

# dolt: git for data

The Data as Code Series

Ernest Guevarra

[ernest@guevarra.io](mailto:ernest@guevarra.io)

17 March 2026

# Learning outcomes

At the end of this session, participants should be able to:

- describe the similarities between `git` and `dolt` and `GitHub` and `DoltHub`
- use concepts of `dolt`-based version control of data in relation to `git`-based version control of code
- identify the advantages and limitations of using `dolt` and `DoltHub` for version control of data
- describe some relevant use cases for `dolt` and `DoltHub` for version control of data

# Session outline

1. What is `do1t` and `DotHub`
2. Demonstration: How to initiate and manage a version-controlled database using `do1t` and `Do1tHub`
3. Use cases for version-controlled database using `do1t` and `Do1tHub`

# What is **do1t** and **Do1tHub**

# dolt

- an SQL database that can be forked, cloned, branched, merged, pushed, and pulled just like a `git` repository

The logo for dolt, featuring the word "dolt" in a stylized, teal-colored font with a monospace-like appearance, set against a dark blue rounded rectangular background.

# Do1tHub

The logo for Do1tHub, featuring the text "d01thub" in a stylized, pink, lowercase font with a monospace-like appearance, set against a dark blue rounded rectangular background.

d01thub

- a place to share `do1t` databases
- public data are hosted for free
- adds a modern, secure, always on database management web GUI to the `do1t` ecosystem.

**do1t is to Do1tHub as git is to GitHub**

# Creating and managing a version-controlled database using **Do1tHub**

# Create and login to your **DoItHub** account

# Go to the *Databases* tab and click on *Create Database* button

The screenshot displays the 'Databases' interface. At the top left, the title 'Databases' is shown. To its right is a search bar with the placeholder text 'Search databases'. Further right is a blue button labeled '+ Create Database'. Below the search bar, there are three tabs: 'My Databases' (highlighted with a pink underline), 'Starred', and 'Discover'. A 'Filter' dropdown menu is located on the right side of the tab bar.

The main content area shows a database entry for 'oxford-ihealth / seystats'. The entry includes a profile icon, the name 'oxford-ihealth / seystats', and a globe icon. Below the name, it says 'Admin' and '441 kB'. At the bottom of the entry, there is a 'Star' button with a count of '1' and a 'Clone' button with a dropdown arrow.

To the right of the database entry, there is a 'Tables' section listing three tables: 'deaths\_endyear\_by\_age\_sex', 'population\_midyear\_by\_age\_sex', and 'population\_midyear\_by\_district'. Below the tables, there is a section titled 'Activity over the last 30 days'. This section contains two progress bars. The first progress bar is for '0 recent pull requests', showing '0 merged' and '0 open'. The second progress bar is for '0 recent issues', showing '0 closed' and '0 open'. At the bottom of the activity section, there is a bullet point indicating the database was 'updated 8 months ago'.

