MOVING MOUNTAINS OF PLAYER DATA

SCALABLE INTERNET SERVICES
UCLA/UCSB - NOV 2016

SEAN MALONEY
RIOT GAMES
@SEAN_SEANNERY
SMALONEY
@riotgames.com

WHO IS THIS GUY?

Lead developer on Riot's ETL and real-time services

FUN FACT:
Was a student in this class 5 years ago
Intern at Appfolio

SEAN MALONEY
BIG DATA ENGINEER

MOVING MOUNTAINS OF DATA

1. INTRODUCTION
2. THE GAME PLATFORM: OUR MAIN DATA SOURCE
3. HOW WE INGEST AND QUERY DATA
4. HOW WE SCALE IN AWS
5. CONCLUSION - SEAN'S PRO TIPS

INTRODUCTION

WHAT IS LEAGUE OF LEGENDS?

2009
LAUNCH

ONLINE
MULTIPLAYER

WINDOWS
/ OSX

40-50 MIN
GAMES

YOUR CHAMP

THE TEAM

THE BATTLE GROUND

REPLAY

12 BILLION

GAME RELATED EVENTS

0.5 TRILLION

DATA POINTS

50 TB

STORAGE

DAILY

26 PETABYTES

PLAYER DATA

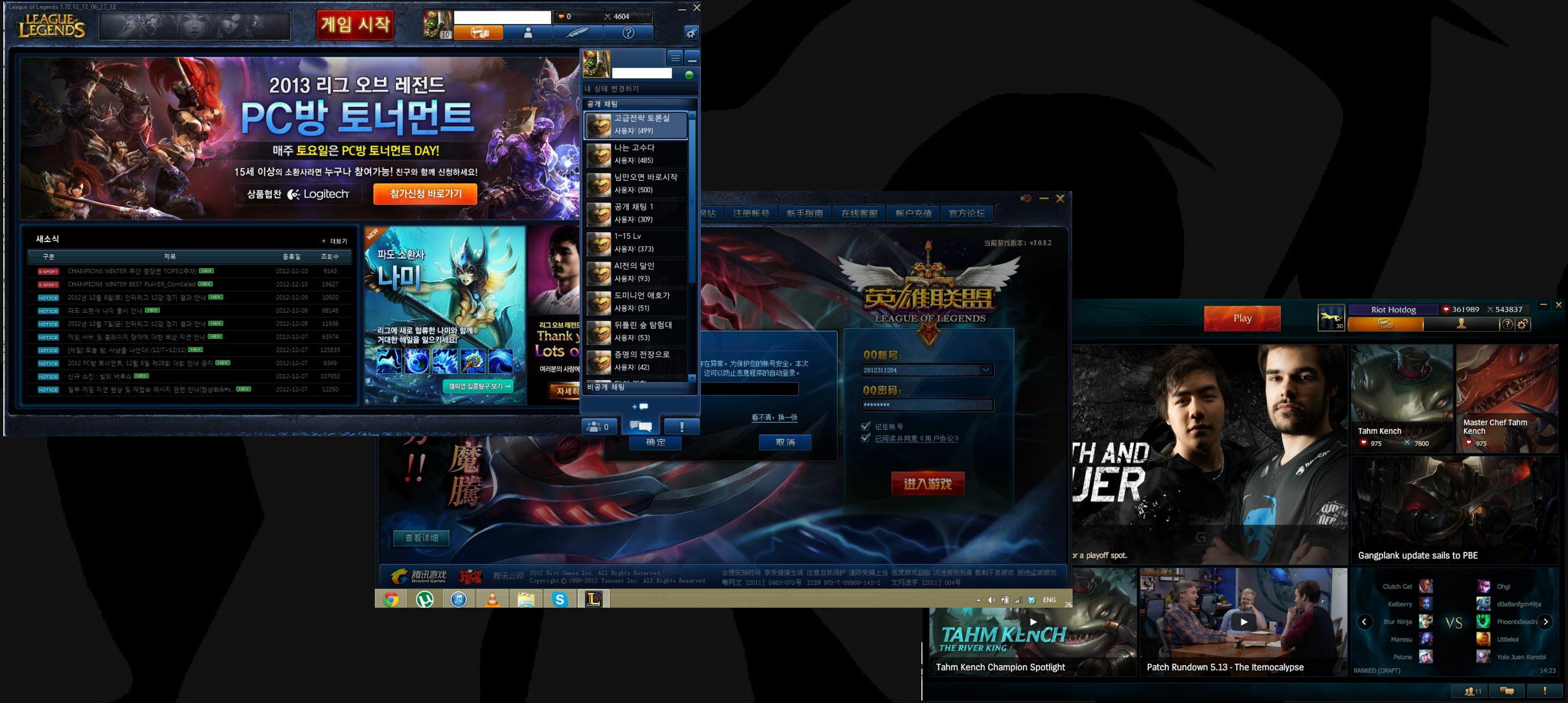
SINCE BETA

OUR MISSION

WE ASPIRE
TO BE THE MOST
PLAYER
FOCUSED
GAME COMPANY IN THE
WORLD

THE GAME PLATFORM

THE CLIENT.



League of Legends 1.72.12_12_06_17_12
LEAGUE of LEGENDS
게임 시작
0
4604
10
2013 리그 오브 레전드
PC방 토너먼트
매주 토요일은 PC방 토너먼트 DAY!
15세 이상의 소환사라면 누구나 참여가능! 친구와 함께 신청하세요!
상품협찬 Logitech
참가신청 바로가기
NEW
파도 소환사
나미
리그에 새로 합류한 나미와 함께 거대한 해일을 일으키세요!
챔피언 집중탐구 보기 →
Thank y
Lots o
여러분의 사랑에
자세히
내 상태 변경하기
공개 채팅
고급전략 토론실 사용자: (499)
나는 고수다 사용자: (485)
님만오면 바로시작 사용자: (500)
공개 채팅 1 사용자: (309)
1~15 Lv 사용자: (373)
AI전의 달인 사용자: (93)
도미니언 애호가 사용자: (51)
뒤틀린 숲 탐험대 사용자: (53)
증명의 전장으로 사용자: (42)
비공개 채팅
0
网站
注册帐号
新手指南
在线客服
帐户充值
官方论坛
当前游戏版本：v3.0.8.2
英雄联盟
LEAGUE OF LEGENDS
魔
腾
查看详细
看不清，换一张
确 定
取 消
进入游戏
腾讯游戏 Tencent Games
RIOT
腾讯公司
2012 Riot Games Inc. All Rights Reserved
Copyright © 1998-2012 Tencent Inc. All Rights Reserved
合理安排时间 享受健康生活 注意自我保护 谨防受骗上当 适度游戏益脑 沉迷游戏伤身 抵制不良游戏 拒绝盗版游戏
粤网文［2011］0483-070号 ISBN 978-7-89989-145-2 文网进字［2011］004号
ENG
Play
Riot Hotdog
361989
543837
30
TH AND
JER
or a playoff spot.
Tahm Kench
975
7800
Master Chef Tahm Kench
975
Gangplank update sails to PBE
TAHM KENCH
THE RIVER KING
Tahm Kench Champion Spotlight
Patch Rundown 5.13 - The Itemocalypse
Clutch Cat
Kelberry
Blur Ninja
Maresu
Pslune
VS
Ohgl
d0a8snfgm49ja
PhoenixSeadr
Littlekoi
Yolo Juan Kenobi
RANKED (DRAFT)
14:23
11

