

Migrating GitHub's Global IDs

Andrew Hoglund
GitHub



@andrewhoglund

Brief Overview

Introduced in 2016

- ✓ Shortly after Facebook's public release of GraphQL.
- ✓ Originally intended as v4.
- ✓ Now considered another API, alongside REST.

300,000,000+

Request Per Day

3-6K

Request Per Second

400+

Objects

350+

Mutations

Today's Talk

- ✓ Current ID Format
- ✓ Motivation for Change
- ✓ Rollout Strategy
- ✓ Pitfalls/Challenges

Current ID Format

Current Format

- **Base64 Encoded**
- **Contains Object Type and Object Id**
- **Intended to be Opaque**

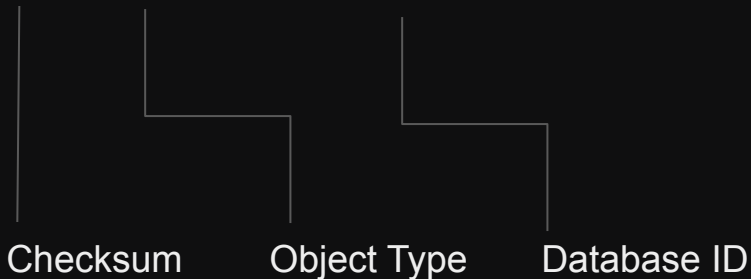
Current Format - Example

```
{  
  viewer {  
    repository(name: "test") {  
      id  
    }  
  }  
}
```

```
{  
  "data": {  
    "viewer": {  
      "repository": {  
        "id": "MDEwO1JlOG9zaXRvcnkxODY4OTAzOTU="
```

Current Format - Example

```
$ echo MDEwO1JlcG9zaXRvcnkxODY4OTAzOTU= | base64 -d  
010:Repository186890395
```



Current Format

- Simple loading logic.
- Works well with single database.

Why Change?

Motivation for Change

- Performant queries to sharded data.
- Flexibility for future architectures.
- Eventually migration to multi-region.

“

This id should be a globally unique identifier for an object, and given just this id, the server should be able to refetch the object.

<https://graphql.org/learn/global-object-identification/>

A Bit about Sharding

Using Vitess with unique VIndexes

```
> SELECT * from pull_requests WHERE id = 1;
```



pull_requests

The diagram illustrates a database query. A yellow rectangular box at the top contains the SQL query: `> SELECT * from pull_requests WHERE id = 1;`. A vertical yellow line connects the bottom center of this box to the top center of a yellow cylinder. The cylinder is labeled 'pull_requests' in the center, representing the database table being queried.


```
> SELECT * from pull_requests WHERE id = 1;
```

VtGate

pull_requests

pull_requests

pull_requests

*Sharded by Repository id

```
> SELECT * from pull_requests WHERE id = 1 AND repository_id = 1;
```

