# Minerva SCPMT 2020

# Experiment Organization Chart for Offline Computing

- **Computing**:
  - Offline Infrastructure Coordinator: Andrew Olivier (Rochester)
  - Team Members: Satyajit Jena (IISER Mohali), Hang Su (Pittsburgh), David Last (UPenn)
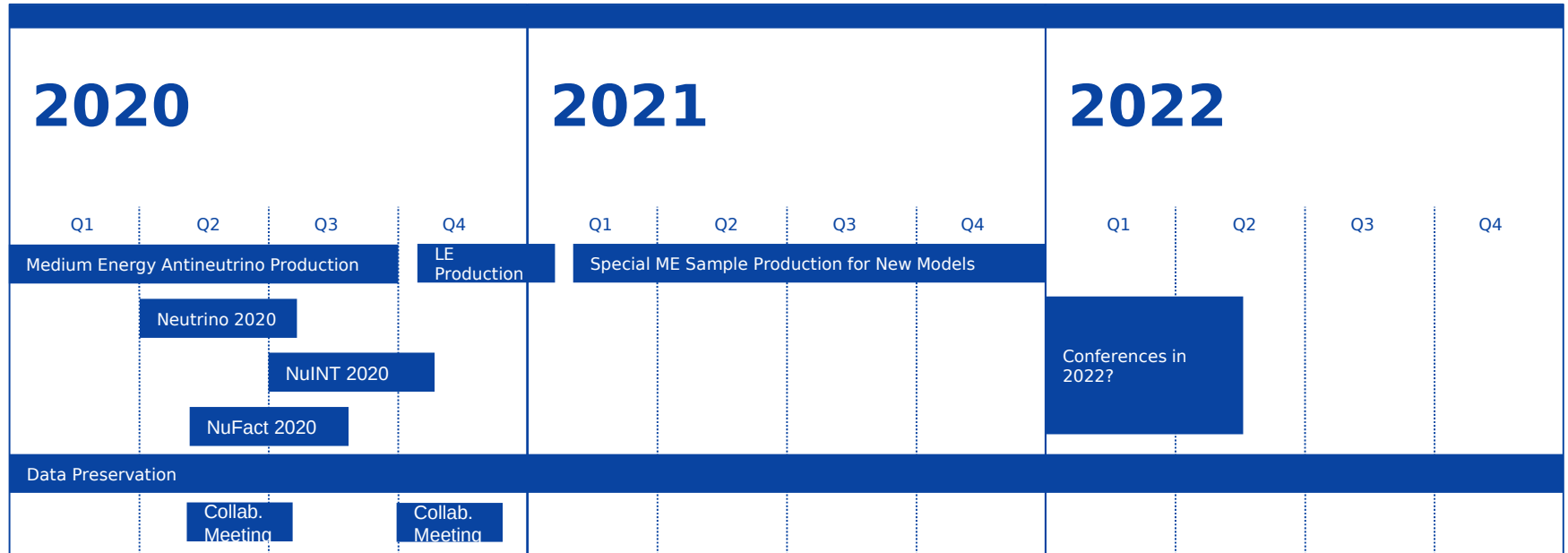- **Production**:
