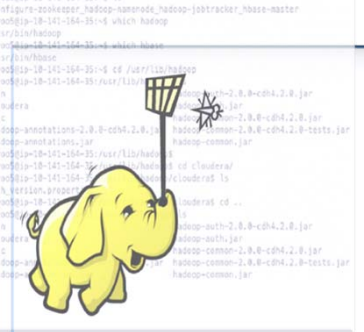


## Data Spillage in Hadoop Clouds: An Overview

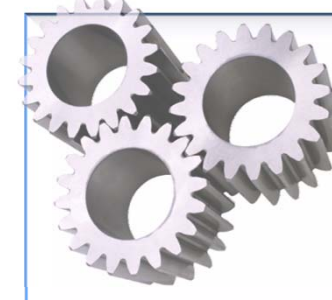
Adviser: Dr. Brandeis Marshall

Student: Dri Torres



### Problem Statement

What are possible Data Spill **solutions** in Hadoop, and can they **integrate** with current Data Spill procedure?



### Motivation

**Economic**, Data Spill recovery costs millions.  
**Address Security** downfalls in Hadoop



### Significance

To harness the full **potential** of Big Data, Data Spillage must be addressed

#### What is Data Spillage?

The transfer of classified data onto an unclassified information systems

Once classified data has leaked it is **dirty**

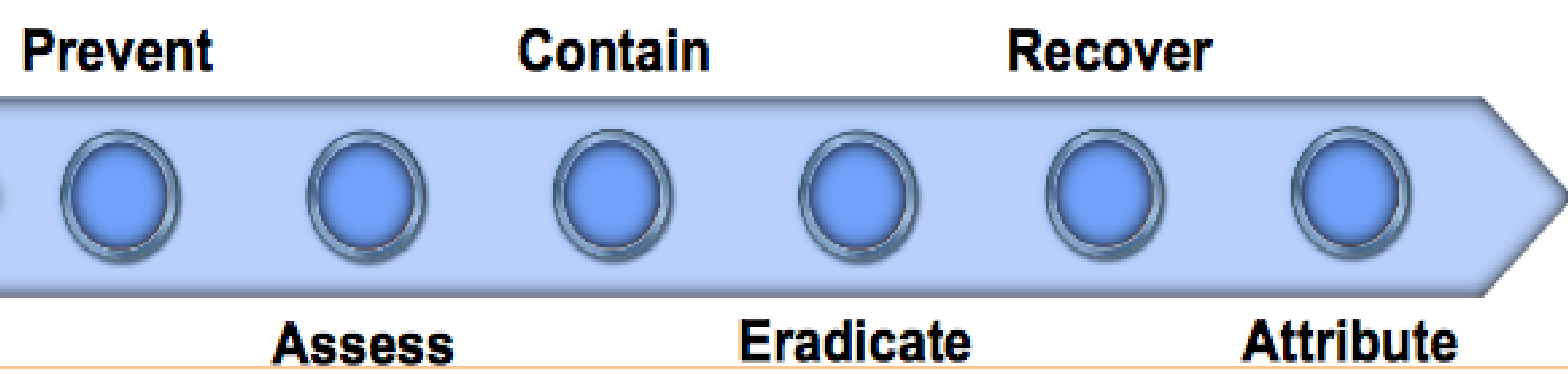
#### What is Big Data?