새소식 + 더보기

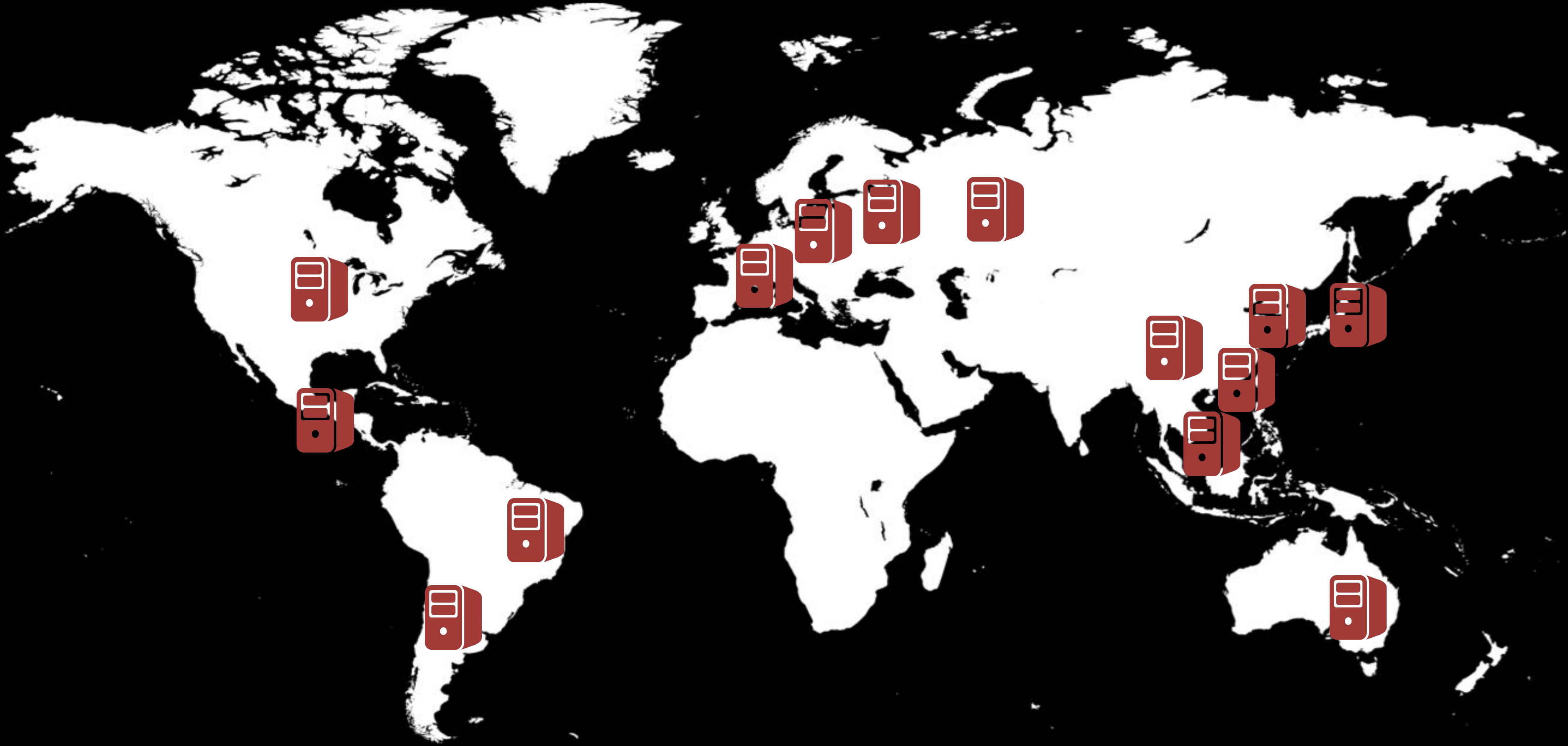
구분	제목	등록일	조회수
E-SPORT	CHAMPIONS WINTER 주간 명장면 TOP5(2주차) NEW	2012-12-10	9143
E-SPORT	CHAMPIONS WINTER BEST PLAYER_CornSalad NEW	2012-12-10	19627
NOTICE	2012년 12월 8일(토) 인터리그 12강 경기 결과 안내 NEW	2012-12-09	10920
NOTICE	파도 소환사 나미 출시 안내 NEW	2012-12-08	68148
NOTICE	2012년 12월 7일(금) 인터리그 12강 경기 결과 안내 NEW	2012-12-08	11536
NOTICE	게임 서버 및 홈페이지 장애에 대한 보상 지연 안내 NEW	2012-12-07	63574
NOTICE	[세일] 오늘 밤, 사냥을 나선다! (12/7~12/11) NEW	2012-12-07	125833
NOTICE	2012 PC방 토너먼트, 12월 8일 제28회 대회 안내 공지 NEW	2012-12-07	6349
NOTICE	신규 스킨 : 빛의 바루스 NEW	2012-12-07	107650
NOTICE	일부 게임 지연 현상 및 재접속 메시지 관련 안내(정상화&#x... NEW	2012-12-07	12250