# Give the database a name and a description and then create database

## Databases

+ Create Database

+ Create Database My Databases Starred Discover

**Owner**

**Database Name**

**Database Description** 96/2048

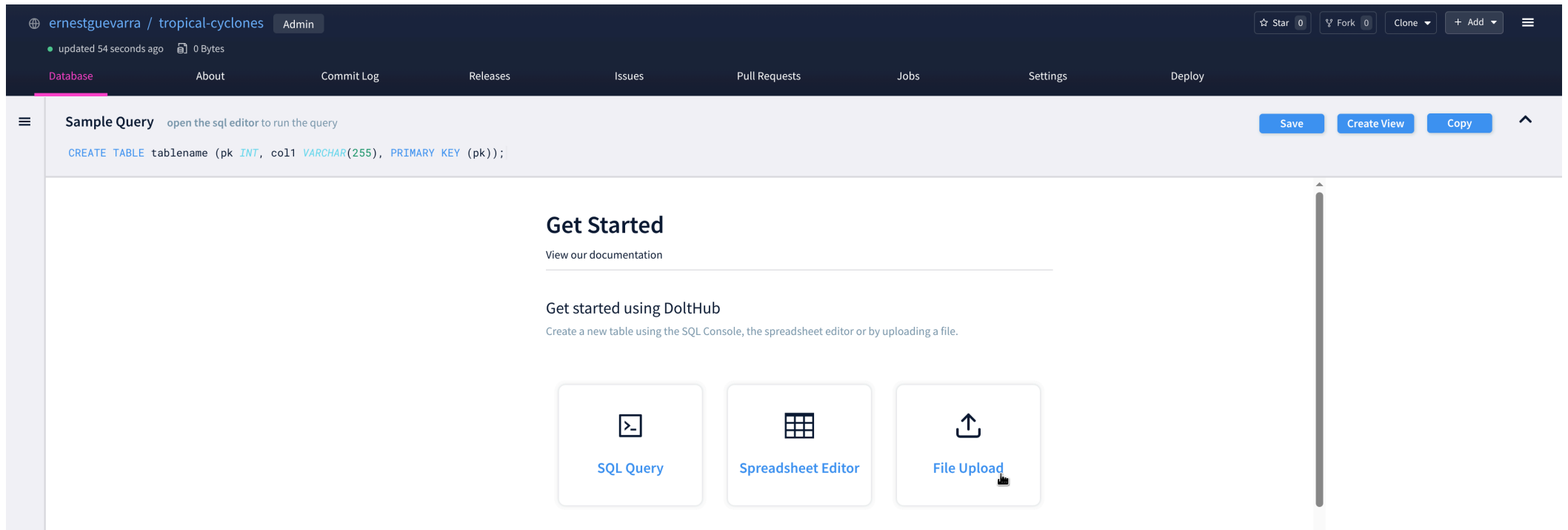
**Public**

**Private**

Want to make private databases? [Sign up for DoltHub Pro!](#)

Create database cancel

# Click on *File upload* option



The screenshot shows the DoltHub interface for a repository named 'tropical-cyclones' by user 'ernestguevarra'. The interface includes a dark header with navigation links (Database, About, Commit Log, Releases, Issues, Pull Requests, Jobs, Settings, Deploy) and action buttons (Star, Fork, Clone, Add). Below the header is a 'Sample Query' editor with a SQL query: `CREATE TABLE tablename (pk INT, col1 VARCHAR(255), PRIMARY KEY (pk));`. The main content area is titled 'Get Started' and contains three options: 'SQL Query', 'Spreadsheet Editor', and 'File Upload'. The 'File Upload' option is highlighted with a mouse cursor, indicating the action to be taken.

ernestguevarra / tropical-cyclones Admin

updated 54 seconds ago 0 Bytes

Database About Commit Log Releases Issues Pull Requests Jobs Settings Deploy

Star 0 Fork 0 Clone + Add

Sample Query open the sql editor to run the query

Save Create View Copy

```
CREATE TABLE tablename (pk INT, col1 VARCHAR(255), PRIMARY KEY (pk));
```

## Get Started

View our documentation

### Get started using DoltHub

Create a new table using the SQL Console, the spreadsheet editor or by uploading a file.

- SQL Query
- Spreadsheet Editor
- File Upload

# Give the new table a name

ernestguevarra / tropical-cyclones File Importer Exit

1. Branch 2. Table 3. Upload 4. Import

[back](#) Choose an option

Upload a SQL file ?

---

Upload a Workflow YAML file ?

---

Create a new table ?

Table name

---

Update an existing table ?

---

Upload a non-supported file type ?