  - Production Coordinator: Deepika Jena (Fermilab)
  - Team Members: Luis Zazueta (William and Mary), Jennica LeClerc (University of Florida)
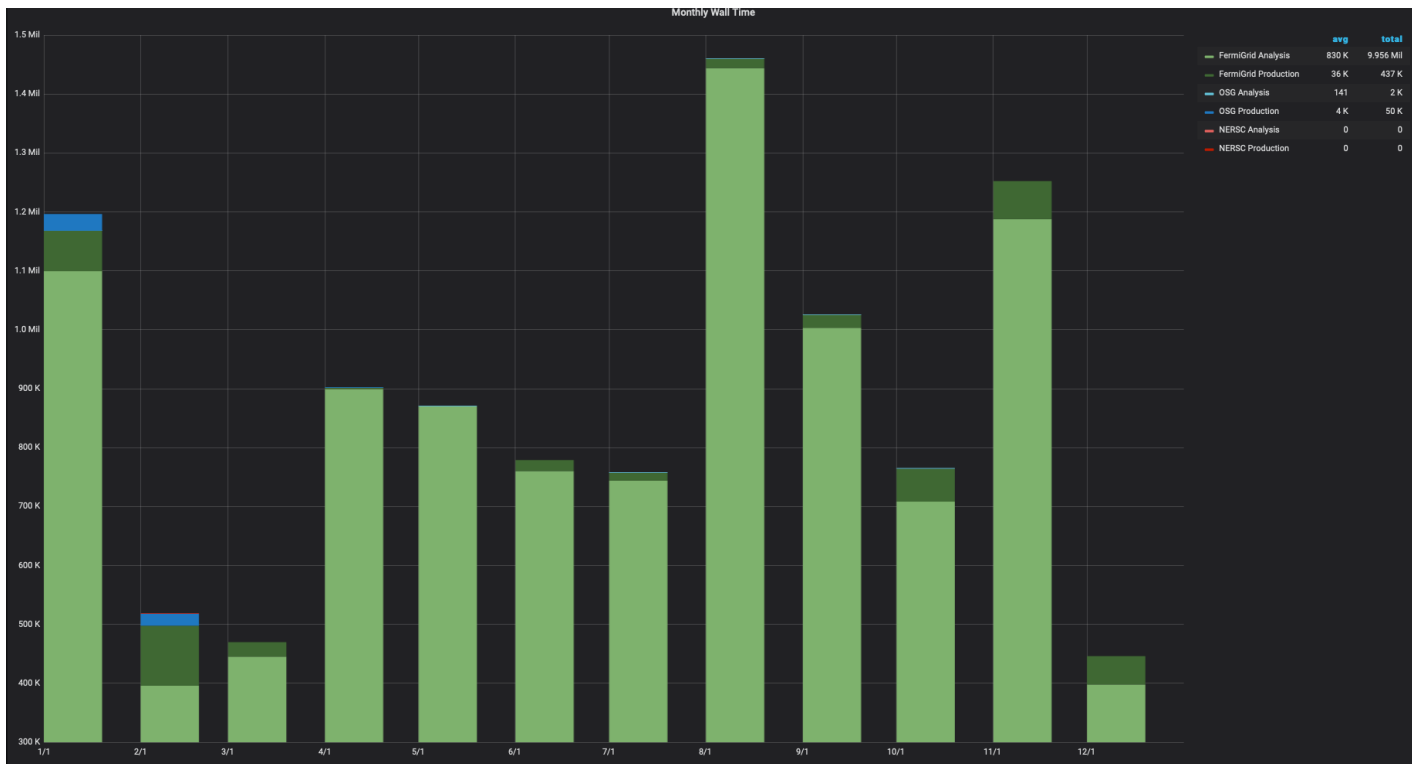- **Analysis Coordinator**: Dan Ruterbories (Rochester)
- Release management coordinated by computing team
- Priority for computing team is production

