

GROUP 1

1. WAN-LIKE services as bridge
↳ tools to help debug
2. MOVE TO PYTHON ECOSYSTEM
↳ benefit
↳ bring own expertise
3. CONTAINERS not yet enough
↳ how to capture work in sustainable way
↳ who does it
↳ long-term

Group 2

- ① Storage - use virtual data
↳ more CPU - storage trade off (regenerate on disk)
↳ Do a \$ analysis
- ② Batch systems → workflows
↳ Default way users approach analysis
↳ "sociology"
- ③ Jup. notebooks: Use for publications
↳ constantly factor out bits to keep the notebook size under control

GROUP 4

1. METADATA
↳ IN CONTAINERS
↳ TOO BIG?
↳ IN CENTRAL DB
↳ MANAGE FOREVER
• DIFF. VS EVENT DATA
→ UNIFY / STANDARDIZE
2. INTERFACE FOR ANAL.
↳ INTERNAL + EXTERNAL

virtual data
→ means this is in both places - not fully preserved

GROUP 5

1. GROUP ANALYSIS PRODUCTION
↳ users need to buy into analysis
↳ latency → run jobs overnight
↳ run CENTRALLY (organized)
↳ how to do this at scale?
↳ avoid having to debug container pods
↳ of the intermediate data?
↳ Competition in facilities?
2. ANALYSIS PRESERVATION
↳ short → industry std (containers)
↳ long term → high-level description language
↳ why we do things this way (human knowledge)
3. + 1 to

Group 6

- ① Column extraction from datasets by index (add a missed variable)
↳ doesn't work w/ some provenance
↳ Network / infrastructure is not reliable enough

② Virtual data set

↳ latency is ok (today) as long as the user doesn't need to do anything
↳ if you make this too easy, then they may overwhelm the system

③ Virtual Data

↳ Evolution away from the ~~job~~ job to analysis workflow
↳ train people starting w/ workflow
↳ workflow - work towards understandable physics language
↳ analysis language
↳ syntax
↳ even