[Show transform file settings](#)

Next

# Click on *Browse files* and then select the data to upload

The screenshot shows a web application interface for a file importer. The main header is dark blue with the text "ernestguevarra / tropical-cyclones" on the left and an "Exit" button on the right. The main content area is light gray and contains a "File Importer" section. In the center, a "File Importer" dialog box is open, showing a file selection interface. The dialog has a "Cancel" button on the top left and a "Select" button on the top right. The main area of the dialog is titled "Open File" and shows the user's desktop. The desktop contains several files and folders, including "FenimoreCoopersDefense-1930.pdf", "github\_issues\_images", "image-2.png", "interview\_links.txt", "llm.txt", "ors\_setup", "sc\_d4dm\_email.txt", "sleacr-app", "top", and "tropical\_cyclones.csv". The "tropical\_cyclones.csv" file is highlighted. Below the file list, there is a checkbox for "Open files read-only" and a dropdown menu set to "(None)". Below the dialog, there are two buttons: "Browse files" and "Open editor". Below these buttons, there is a text label: "File types: .csv, .psv, .xlsx, .json\*, .sql, .yaml, .yml". At the bottom of the main content area, there is a "Next" button.

ernestguevarra / tropical-cyclones

File Importer

Cancel

Open File

Recent

Home

Desktop

Documents

Downloads

Music

Pictures

File types: .csv, .psv, .xlsx, .json\*, .sql, .yaml, .yml

Next

or

Browse files

Open editor

Name	Size	Type	Modified
FenimoreCoopersDefense-1930.pdf	1.2 MB	Document	17 Oct 2025
github_issues_images			3 Feb
image-2.png	221.6 kB	Image	2 Feb
interview_links.txt	464 bytes	Text	16 Dec 2025
llm.txt	5.4 kB	Text	6 Nov 2025
ors_setup			25 Sep 2025
sc_d4dm_email.txt	604 bytes	Text	20 Jun 2025
sleacr-app			28 Dec 2024
top			8 Mar
tropical_cyclones.csv	458 bytes	Text	12:56

# Choose primary keys for the table you uploaded

ernestguevarra / tropical-cyclones

File Importer

Exit

1. Branch 2. Table 3. Upload 4. Import

< back

Upload file or open spreadsheet editor

Create table cyclones

✓ Upload successful  
tropical\_cyclones.csv  
459 Bytes  
remove

or

Create spreadsheet  
Open editor

Keyless table ? or Choose primary keys

Select at least one primary key column.

