Amazon Conference Call #2:

Team: Abbreviated (T)   AWS: Abbreviated (A)
— — — — — — — — — — — — — — — —

AWS Representative: David Gruley (Architect)
David Lurie (Sales)

David Gruley: — Missed call — needs to backup.
— — — — — — — — — — — — — — — —

- o— Big Data Problem that needs solving and provide a platform for us to work with.

- 0 — 13 windfarms — Minute × Minute
    - ↳ 3 lines × 13, ← per minute
        - ↳ .csv file

- 0 — 5TB for storage

- 0 — Spark Cluster (Virtual Nodes)

- 0 — Hadoop Backbone — { specific versions of Hadoop only. }

o – Hadoop Backbone ⌉ specific versions
of Hadoop only.

2.3 → 2.5

o Versions of Spark?

minimum: 1.4.1 → R capabilities
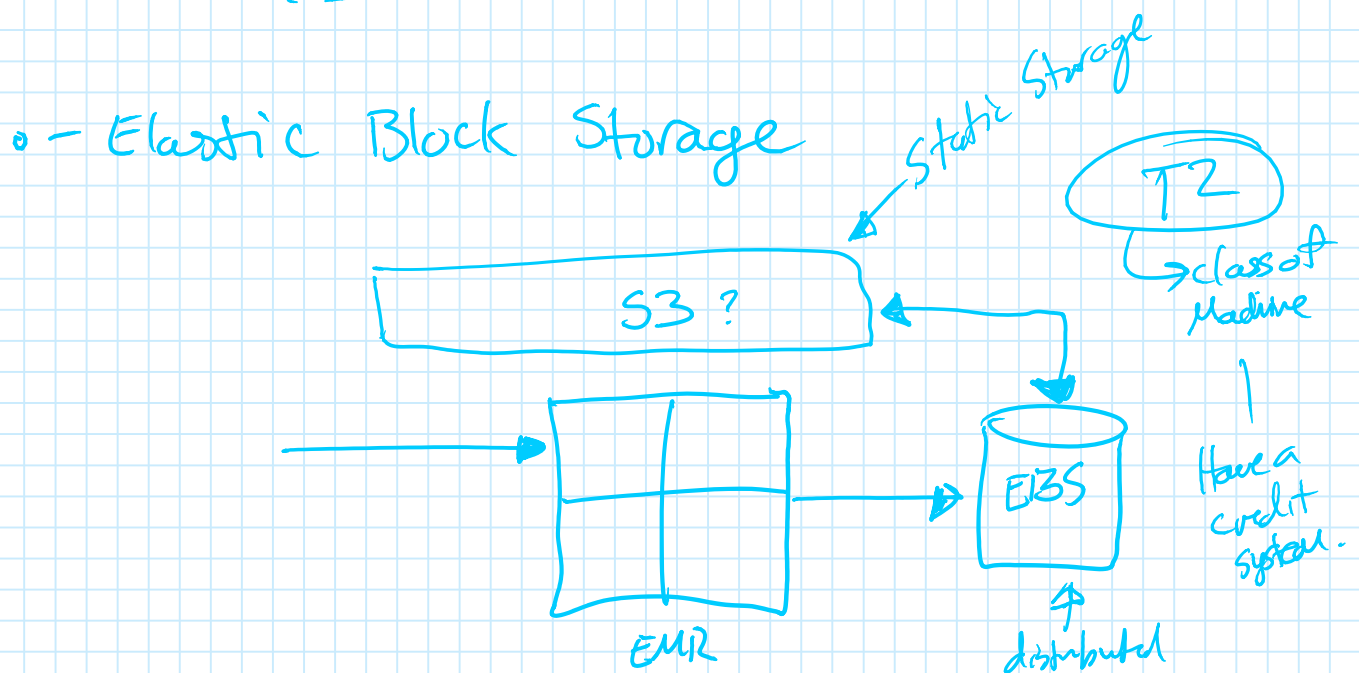ideally: 1.51 → newest.

1.5.x

o – Amazon Linux? ⌉
⌉ OS used
o – CentOS? ⌋ on VM,

o – EMR ??

o o Amazon Linux – Compatable w/ RHEL

o "Spot Instance" ← Bid System

- We cannot clone Amazon EMR
  but we can export

- — Job scheduled ← Data Pipeline
       ↳ via cron.

- — Pick different Region to host our data
       ↳ Seattle? ← - ✳ ✳
       ↳ San Fran?

- — Elastic Block Storage

Static Storage

S3 ?

T2 → class of Machine

EBS

Have a credit system.

EMR

↑ distributed

- Tech Support —| Basic

        —| Email-only Support

        =| 12-24 Hour response ($50/month)

→ 24/7 — ($100/month) ≈ min or more ↑
       └▶ credits?

Next Steps: — Micro instances are free
              — Deployment is _Self-Serve_

              — Business Costs ....