# Imbalance Factor with Load

- Flaws of CoV
  1. Dispersion only
  2. Unfixed range

- Target:
  1. Represent Imbalance
  2. A percentage
  3. Consider the urgency

- Component:
  1. Imbalance of MDSs load
  2. Urgency of imbalance

# IF: Imbalance

Target1: Represent Imbalance

Intuitive Solution: CoV

# IF: Imbalance

Target1: Represent Imbalance

Intuitive Solution: CoV

Target2: Normalization CoV

Improved solution: Divided by the worst value($\sqrt{n}$)

Imbalance of MDSs load: $\frac{\sqrt{\sum_{i=1}^{n}(l_i - \bar{l})^2/(n-1)}}{\sqrt{n} \cdot \sum_{i=1}^{n} l_i/\mathrm{n}}$

# IF: Urgency

Target:

1. Represent urgency of heaviest MDS
2. Close to 0 at low load
3. Close to 1 at high load

Solution:

1. Represent urgency of heaviest MDS
2. Close to 0 at low load
3. Close to 1 at high load

# IF: Urgency

Intuitive Solution:

- Get a intuitive urgency(u):$\frac{\underset{\forall \vec{l}}{MAX}\, l}{Capacity}$
- Capacity
  - A preset value
  - Updating while runtime

# IF: Urgency

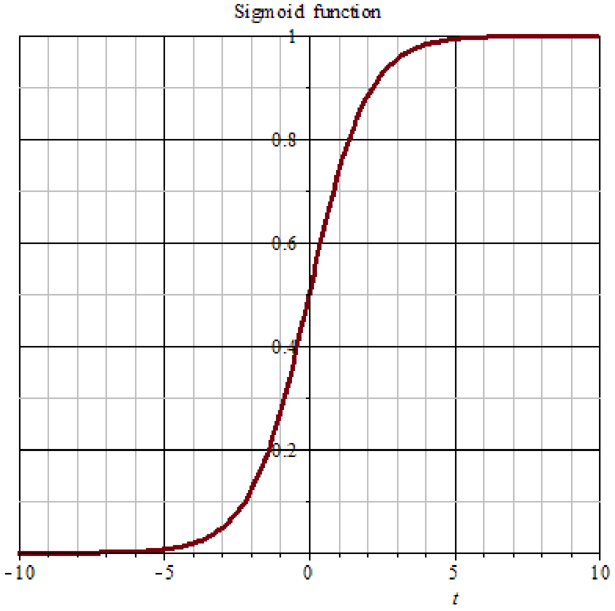
Intuitive Solution:

- Get a intuitive urgency(u): $\frac{\underset{\forall \vec{l}}{MAX_{l}}}{Capacity}$
- Capacity
  - A preset value
  - Updating while runtime

Improved solution:

- Introduce transfer function

U=Sigmoid(u)

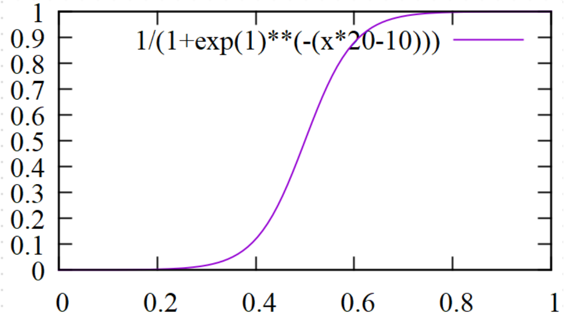


# Formulation

- Coefficient of Variation--CoV: $\frac{\sqrt{\sum_{i=1}^{n}(l_i - \bar{l})^2/(n-1)}}{\sum_{i=1}^{n} l_i/\mathrm{n}}$

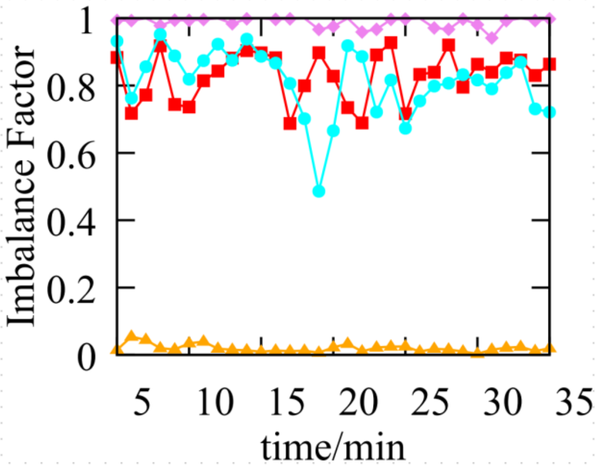
- Urgency--u: $\frac{\underset{\forall \vec{l}}{MAX}\, l}{Capacity}$

