



# Who am I

- **Scott Miao**
- RD, SPN Infra., Trend Micro
- Worked on Hadoop ecosystem since 2011
- Expertise in HDFS/MR/HBase
- [@takeshi.miao](https://twitter.com/takeshi.miao)



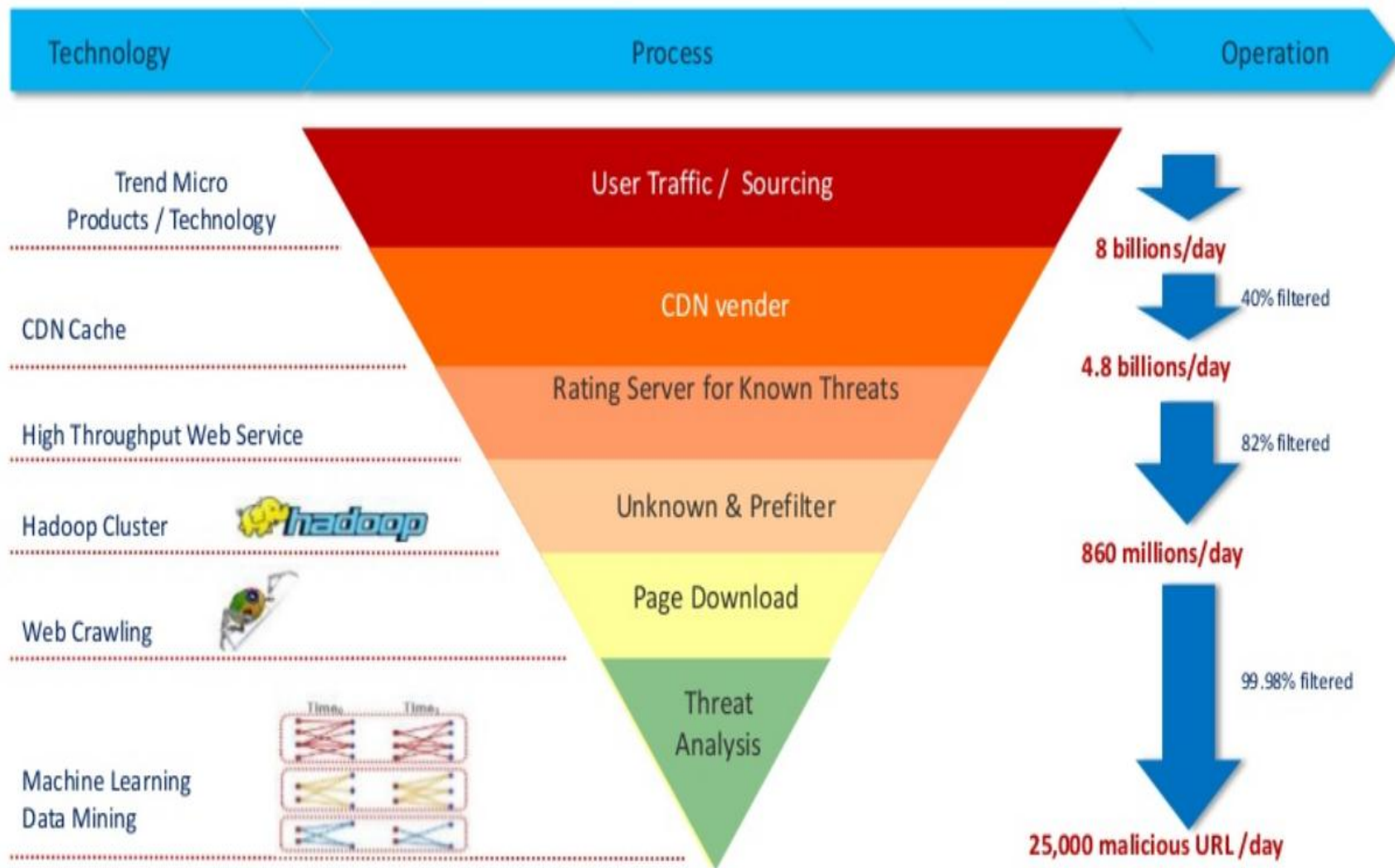
# Community Activities

- Contributor for Apache HBase/HDFS
- Speaker in conference
  - '[Zero-downtime Hadoop/HBase Cross-datacenter Migration](#)' at Hadoop Summit 2015 in Taipei
  - '[A Graph Service for Global Web Entities Traversal and Reputation Evaluation based on HBase](#)' at HBaseCon 2014 in San Francisco
  - '[A visualized cyber-threats entity reporting system backed with Hadoop ecosystem](#)' at China Hadoop Summit 2013 in Beijing
  - '[Rowkey design for Apache HBase](#)' in Hadoop Summit 2012 in Taipei

