

# Tool Comparison – lakeFS vs DVC

Feature	lakeFS	DVC
Installation	Docker setup, UI-heavy	Pip install, CLI-based
Data Versioning	Git-like branches over S3/Lake	Git-like tracking over local/remote
Switching Between Versions	Easy via <a href="#">lakeFS UI</a> or CLI	Easy via <a href="#">dvc checkout</a> , but less visual
Versioning Granularity	File + object-level (entire lake)	File + directory level
Integration with ML Workflow	Great for collaborative pipelines	Tighter local workflow control
DP Impact on Accuracy (v2 only)	DP-SGD model ran on lakeFS-managed v2	DVC tracked v2 showed same metrics
Overall Experience	Great for large teams, cloud-native	Great for small teams, Git-style local

## Summary:

- **lakeFS** excels in **cloud-native versioning**, ideal for large, collaborative workflows.
- **DVC** is simple, **dev-friendly**, and better for fast iteration in local ML projects.
- **Both tools** enabled clean versioning, reproducible training, and smooth DP experimentation.

## DP vs Non-DP Model Comparison (v2 dataset)

Model Type	MSE	R <sup>2</sup> Score	$\epsilon$ (Privacy Loss)
Non-DP	43,620.04	0.44	—
DP-SGD	40,577.72	0.48	<b>0.97</b> ( $\delta=1e-5$ )

### Observations:

- DP model achieved **comparable accuracy** to the non-DP model.
- A slight **performance gain** (higher R<sup>2</sup>, lower MSE) may be due to regularization effect of DP noise.
- **Privacy budget  $\epsilon=0.97$**  provides a **reasonable privacy guarantee** while retaining utility.