- Transfer function: Sigmoid function

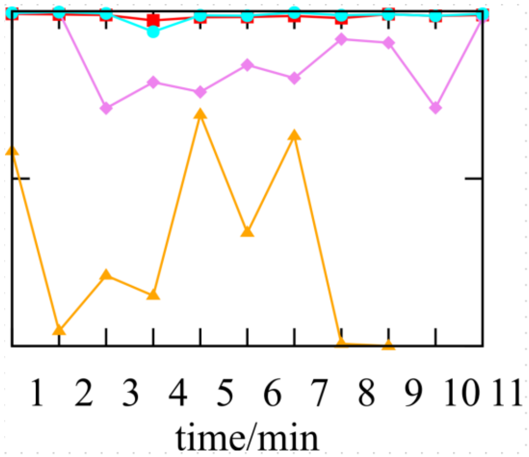


- IF: $\frac{CoV}{\sqrt{n}} \cdot \frac{1}{1+e^{10-20\mathrm{u}}} \cdot 100\%$

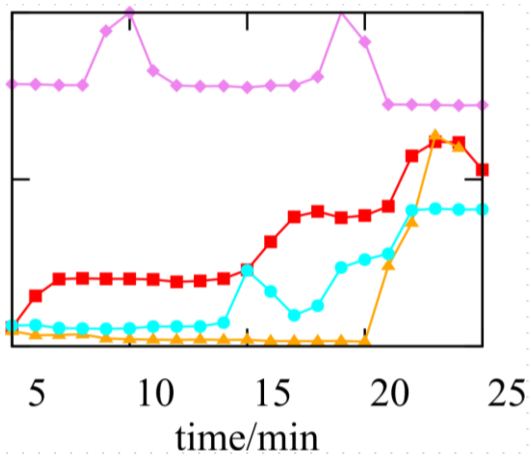
# New IF result



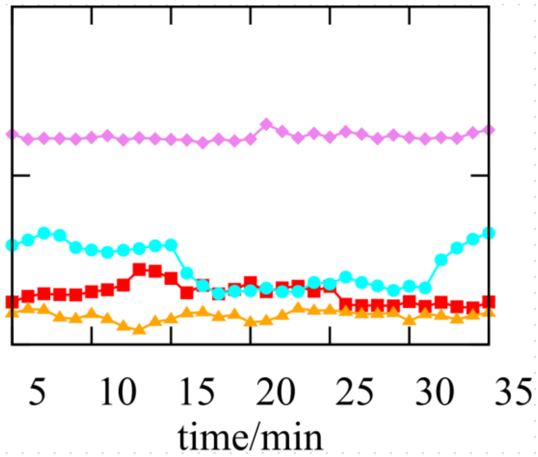
AI pre-training



Tar



Zipfian



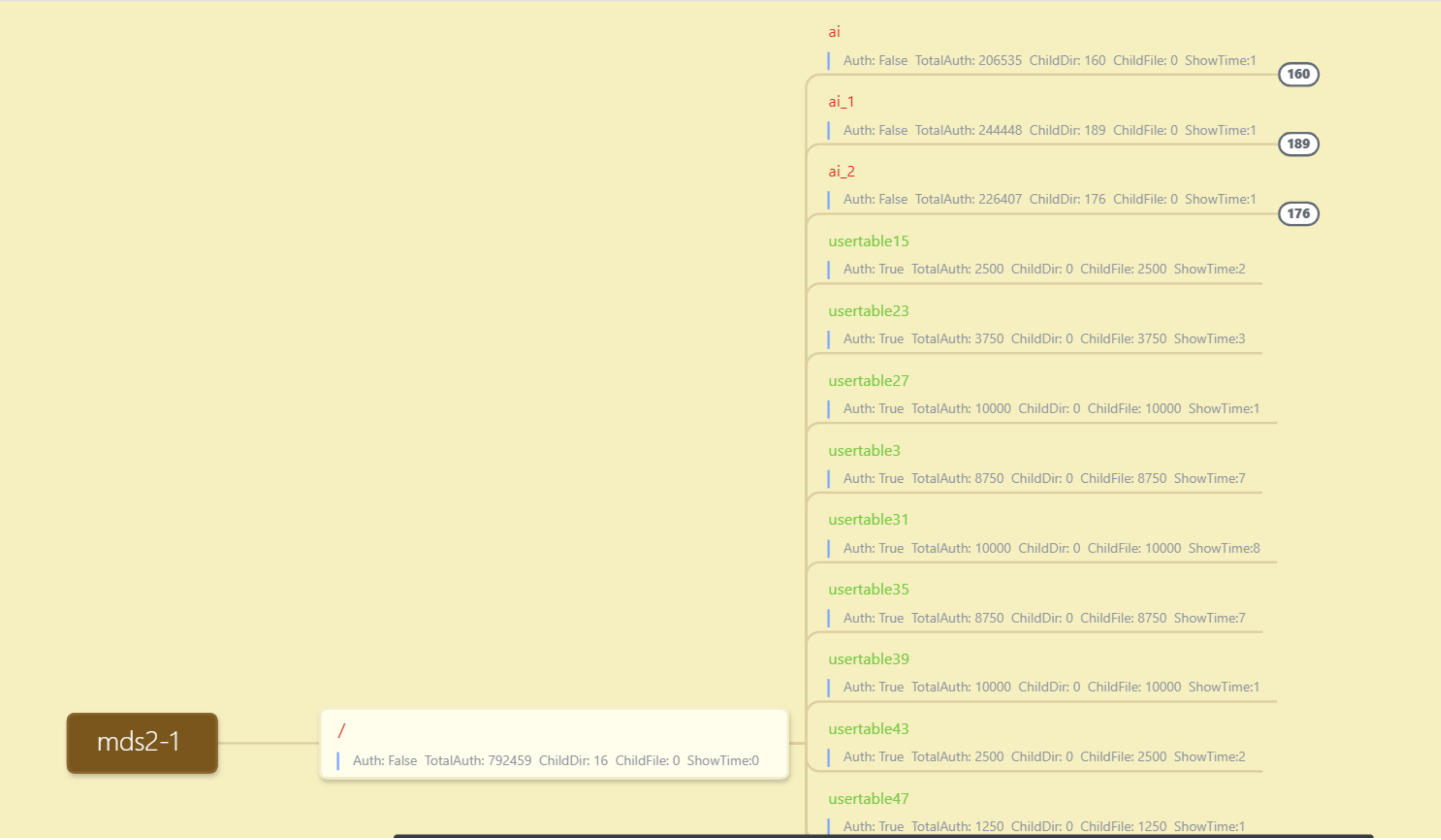
Web Access

# Construct Namespace

- Difficulties
  1. Positioning from multilayer
  2. Implicit directories
  3. Split directories

- Next Step
  - Visualization?
    - Mds1-1: https://mubu.com/doc/86Af5rma6C#mindmap
    - Mds2-1: https://mubu.com/doc/1ObDKdNFhuC#mindmap
    - Mds3-1: https://mubu.com/doc/3USfsiVACSC#mindmap
    - Mds4-1: https://mubu.com/doc/4WZSqggrp6C#mindmap
    - Mds5-1: https://mubu.com/doc/5dPi4HQ7GmC#mindmap

# Construct Namespace



mds2-1

/
Auth: False  TotalAuth: 792459  ChildDir: 16  ChildFile: 0  ShowTime:0

ai
Auth: False  TotalAuth: 206535  ChildDir: 160  ChildFile: 0  ShowTime:1