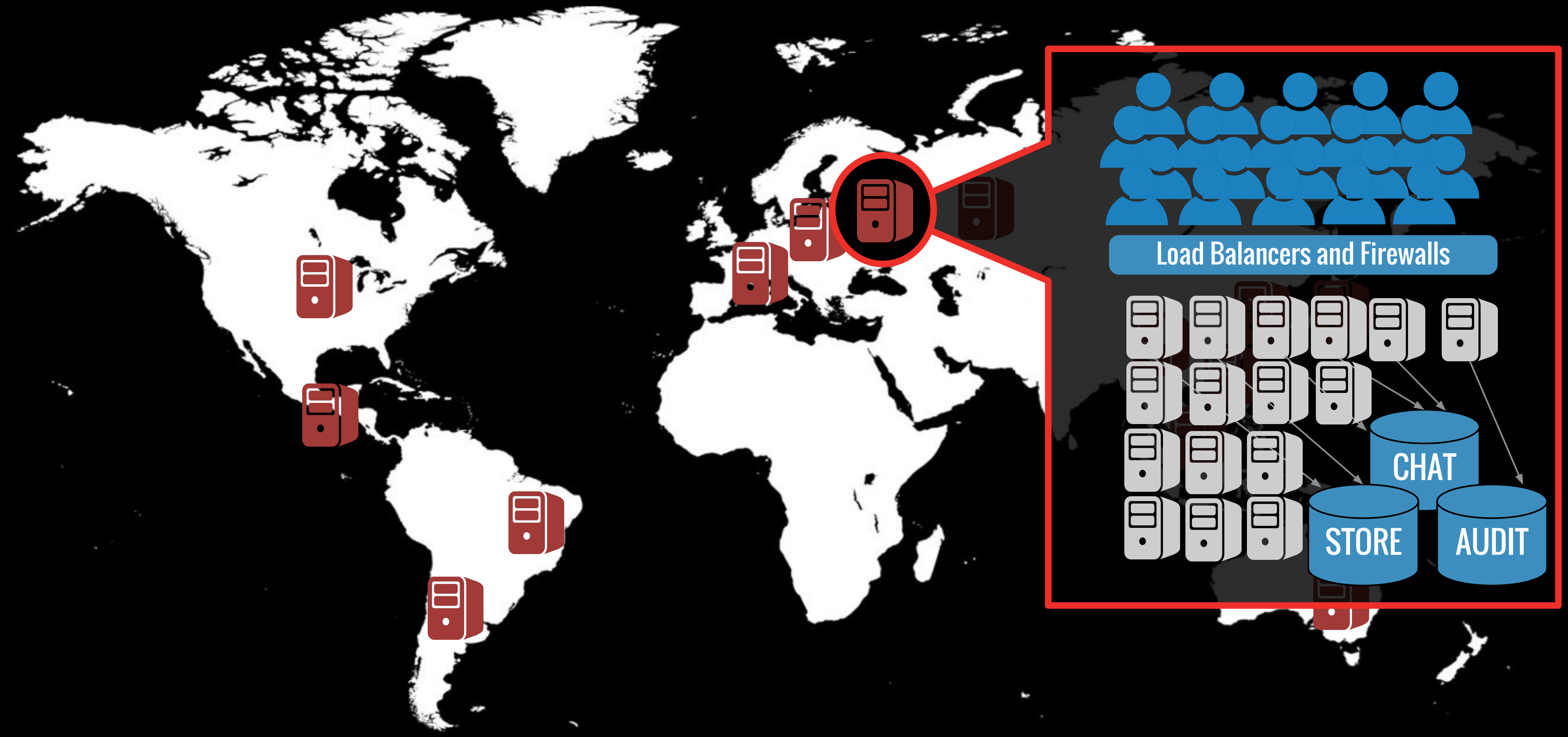
存在异常。为保护您的帐号安全，本次
这可以防止恶意程序的自动登录。

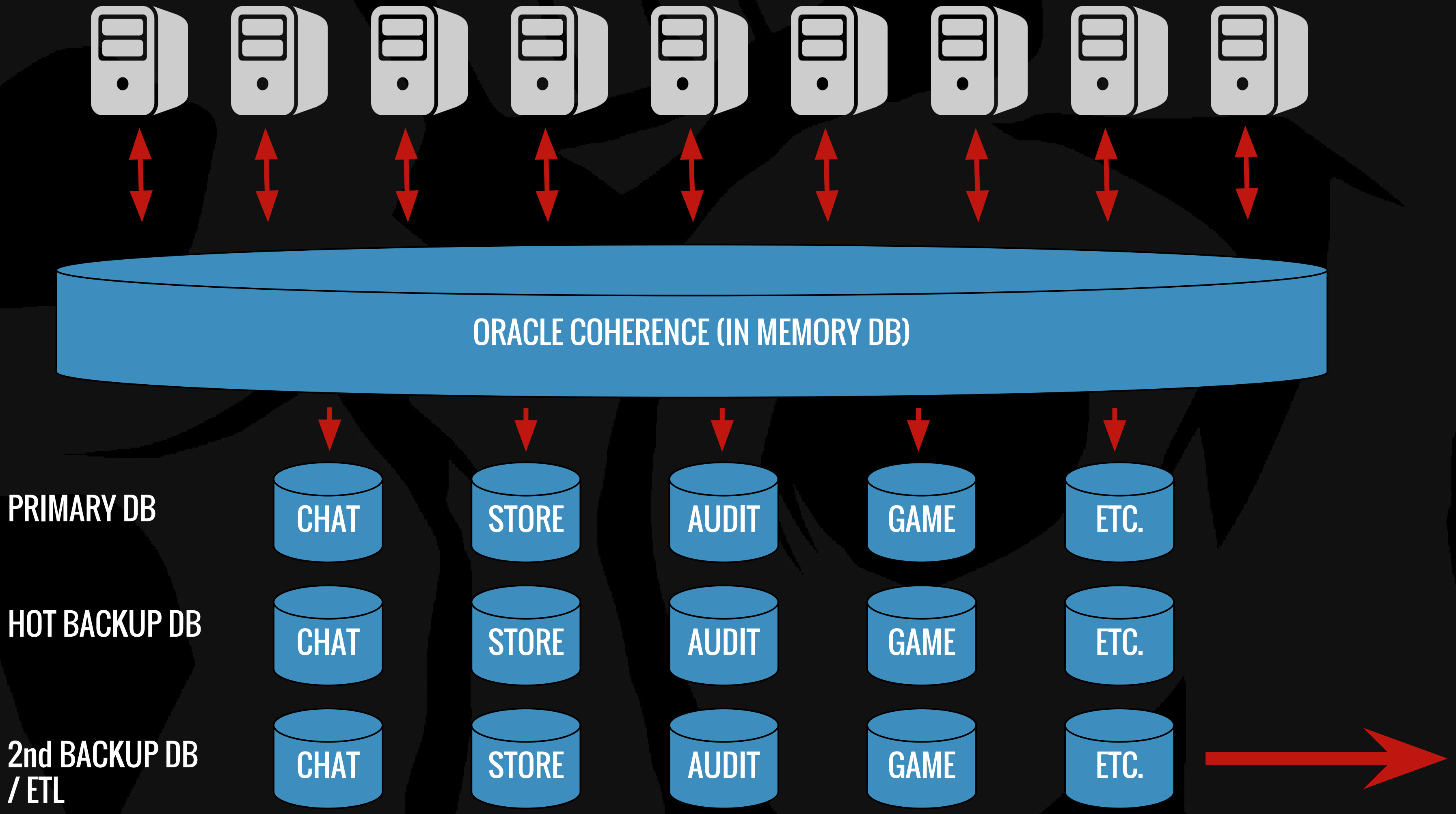
QQ帐号：2812311284

QQ密码：*********

☑ 记住帐号
☑ 已阅读并同意《用户协议》

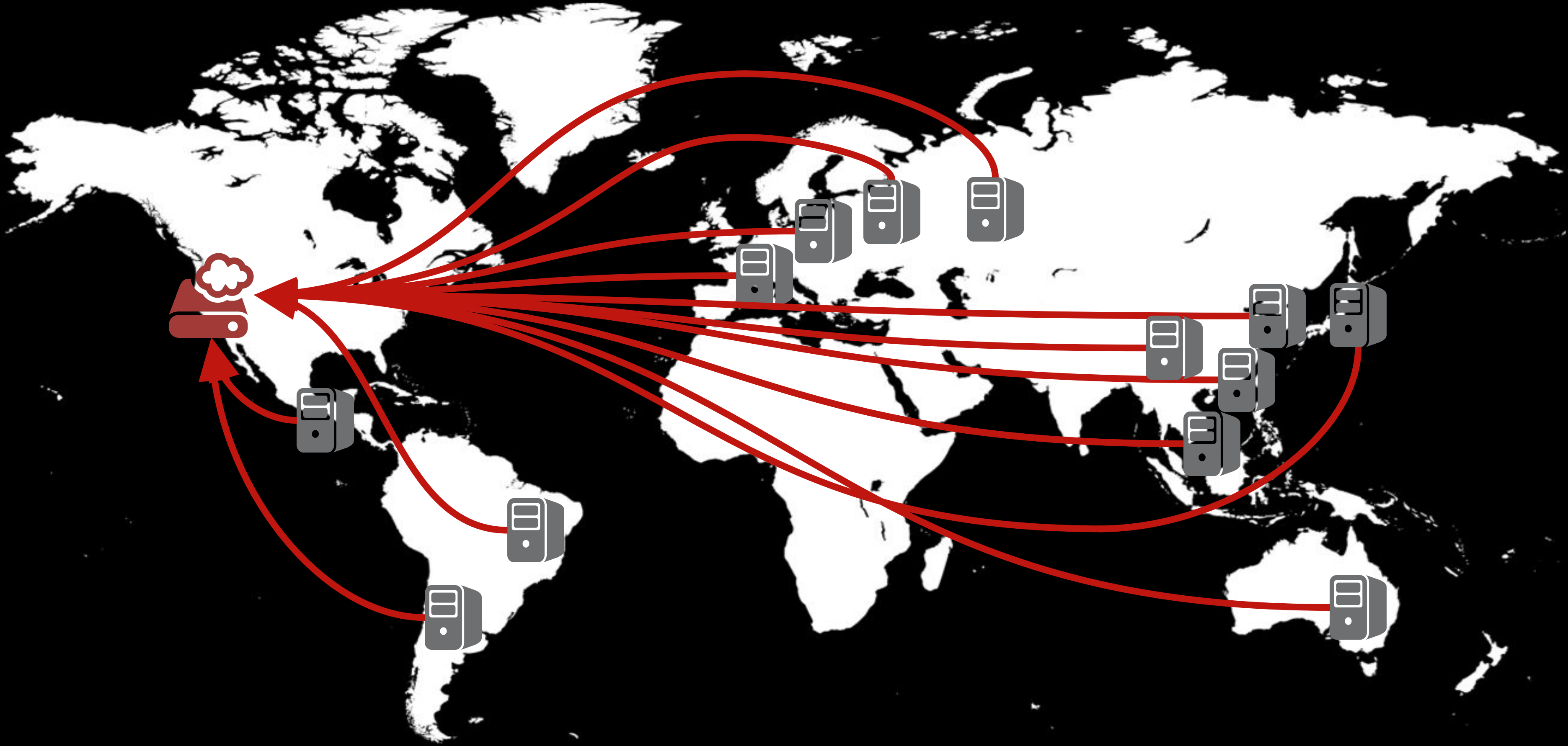



Load Balancers and Firewalls
CHAT
STORE
AUDIT



ORACLE COHERENCE (IN MEMORY DB)

PRIMARY DB	CHAT	STORE	AUDIT	GAME	ETC.
HOT BACKUP DB	CHAT	STORE	AUDIT	GAME	ETC.
2nd BACKUP DB / ETL	CHAT	STORE	AUDIT	GAME	ETC.



DATA INGESTION

INGESTION

PULL-BASED / ETL

FU ETL

FuETL
- OLTP game data
- External Data Sources

PUSH-BASED

kafka

- Anything pushed to it
- Server logs

STORAGE

MASTER WAREHOUSE

amazon web services™
S3 Simple Storage Service

DATA AUDITING

Wasp

QUERY / VIEWS

AGGREGATE QUERIES

VERTICA
An HP Company