<Minerva>

# Important Dates to Remember

| | 2020 | | | | 2021 | | | | 2022 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |

Medium Energy Antineutrino Production | LE Production | Special ME Sample Production for New Models

Neutrino 2020

NuINT 2020

NuFact 2020

Conferences in 2022?

Data Preservation

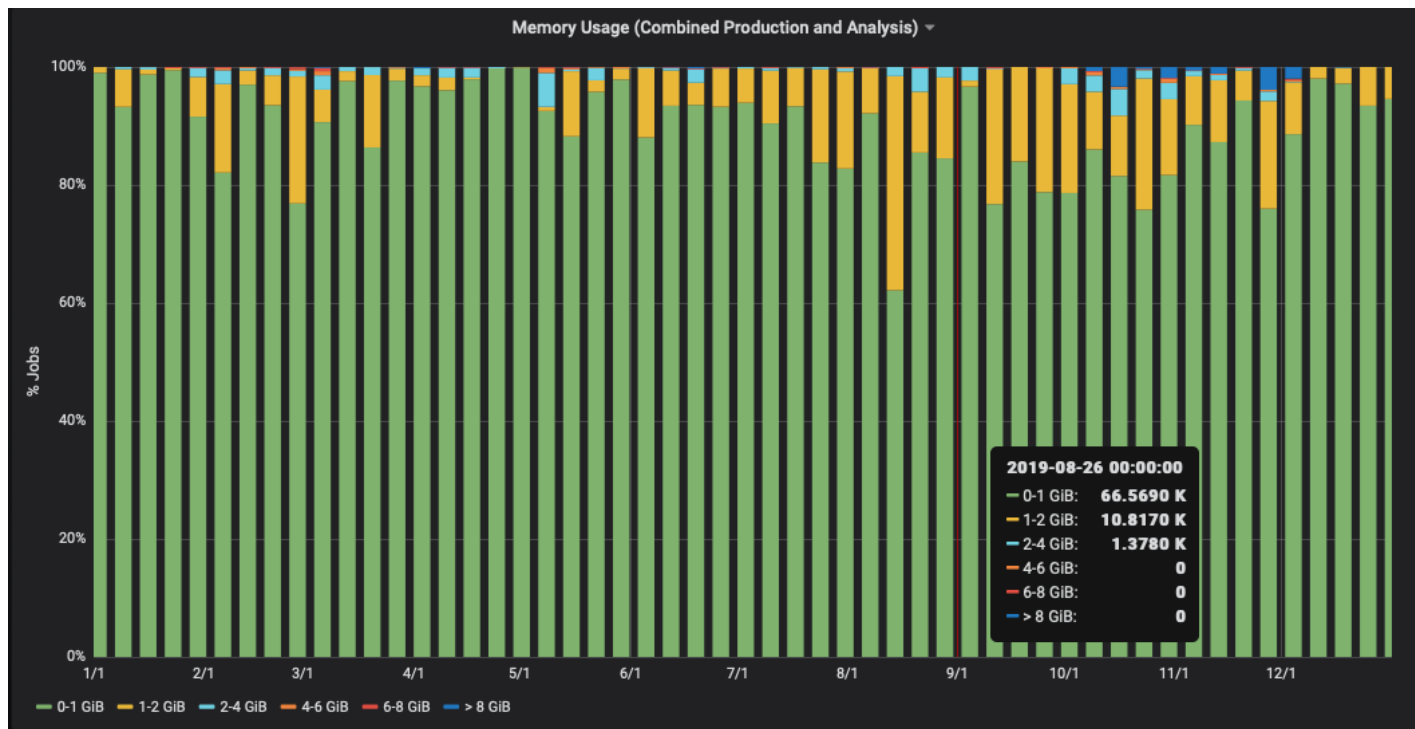Collab. Meeting

Collab. Meeting

<Minerva>

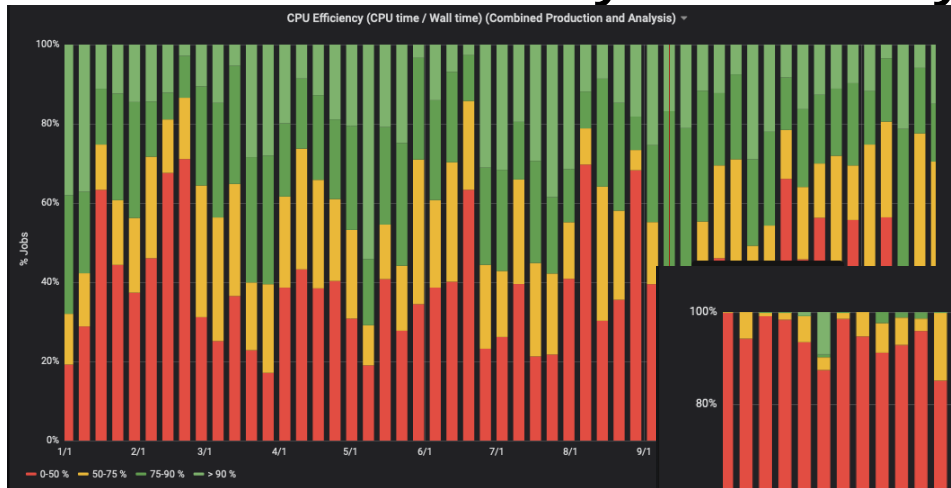# CPU  - Experiment Usage Over the Last Year



