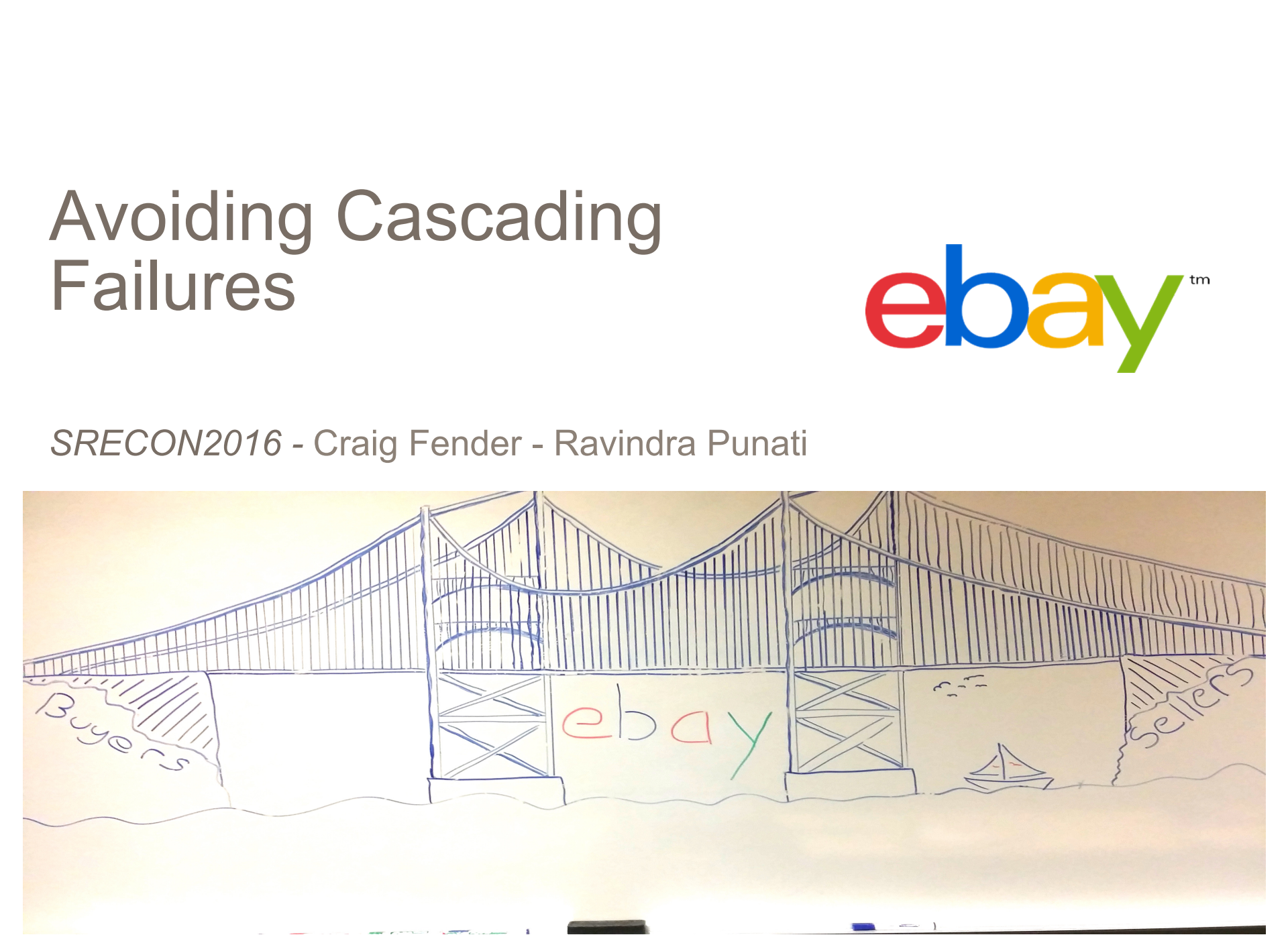
Avoiding Cascading Failures



SRECON2016 - Craig Fender - Ravindra Punati

Avoiding Cascading Failures

Craig Fender

Craig is presently a Senior Technical Duty Officer at eBay and is responsible for commanding all types of large scale site incidents. In addition to an undergraduate degree Craig holds numerous professional certifications related to the computer industry (RHCE, SSCA, ITIL and et cetera). Craig has held several roles at multiple start-up and fortune 500 companies such as Senior Systems Engineer, Project Manager, Presenter and Major Incident Commander.

Ravindra Punati

Ravindra Punati is leader of the Site Reliability Engineering team. In other roles at eBay Ravi has been responsible for the infrastructure automation initiatives and cloud operations. Ravi brings extensive expertise in the fields of database engineering, application development and software as a service product lines. In addition to holding multiple degrees in computer science Ravi has held several roles as an engineer, architect, manager and executive in various silicon valley start-ups.

EBAY AT A GLANCE

.

OUR BUSINESS

EBAY AT A GLANCE

$8.6B
Revenue in 2015

162M
Global Active Buyers *

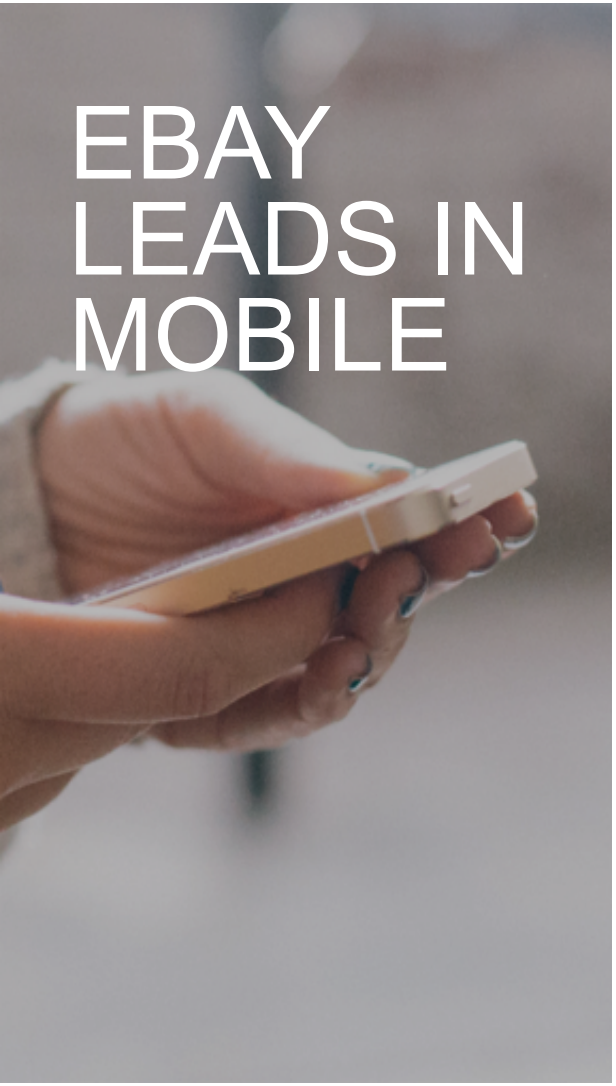
800M
Live Listings *

57%
Percentage of eBay Inc. revenue that is international *

190
Markets eBay apps are available in*

$82B
GMV in 2015

*Q4 2015 data



EBAY LEADS IN MOBILE

304M

eBay Application downloads *

43%

Percentage of GMV closed on Mobile *

$33B

2015 MCV across eBay's portfolio of apps

Every

28 sec.

a tablet is bought via mobile in the U.S. *

Every

10 sec.

a ladies handbag is bought via mobile in the U.S. *

9.2M

new listings are added via mobile every week *

*Q4 2015 data

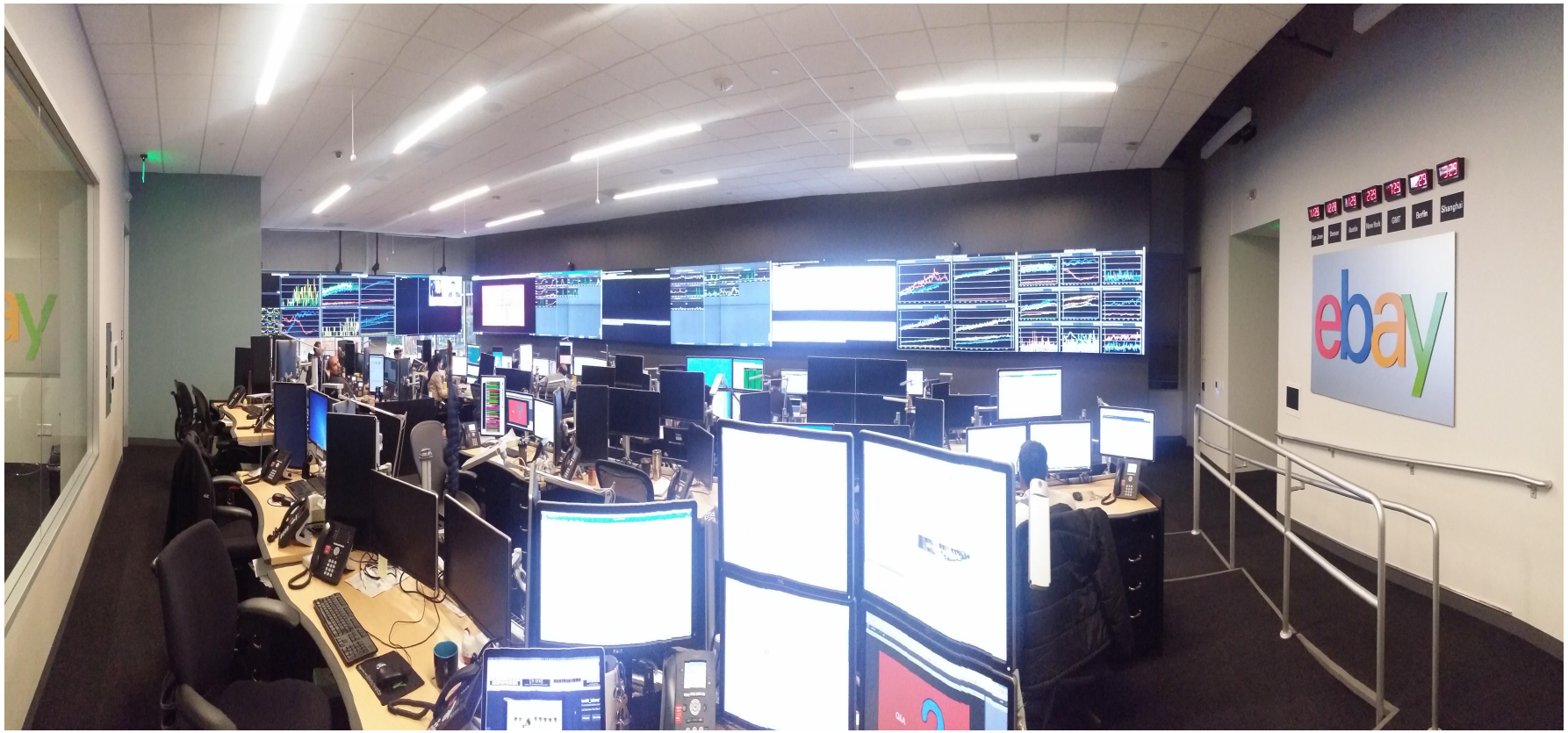
EBAY OPERATIONS

.

ebay

Background on eBay Operations

Background on eBay Operations



Background on eBay Operations

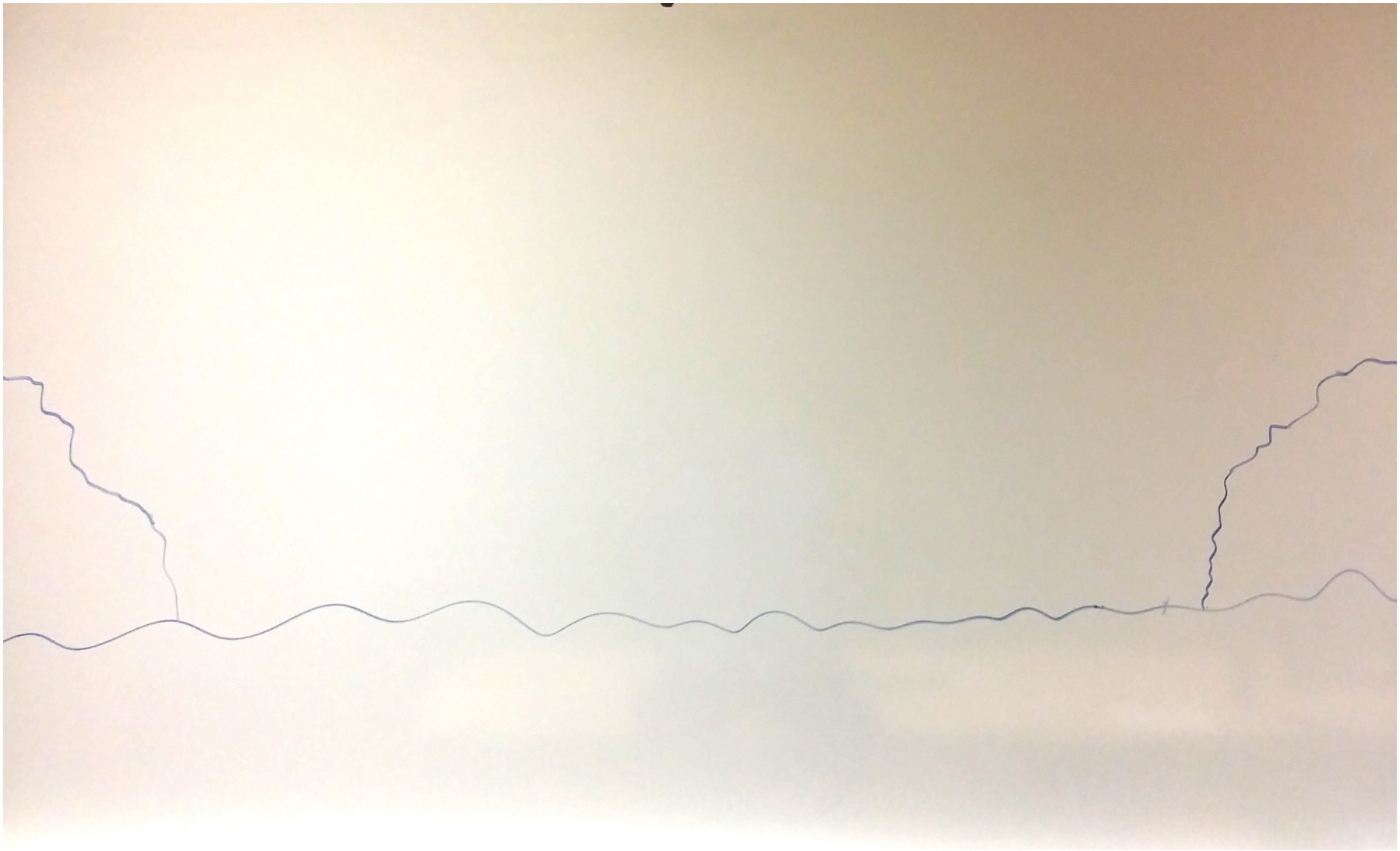
- In any mission critical real-time environment where one is highly incentivized toward maximum uptime one must have:
  - redundancy ,
  - resiliency and
  - robustness.
- These needs give rise to complexity.
- Complexity can lead to fragility.
  - Multilayered dependencies can cause cascading failure.
- To manage that complexity certain behaviors, culture, principles and technology emerge as the most successful.

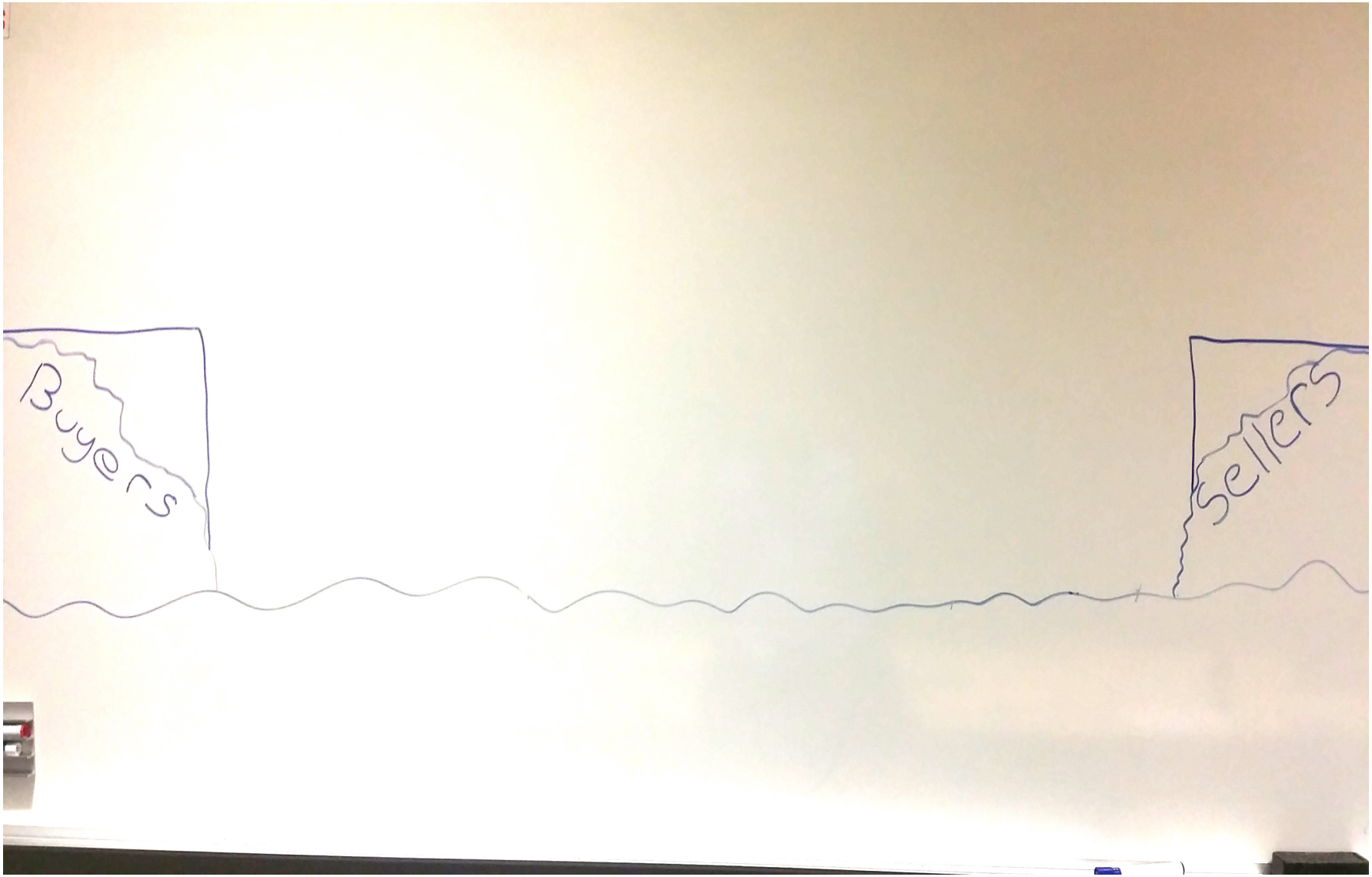
Background on eBay Operations

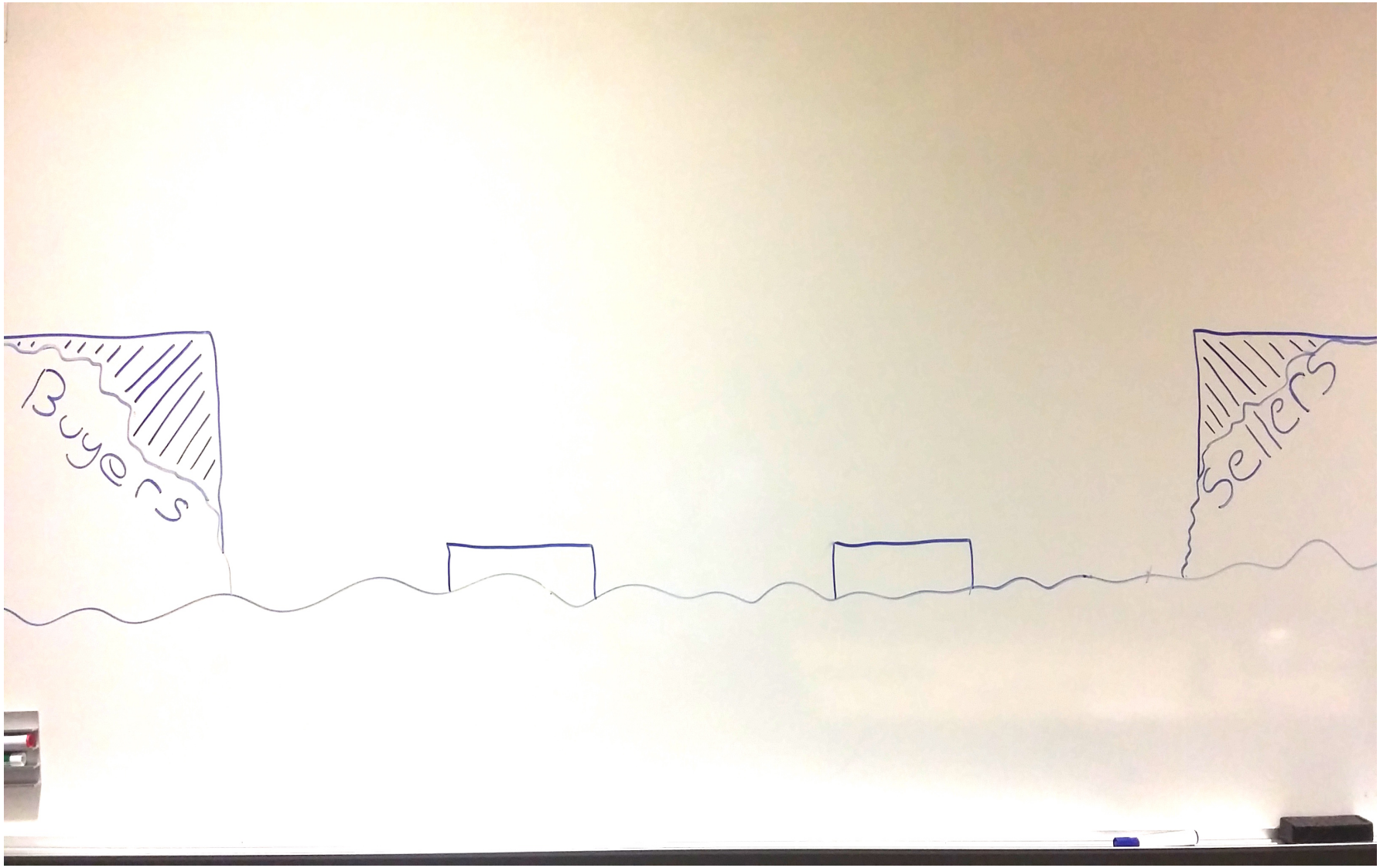
- eBay itself is like a bridge between buyers . . .