BATCH QUERIES

HIVE

SINGLE-ROW QUERIES

amazon DynamoDB

VIZ. TOOLS

+ a b | e a u

platfora
You should know.

INGESTION

PULL-BASED / ETL

FuETL

FuETL
- OLTP game data
- External Data Sources

PUSH-BASED

kafka

- Anything pushed to it
- Server logs

STORAGE

MASTER WAREHOUSE

amazon webservices™
S3 Simple Storage Service

DATA AUDITING

Wasp

QUERY / VIEWS

AGGREGATE QUERIES

VERTICA
An HP Company

BATCH QUERIES

HIVE

SINGLE-ROW QUERIES

amazon DynamoDB

VIZ. TOOLS

+ a b | e a u

platfora
You should know.

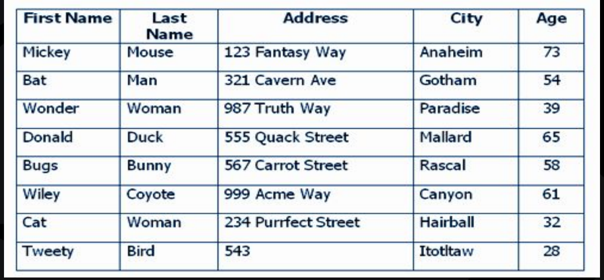
Distributed ETL Software written in Ruby.

Same ETL applied to multiple regions / datacenters

Scales Horizontally

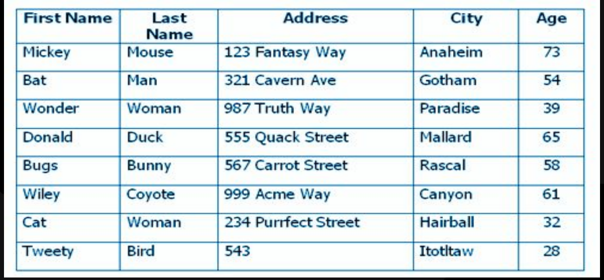
BEST LOGO EVER!