##### A Buzz Word

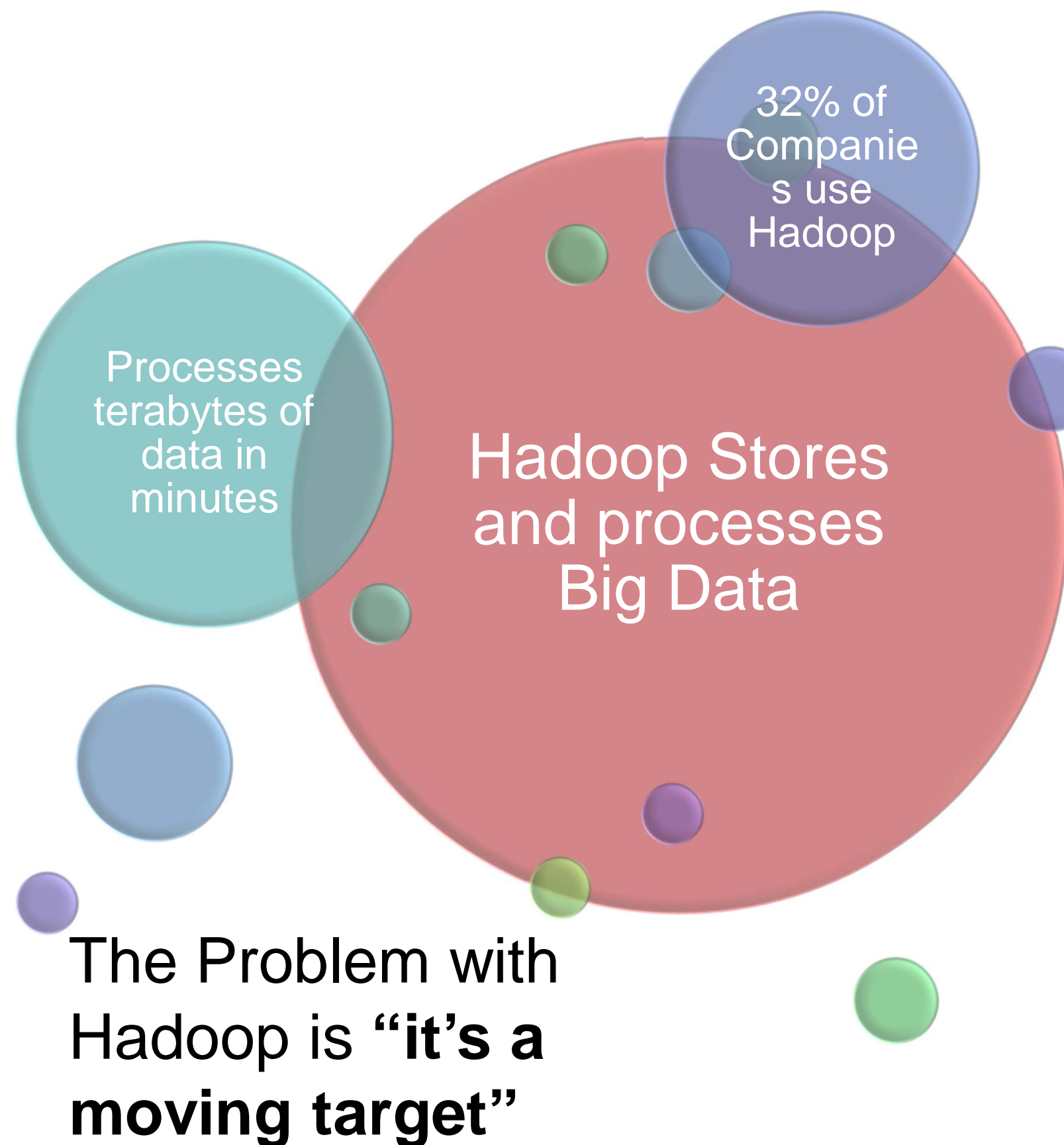
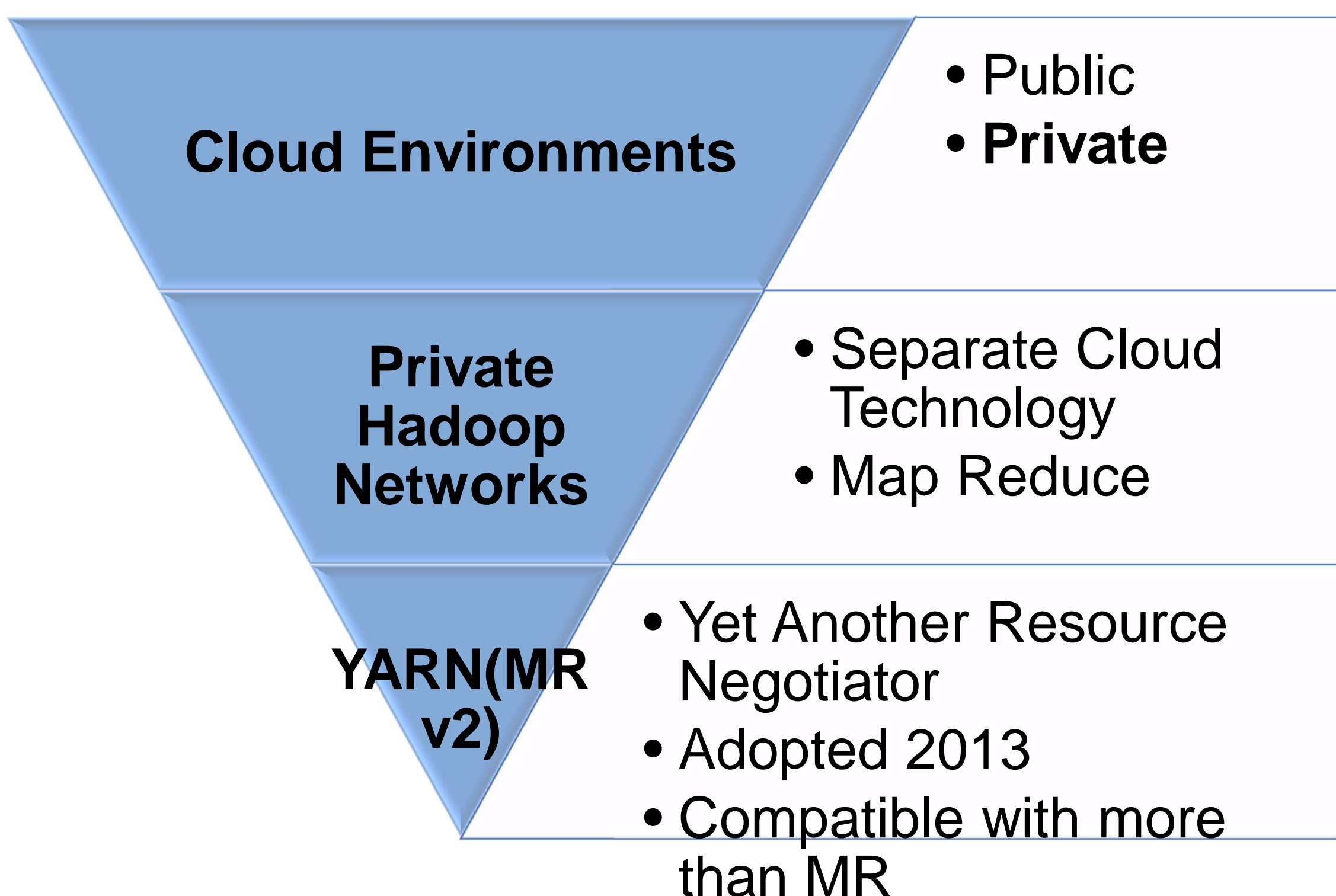
Less than 50% think it is clearly defined