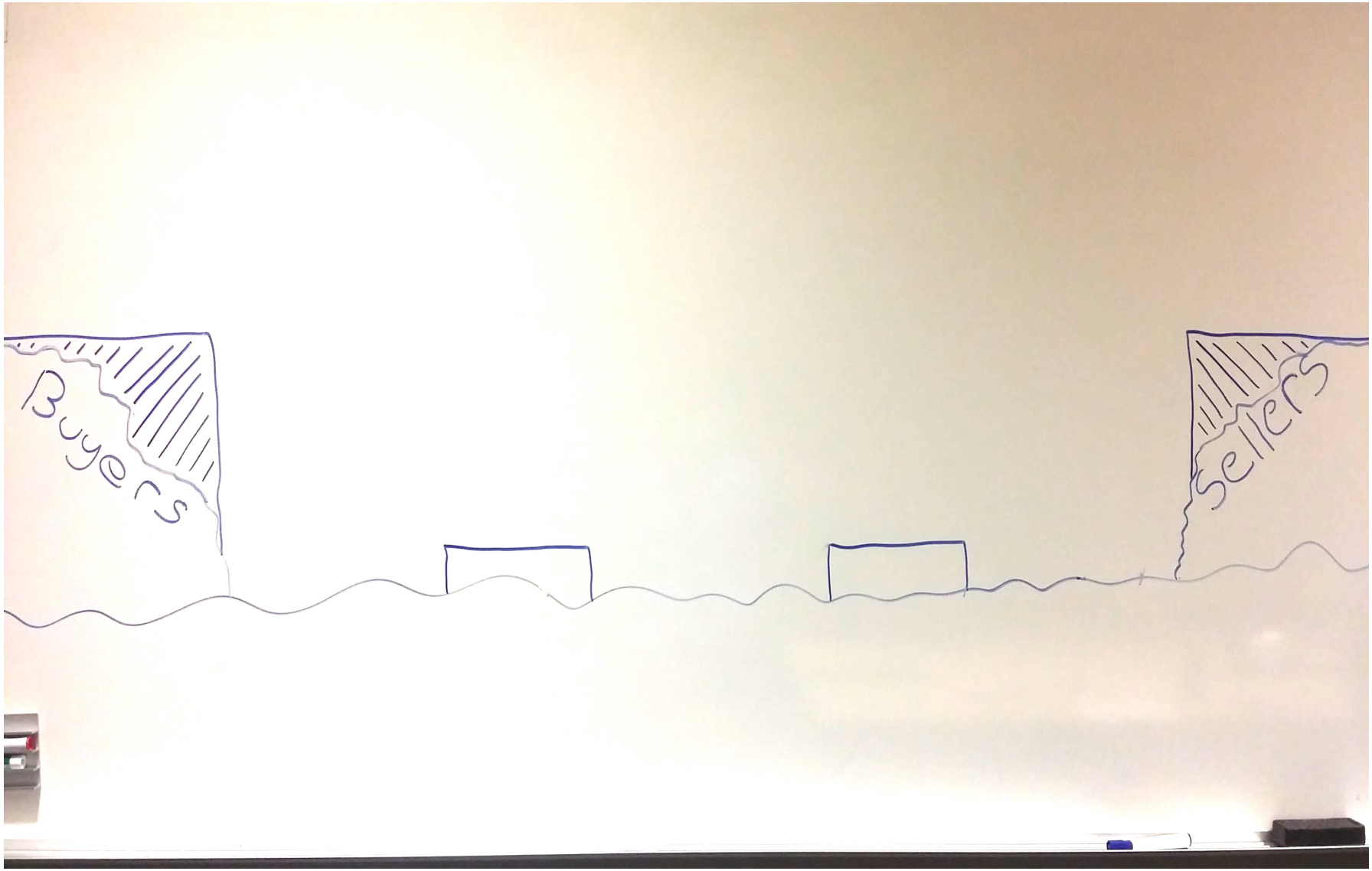
Background on eBay Operations

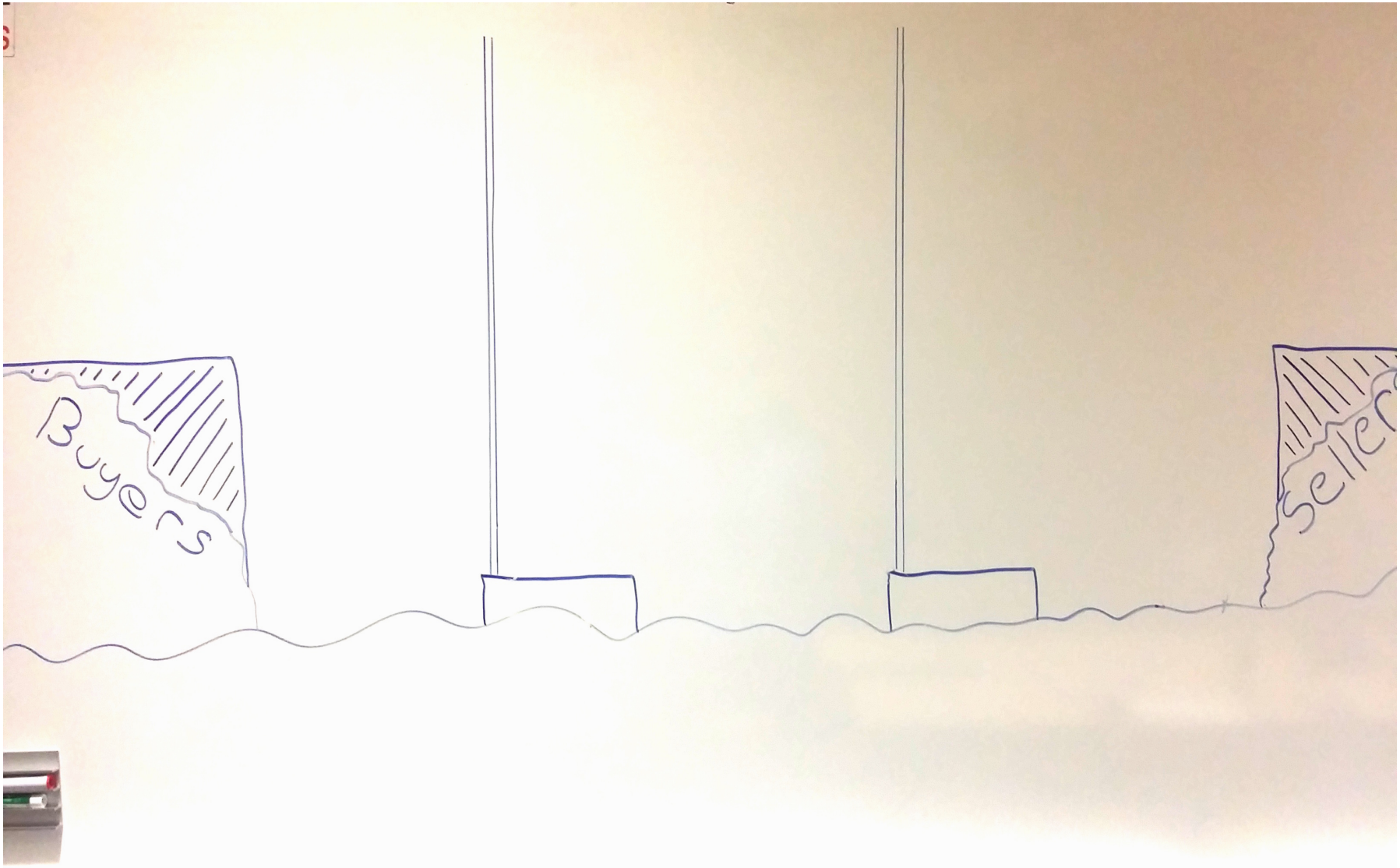
- . . .and sellers

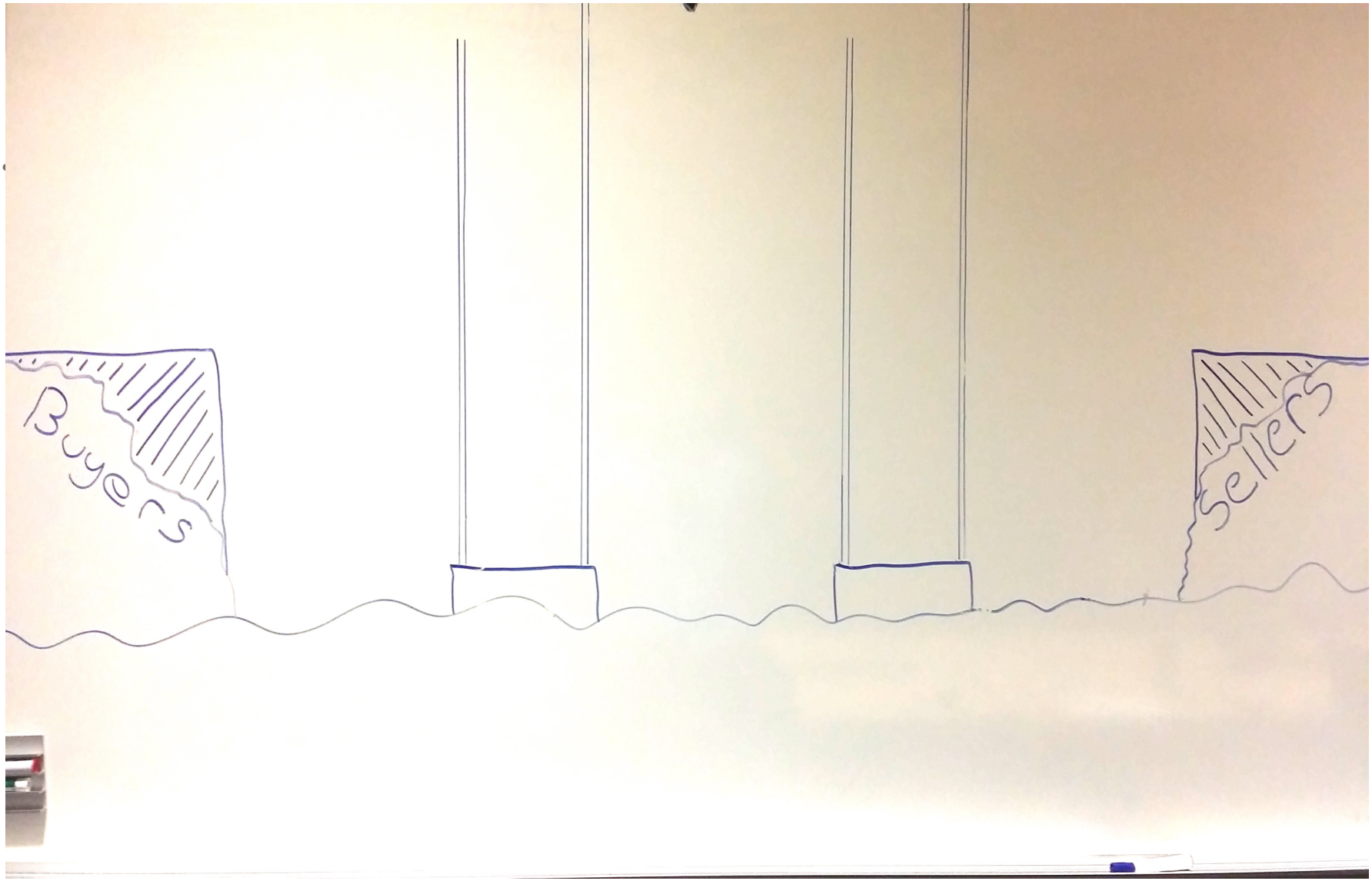


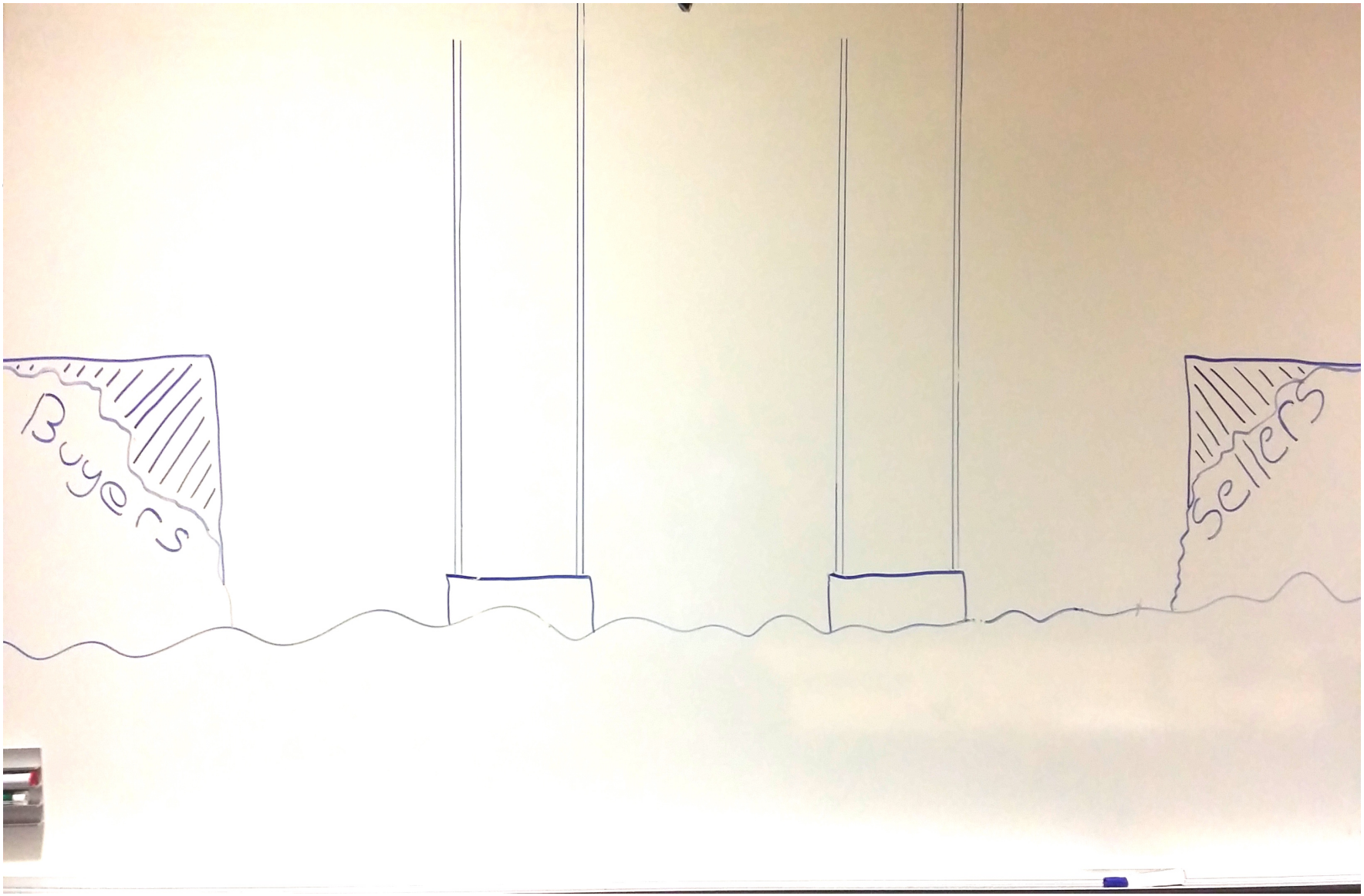

Buyers
Sellers

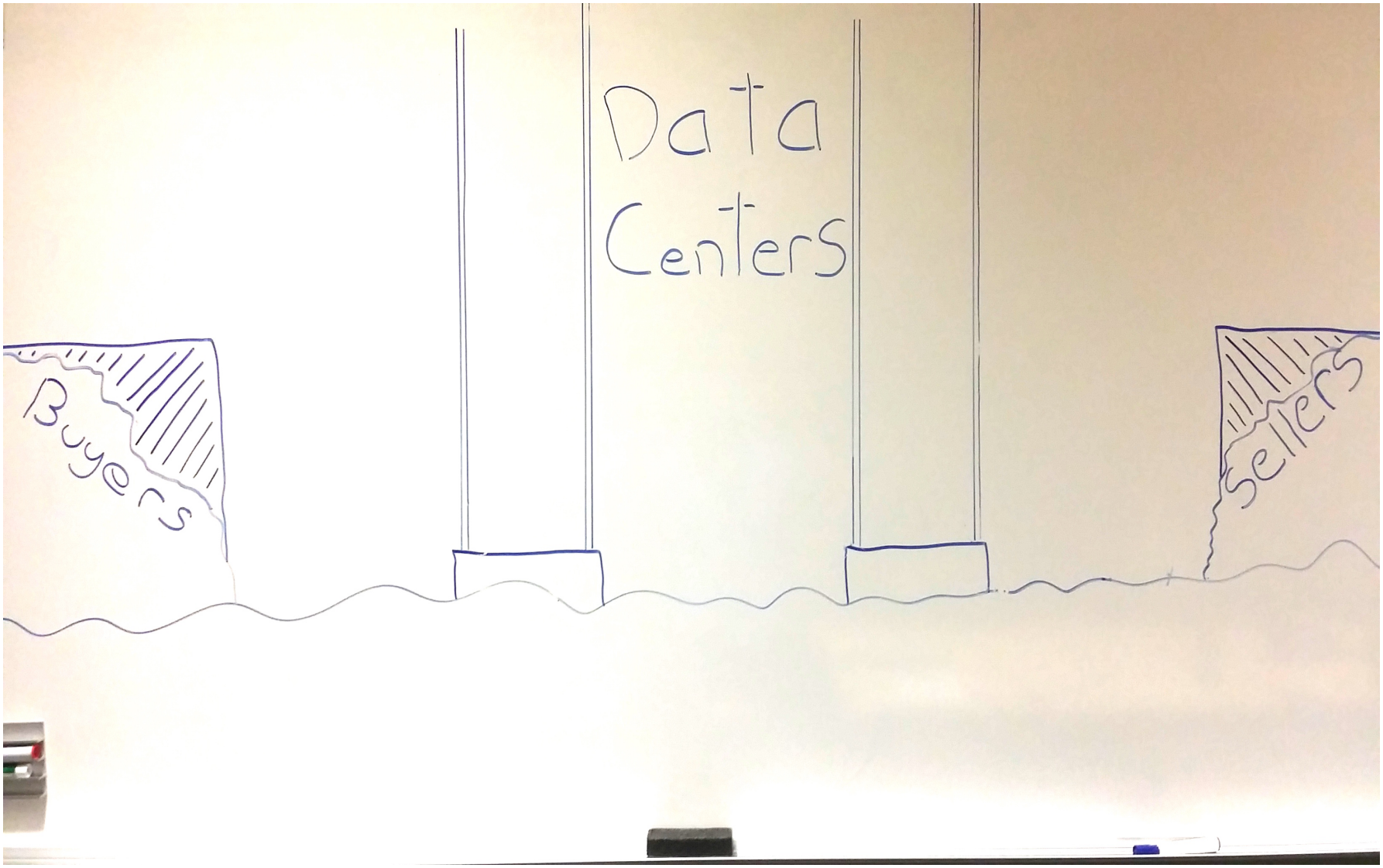

Buyers
Sellers

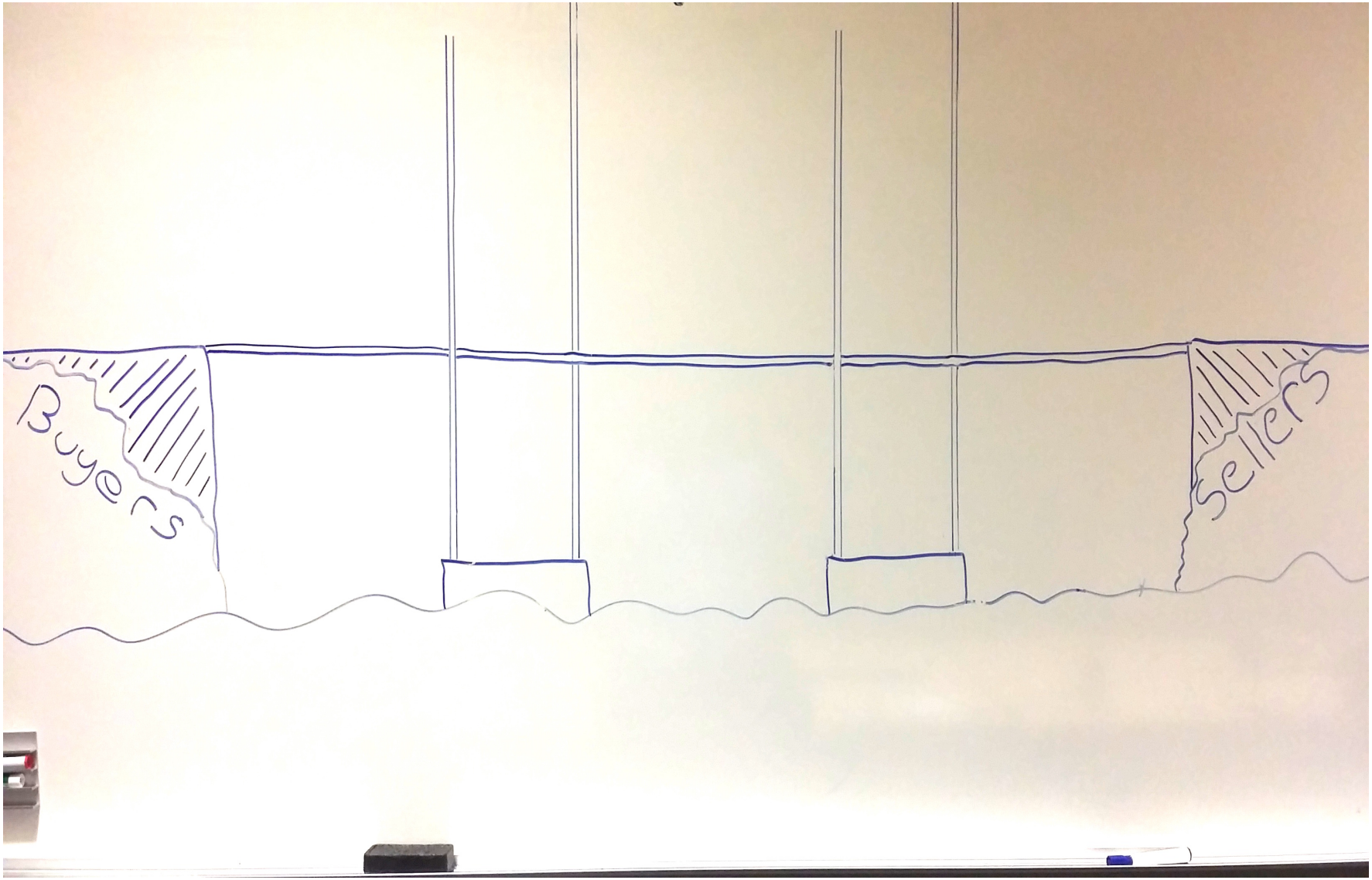

Buyers
Sellers

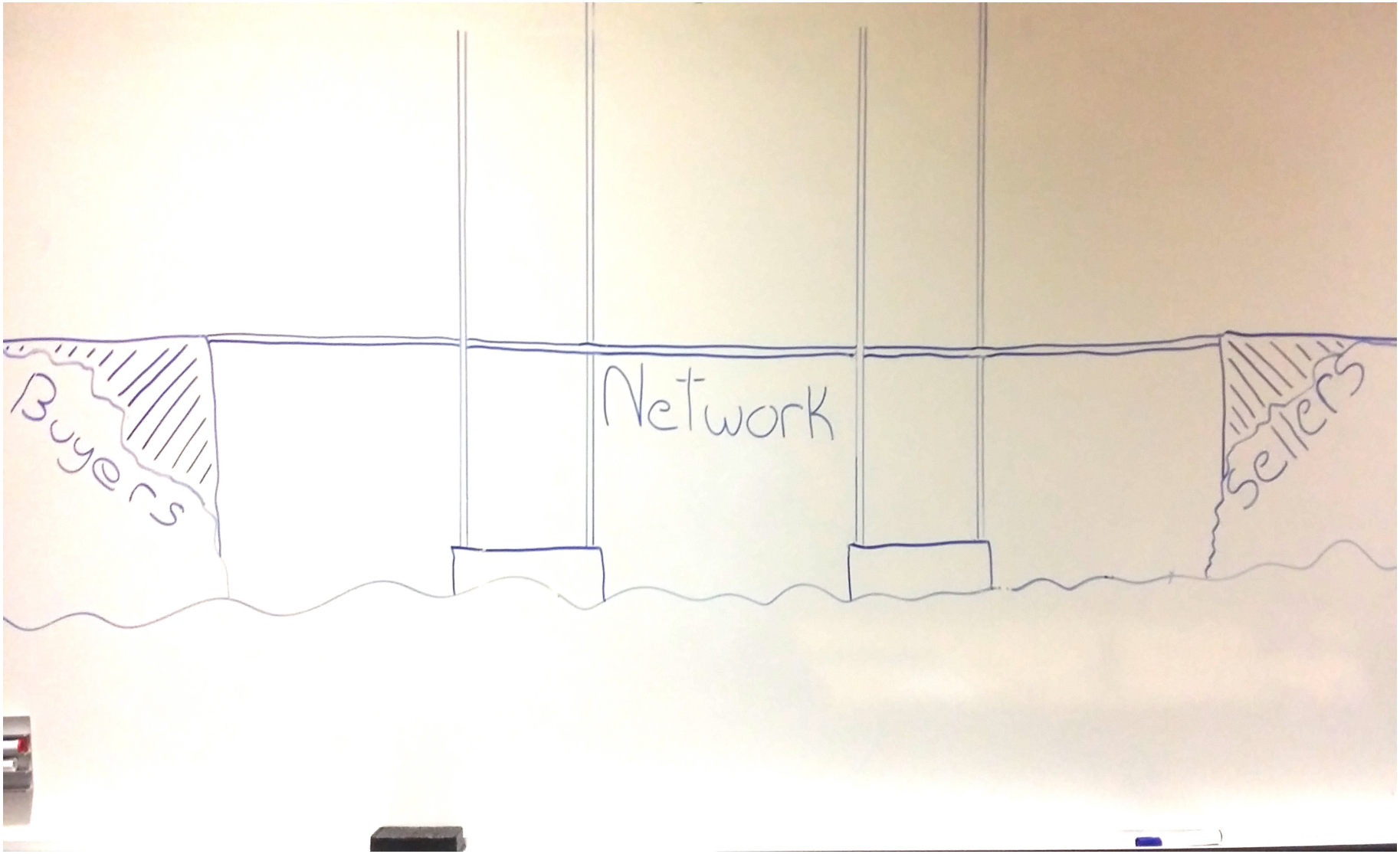

Buyers

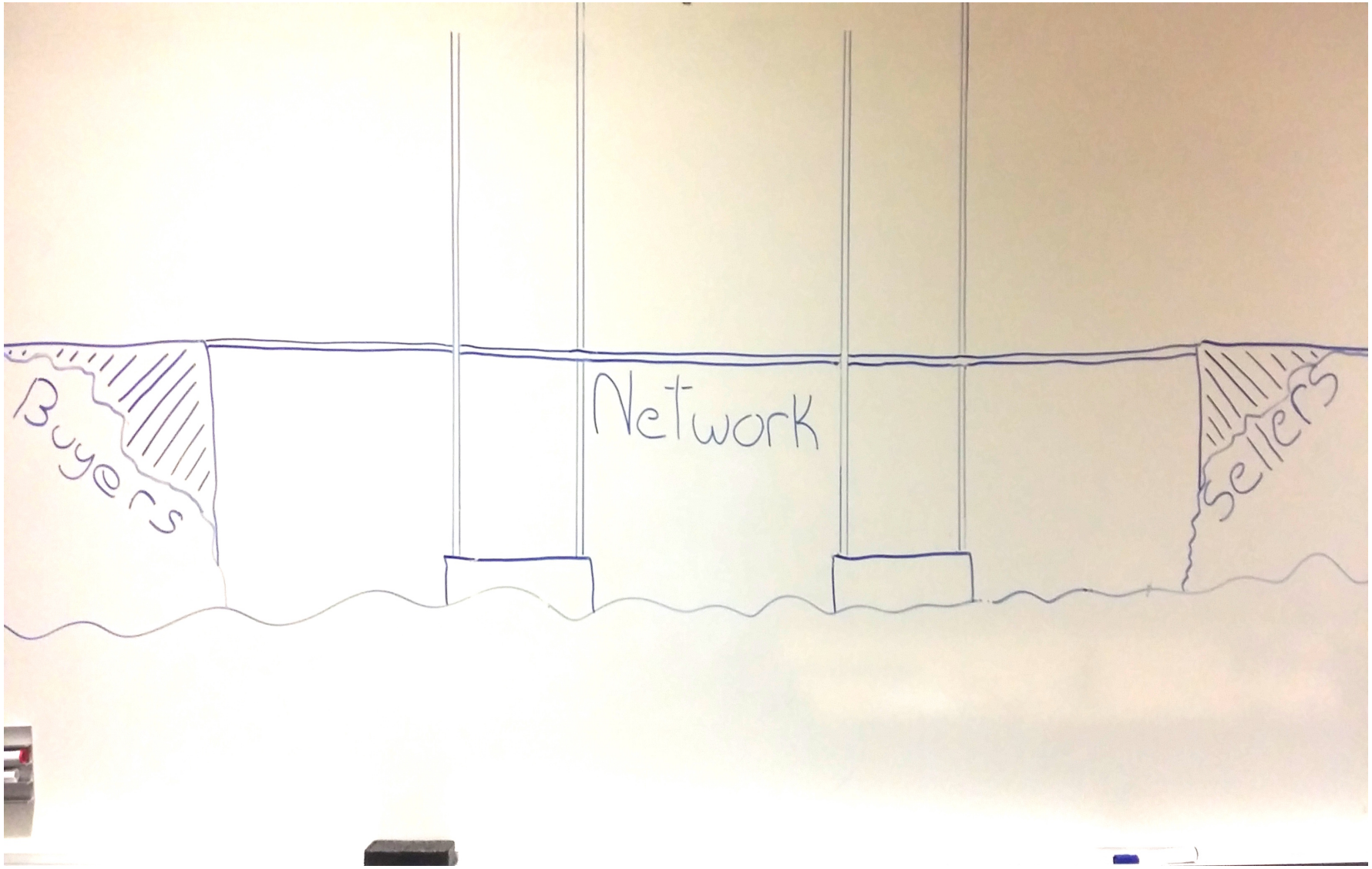

Buyers
Sellers