NA



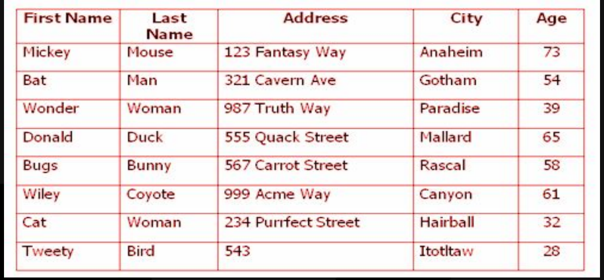
First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

Korea

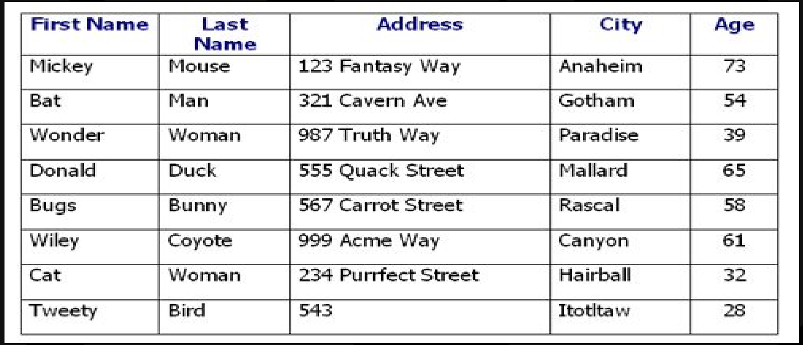


First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

Russia



First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28



First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

OTHER DATA SOURCES

LOG

<REST>

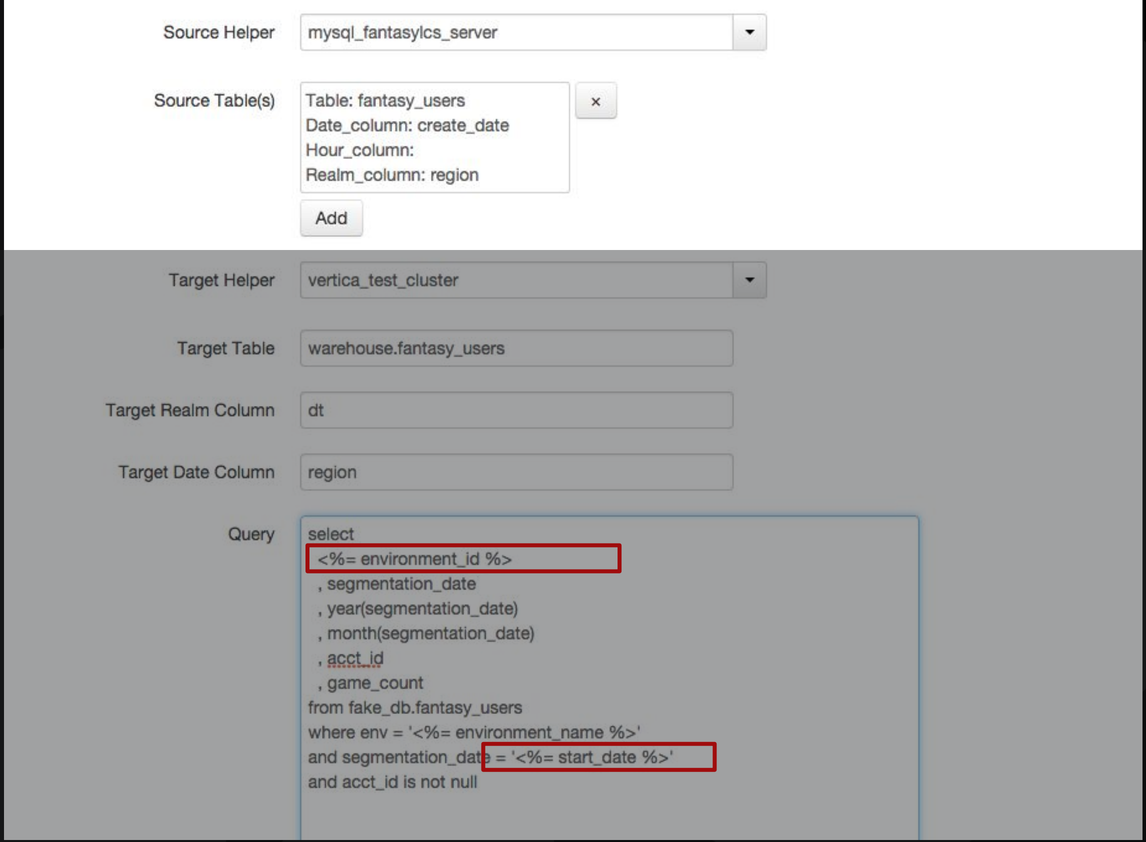
twitch

You Tube

FUETL CAN CONNECT TO

- Amazon S3
- SQS
- (S)FTP
- Hive
- Microsoft SQL Server
- MySQL
- DynamoDB
- Vertica
- Redshift
- REST websites

Create an ETL



Source Helper: mysql_fantasylcs_server

Source Table(s):
Table: fantasy_users
Date_column: create_date
Hour_column:
Realm_column: region