ai_1
Auth: False  TotalAuth: 244448  ChildDir: 189  ChildFile: 0  ShowTime:1

ai_2
Auth: False  TotalAuth: 226407  ChildDir: 176  ChildFile: 0  ShowTime:1

usertable15
Auth: True  TotalAuth: 2500  ChildDir: 0  ChildFile: 2500  ShowTime:2

usertable23
Auth: True  TotalAuth: 3750  ChildDir: 0  ChildFile: 3750  ShowTime:3

usertable27
Auth: True  TotalAuth: 10000  ChildDir: 0  ChildFile: 10000  ShowTime:1

usertable3
Auth: True  TotalAuth: 8750  ChildDir: 0  ChildFile: 8750  ShowTime:7

usertable31
Auth: True  TotalAuth: 10000  ChildDir: 0  ChildFile: 10000  ShowTime:8

usertable35
Auth: True  TotalAuth: 8750  ChildDir: 0  ChildFile: 8750  ShowTime:7

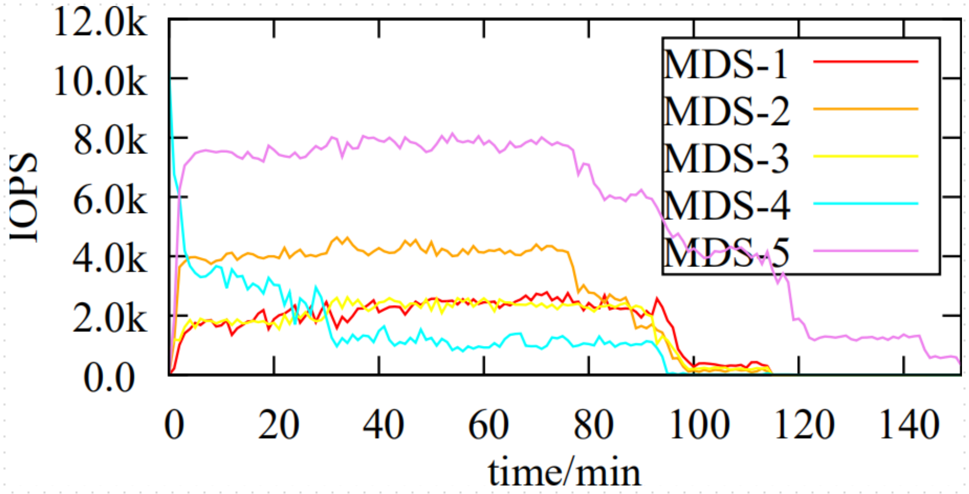
usertable39
Auth: True  TotalAuth: 10000  ChildDir: 0  ChildFile: 10000  ShowTime:1

usertable43
Auth: True  TotalAuth: 2500  ChildDir: 0  ChildFile: 2500  ShowTime:2

usertable47
Auth: True  TotalAuth: 1250  ChildDir: 0  ChildFile: 1250  ShowTime:1