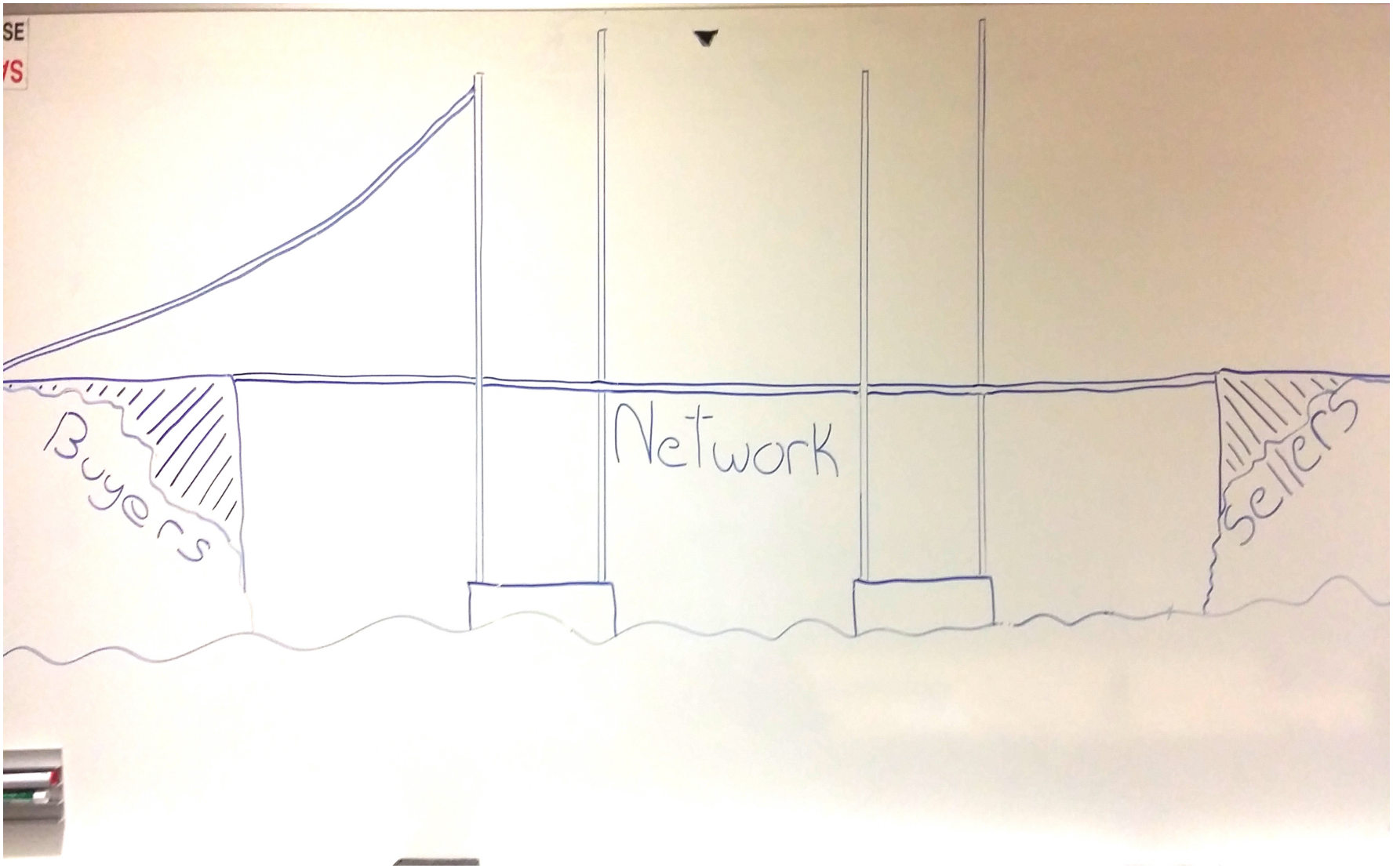

Buyers
Sellers

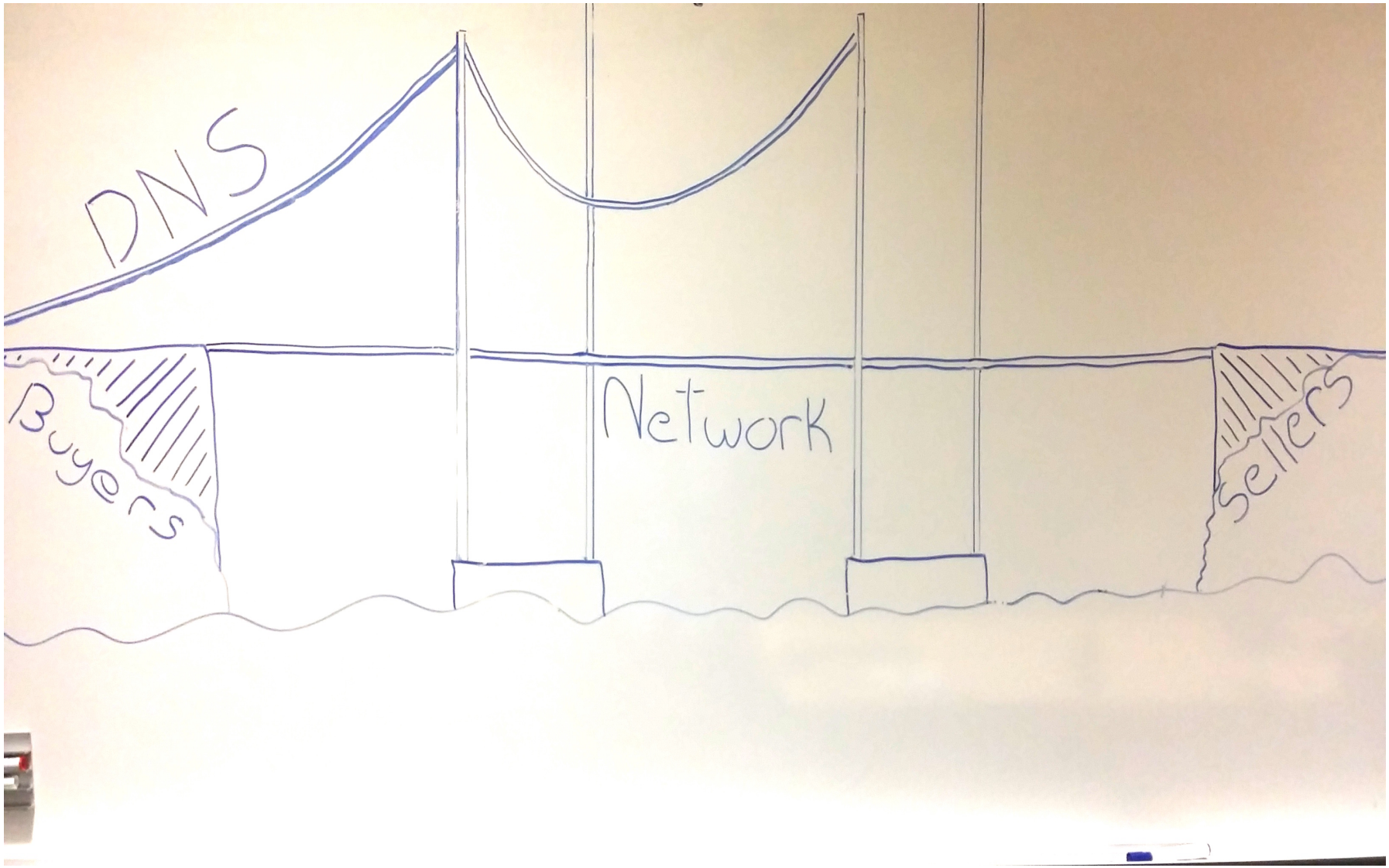

Data
Centers
Buyers
Sellers

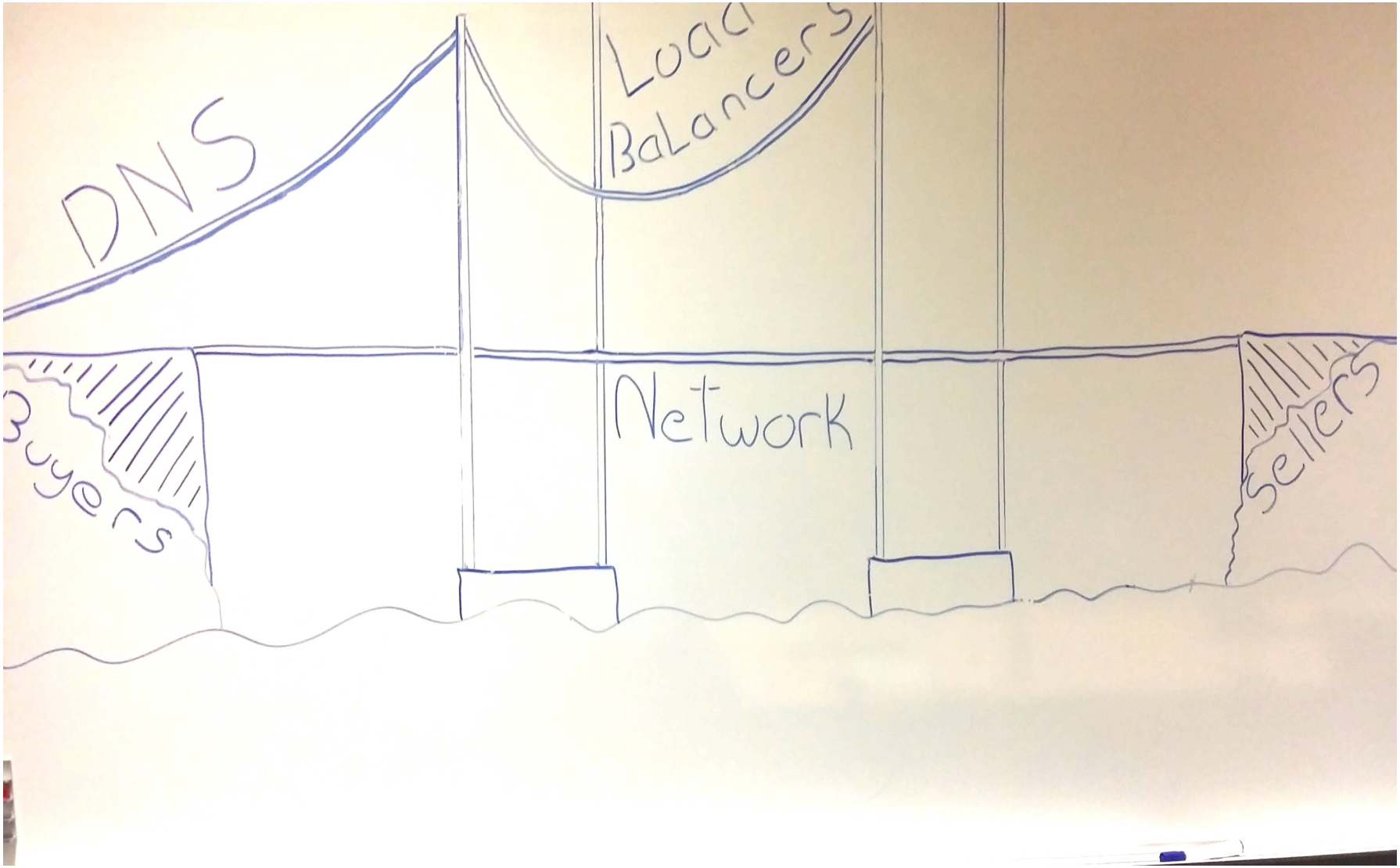

Buyers
Sellers

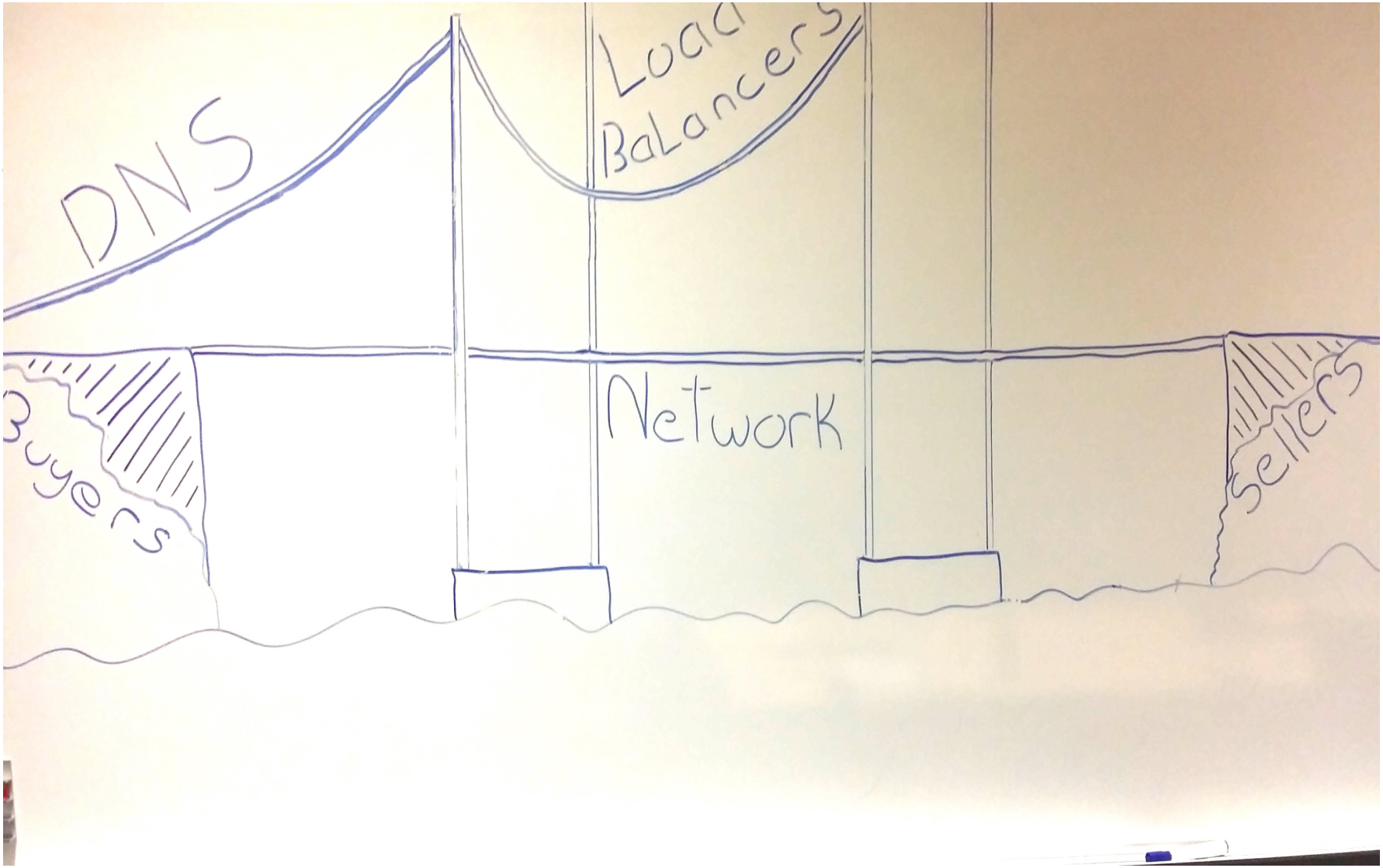

Buyers
Network
Sellers

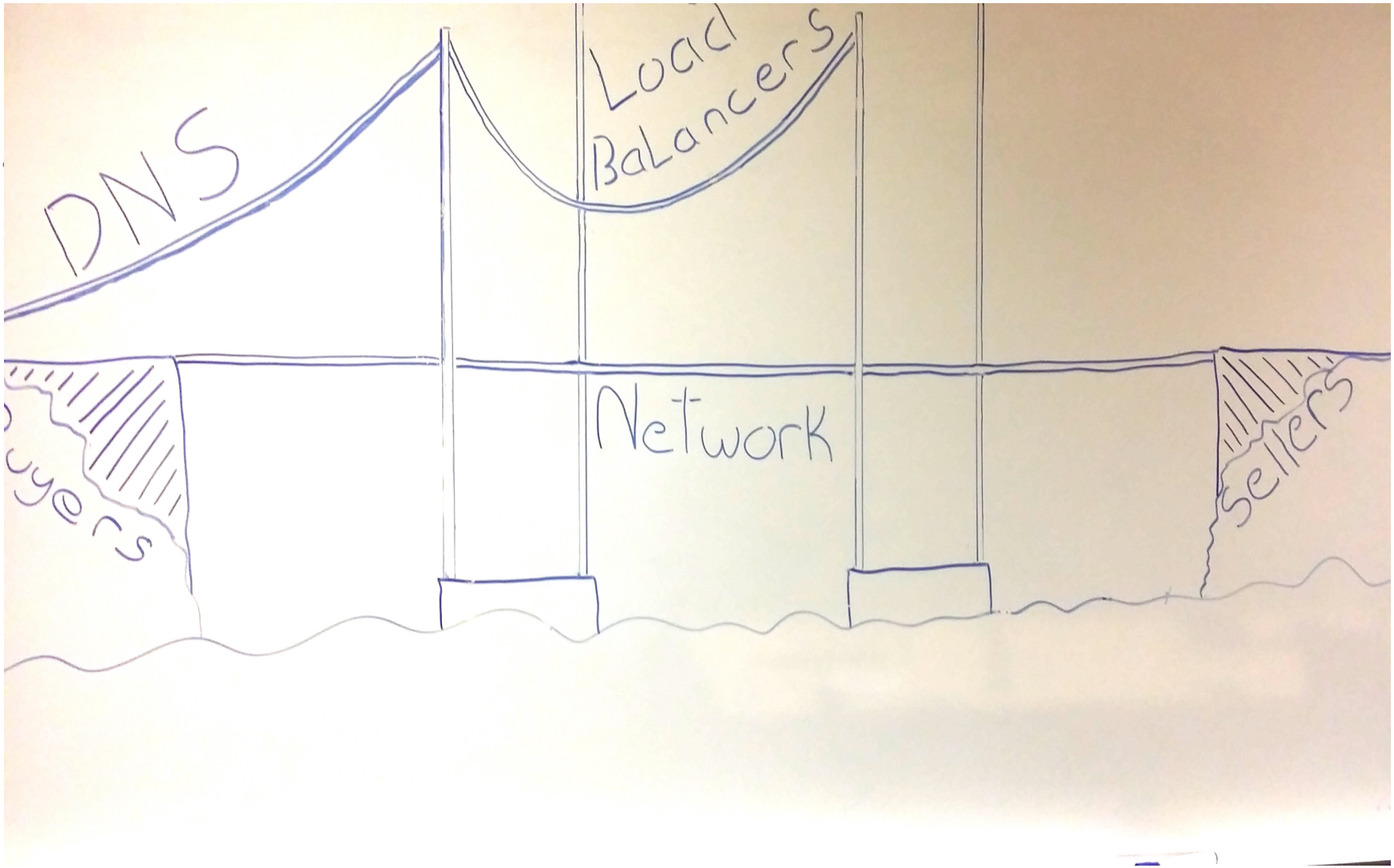

Buyers
Network
Sellers

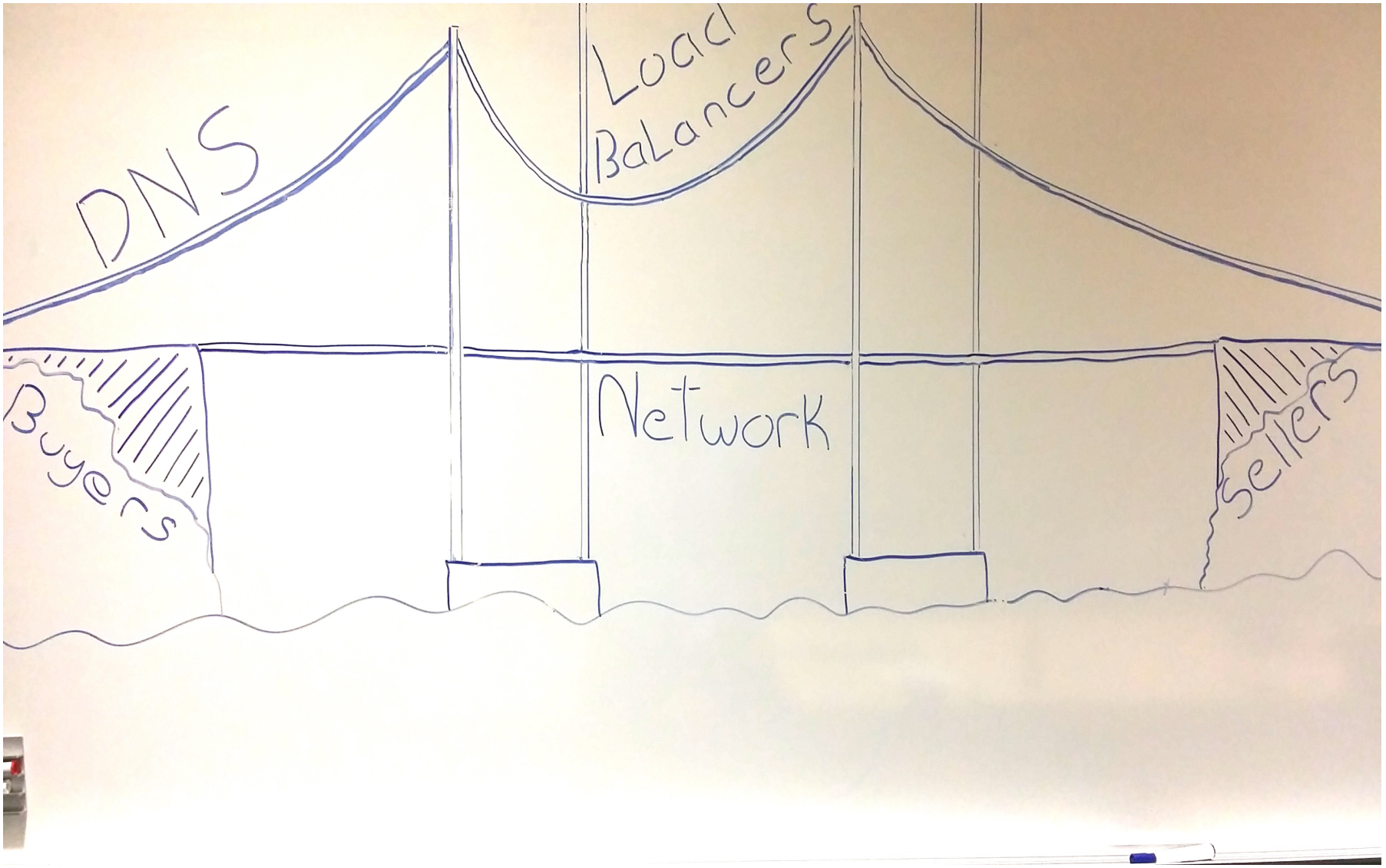

Buyers
Network
Sellers

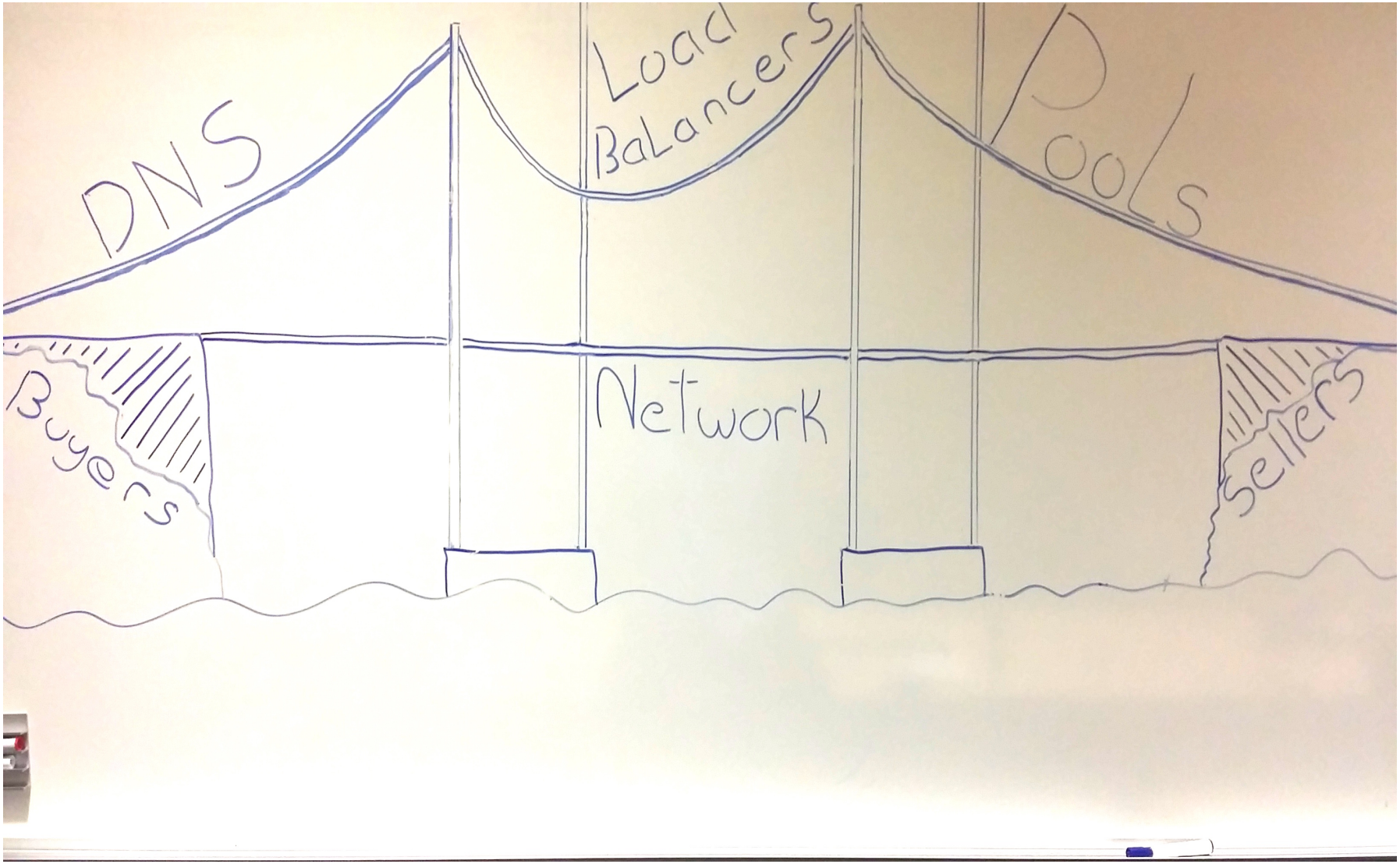

DNS
Buyers
Network
Sellers