id  
 year

# Add a commit message and specify a branch (optional)

< ernestguevarra / tropical-cyclones

File Importer

Exit

1. Branch 2. Table 3. Upload 4. Import

< back

## Start import

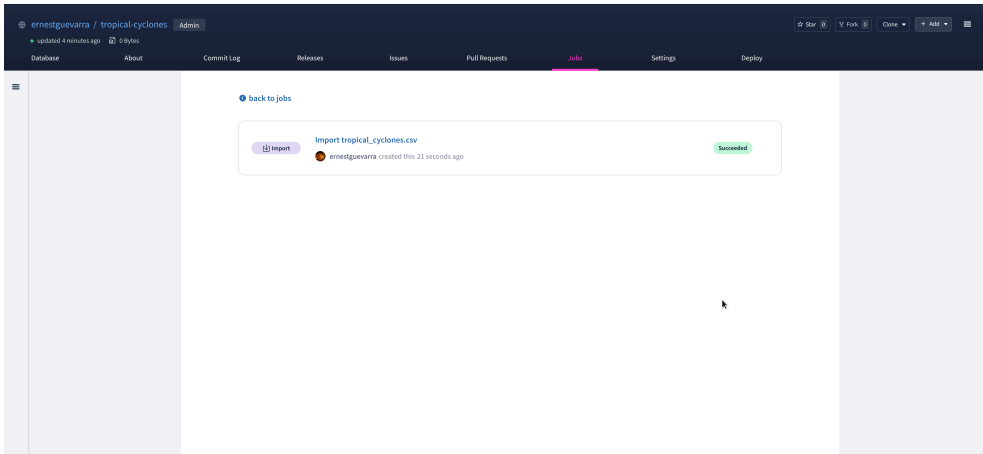
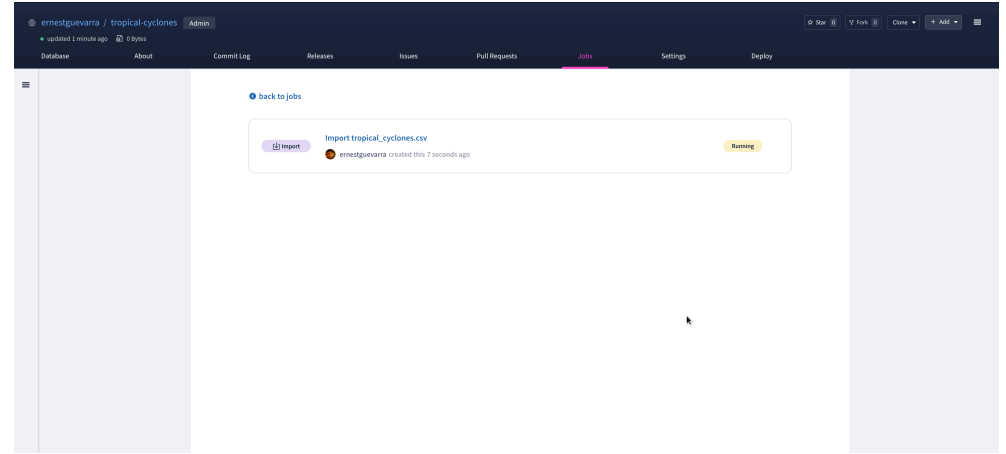
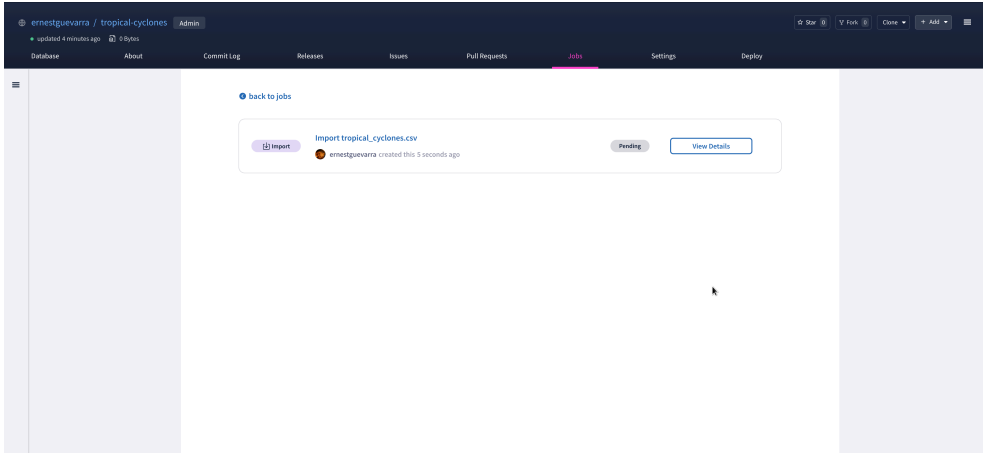
**Optional commit message**  
Provide a custom commit message for these changes. If none is provided, we will use a default commit message.

create cyclones database

**New branch name**

dev

Start import



# Review pull request and merge

The screenshot shows the GitHub interface for a pull request. At the top, the repository name 'ernestguevarra / tropical-cyclones' is displayed with an 'Admin' button. The top right corner contains 'Star 0', 'Fork 0', 'Clone', and '+ Add' buttons. A navigation bar includes 'Database', 'About', 'Commit Log', 'Releases', 'Issues', 'Pull Requests' (highlighted), 'Jobs', 'Settings', and 'Deploy'. The main content area is titled 'Pull Requests' and features a 'Create Pull Request' button. Below this is a search bar 'Search by title or creator' and two filter dropdowns: 'Filter by pull state' and 'Filter by review status'. A single pull request is listed: 'cyclones created with data from tropical\_cyclones.csv' by user 'ernestguevarra', opened 29 seconds ago. A 'View diff >' link is visible to the right of the pull request entry. A URL bar at the bottom shows 'https://www.dolthub.com/repositories/ernestguevarra/tropical-cyclones/pulls?refName=main'.