Volume, Velocity, Variety (The 3 Vs)

#### NIST Procedure for Data Spillage



### Spillage and Cloud Computing



### Sample Cases for Analyzing Spills

- Long Term Storage
- Big Data Analytics
- Knowledge Systems for Metadata
- Inter-Agency Collaboration

### Solutions

#### A Top-Down Approach

Governments should establish an advanced Analytics Agency

Establish Big Data control centers

#### Standardization

Standardized Software and Hardware

Standard skilled and professional staff

#### Information Sharing

Leads to Information Securing

Always report a spill. Fully document and publish work

#### References

Gang-Hoon Kim, Silvana Trimi, and Ji-Hyong Chung. 2014. Big-data applications in the government sector. *Commun. ACM* 57, 3 (March 2014), 78-85. DOI=10.1145/2500873 <http://doi.acm.org/10.1145/2500873>

Lindner, Felix FX, and Sandro Gaycken. "Back to Basics: Beyond Network Hygiene." *Best Practices in Computer Network Defense: Incident Detection and Response* (2014): 54-64. Security Labs. IOS Press, 10 Feb. 2014. Web. 17 July 2014. <<http://recurity-labs.com>>.

Michael Stonebraker and Judy Robertson. 2013. Big data is 'buzzword du jour;' CS academics 'have the best job'. *Commun. ACM* 56, 9 (September 2013), 10-11. DOI=10.1145/2500468.2500471