- No plan for HEPCloud. Focusing on stability

- Delay publication if peak not attainable

<Minerva>

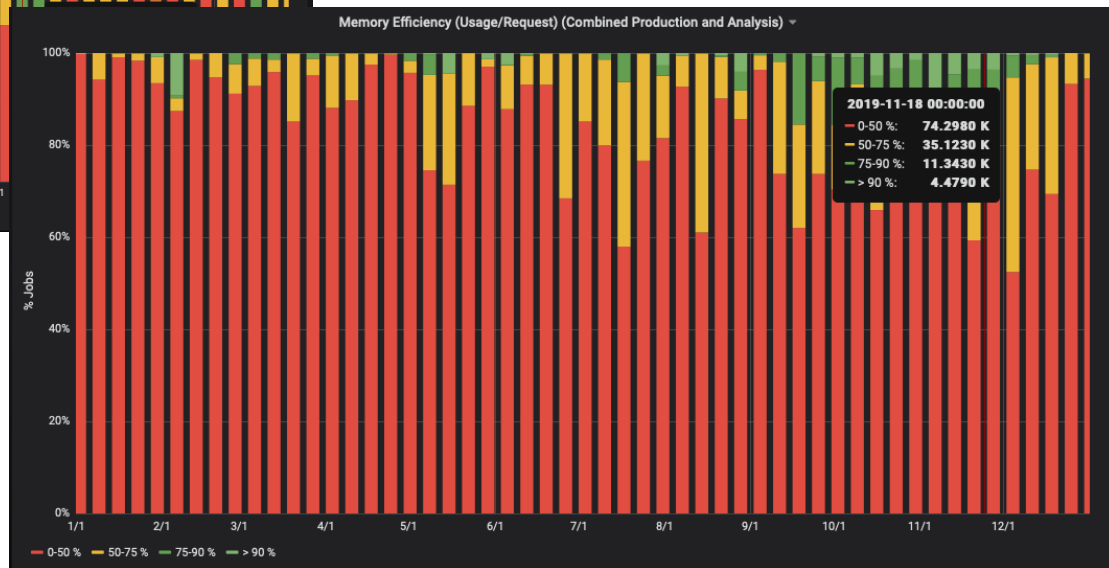# Memory Footprint Over the Last Year

<Minerva>

# CPU and Memory Efficiency Over the Last Year



- Are CPU-inefficient jobs waiting on streamed input files?  Database access?

- No planned improvements

- Low hanging fruit: reduce default memory request?

<Minerva>

# CPU - Prediction Going Forward and Accuracy of Your Predictions  [units of Million (1 CPU, 2GB) wall hours per CY]

|  | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** |
|---|---|---|---|---|---|---|
| **Requested** |  | 18.31 | 16.7 | 16.24 | 14 | 12 |
| **Actual Used** | 10.47 | 14.02 | 10.39 | ? thru Feb | N/A | N/A |
| **Efficiency** | ? | 77% | 62% | To date: | N/A | N/A |

<Minerva>

# CPU Adaptations Going Forward

How can MINERvA use OSG/HPC/HEPCloud going forward?

- GENIE jobs for production can use OSG

- MINOS reconstruction blocks other production stages because of old libraries

- 2 analysis stages:

  - Gaudi stage: More library requirements → not feasible on OSG with current human power

  - Macro stage: ROOT without framework
    - Far fewer library requirements
    - Higher dimensional analyses need > 2GB
    - → investigate using OSG for 1D analyses?