ernestguevarra / tropical-cyclones Admin

updated 5 minutes ago 0 Bytes

Database About Commit Log Releases Issues **Pull Requests** Jobs Settings Deploy

back to pull requests

### cyclones created with data from tropical\_cyclones.csv edit #1 View Diff >

OPEN ernestguevarra wants to merge 1 commit into main from ernestguevarra/import-torpid-blackbird

ernestguevarra commented 51 seconds ago ...  
cyclones created with data from tropical\_cyclones.csv

ernestguevarra added 1 commit 51 seconds ago

create cyclones database k0bk3dL  
17/03/2026, 13:06

ernestguevarra opened this 50 seconds ago

Ready to merge. Merge

View manual merge command line instructions.  
Want to run tests on your pull request before merging? Learn how to set up CI here.

Reviewers  
select or type to add  
No Reviews

# Database is now created in main branch ...

ernestguevarra / tropical-cyclones Admin

updated 2 minutes ago 3.26 kB

Database About Commit Log Releases Issues Pull Requests Jobs Settings Deploy

main +

Query

```
SELECT * FROM `cyclones`
```

Save Create View Copy

Tables Views Queries Schemas

^ cyclones Viewing

id	year	category_code	category_name	name
1	2017	TD	Tropical Depression	Auring
2	2017	TD	Tropical Depression	Bising
3	2017	TD	Tropical Depression	Crising
4	2017	TS	Tropical Storm	Dante
5	2017	STS	Severe Tropical Storm	Emong
6	2017	TD	Tropical Depression	Fabian
7	2017	TY	Typhoon	Gorio
8	2017	TS	Tropical Storm	Huaning
9	2017	STS	Severe Tropical Storm	Isang
10	2017	TS	Tropical Storm	Jolina

ER Diagram

+ Add new table

# ... and it is the same as in dev branch

The screenshot shows a database management interface for a repository named 'ernestguevarra / tropical-cyclones'. The interface is dark-themed and includes a navigation bar with links for 'Database', 'About', 'Commit Log', 'Releases', 'Issues', 'Pull Requests', 'Jobs', 'Settings', and 'Deploy'. The 'Database' section is active, showing a 'dev' branch. A query is executed: 'SELECT \* FROM `cyclones`'. The results are displayed in a table with columns: 'id', 'year', 'category\_code', 'category\_name', and 'name'. The table contains 10 rows of data. On the left, there is a sidebar with 'Tables', 'Views', 'Queries', and 'Schemas' tabs. The 'cyclones' table is selected, and its schema is shown: 'id' (INT NOT NULL), 'year' (INT), 'category\_code' (VARCHAR(200)), 'category\_name' (VARCHAR(200)), and 'name' (VARCHAR(200)). There is also an 'ER Diagram' button and an 'Add new table' option.

id	year	category_code	category_name	name
1	2017	TD	Tropical Depression	Auring
2	2017	TD	Tropical Depression	Bising
3	2017	TD	Tropical Depression	Crasing
4	2017	TS	Tropical Storm	Dante
5	2017	STS	Severe Tropical Storm	Emong
6	2017	TD	Tropical Depression	Fabian
7	2017	TY	Typhoon	Gorio
8	2017	TS	Tropical Storm	Huaning
9	2017	STS	Severe Tropical Storm	Isang
10	2017	TS	Tropical Storm	Jolina

# Making versioned changes to the database

# On **dev** branch, add a new row of data to the cyclones table

ernestguevarra / tropical-cyclones Admin Star 0 Fork 0 Clone + Add

updated 2 minutes ago 3.26 kB

Database About Commit Log Releases Issues Pull Requests Jobs Settings Deploy

dev + ≡

Query Save Create View Copy

`SELECT * FROM `cyclones``

Tables Views Queries Schemas

^ cyclones Viewing

id INT NOT NULL  
year INT  
category\_code VARCHAR(200)  
category\_name VARCHAR(200)  
name VARCHAR(200)