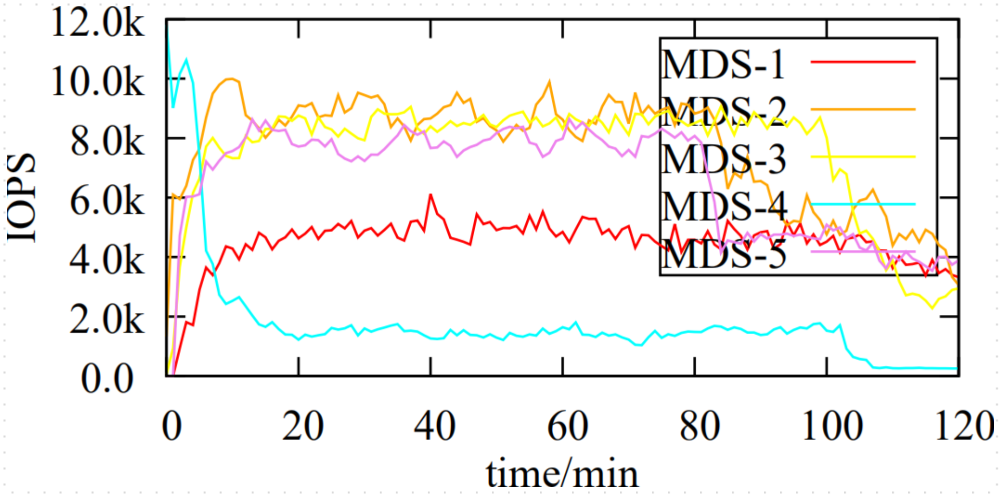
# Revision of last meeting

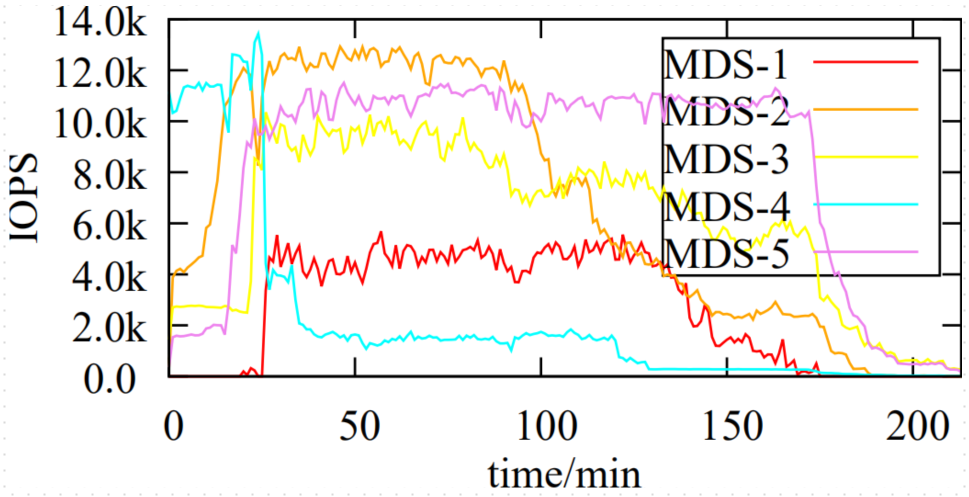
- Impact of workload client size on cluster load