Target Helper: vertica_test_cluster

Target Table: warehouse.fantasy_users

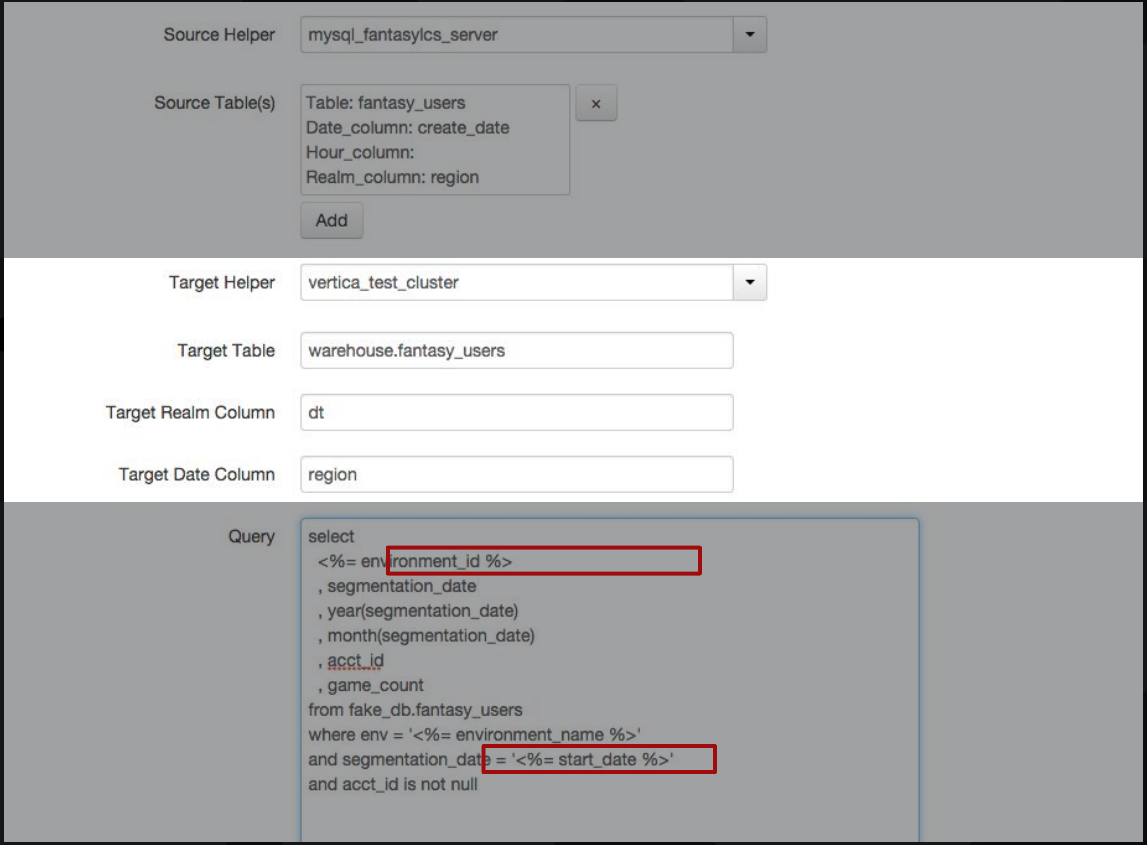
Target Realm Column: dt

Target Date Column: region

Query:

```
select
 <%= environment_id %>
 , segmentation_date
 , year(segmentation_date)
 , month(segmentation_date)
 , acct_id
 , game_count
from fake_db.fantasy_users
where env = '<%= environment_name %>'
and segmentation_date = '<%= start_date %>'
and acct_id is not null
```

Create an ETL

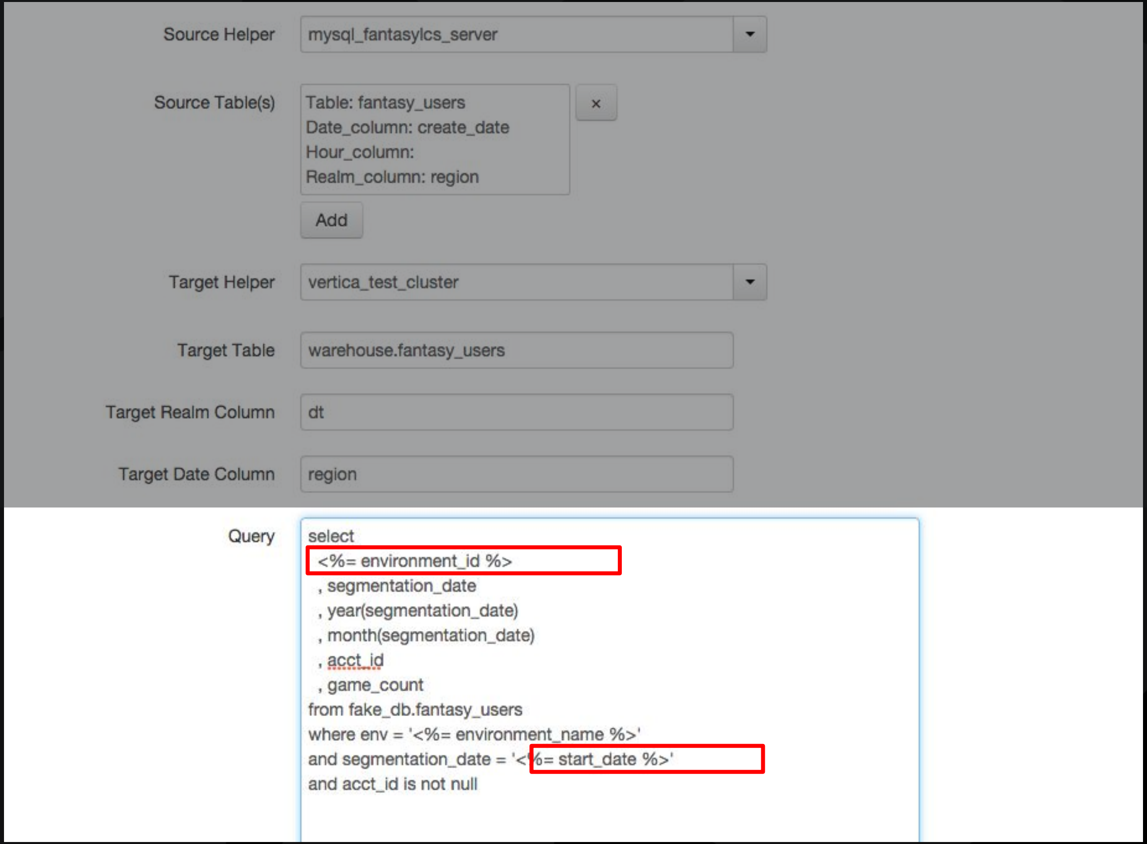


Source Helper	mysql_fantasylcs_server
Source Table(s)	Table: fantasy_users Date_column: create_date Hour_column: Realm_column: region
Target Helper	vertica_test_cluster
Target Table	warehouse.fantasy_users
Target Realm Column	dt
Target Date Column	region

Query

```
select
  <%= environment_id %>
  , segmentation_date
  , year(segmentation_date)
  , month(segmentation_date)
  , acct_id
  , game_count
from fake_db.fantasy_users
where env = '<%= environment_name %>'
and segmentation_date = '<%= start_date %>'
and acct_id is not null
```

Create an ETL



Source Helper: mysql_fantasylcs_server

Source Table(s):
Table: fantasy_users
Date_column: create_date
Hour_column:
Realm_column: region