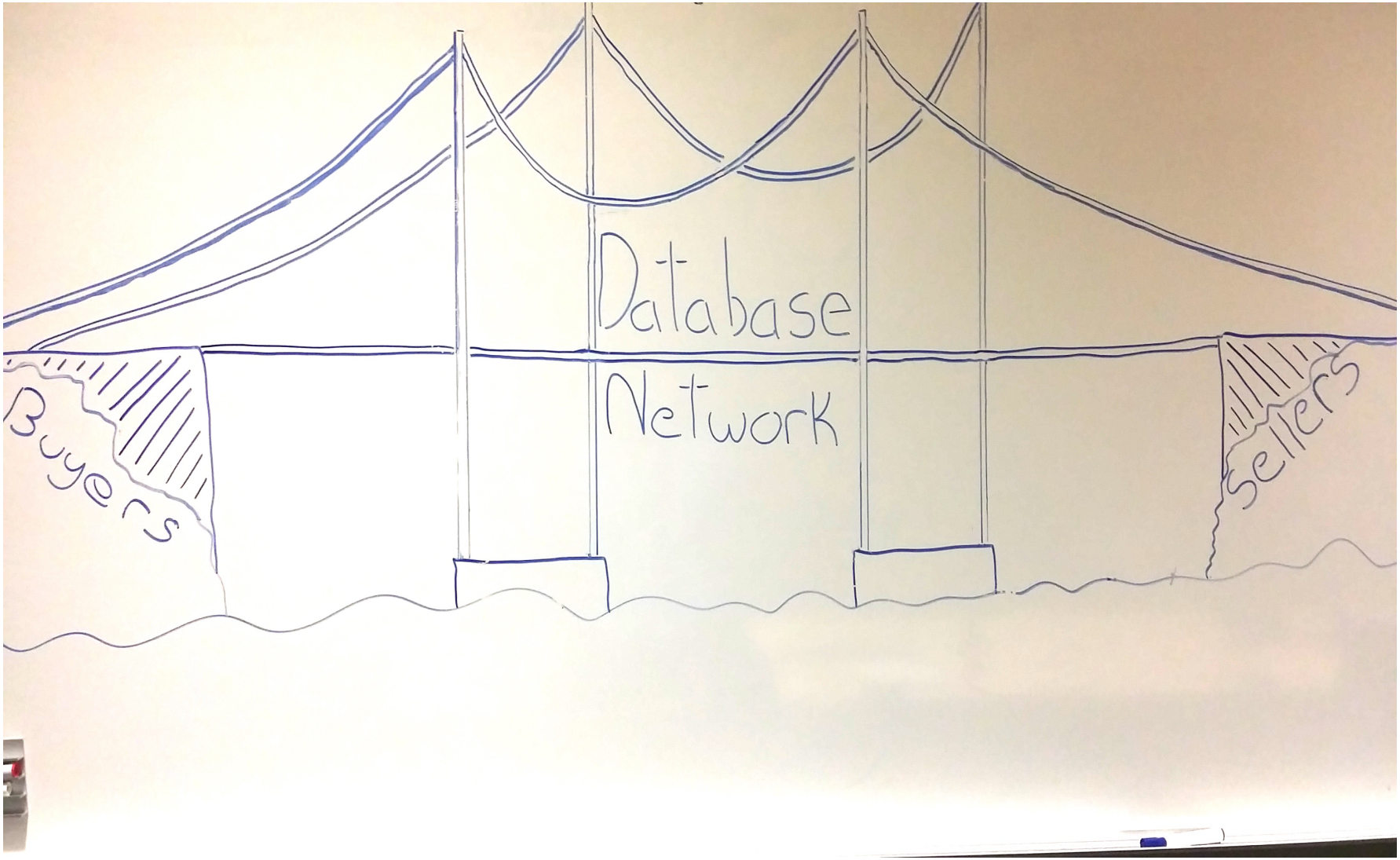

DNS
Load
BaLancers
Network
Buyers
Sellers

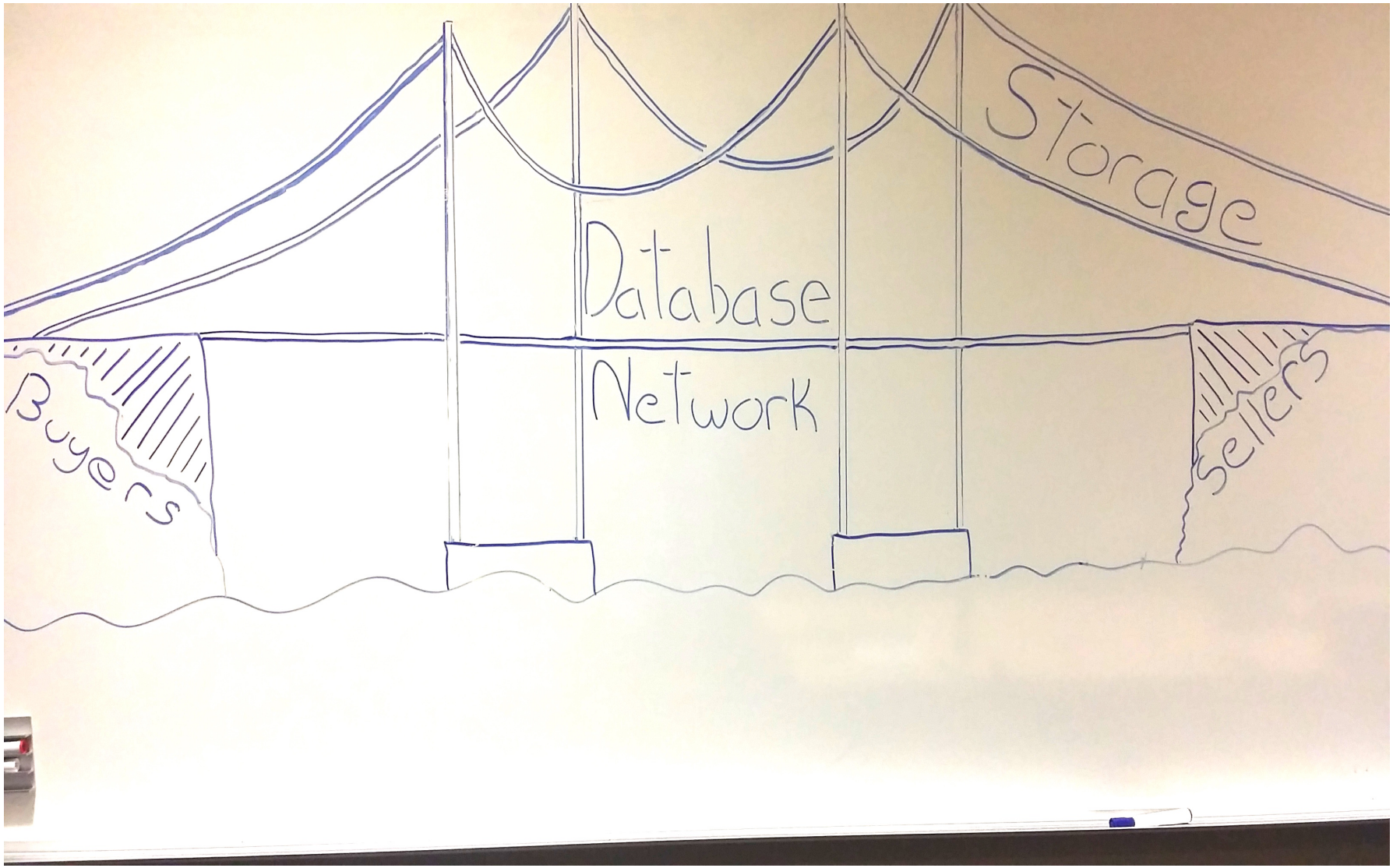

DNS
Load
Balancers
Network
Buyers
Sellers

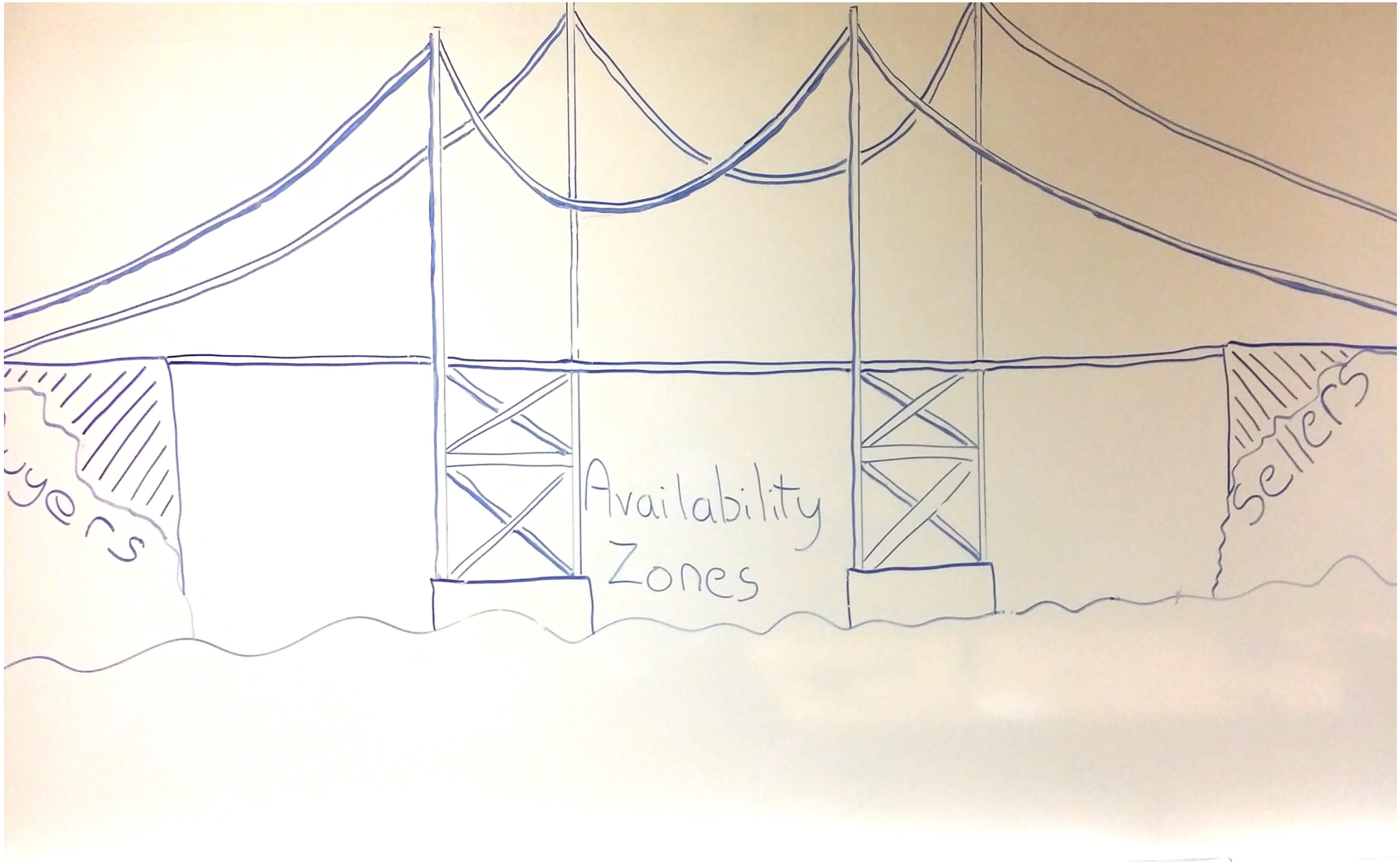

DNS
Load
BaLancerS
Network
sellers

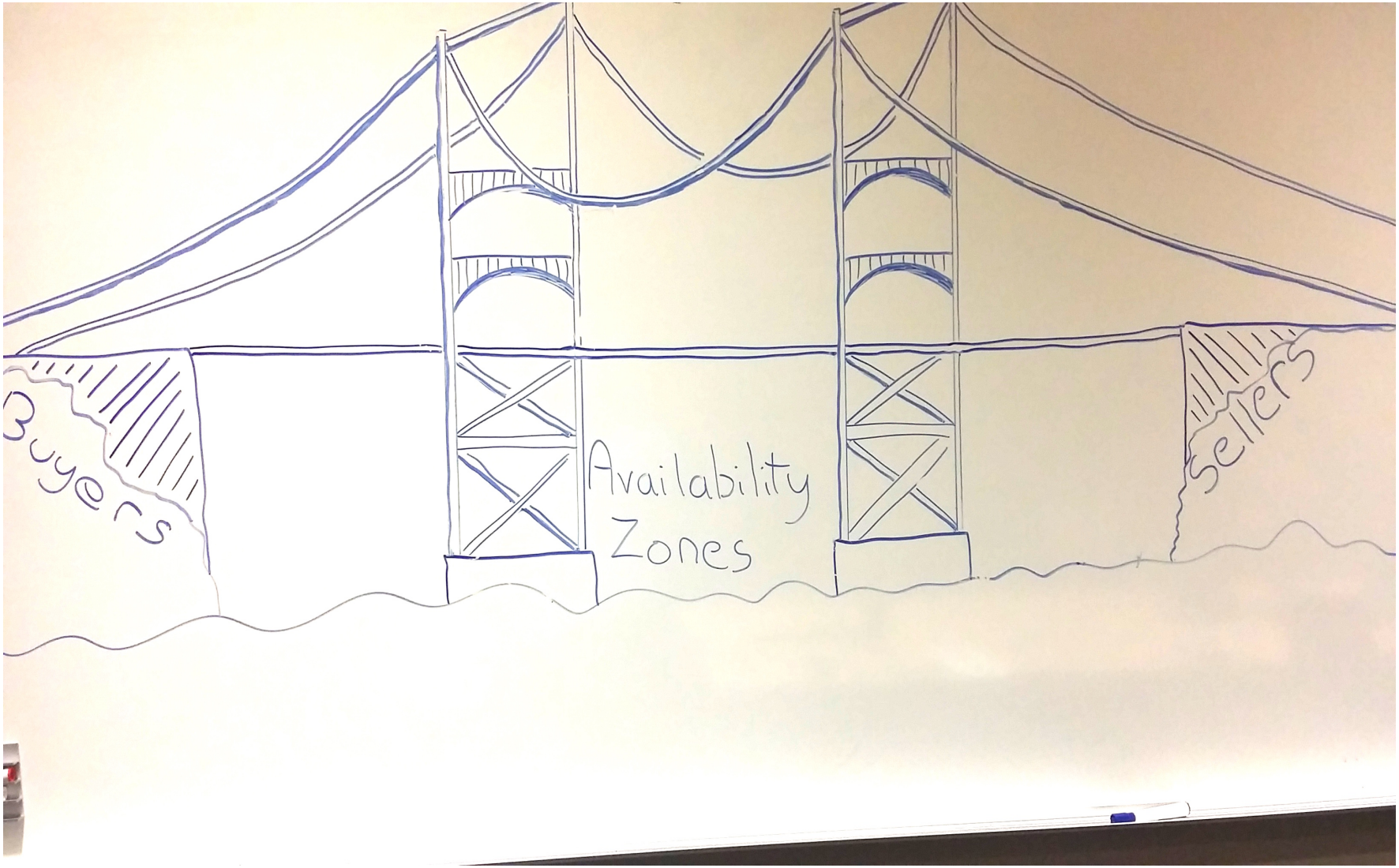

DNS
Load Balancers
Buyers
Network
Sellers

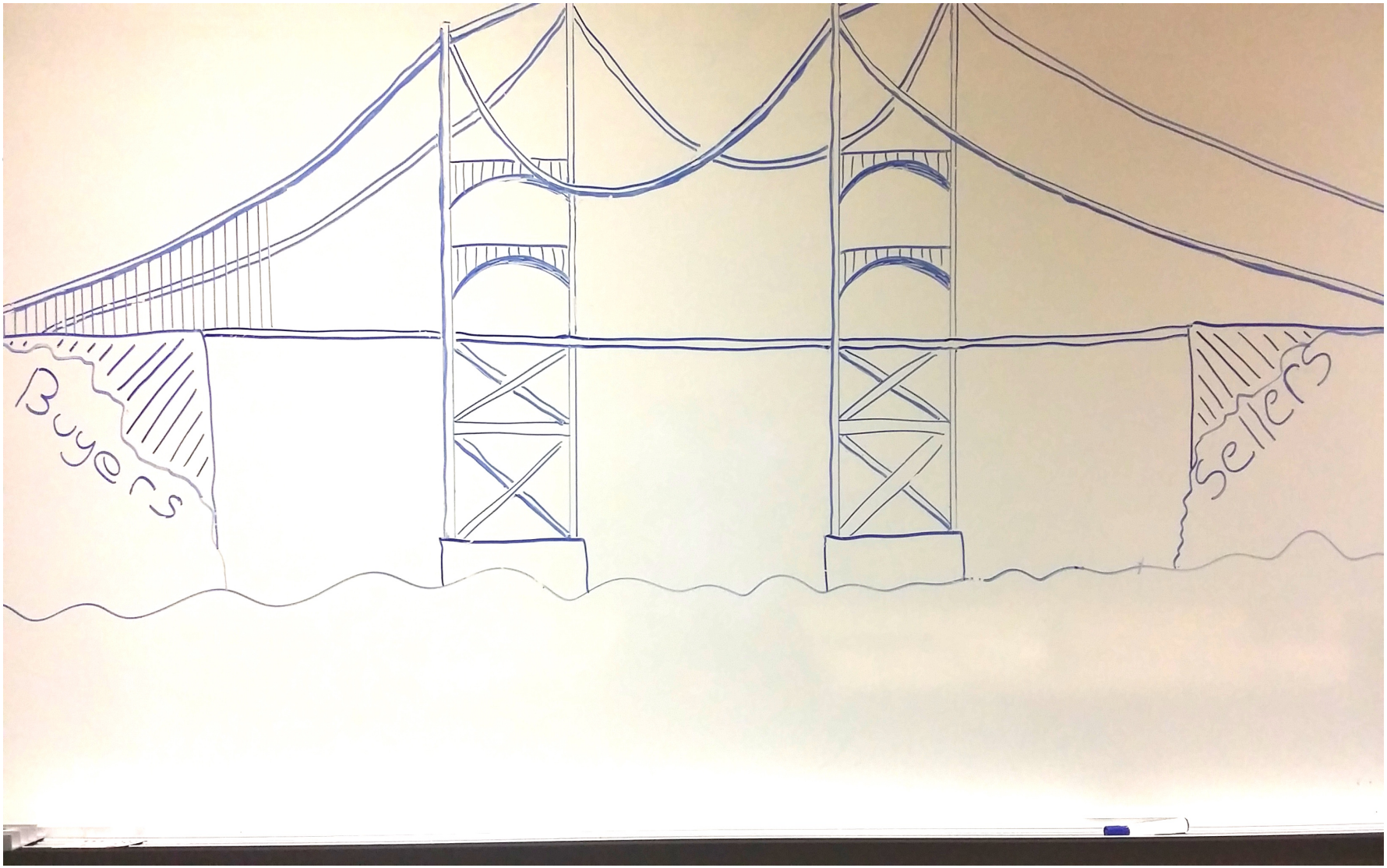

DNS
Load
BaLancers
Pools
Buyers
Network
Sellers

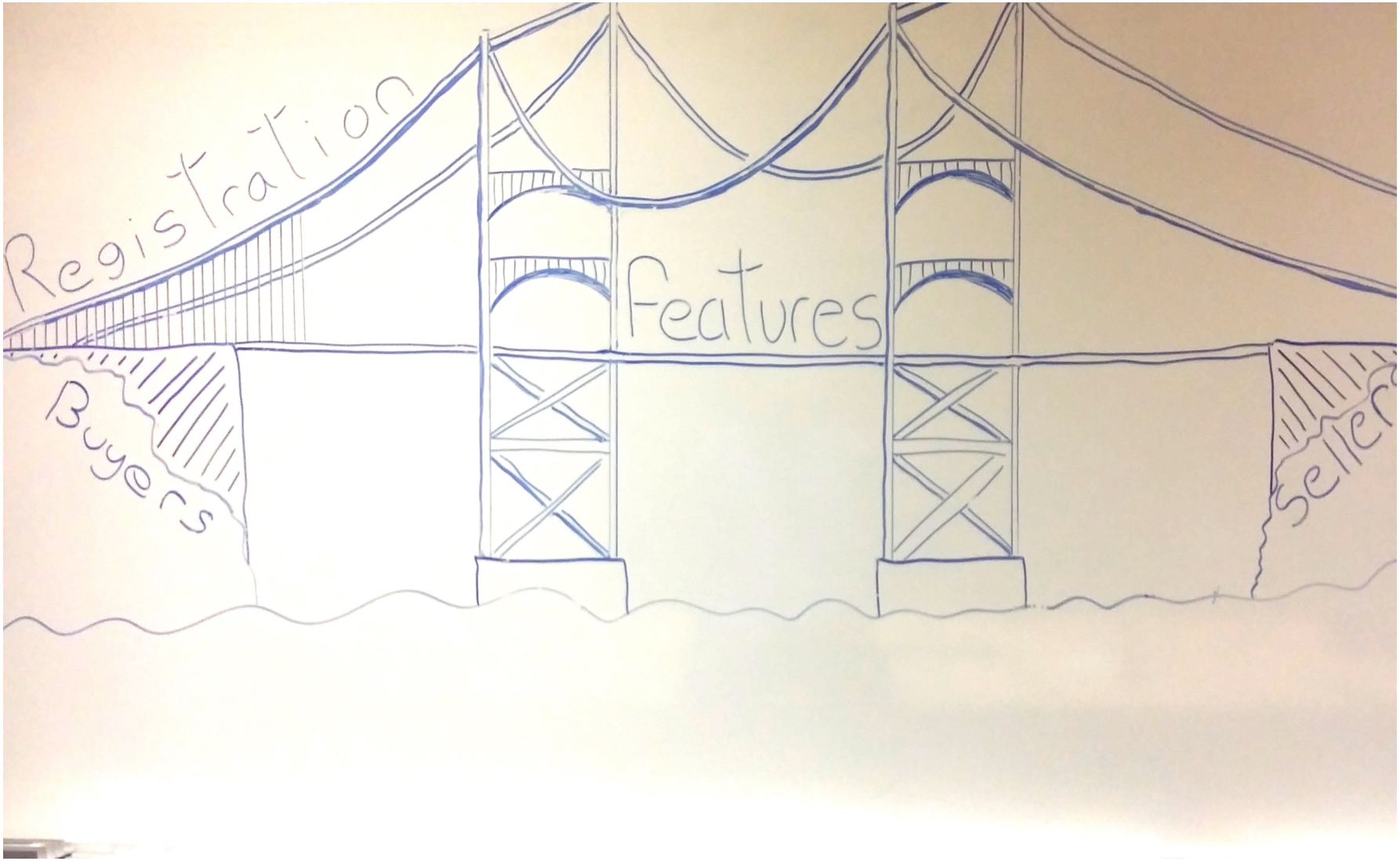

Database
Network