ER Diagram

+ Add new table

Table SQL API

id	year	category_code	category_name	name
1	2017	TD	Tropical Depression	Auring
2	2017	TD	Tropical Depression	Bising
3	2017	TD	Tropical Depression	Crising
4	2017	TS	Tropical Storm	Dante
5	2017	STS	Severe Tropical Storm	Emong
6	2017	TD	Tropical Depression	Fabian
7	2017	TY	Typhoon	Gorio
8	2017	TS	Tropical Storm	Huaning
9	2017	STS	Severe Tropical Storm	Isang
10	2017	TS	Tropical Storm	Jolina

Options

- + Add a row
- Download current results as CSV
- Start a job to download unlimited results as CSV
- Download cyclones as CSV
- Download cyclones as CSV for Excel

ernestguevarra / tropical-cyclones Admin

updated 3 minutes ago 3.26 kB

Database About Commit Log Releases Issues Pull Requests Jobs Settings Deploy

dev

Query

SELECT \* FROM `cyclones`

Save Create View Copy

Tables Views Queries Schemas

^ cyclones Viewing

id INT NOT NULL  
year INT  
category\_code VARCHAR(200)  
category\_name VARCHAR(200)  
name VARCHAR(200)

ER Diagram

+ Add new table

Table SQL API

id	year	category_code	category_name	name
11	2017	TD	Tropical Depression	Kidlat
1	2017	TD	Tropical Depression	Auring
2	2017	TD	Tropical Depression	Bising
3	2017	TD	Tropical Depression	Crising
4	2017	TS	Tropical Storm	Dante
5	2017	STS	Severe Tropical Storm	Emong
6	2017	TD	Tropical Depression	Fabian
7	2017	TY	Typhoon	Gorio
8	2017	TS	Tropical Storm	Huaning
9	2017	STS	Severe Tropical Storm	Isang
10	2017	TS	Tropical Storm	Jolina

submit

# Create a pull request from **dev** to **main** with the new row of data

The screenshot shows a database management interface for a repository named 'ernestguevarra / tropical-cyclones'. The interface is dark-themed and includes a navigation bar with options like 'Database', 'About', 'Commit Log', 'Releases', 'Issues', 'Pull Requests', 'Jobs', 'Settings', and 'Deploy'. The 'Database' section is active, showing a workspace named 'dev' with '1 open workspace'. A SQL query is displayed: `INSERT INTO `cyclones` (`id`, `year`, `category_code`, `category_name`, `name`) VALUES (('11', '2017', 'TD', 'Tropical Depression', 'Kidlat'))`. Below the query, it states 'Query OK, 1 row affected.' and provides tabs for 'Query History', 'Cumulative Diff', and 'Current Diff'. A 'Create pull request' button is visible in the top right corner of the workspace area. The 'Query History' tab is selected, showing a run SQL query with the same code as above, executed by 'ernestguevarra' on '17/03/2026, 13:22'.