Target Helper: vertica_test_cluster

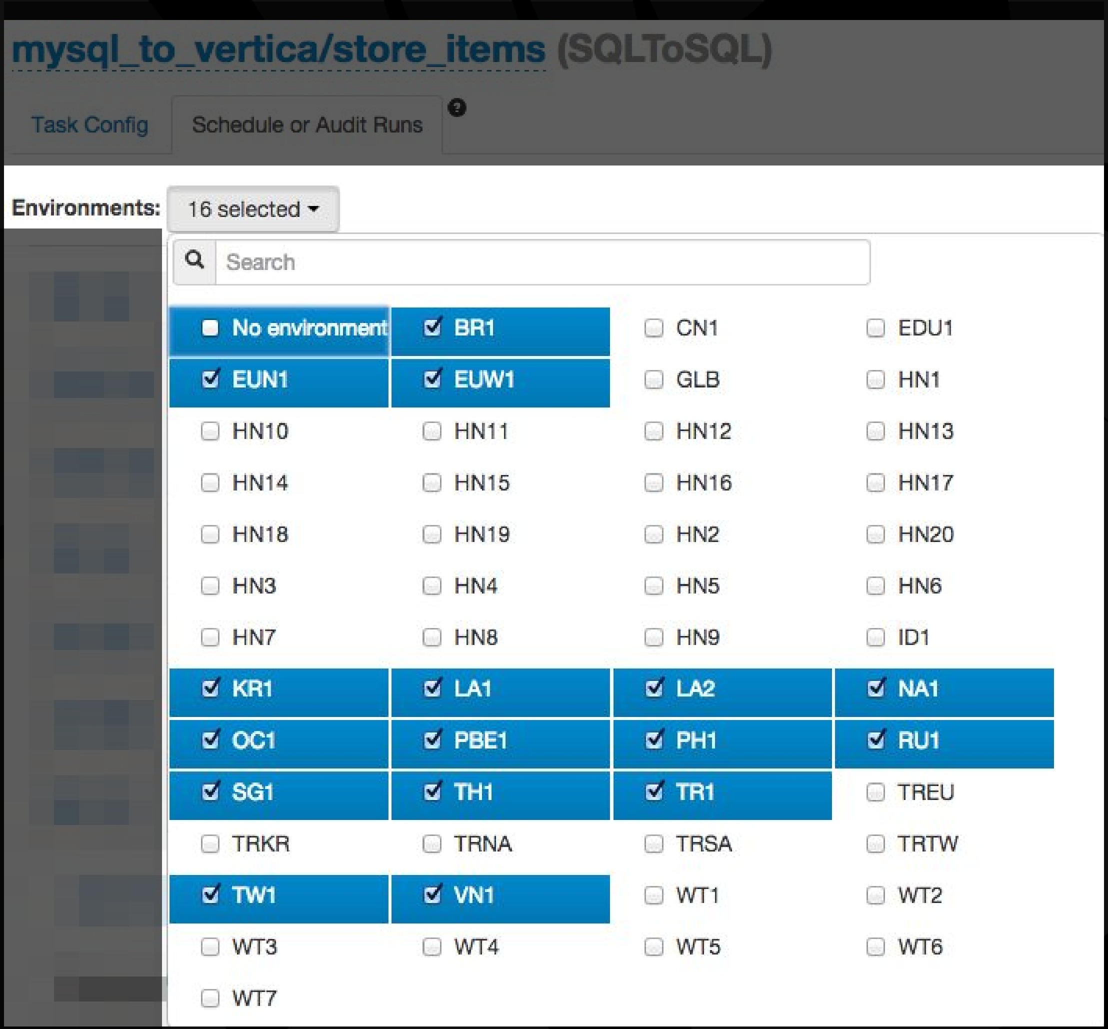
Target Table: warehouse.fantasy_users

Target Realm Column: dt

Target Date Column: region

Query:

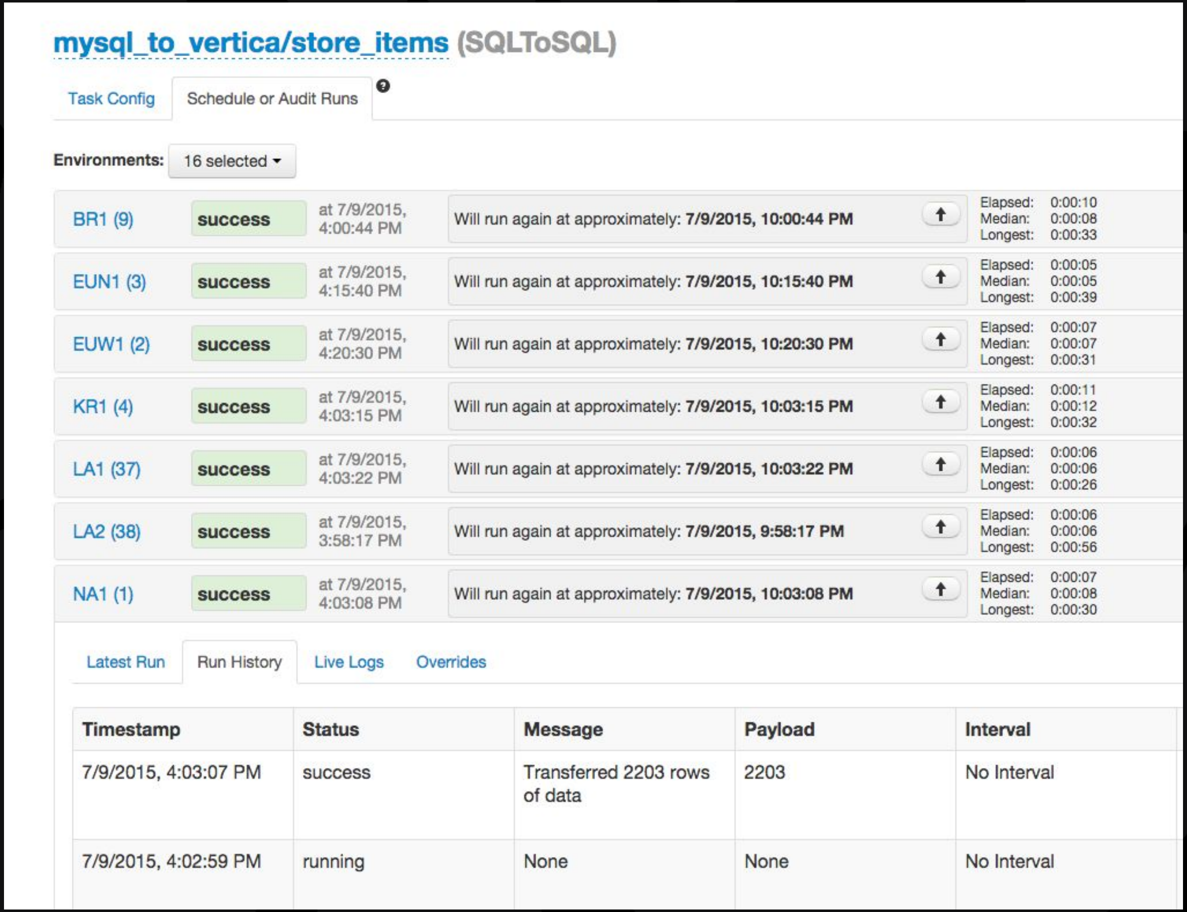
```
select
 <%= environment_id %>
 , segmentation_date
 , year(segmentation_date)
 , month(segmentation_date)
 , acct_id
 , game_count
from fake_db.fantasy_users
where env = '<%= environment_name %>'
and segmentation_date = '<%= start_date %>'
and acct_id is not null
```



mysql_to_vertica/store_items (SQLToSQL)