VtGate

pull_requests

pull_requests

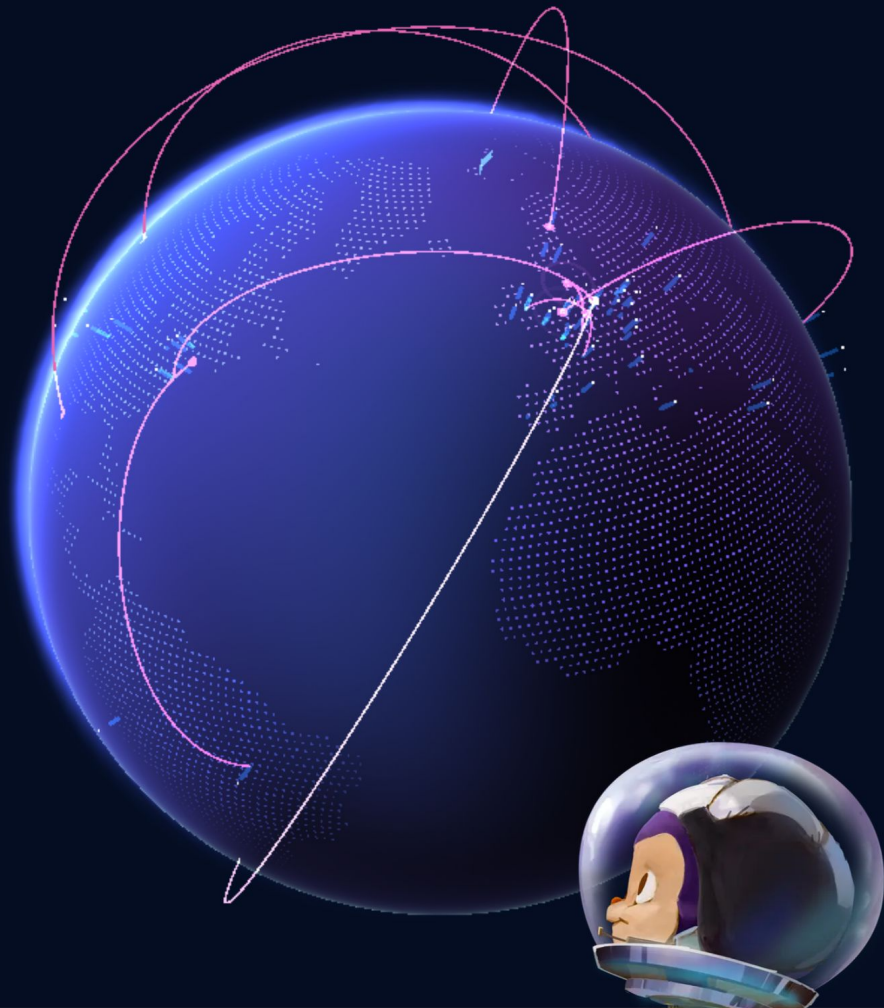
pull_requests

*Sharded by Repository id

Multi-Region

Geo-located data.

- Global Routing
- Multi Data Center
- Complete data isolation



New Format

- Encoded Ownership Information.
- Compressed to reduce size.
- Backwards compatible for users already storing strings.

New Format

PR_kwABAQ



New Format

- PR_ - Type “hint”
- kwABABQ => [0, 1, 1] - ownership/id

Rollout Strategy



Goals

Cannot break users.

Continue to support those who store IDs.

Progressive.

Avoid rollout out all IDs at one time.

Deprecation Period.

Provide time and tooling for users to be able to migrate.

Avoid Breaking Changes

- Support both ID formats.
- Advanced warning via communication channels.
- Backwards compatible format.

“

When building integrations that use either the REST API or the GraphQL API, it's best practice to persist the global node ID so you can easily reference objects across API versions.

Progressive Rollout

- Determine implementation for each object.
- Ensure the correct ownership data.
- Set dates in code for when an object will be “ready”.

Progressive Rollout

```
{
  organization(login: "github") {
    old_format: teams(first: 1) {
      nodes {
        id
      }
    }
    new_format: teams(last: 1) {
      nodes {
        id
      }
    }
  }
}
```

```
{
  "data": {
    "organization": {
      "old_format": {
        "nodes": [
          {
            "id": "MDQ6VGVBTEw"
          }
        ]
      },
      "new_format": {
        "nodes": [
          {
            "id": "T_kwDNJr_OAEz16Q"
          }
        ]
      }
    }
  }
}
```

Deprecation Period

- Time to migration caches to new ID format.
- Tooling to perform ID conversion.
- Warning messages when using old format.
- Eventual brownout and sunset of old format.

Challenges

Some pitfalls encountered along the way.

Challenges

- Decoding IDs.
- Delegating work to first responder.
- Missing/unknown ownership information.



More Challenges

- Initial ID format too long.
- Unstable IDs due to ownership changes.



Thank you for joining.

Andrew Hoglund
GitHub

29th Sep 2021 • graphqlconf.org