Buyers
Sellers

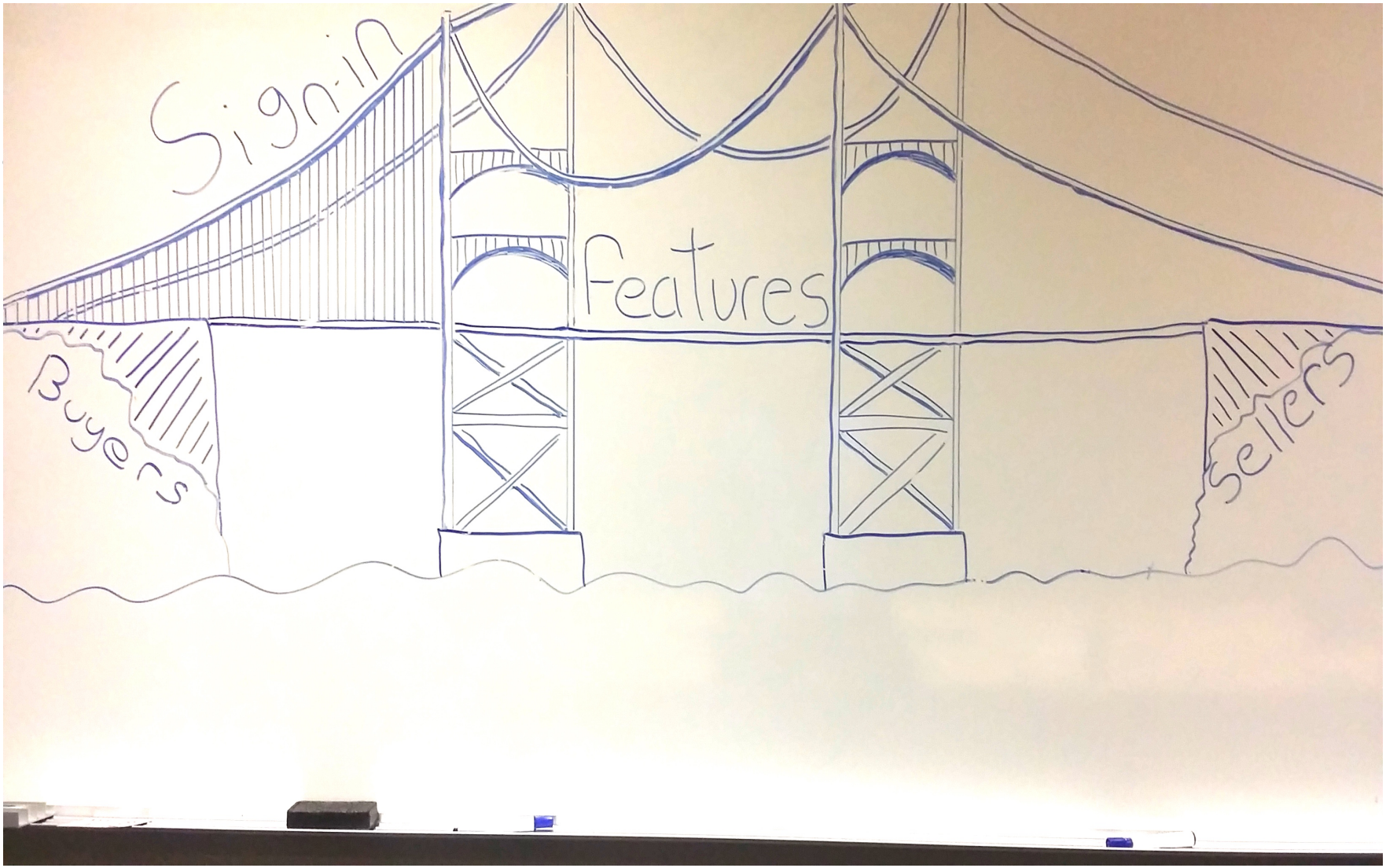

Storage
Database
Network
Buyers
Sellers

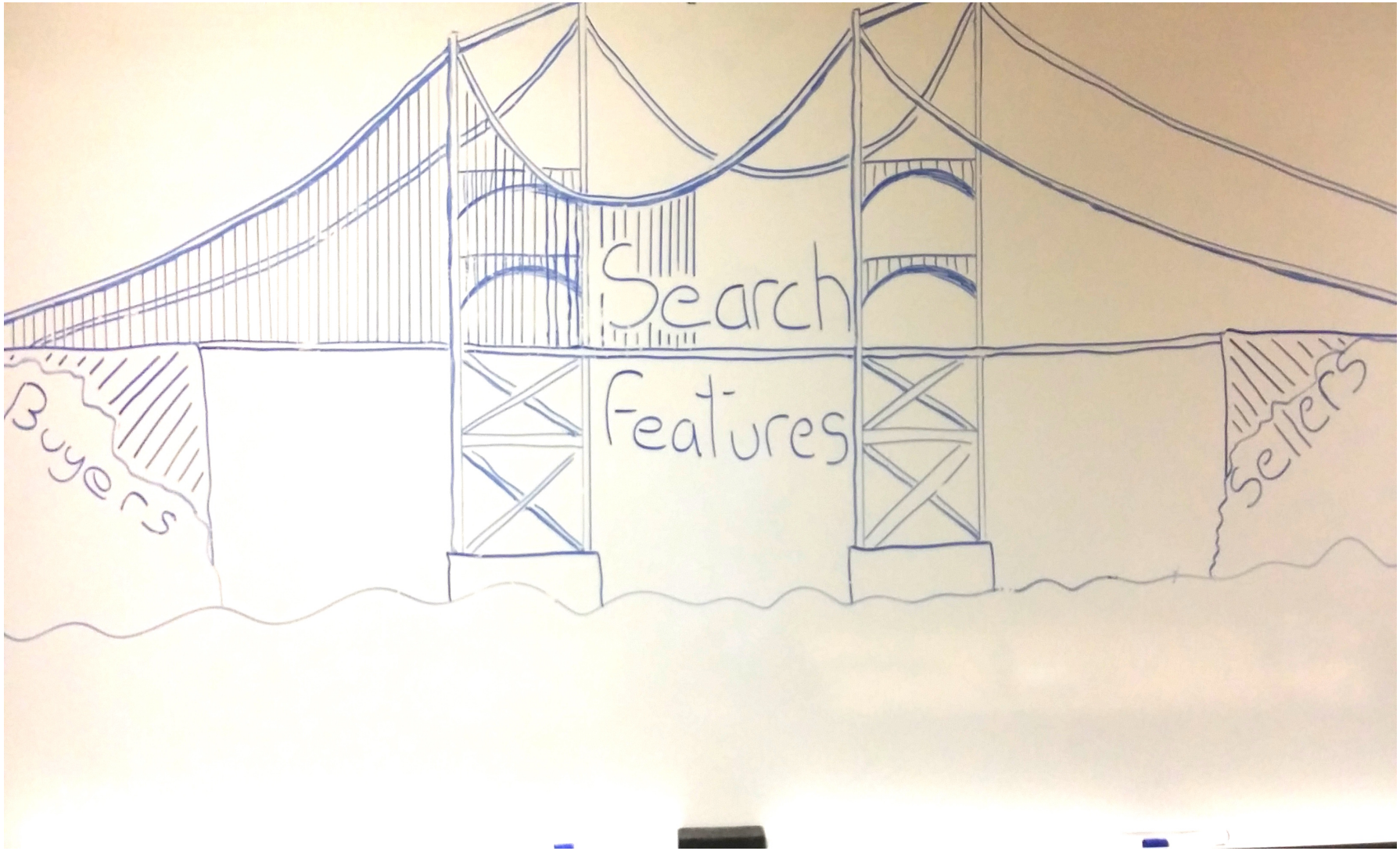

yers
Availability
Zones
Sellers

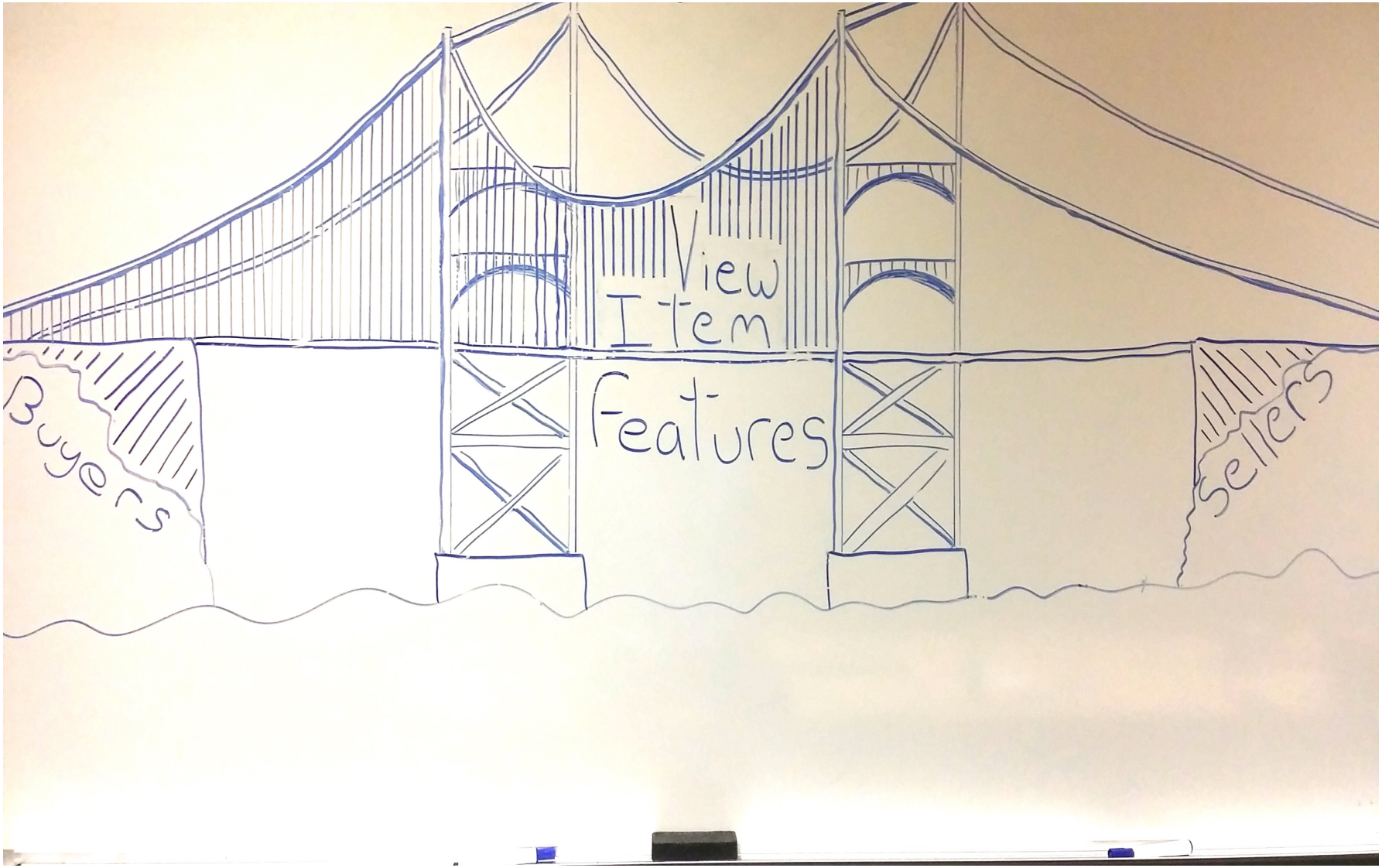

Buyers
Availability
Zones
Sellers

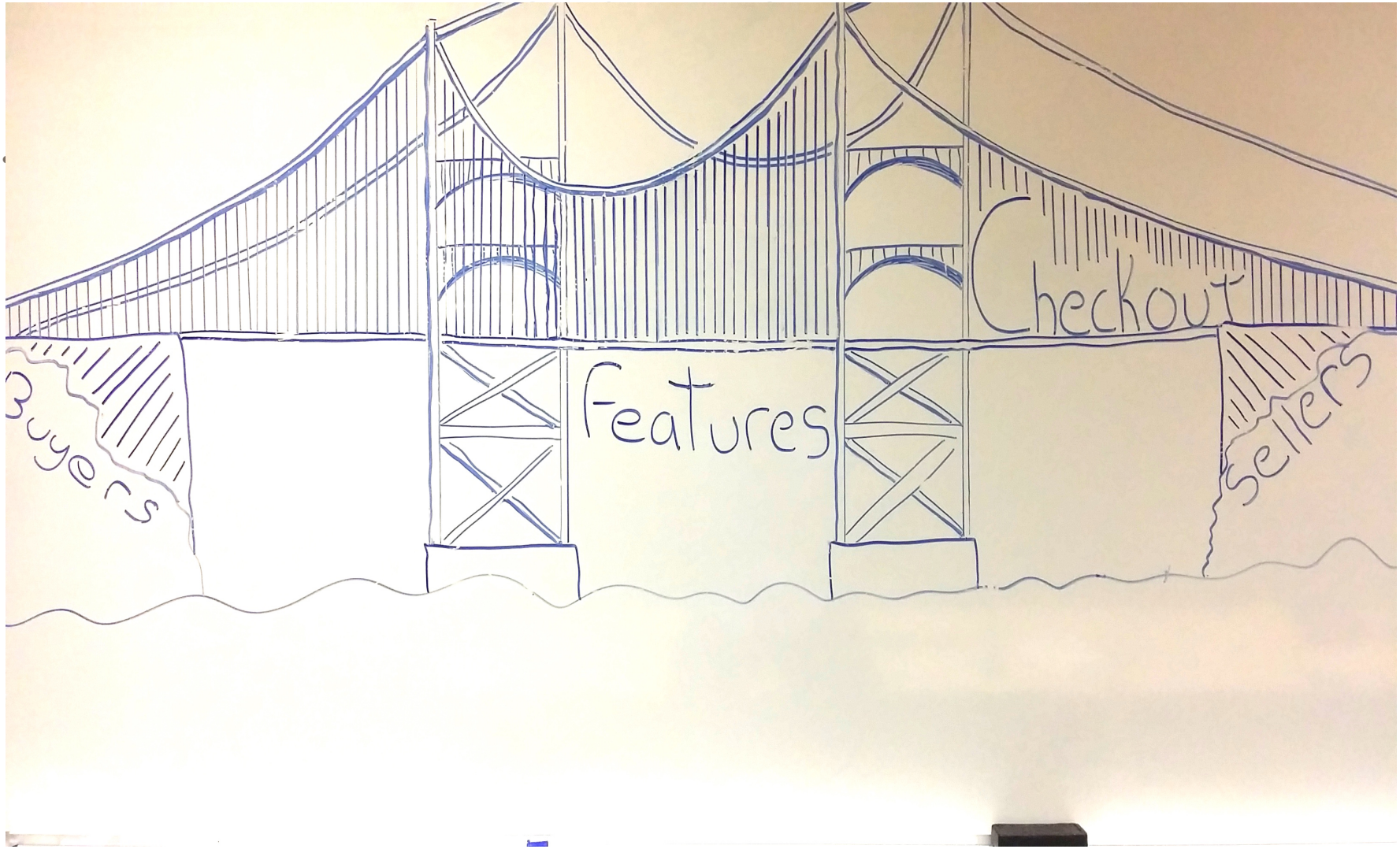

Buyers
Sellers

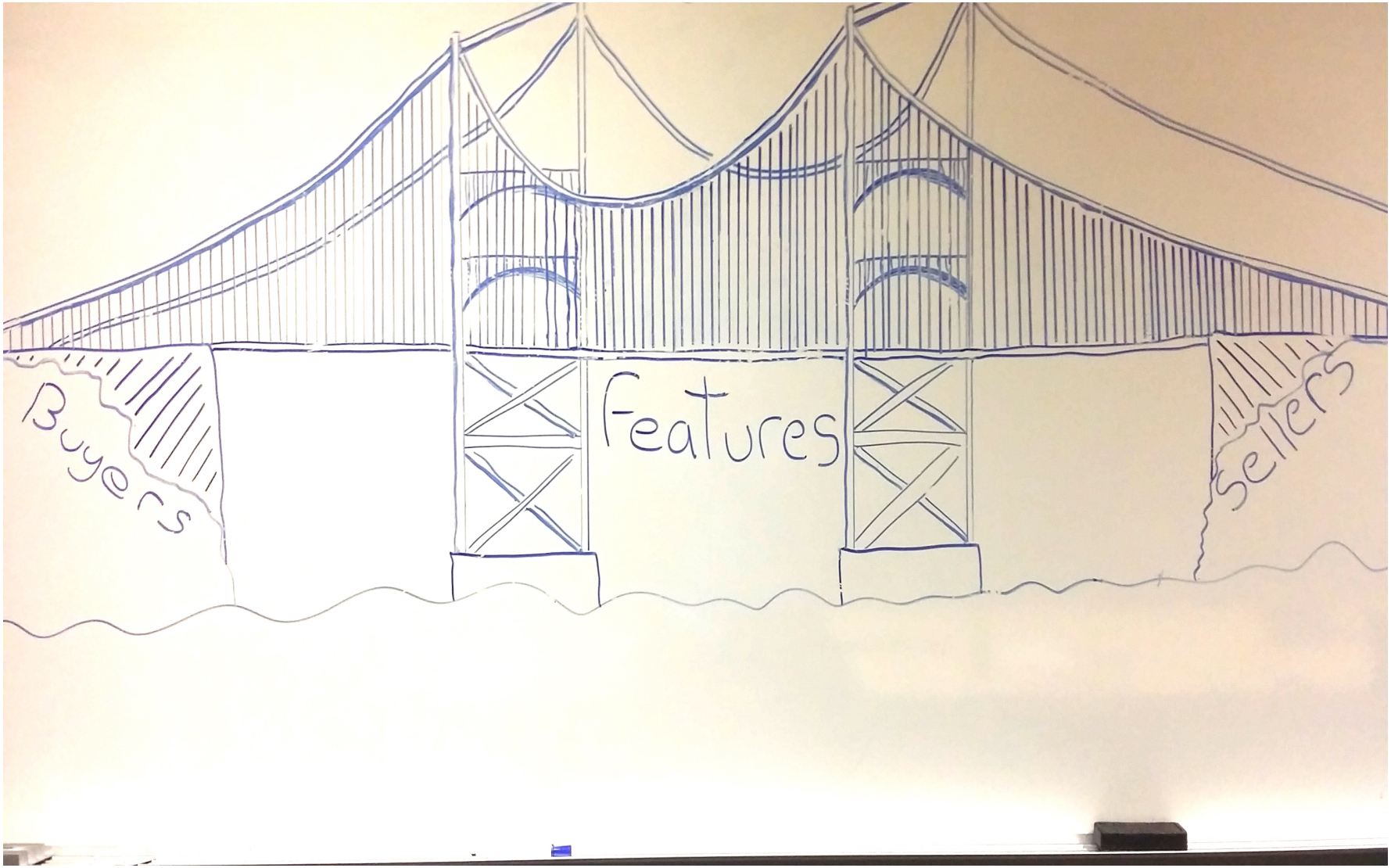

Registration
Features
Buyers
Seller

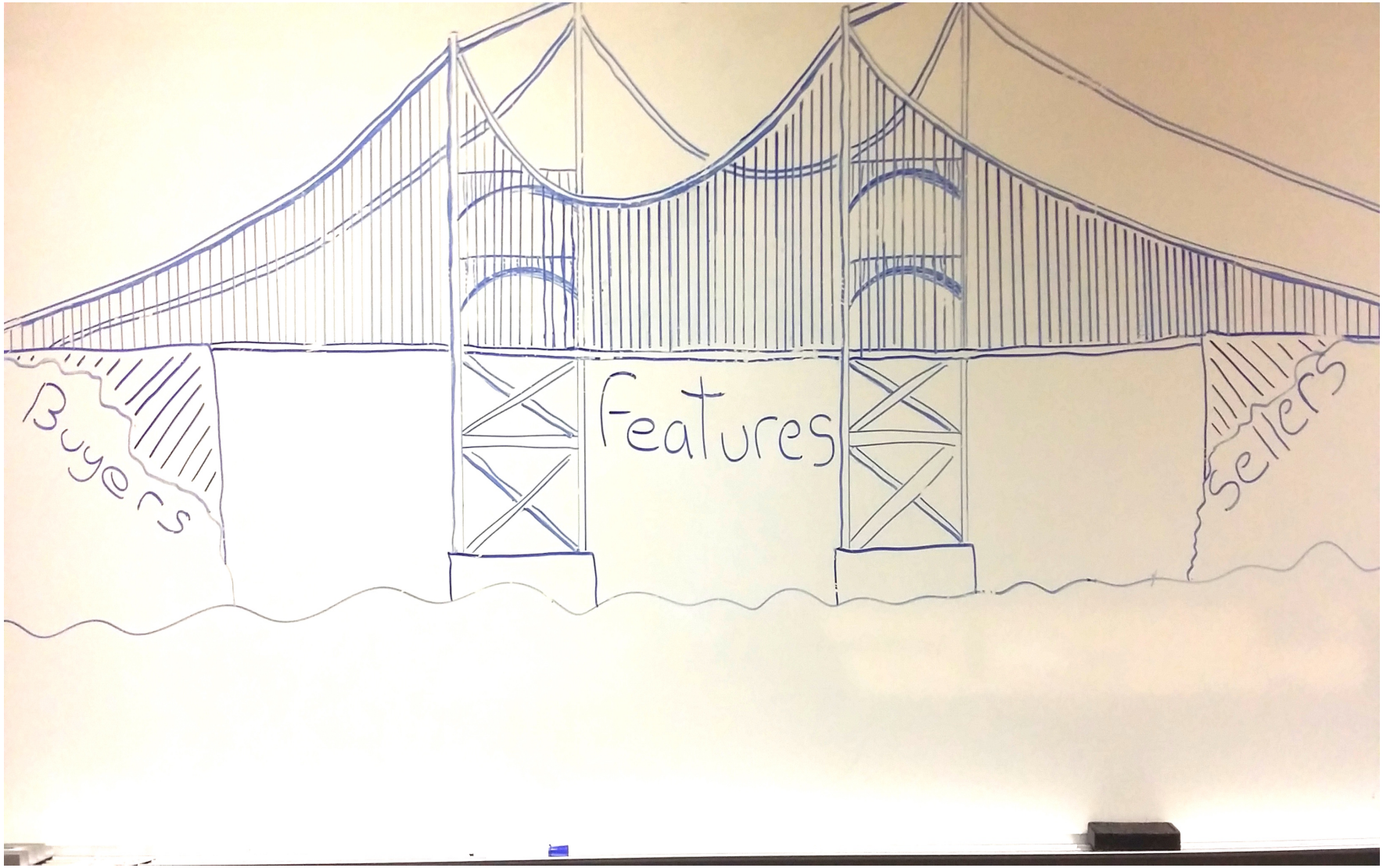

Sign-in
Features
Buyers
Sellers


Search