25

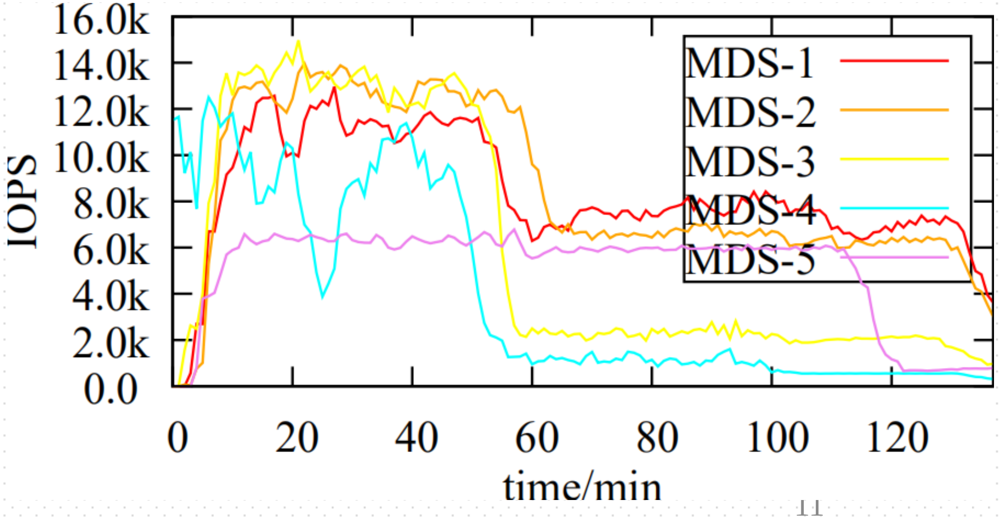
50



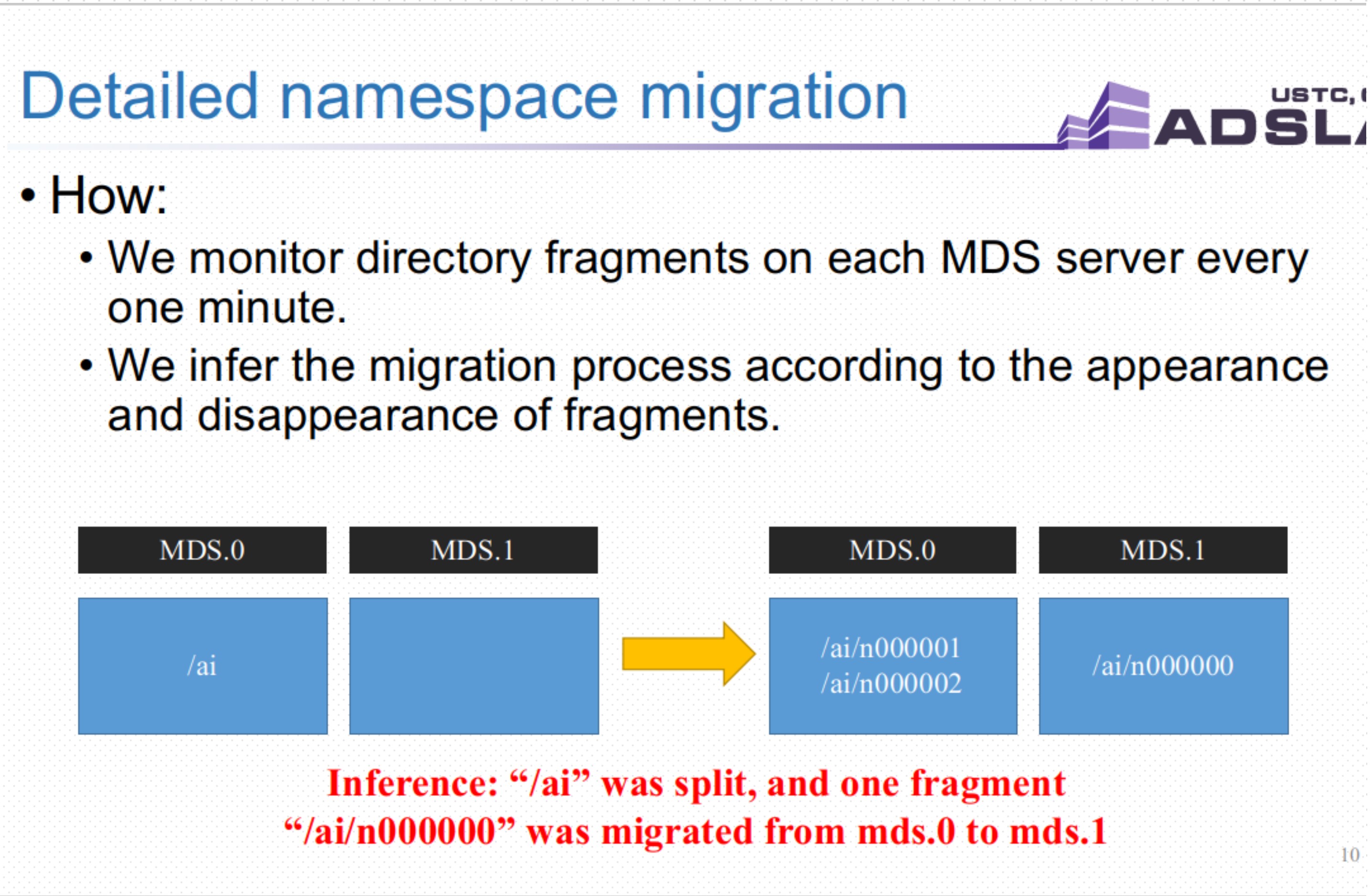


75

100



# Primitive method of monitoring migration



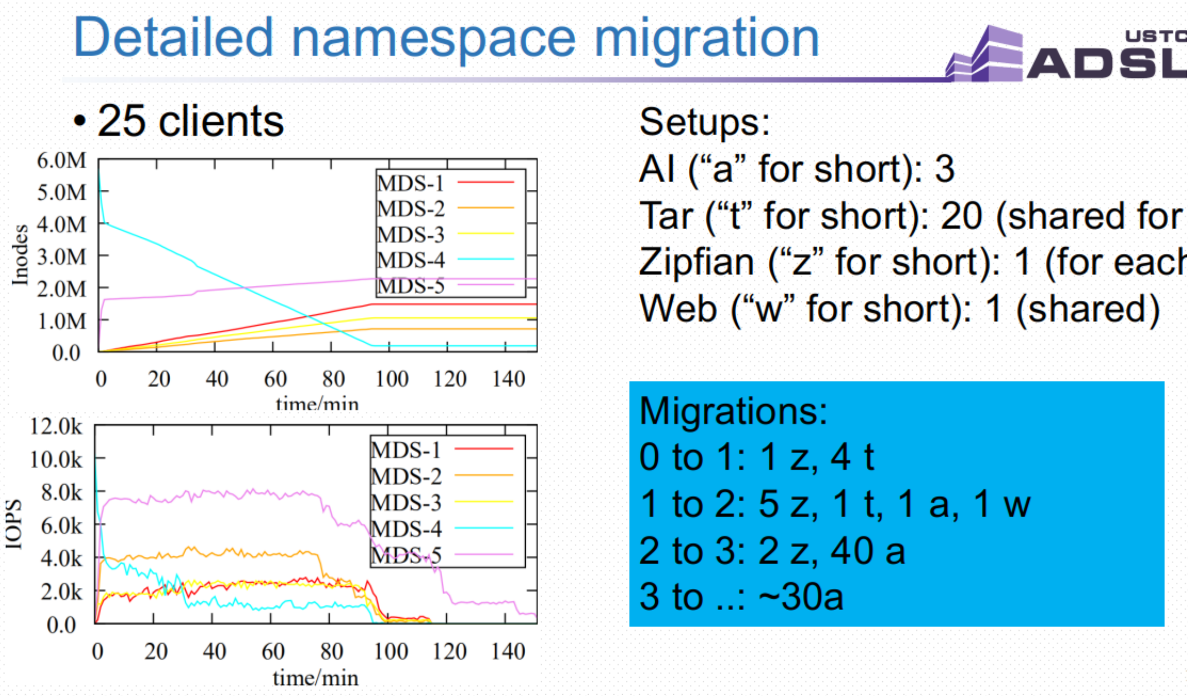
## Detailed namespace migration

- How:
  - We monitor directory fragments on each MDS server every one minute.
  - We infer the migration process according to the appearance and disappearance of fragments.