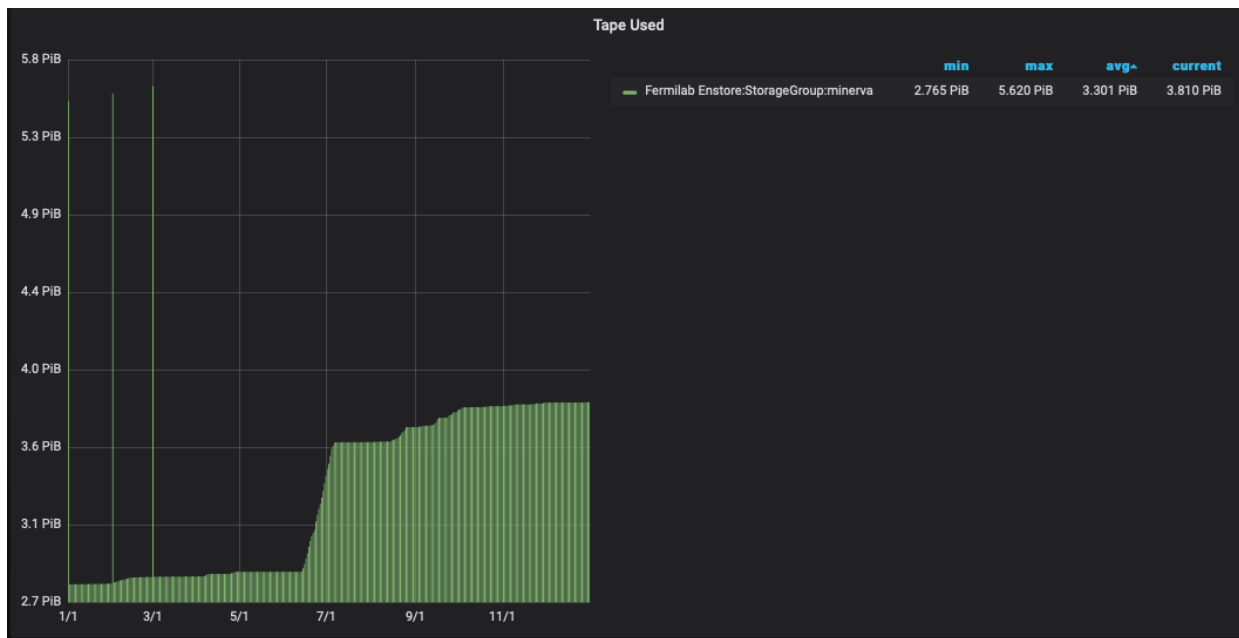<Minerva>

# Disk: dCache Usage and Predictions (in TB)
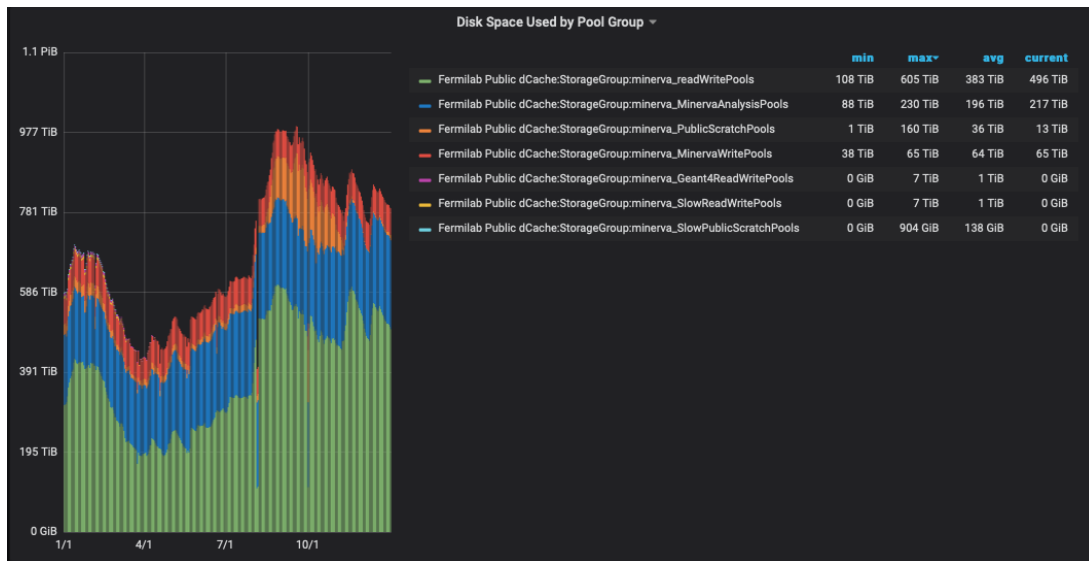


Total r/w (tape backed): 6264 TB
Total scratch: 2333 TB
Total persistent: 2576 TB
Total other: 2131 TB