ernestguevarra / tropical-cyclones Admin

updated 7 minutes ago 3.26 kB

Database About Commit Log Releases Issues **Pull Requests** Jobs Settings Deploy

## New Pull Request

**Base database:** ernestguevarra/tropical-cyclones

**From database:** ernestguevarra/tropical-cyclones

**Base branch:** main

**From branch:** dev

**Title**  
added most recent cyclone to database

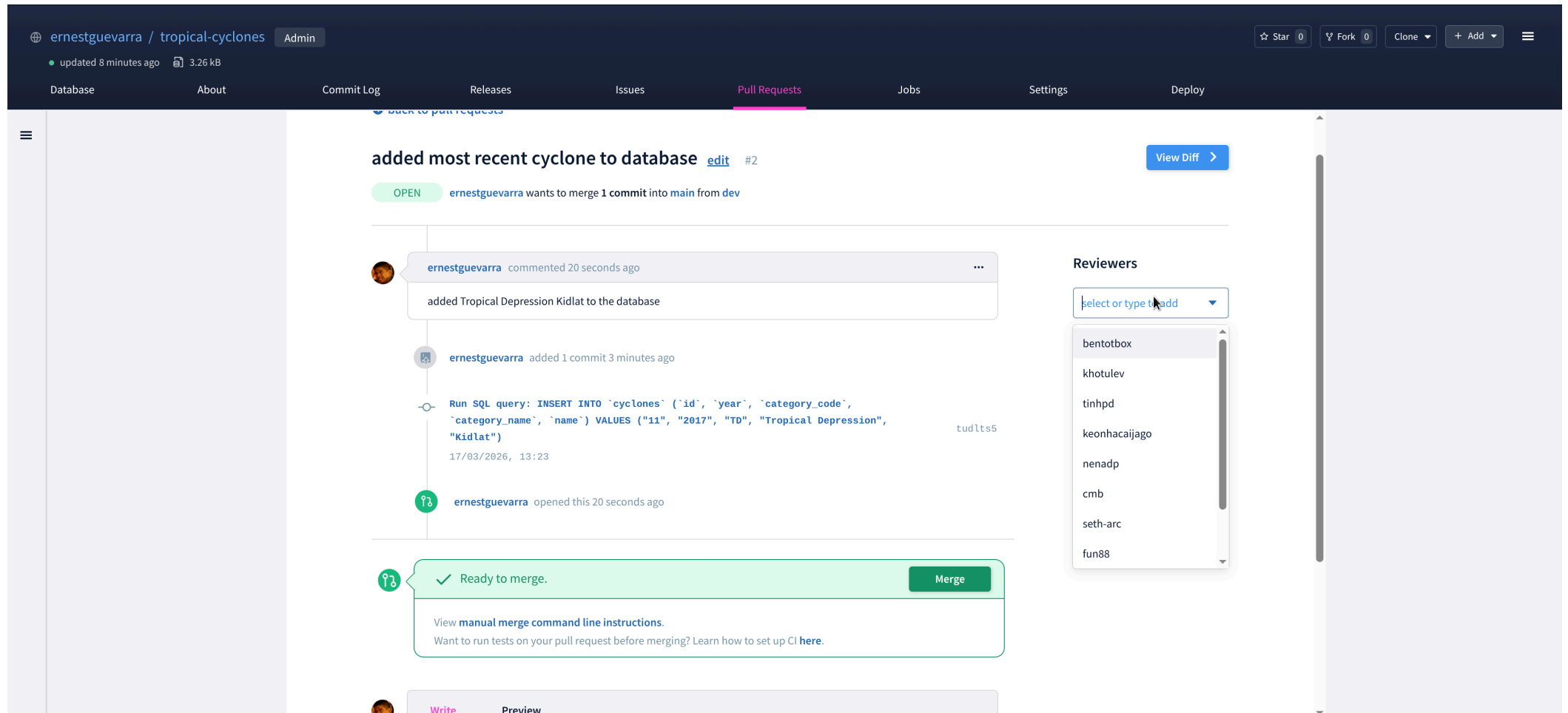
**Description**

Write Preview 48/2048

added Tropical Depression ~~Kidlat~~ to the database

Create pull request cancel

# Select a reviewer of the pull request



The screenshot shows a GitHub pull request for the repository 'ernestguevarra / tropical-cyclones'. The pull request title is 'added most recent cyclone to database' and it is currently in an 'OPEN' state. A comment from 'ernestguevarra' states 'added Tropical Depression Kidlat to the database'. The commit message is 'Run SQL query: INSERT INTO `cyclones` (`id`, `year`, `category\_code`, `category\_name`, `name`) VALUES ("11", "2017", "TD", "Tropical Depression", "Kidlat")' with a commit hash of 'tudlts5' and a timestamp of '17/03/2026, 13:23'. A status bar at the bottom indicates 'Ready to merge' with a 'Merge' button. On the right side, the 'Reviewers' dropdown menu is open, showing a list of users: bentotbox, khotulev, tinhpd, keonhacajago, nenadp, cmb, seth-arc, and fun88.