**Inference: "/ai" was split, and one fragment "/ai/n000000" was migrated from mds.0 to mds.1**

- Discarded
- New method adopted: logging in Ceph source code.

# AI shadows other 3 workloads

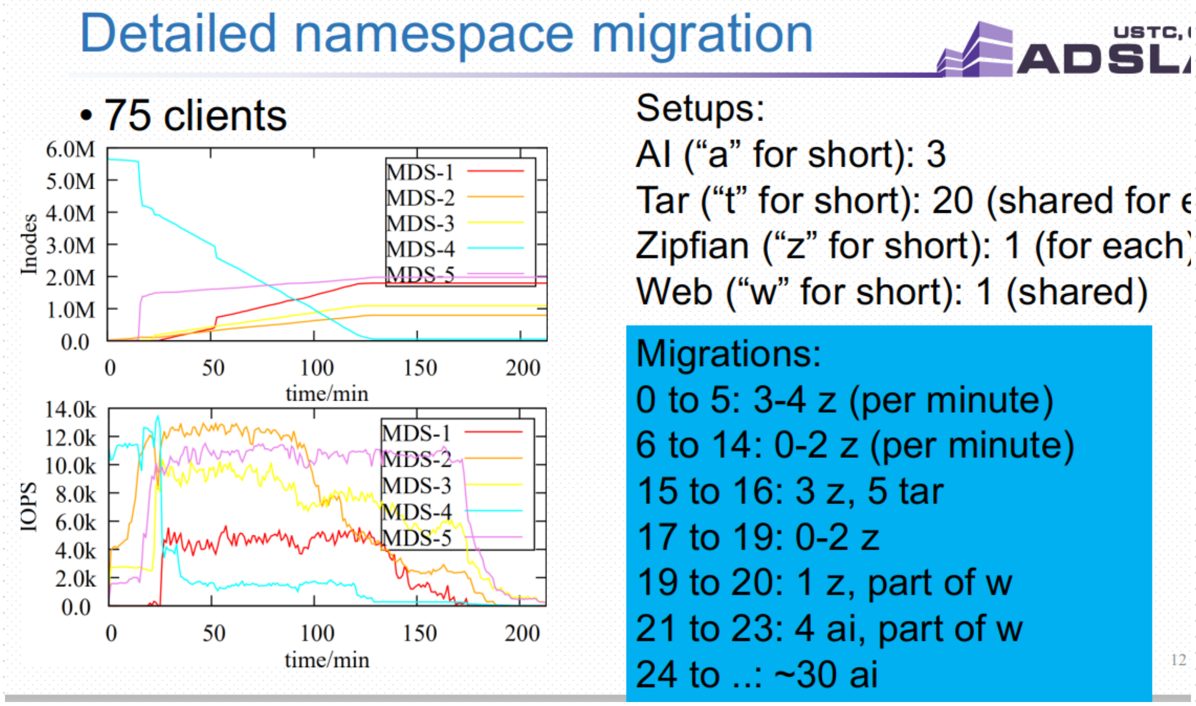


## Detailed namespace migration

- 25 clients

Setups:
AI (“a” for short): 3
Tar (“t” for short): 20 (shared for
Zipfian (“z” for short): 1 (for each
Web (“w” for short): 1 (shared)

Migrations:
0 to 1: 1 z, 4 t
1 to 2: 5 z, 1 t, 1 a, 1 w
2 to 3: 2 z, 40 a
3 to ..: ~30a



## Detailed namespace migration

- 75 clients

Setups:
AI (“a” for short): 3
Tar (“t” for short): 20 (shared for e
Zipfian (“z” for short): 1 (for each)
Web (“w” for short): 1 (shared)