# What SPN team doing in Trend ?

# Web Reputation 8+ billions URL process daily

Clip slide



*Block malicious URL within 15 minutes once it goes online!*

# SPN Solution Architecture

Clip slide

Sourcing

Processing  
& Analysis

Validate &  
Create Solution

Quality Assurance

Solution  
Distribution

Solution  
Adoption

File

Web / URL

Email

Domain

IP

File Reputation Service

Web Reputation Service

Email Reputation Service

Smart Protection

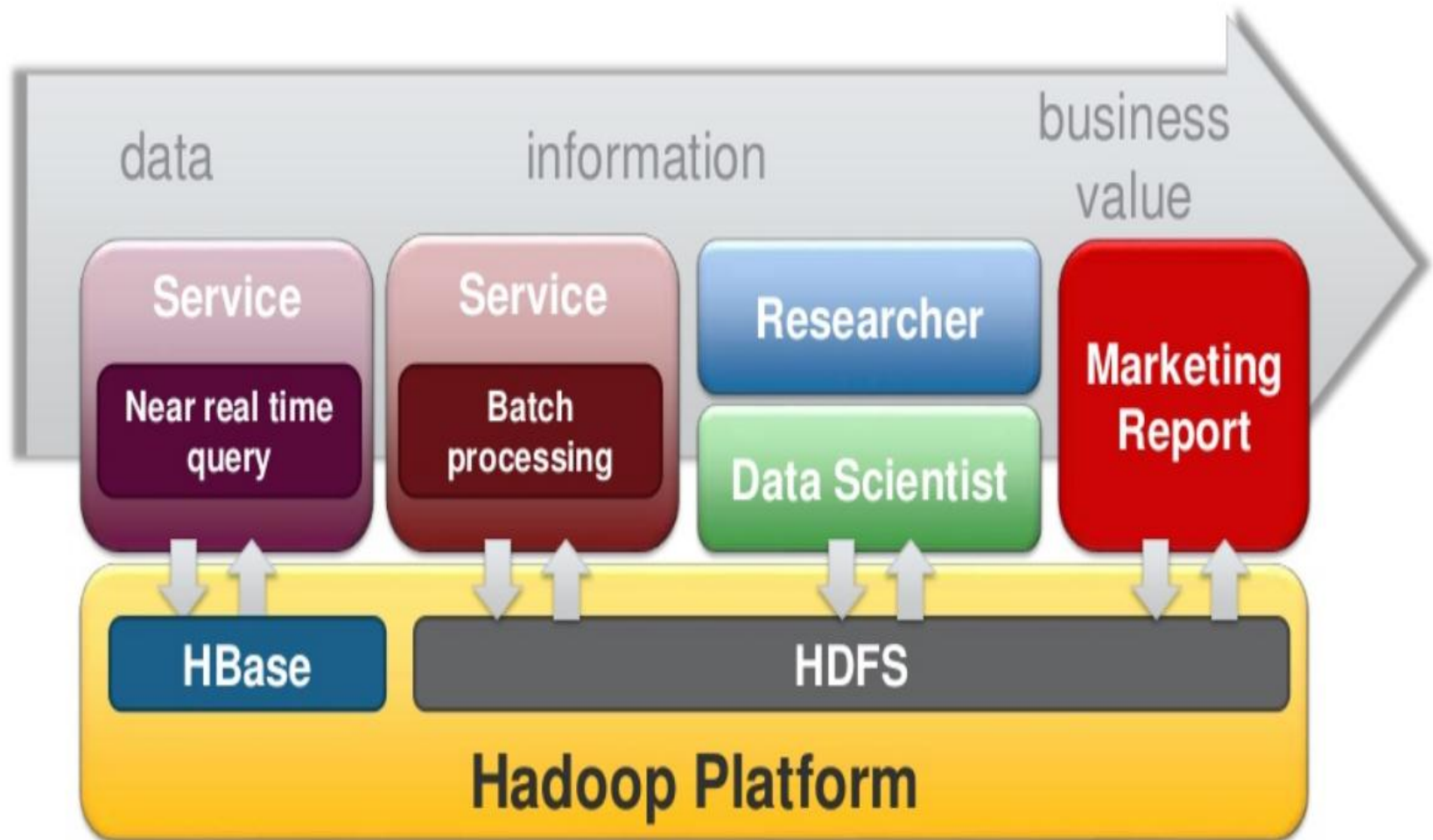
Customer

SPN Correlation



Community Intelligence  
(Feedback loop)