<Minerva>

| | Analysis (Persistent) | Other Dedicated (Write) |
|---|---|---|
| **Current** | 248 TB (actual) | 126 TB (actual) |
| **2020** | 250 TB | 125 TB |
| **2021** | 250 TB | 125 TB |
| **2022** | 250 TB | 125 TB |

# Tape - Usage and Predictions (in PB)



Tape Used

Fermilab Enstore:StorageGroup:minerva — min 2.765 PiB, max 5.620 PiB, avg 3.301 PiB, current 3.810 PiB

<Minerva>

| | Total Added By End of Year |
|---|---|
| **At end 2019** | +1190 TB (actual) |
| **2020** | +400TB |
| **2021** | +400TB |
| **2022** | +400TB |

# Disk: NAS Usage and Predictions (in TB Units)
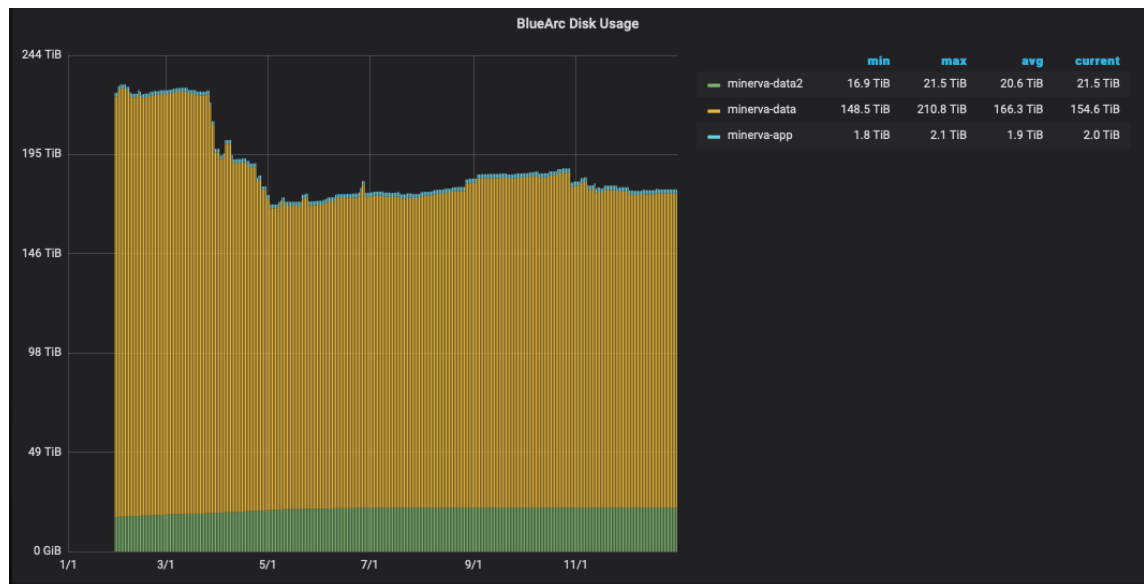


| | App | Data |
|---|---|---|
| **2020** | **2 TB** | **300 TB** |
| **2021** | **2 TB** | **300 TB** |
| **2022** | **2 TB** | **300 TB** |

Reallocated unused minerva-data space to minerva-app in 2020 to make room for new analyzers