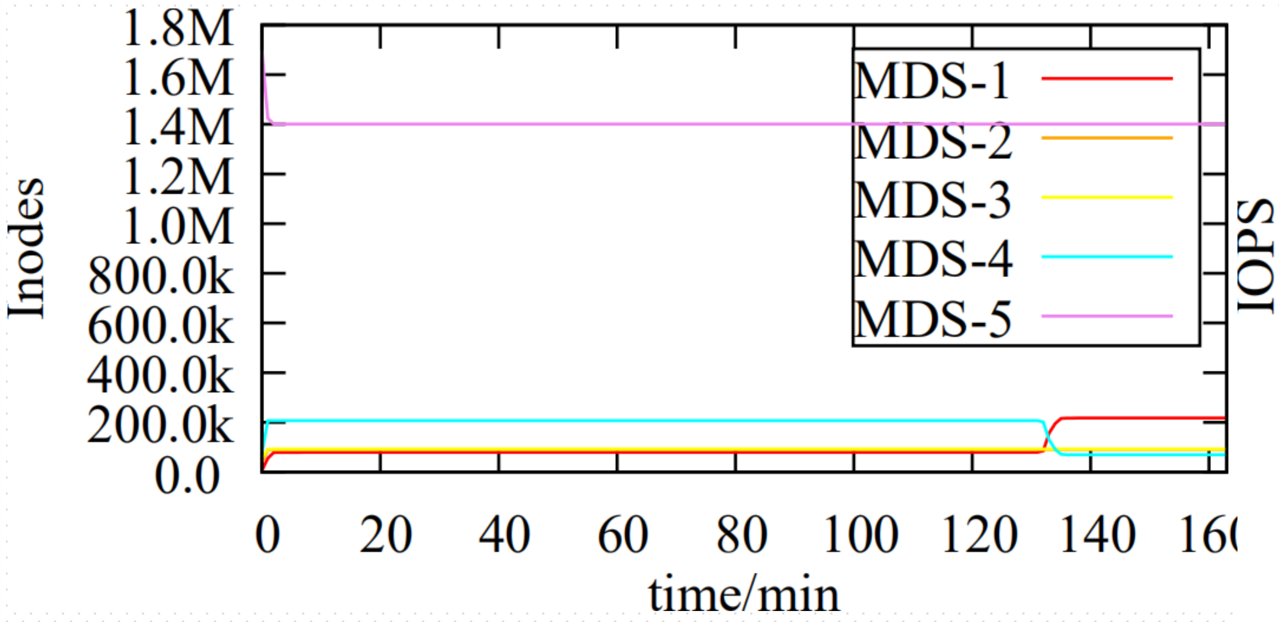
Migrations:
0 to 5: 3-4 z (per minute)
6 to 14: 0-2 z (per minute)
15 to 16: 3 z, 5 tar
17 to 19: 0-2 z
19 to 20: 1 z, part of w
21 to 23: 4 ai, part of w
24 to ..: ~30 ai

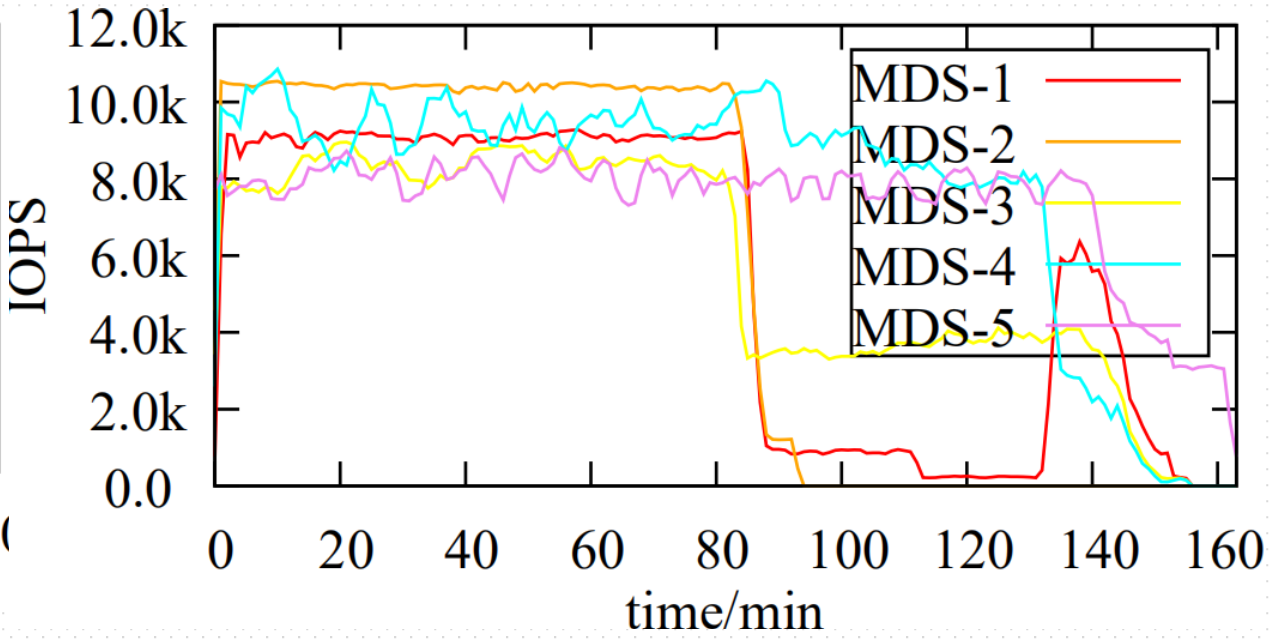
12

Solution: try workloads without AI

# Mixed workload without AI

- Setup:
  - 25 clients per workload (tar, zipfian, web)
  - start at the same time



