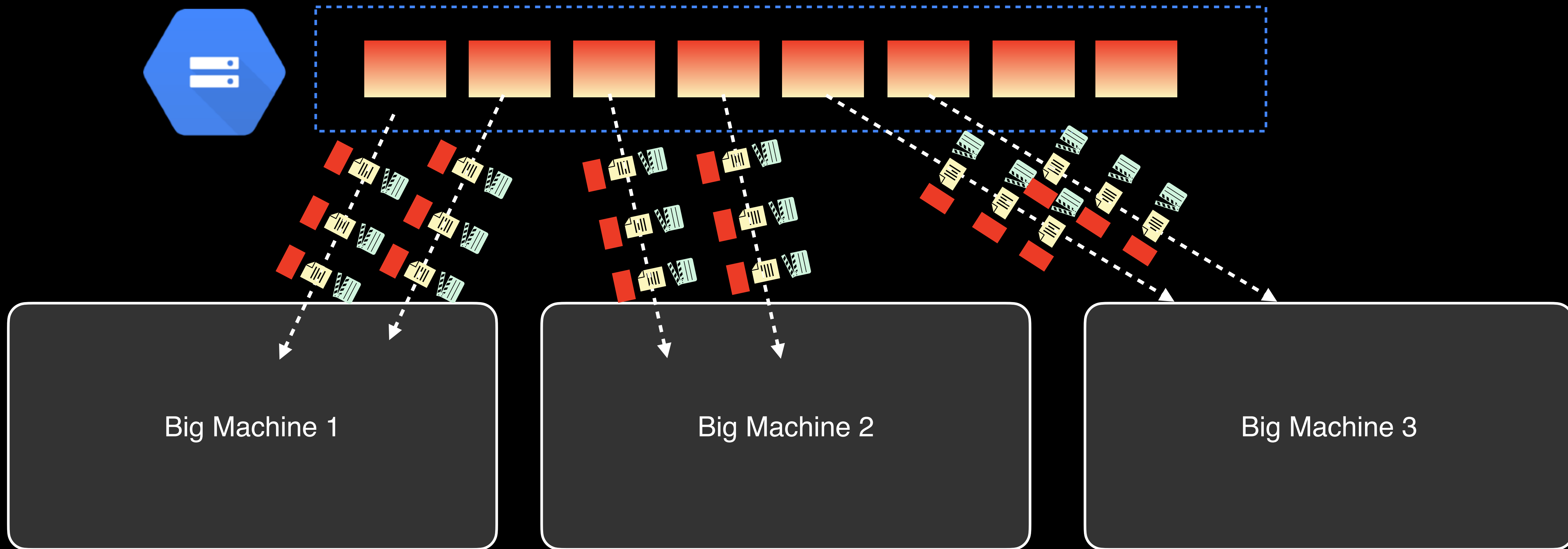


Webdataset (aka tarballs + API)



Taking Stock — Webdataset

- ✓ 1. Massive unstructured data
- ✓ 2. Structured data
- ✗ 3. Large column appends w/ backfills
- 😐 4. Dynamic partitioning across workers
- ✗ 5. EDA