# SPN Hadoop Use Cases





# Yesterday

- ~40 Hadoop nodes
- ~15 Service/user accounts
- 3 Teams
- <50 TB storage
- <100 Jobs per day

# Today

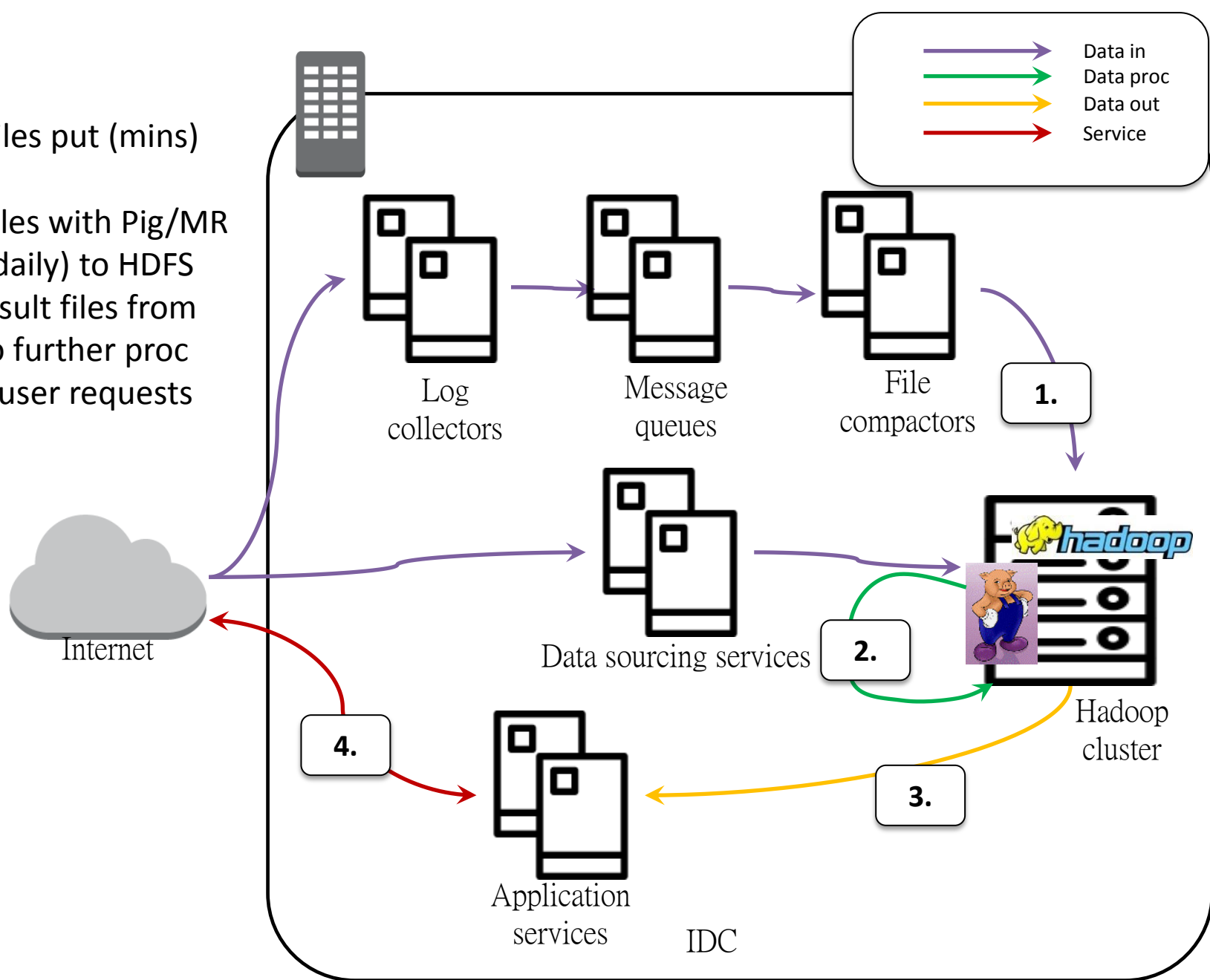
hundreds Hadoop nodes  
>170 Service/user accounts  
>13 Teams  
~1.5 PB storage  
>16000 Jobs per day



# What we doing in daily job ?

On HDFS

1. New files put (mins) to HDFS
2. Proc files with Pig/MR (hourly/daily) to HDFS
3. Get result files from HDFS, do further proc
4. Serve user requests



On HBase

1. New files put (mins) to HDFS
2. Proc files with Pig/MR (hourly/daily) to **HBase**
3. Random read from **HBase**
4. Serve user requests
5. Random writes to **HBase**