<Minerva>

# Age of files in NAS



Total Usage by Age Bucket (use filter above or in table to limit)
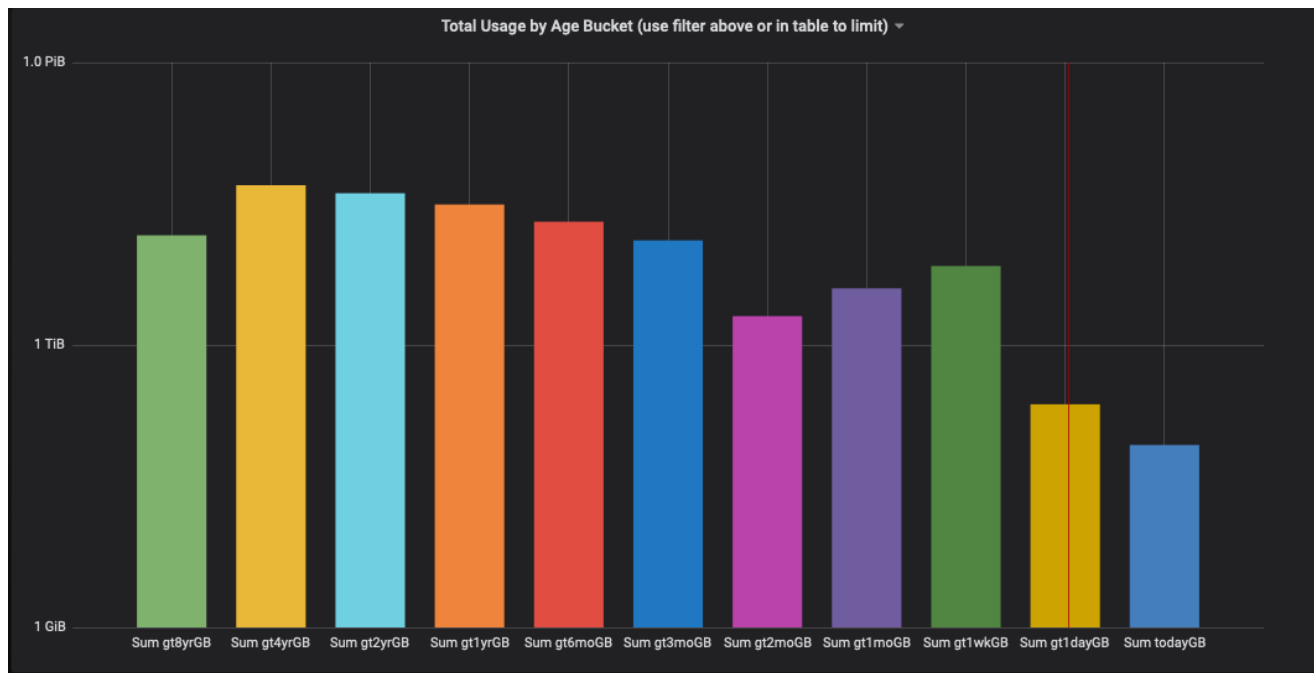
# Data Lifetimes

Data preservation project will drive long-term storage usage.  No estimate yet of final AnaTuple size.

Can archive production after data preservation AnaTuples complete

# What Do You Want to Achieve in Computing Over Next Three Years

| Goals | Where does the experiment need to contribute | Where does SCD need to contribute |
|---|---|---|
| Migrate to SL 7 for Production and Analysis | Testing on SL7 node Monitor issues as they evolve | SCD already provided SL 7 node |
| ML Vertexing Production for Antineutrino Dataset | New production team needs to train an expert | Access to Wilson cluster |
| Deliver Data Preservation Product | Porting existing analyses to data preservation format | Best practices for maintaining a github repository |

<Minerva>

# Anything else?

- Migrating PlotUtils to github

    - Nearly all MINERvA software uses CVS now

    - PlotUtils = core of post-production analysis

    - Part of data preservation product

    - Coordinating with Office of Technology Transfer about publication and license requirements

<Minerva>

## Facility Costs (K$) for MINERVA ("Amazon equivalent" service)

|      | CPU | Tape | dCache Disk | NAS Storage |
|------|-----|------|-------------|-------------|
| 2020 | 106 | 34   | 31          | 30          |
| 2021 | 91  | 37   | 31          | 30          |
| 2022 | 78  | 54   | 31          | 30          |