ernestguevarra / tropical-cyclones Admin

updated 8 minutes ago 3.26 kB

Database About Commit Log Releases Issues Pull Requests Jobs Settings Deploy

added most recent cyclone to database edit #2 View Diff >

OPEN ernestguevarra wants to merge 1 commit into main from dev

ernestguevarra commented 20 seconds ago ...

added Tropical Depression Kidlat to the database

ernestguevarra added 1 commit 3 minutes ago

Run SQL query: INSERT INTO `cyclones` (`id`, `year`, `category\_code`, `category\_name`, `name`) VALUES ("11", "2017", "TD", "Tropical Depression", "Kidlat") tudlts5  
17/03/2026, 13:23

ernestguevarra opened this 20 seconds ago

Ready to merge. Merge

View manual merge command line instructions.  
Want to run tests on your pull request before merging? Learn how to set up CI here.

Write Preview

Reviewers

select or type to add

- bentotbox
- khotulev
- tinhpd
- keonhacajago
- nenadp
- cmb
- seth-arc
- fun88

# Review the changes proposed by the pull request

ernestguevarra / tropical-cyclones / pull requests / 2 / diff

☆ Star 0   🍴 Fork 0   Clone   + Add   ☰

[← Back to Pull Request](#)

all commits

## cyclones

0 Rows Deleted 0.00%   1 Row Added 10.00%   0 Rows Modified 0.00%   0 Cells Modified 0.00%   10 Rows Unmodified 100.00%

Data   Schema (0)

Options   Filter by all changes

	id	year	category_code	category_name	name
+	11	2017	TD	Tropical Depression	Kidlat

Diff Overview

-0 rows deleted   +1 row added   0 rows modified

cyclones   -0 +1 0



# New row of data is now on **main**

ernestguevarra / tropical-cyclones Admin

updated 10 minutes ago 3.26 kB

Database About Commit Log Releases Issues Pull Requests Jobs Settings Deploy

main + ≡

Query 1 open workspace Save Create View Copy

SELECT \* FROM `cyclones`

Tables Views Queries Schemas

^ cyclones Viewing

id <sup>♂</sup> INT NOT NULL

year INT

category\_code VARCHAR(200)

category\_name VARCHAR(200)

name VARCHAR(200)

ER Diagram

+ Add new table

Table SQL API ?

id <sup>♂</sup>	year	category_code	category_name	name
1	2017	TD	Tropical Depression	Auring
2	2017	TD	Tropical Depression	Bising
3	2017	TD	Tropical Depression	Crasing
4	2017	TS	Tropical Storm	Dante
5	2017	STS	Severe Tropical Storm	Emong
6	2017	TD	Tropical Depression	Fabian
7	2017	TY	Typhoon	Gorio
8	2017	TS	Tropical Storm	Huaning
9	2017	STS	Severe Tropical Storm	Isang
10	2017	TS	Tropical Storm	Jolina
11	2017	TD	Tropical Depression	Kidlat

Options

+

# Use cases

- Data collaboration and curation on **Do1tHub**
- Distributing versioned data through **do1t** as an alternative to APIs
- Data model quality control and versioning

# What questions do you have?

# Thank you!

- Slides can be viewed at <https://d4dm.org/dolt-git-for-data>
- PDF version of slides can be downloaded at <https://d4dm.org/dolt-git-for-data/dolt-git-for-data.pdf>
- R scripts for slides available at <https://github.com/data4decisionmakers/dolt-git-for-data>

license (for code) [GPL3.0](#) license (for writing) [CC BY 4.0](#)