Features
Buyers
Sellers


View
Item
Features
Buyers
Sellers

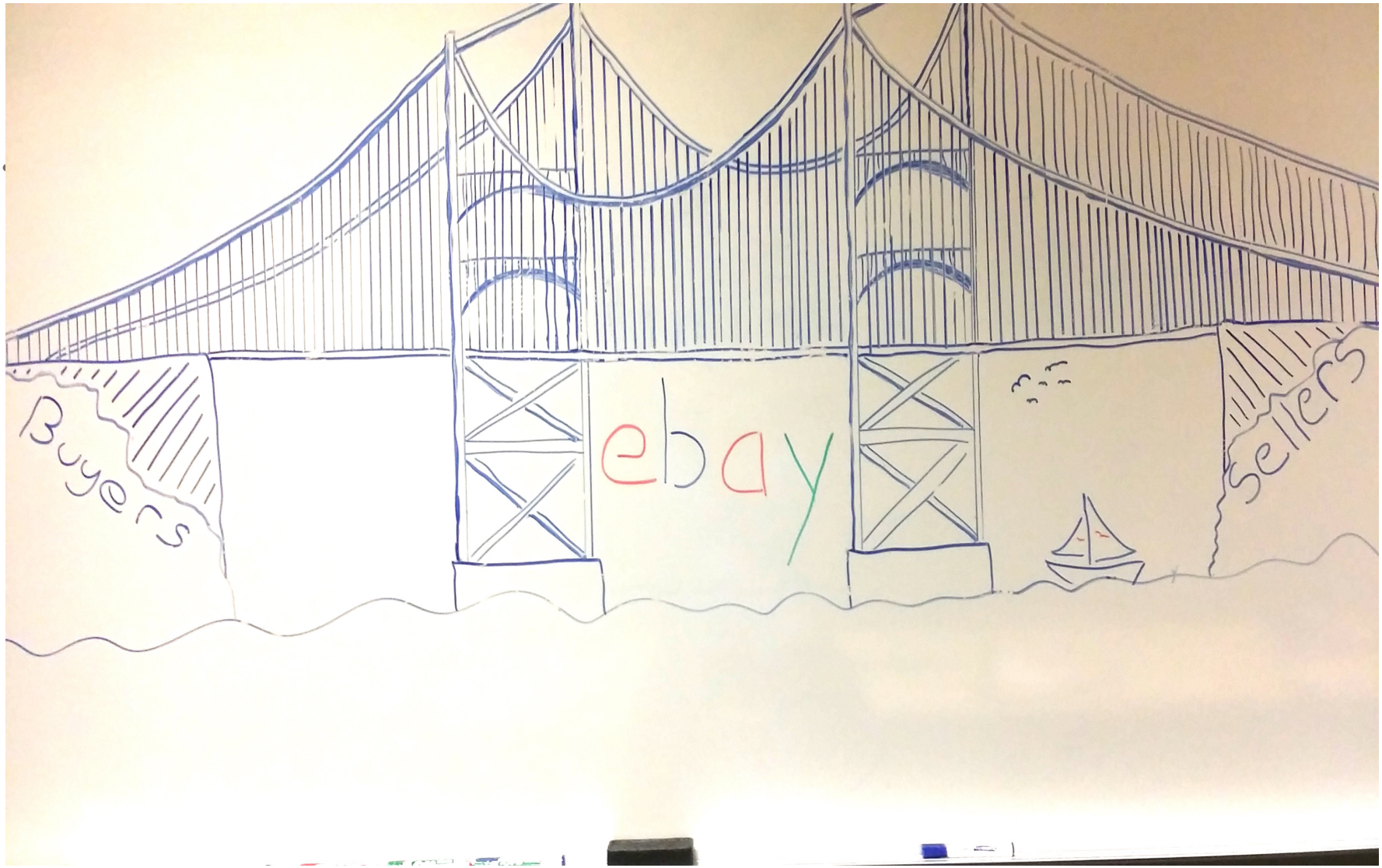

Buyers
Features
Checkout
Sellers


Buyers
Features
Sellers


Buyers
Features
Sellers


Buyers
Features
Sellers


Buyers
ebay
Sellers


Buyers
ebay
Sellers

SO FULL OF FAIL

Creating resiliency through *intelligently* injected failure.

ebay

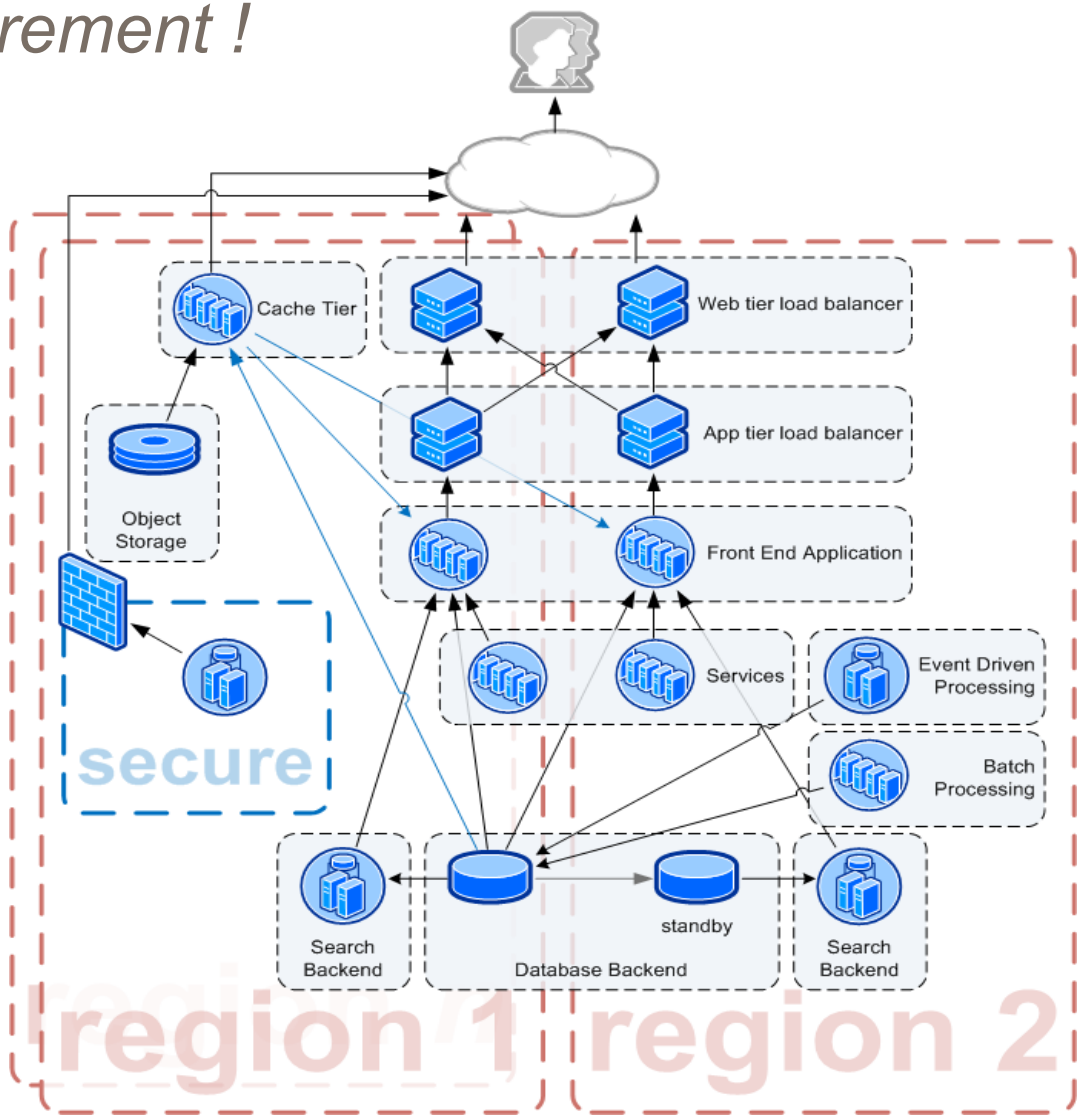

nyugat.hu

So Full of Fail – Failure is NOT an option.
It's a requirement !

