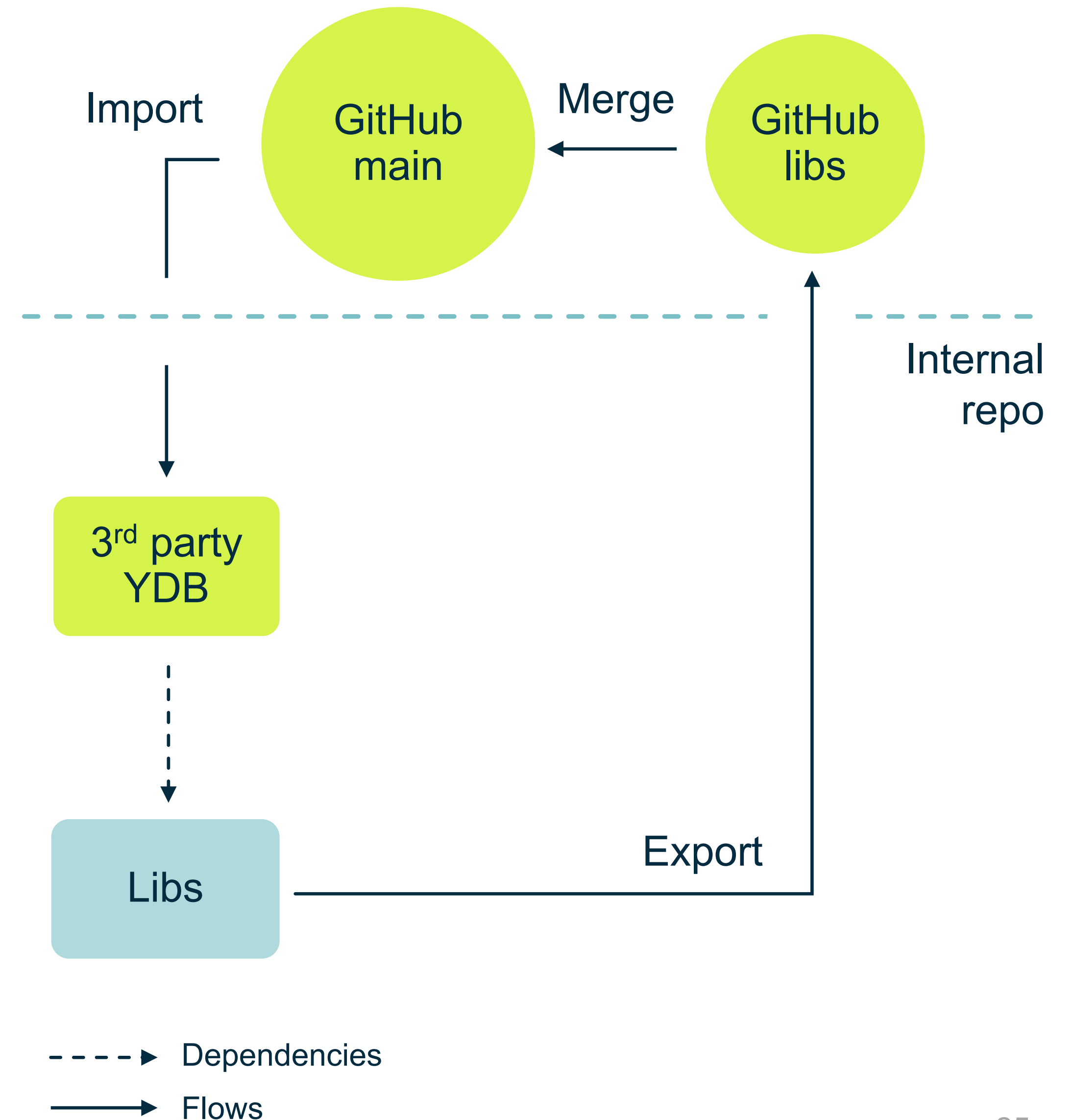
- Switch off YDB sync to GitHub, grant access to merge changes in GitHub
- Development process docs
- Test on a focus group
- A Telegram chat for onboarding



Phase two

Libraries

- We depend on libraries from the parent repository.
- Libraries are exported to a special branch on GitHub.
- Libraries are merged to main once a week.



Achievements

- Full development cycle on GitHub
- One point of truth, one product for many users
- Changes can be verified before submitting
- Infrastructure providers can be replaced

Challenges

- Build for Windows is still CMake-only
- Server logs, core dumps to be collected
- Too many tests run every check because of non-optimal dependencies
- Resolve dependency on libraries

And many more...

Product vision

Sets the goals

Smooth changes

Keep morale high
Mitigate risks one by one

YDB in Nebius

We use it extensively
We contribute to it

YDB - Welcome to contribute



ydb.tech



[github.com/ydb-
platform/ydb](https://github.com/ydb-platform/ydb)

NEBIUS

Thank you!



Alexander Smirnov
Technology Expert, Nebius
alexv-smirnov@nebius.com

YDB - Welcome to contribute



ydb.tech



[github.com/ydb-
platform/ydb](https://github.com/ydb-platform/ydb)