

I Accidentally All the Data

Did we really need any of that?

To Lee, Bobby, and Nicolas:

thank you for not making my blood pressure even worse than it was

Hi! I'm Elizabeth

I'm a data scientist.

I've been doing this a **long** time (11 years!)

But, at one point, I hadn't been doing this a long time.

This is a story about that time.

When I forgot a single  when copying a file.

Accidentally deleted all the data in our datawarehouse.

And then learned very very quickly what to do when you screw up that badly.





What the heck was going on?



A Very Genius plan: move our data to hive!



A quick detour: Linux!

```
cp /path/to/file.csv path/to/target
```


Getting data to Hive back in the day

```
hdfs dfs -copyFromLocal /path/to/file/myfile.csv /path/to/data/myfile.csv
```

And here's what I did because I Knew Linux™

```
hdfs dfs -copyFromLocal /path/to/file/myfile.csv /path/to/data
```

And that's how I overwrote a directory with a file

(my bad!)

What to do when you accidentally all the data

1. Take a deep breath and remember that you didn't set out to wipe the entire database.



(unless you did, in which case: well done!)

What to do when you accidentally all the data

2. Tell your data team immediately.



Data people tend to notice when all the data goes missing

What to do when you accidentally all the data

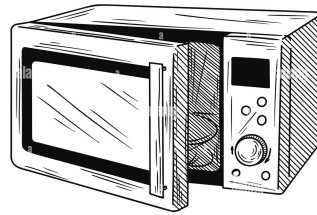
3. Be honest!



It doesn't help anyone to lie, and there's probably server logs anyways

What to do when you accidentally all the data

4. Focus on *how* did this happen and *what* can be done to prevent it in the future.



If your company gives you gas and matches to heat your burrito instead of a microwave, maybe there needs to be a discussion about the budget merits of kitchen appliances

So what happened after all that?

They all queried happily ever after

1. The database team just reloaded everything from the raw server logs. Nothing was lost!
2. Manager ran a post-mortem: turns out “user permissions” weren’t a thing that existed yet!
3. Our database team was actually *happy* it happened. They’d been hounding our research team for weeks to delete old files to free up space. Now there were no more files of any kind!
4. I got a promotion that year! 🤖

Thank you!

Next time: how I almost accidentally spent \$780,000 on ETL

...unless an even bigger mishap happens before then