- The inability of a system or system component to perform a required function within specified limits. A failure may be produced when an error is encountered.

So Full of Fail – Failure is NOT an option.

It's a requirement !



So Full of Fail – Atomic failures – Overview

- **Software**
  - –Database failure
  - –Service failures
- **Hardware**
  - –LB failures
  - –Compute failures
  - –Network failures

So Full of Fail – Cascading failures

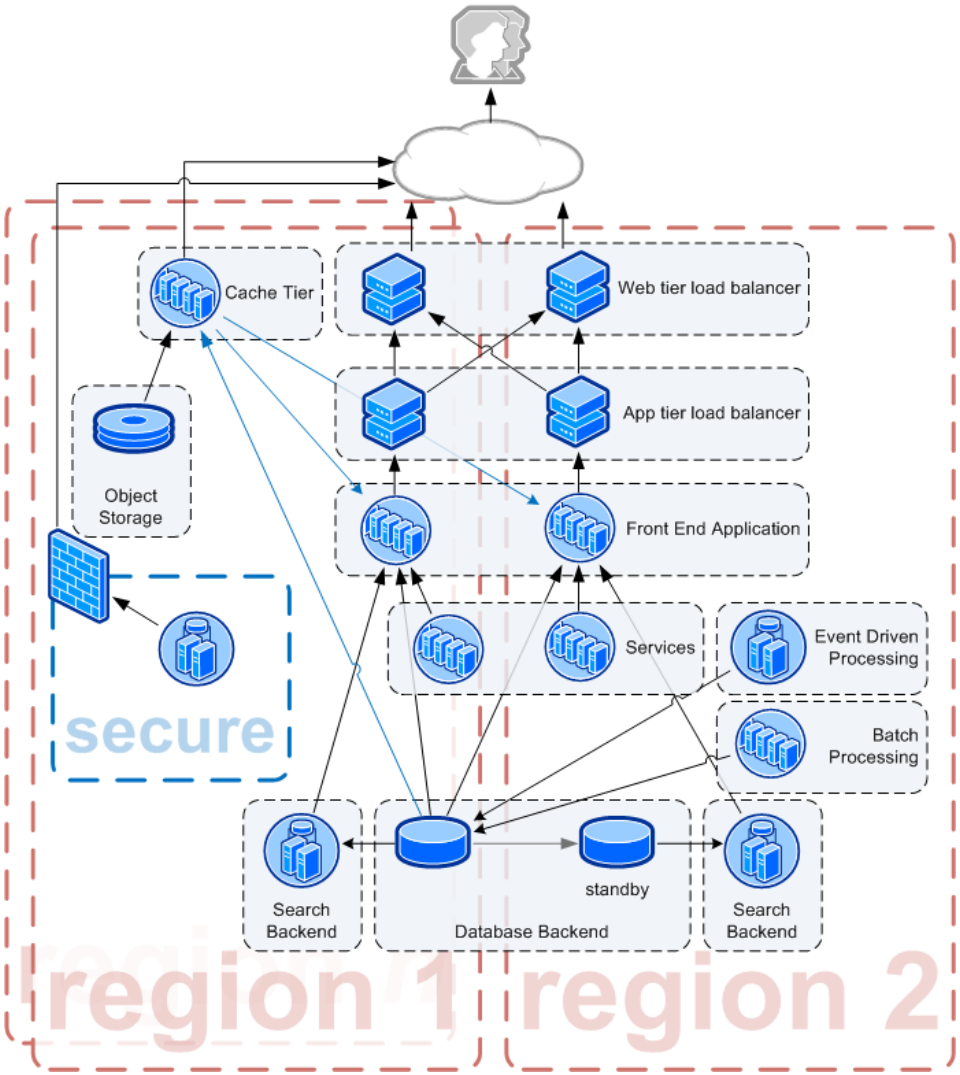
- A **cascading failure** is a failure in a system of interconnected parts in which the failure of a part can trigger the failure of successive parts or the whole system.

So Full of Fail – Cascading failures

So Full of Fail – Cascading failures

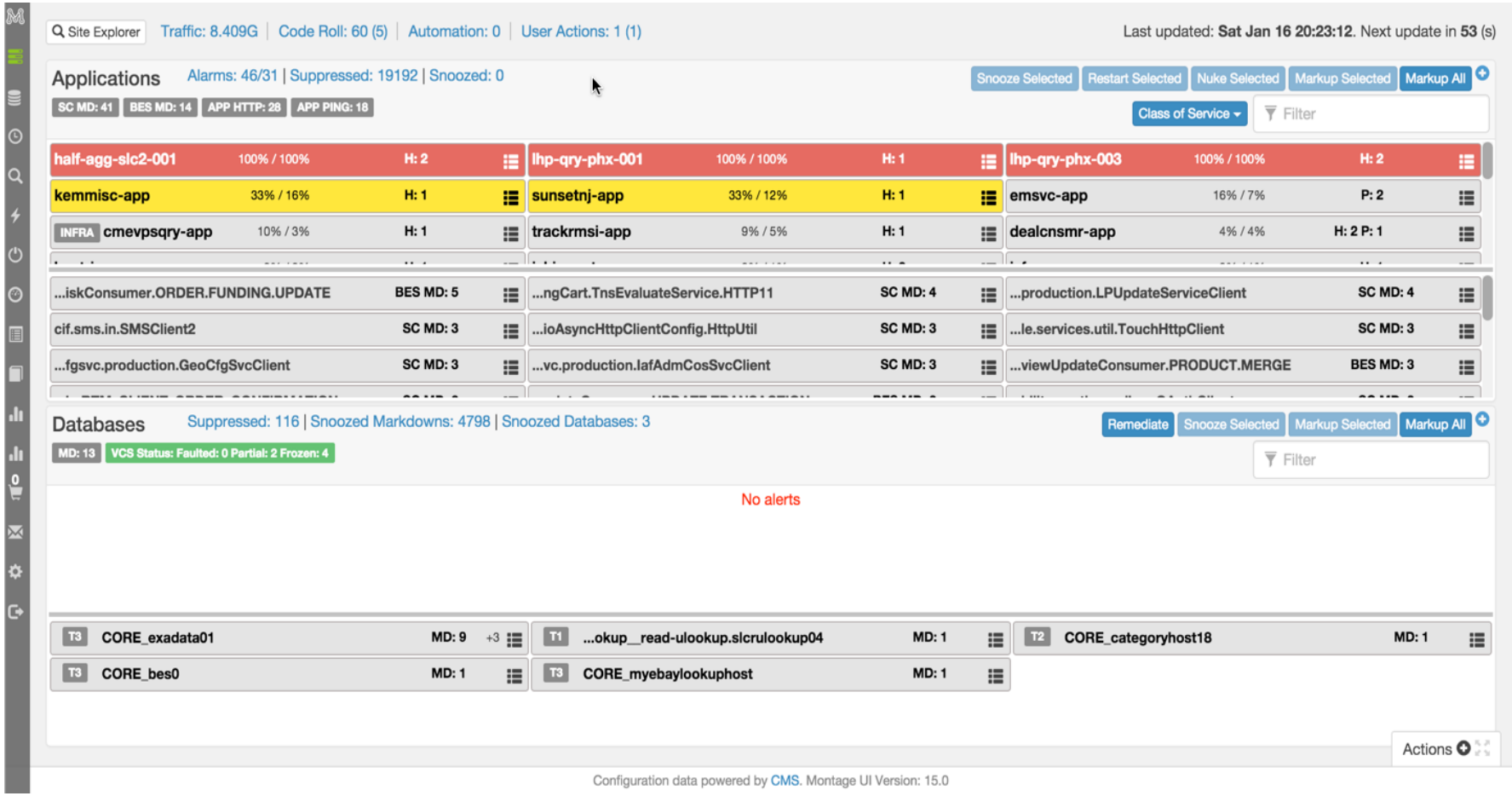
So Full of Fail – Cascading failures – Software

- **Service to Service**
  - –Service A talks to Service B using using load balancer virtual ip
- **Service B to Database**
  - –Service B talks to data store using data access layer

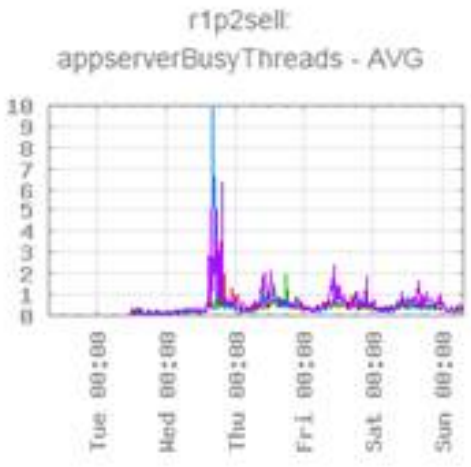


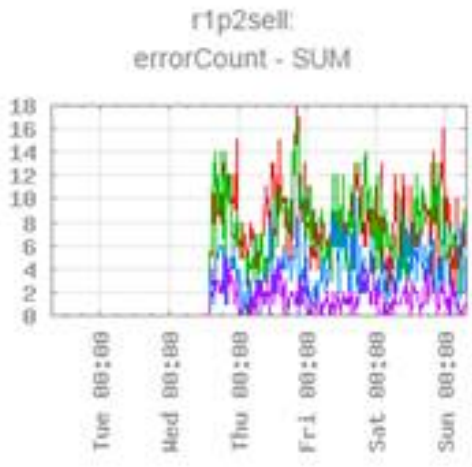
So Full of Fail – Cascading failures – Software

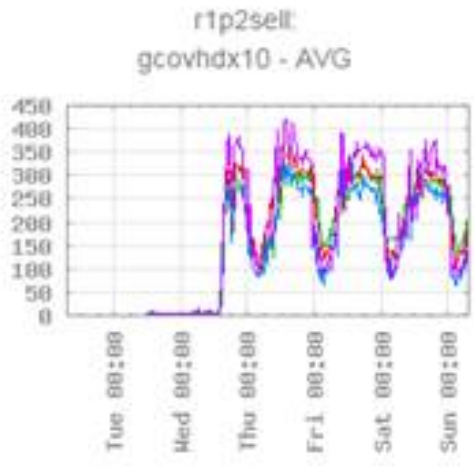
So Full of Fail – Cascading failures – Software

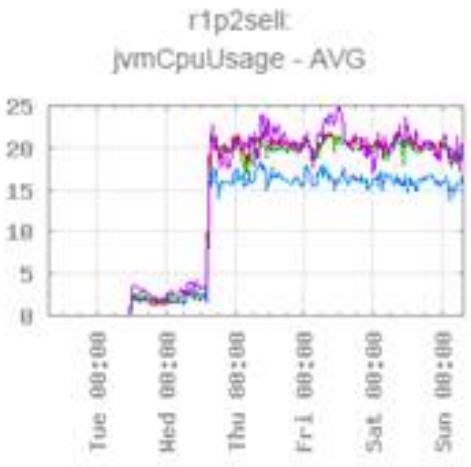


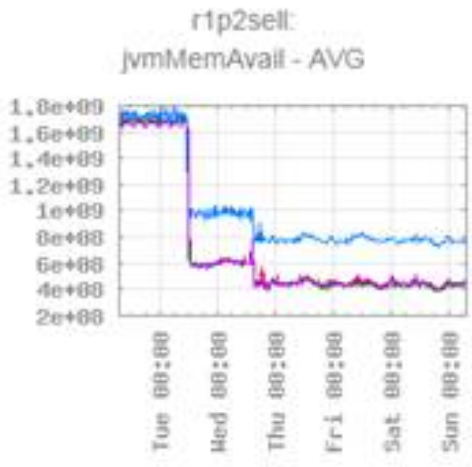
So Full of Fail – Cascading failures– Software

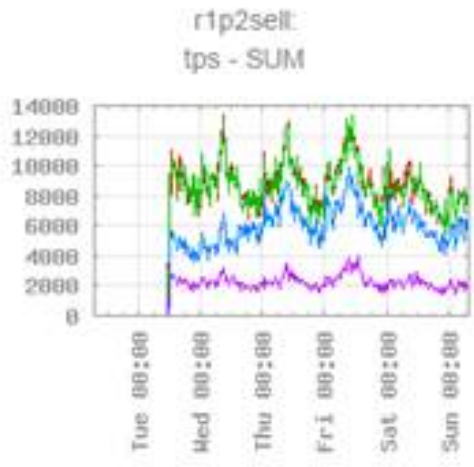












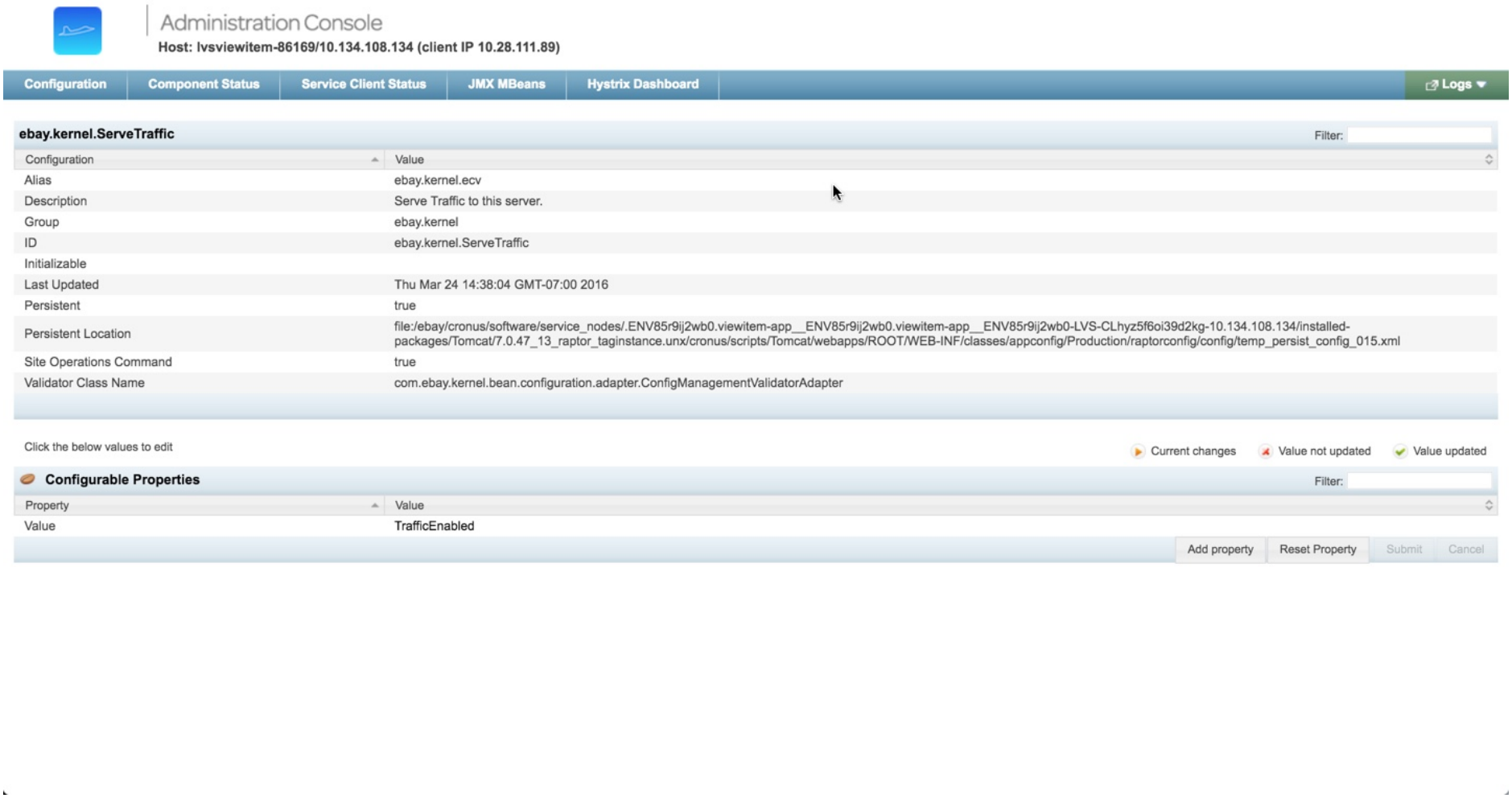
So Full of Fail – Cascading failures – Software - Database

- Connectivity failure
- Query timeouts
- SQL errors

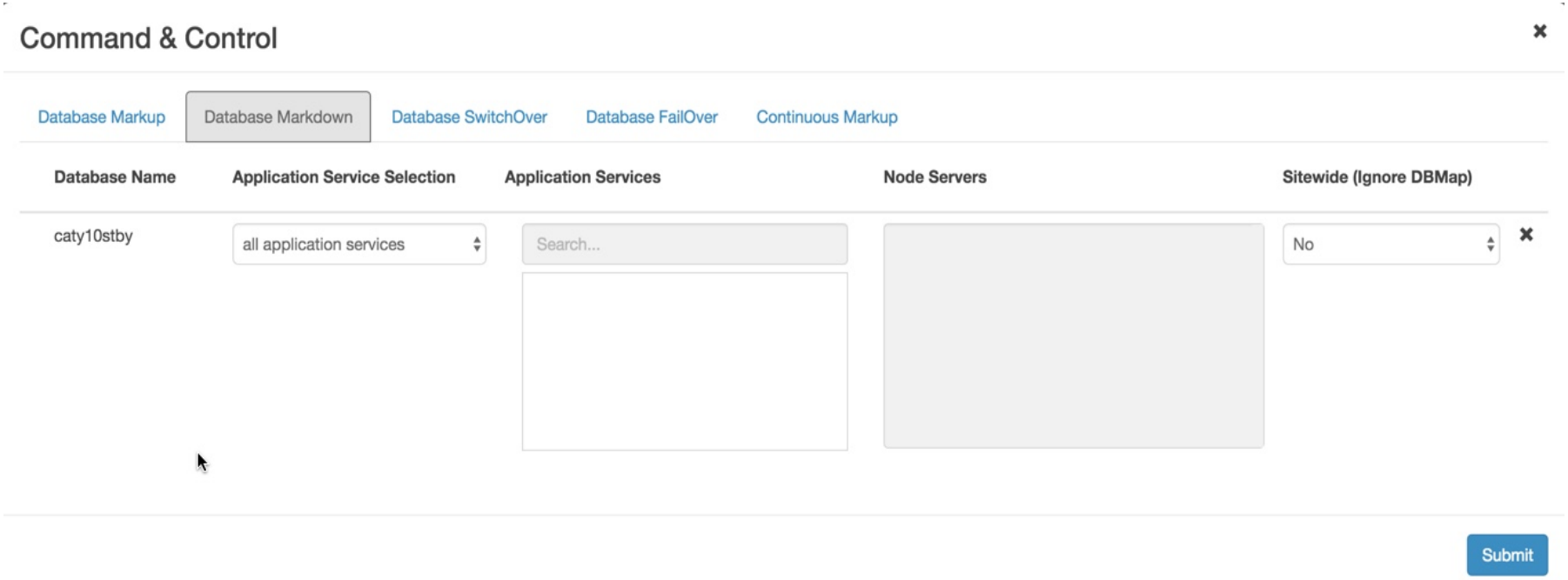
So Full of Fail – Cascading failures – Software