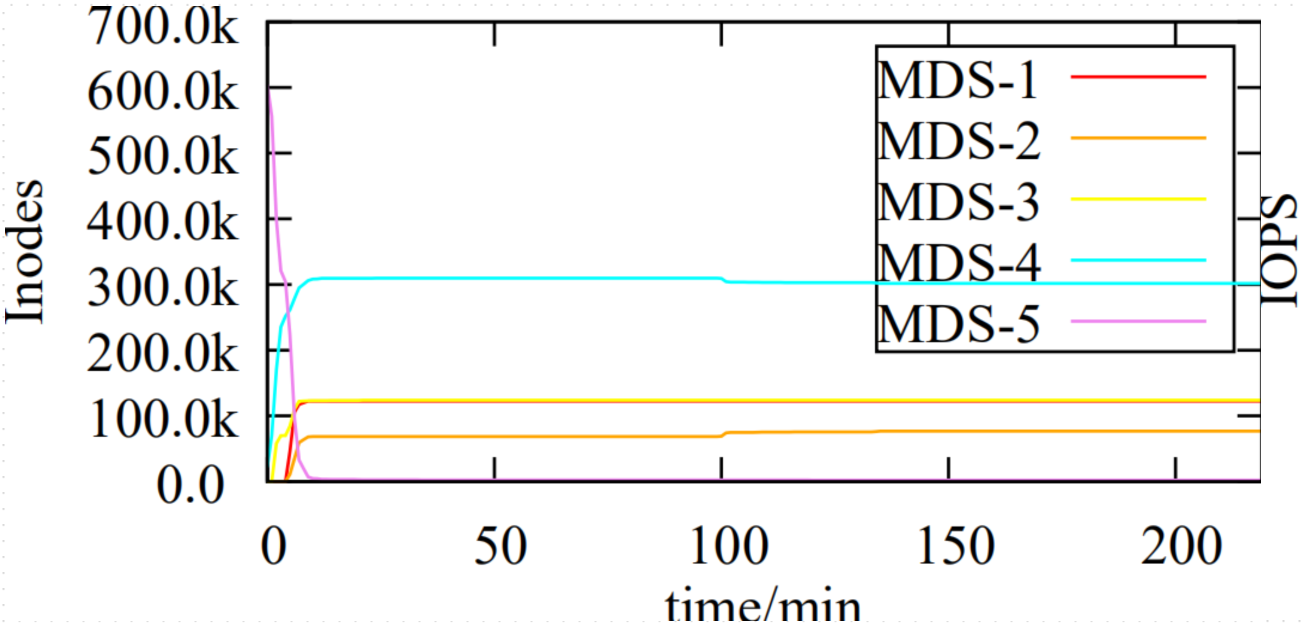


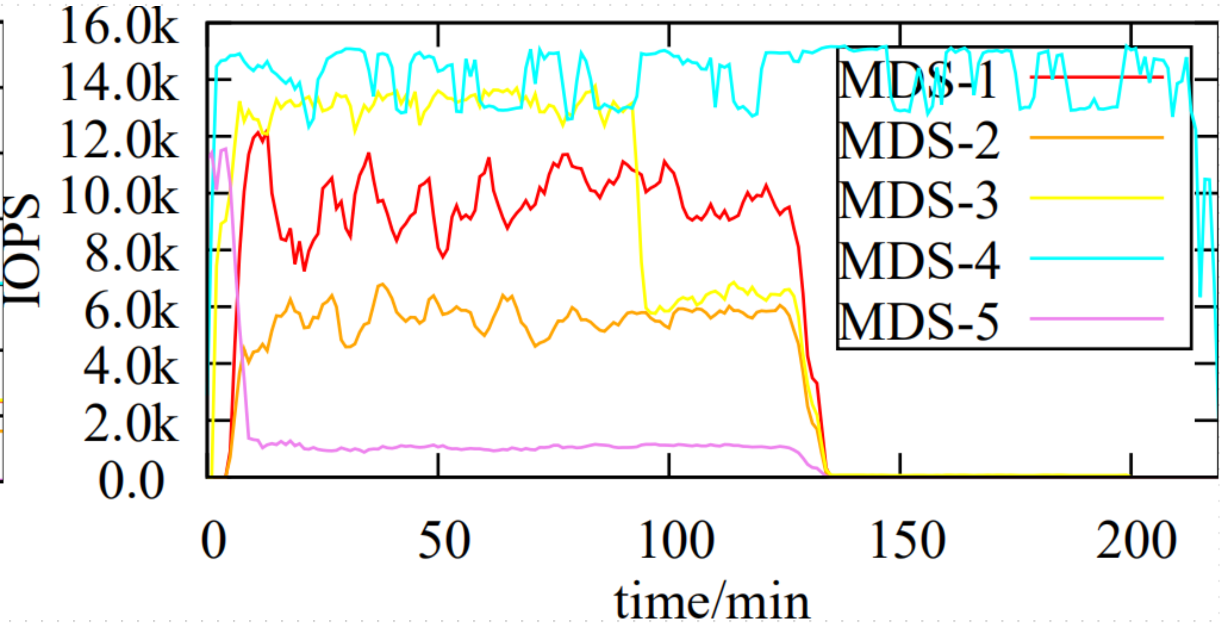
# Mixed workload without AI

- After migration:
  - No tar directories are migrated out from mds-5 (rank 0)
  - Most Zipfian and web directories are shared by other MDSs
  - Zipfian workloads end at 120 minutes.

# Mixed workload: Zipfian + Web

- Setup:
  - 37 clients per workload (zipfian, web)
  - start at the same time





# Mixed workload: Zipfian + Web

- After migration:
  - Nearly nothing on mds-5 (rank 0)
  - Most Zipfian directories are on mds-4 (rank 1).
  - Web files are shared mostly by mds-1 to mds-3 (rank 2-4)

# Animation of Migration with OpenGL