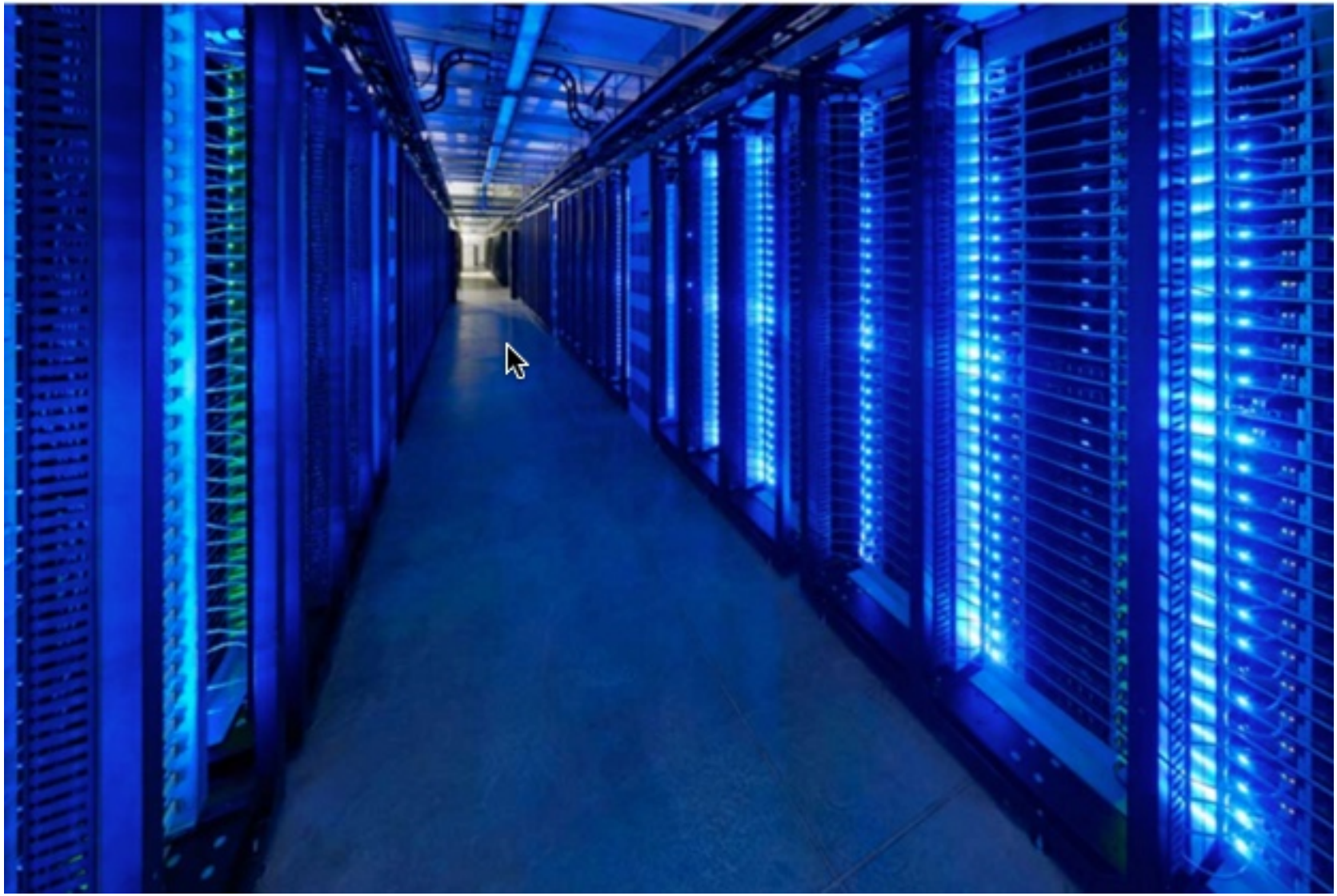
- Data access layer identifies any of the above issues and marks the connection down.
- When the destination is recovered the path is marked active.
- Time to detect the path state is ~15 ms
- Request for connections to a struggling DB can lead to to stacking in upstream
- The fail fast waiter queue logic helps the loaded database

So Full of Fail – Cascading failures – Software



So Full of Fail – Cascading failures – Software



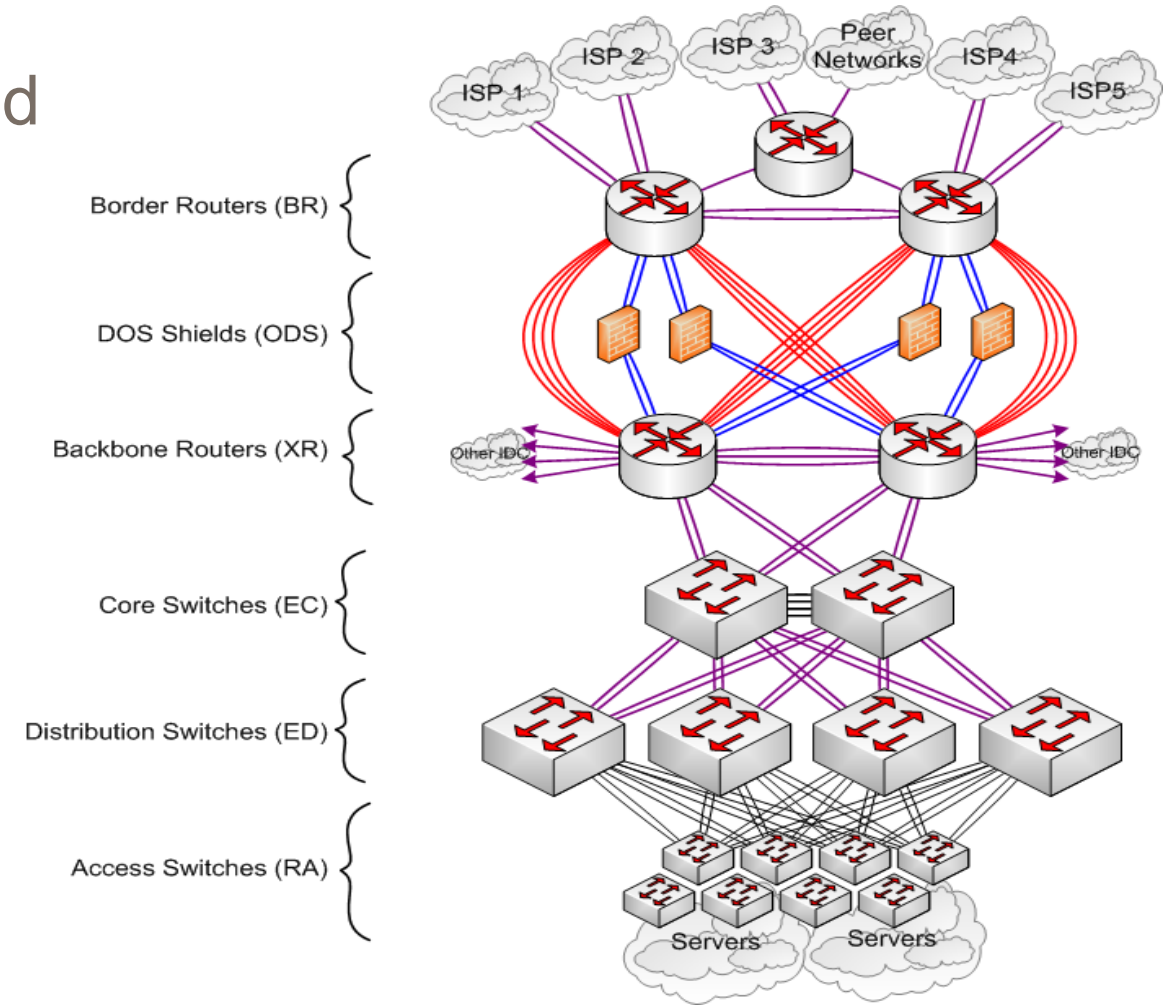


So Full of Fail – Cascading failures – Hardware

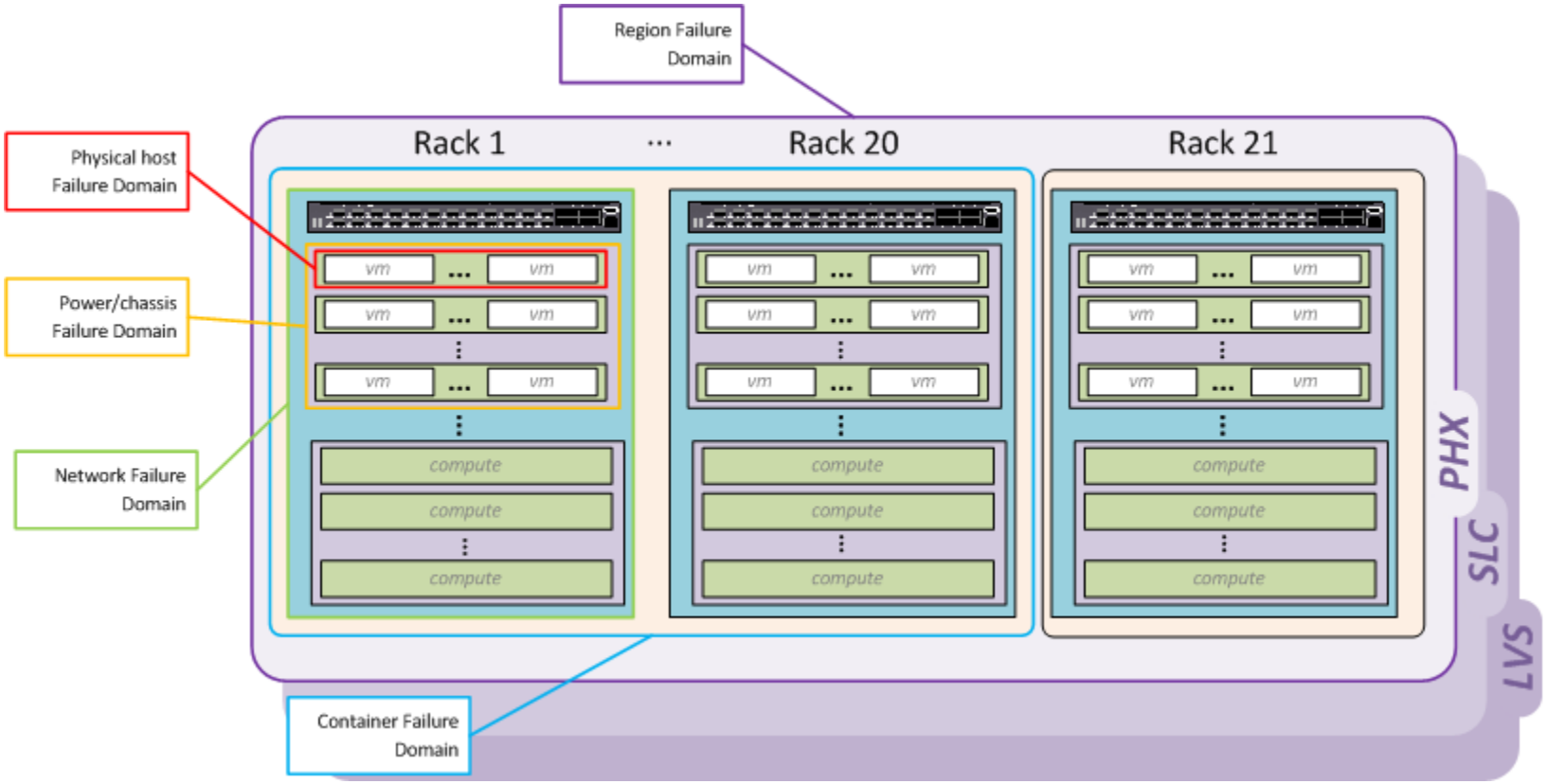
- LB failures
- Compute failures
- Network failures

So Full of Fail – Cascading failures – Hardware - Network

- Traffic organized in layers.
- Redundant interconnected paths.



So Full of Fail – Cascading failures – Hardware



$$P(n) = MAX\left(17\%, \frac{1}{S_n}, \frac{1}{F_n}\right)$$

- n = Failure Domain (e.g. Network)
- S_n = Number of Service Instances in Failure Domain n
- F_n = Number of instances of Failure Domain n

PREVENTION THROUGH PARADIGM

Prevent and remediate atomic failure in order to prevent cascading failure scenarios.

ebay

PREVENTION THROUGH PARADIGM

- Site code is deployed on a common set of platforms which enforce the aforementioned tools.
  - Engineering and development architecture effort is “front-loaded” to ensure that all production services can be restarted in exactly the same way no matter what the underlying command.
  - Uniformity of response
  - Remediation at the node level can be automated.
- State based and event based monitoring of key operating metrics.
  - The 9 or so OS-level metrics we all monitor
- Capacity monitoring for “DR’ compliance
  - Two or more co-locations
  - One feature per “pool”
- Provision on demand
- Regional code roll

PARADIGM BEFORE PROCEDURES.

- For maximum rapid remediation your pool or feature should comply with operational paradigms.
- If the care and feeding of your particular beautiful ‘unicorn’ requires special attentior or procedures:
  - You’re asking someone to go against their established paradigm and training.
  - At a high rate of speed
  - As the rarely seen or remembered exception

DEALING WITH DISASTER DIRECTION AND DECISION

Preventing the cascade failure through redirection and deciding which features to keep.

ebay

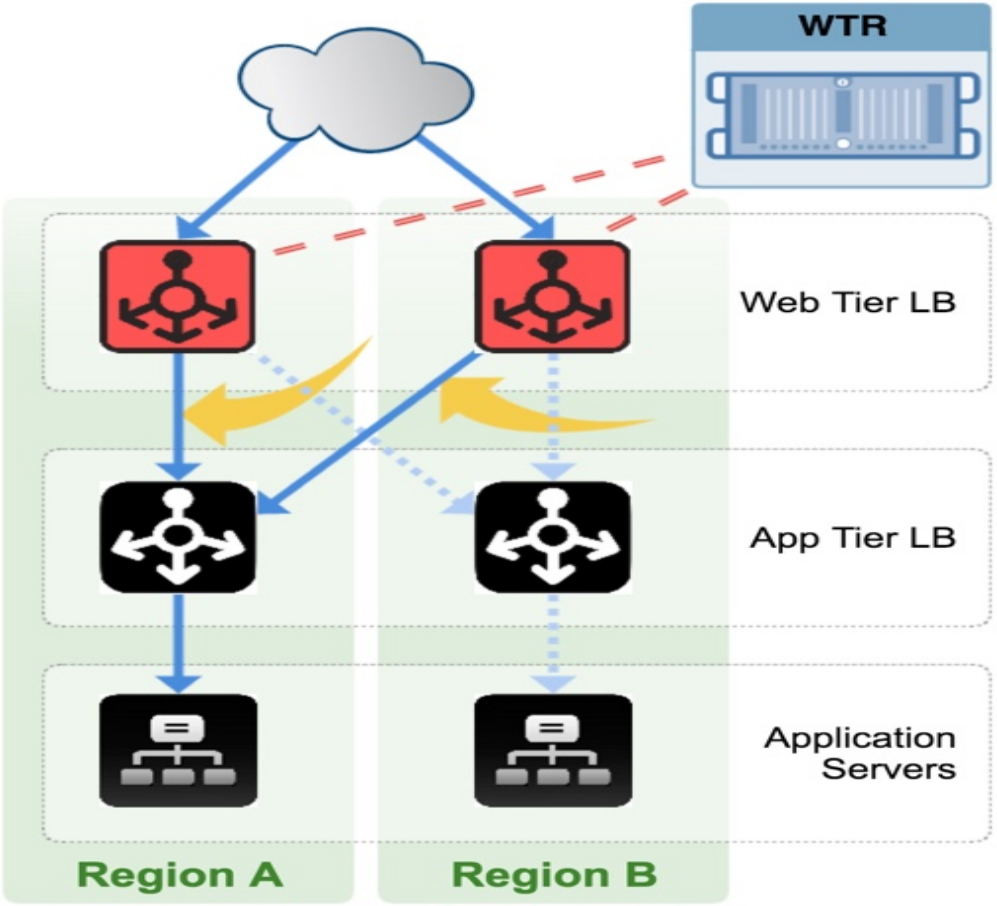
When a lion tamer holds a chair in front of the lion's face, the lion tries to focus on all four legs of the chair at the same time. With its focus divided, the lion becomes confused and is unsure about what to do next. When faced with so many options, the lion chooses to freeze and wait instead of attacking the man holding the chair.

Clyde Beatty taming a lion with a chair. (Image from Harvard Library.)

ebay

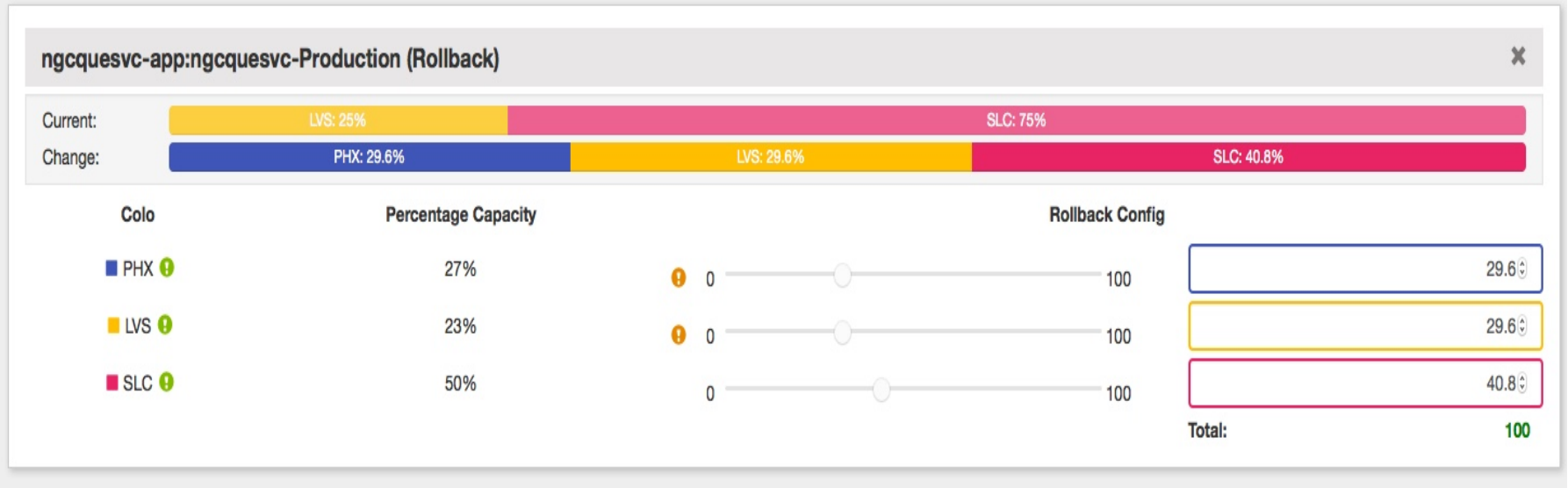

571-7
MISSILE SITE
STATUS
LOCAL
MSA

DEALING WITH DISASTER - DIRECTION

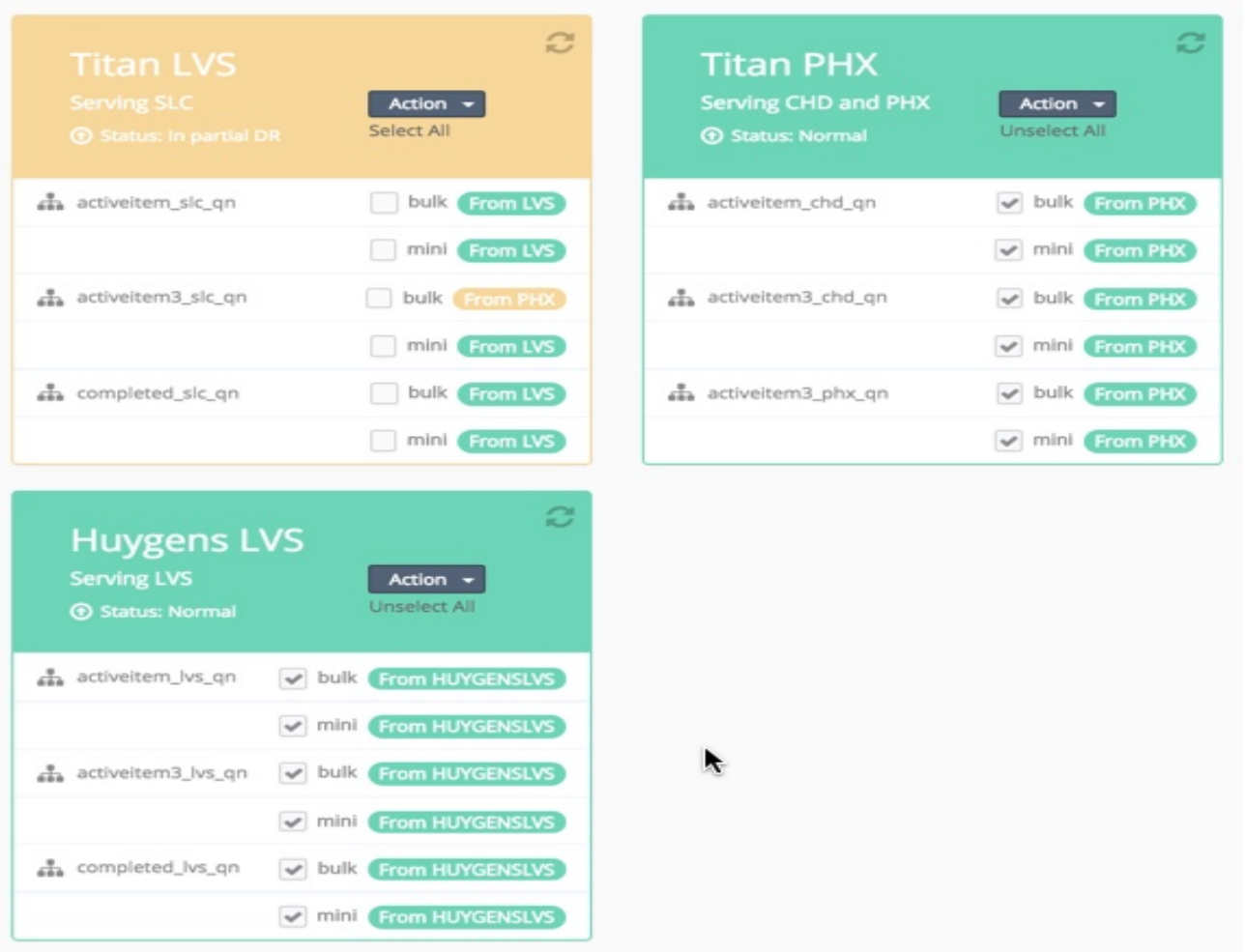


DEALING WITH DISASTER - DIRECTION

DEALING WITH DISASTER - DIRECTION



DEALING WITH DISASTER - DIRECTION



DEALING WITH DISASTER - DIRECTION

- To prevent total failure of the bridge we have to get the kittens to drive over only one lane
- Connection limits would be required
- Blocking bots
- Scale down to minimal serving modes
- Wiring off features like advertising.


1
2
3
4

CONCLUSIONS

ebay

So Full of Fail– Total Failure Avoidance

- Bubble up anomalies for inspection
- Fail Fast
- Mark down the failed paths
- Preserve the user experiences
- Use system approaches to solve