Environments: 16 selected

□ No environment ☑ BR1 □ CN1 □ EDU1
☑ EUN1 ☑ EUW1 □ GLB □ HN1
□ HN10 □ HN11 □ HN12 □ HN13
□ HN14 □ HN15 □ HN16 □ HN17
□ HN18 □ HN19 □ HN2 □ HN20
□ HN3 □ HN4 □ HN5 □ HN6
□ HN7 □ HN8 □ HN9 □ ID1
☑ KR1 ☑ LA1 ☑ LA2 ☑ NA1
☑ OC1 ☑ PBE1 ☑ PH1 ☑ RU1
☑ SG1 ☑ TH1 ☑ TR1 □ TREU
□ TRKR □ TRNA □ TRSA □ TRTW
☑ TW1 ☑ VN1 □ WT1 □ WT2
□ WT3 □ WT4 □ WT5 □ WT6
□ WT7

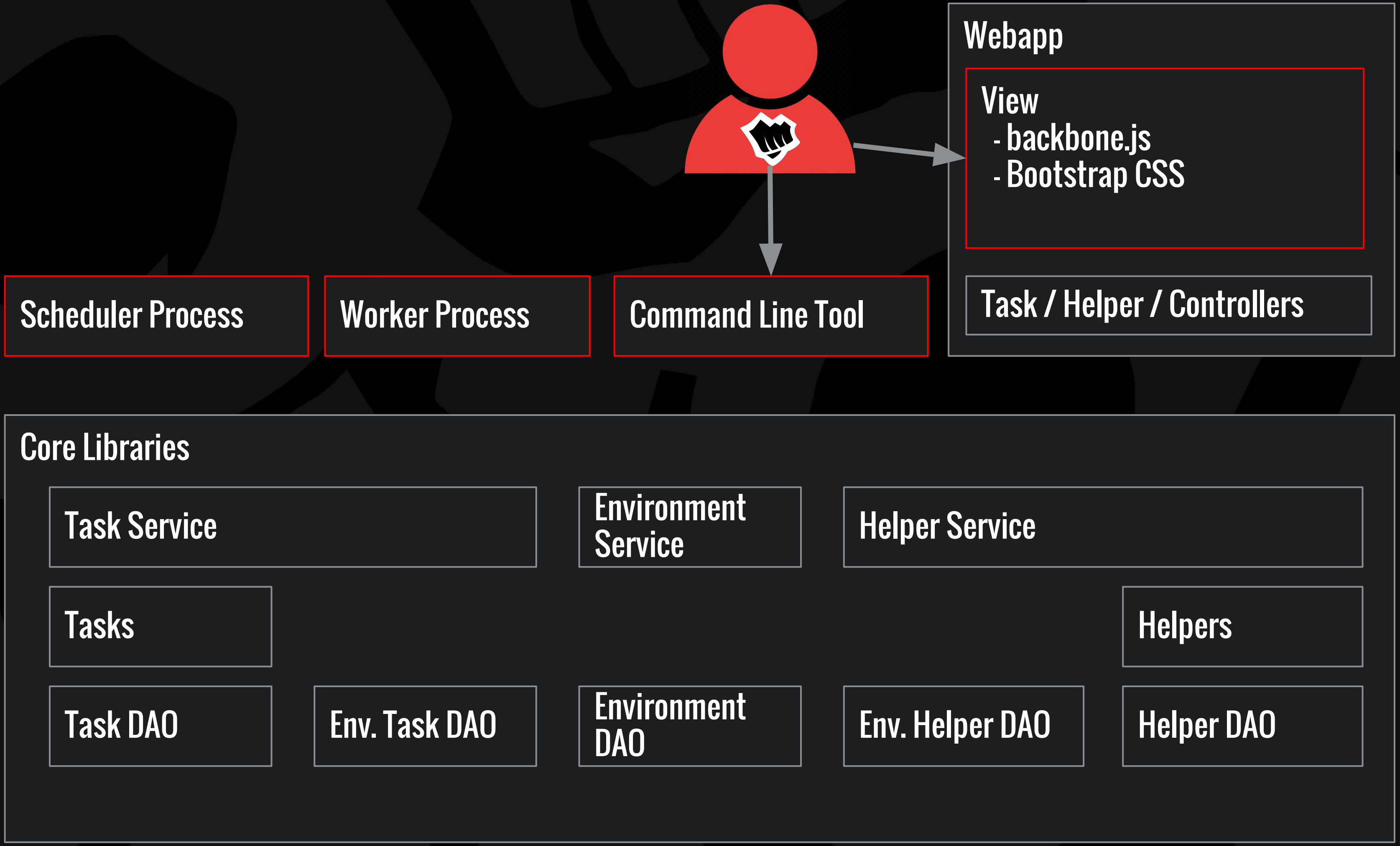


mysql_to_vertica/store_items (SQLToSQL)

Environments: 16 selected

BR1 (9)	success	at 7/9/2015, 4:00:44 PM	Will run again at approximately: **7/9/2015, 10:00:44 PM**	Elapsed: 0:00:10 Median: 0:00:08 Longest: 0:00:33
EUN1 (3)	success	at 7/9/2015, 4:15:40 PM	Will run again at approximately: **7/9/2015, 10:15:40 PM**	Elapsed: 0:00:05 Median: 0:00:05 Longest: 0:00:39
EUW1 (2)	success	at 7/9/2015, 4:20:30 PM	Will run again at approximately: **7/9/2015, 10:20:30 PM**	Elapsed: 0:00:07 Median: 0:00:07 Longest: 0:00:31
KR1 (4)	success	at 7/9/2015, 4:03:15 PM	Will run again at approximately: **7/9/2015, 10:03:15 PM**	Elapsed: 0:00:11 Median: 0:00:12 Longest: 0:00:32
LA1 (37)	success	at 7/9/2015, 4:03:22 PM	Will run again at approximately: **7/9/2015, 10:03:22 PM**	Elapsed: 0:00:06 Median: 0:00:06 Longest: 0:00:26
LA2 (38)	success	at 7/9/2015, 3:58:17 PM	Will run again at approximately: **7/9/2015, 9:58:17 PM**	Elapsed: 0:00:06 Median: 0:00:06 Longest: 0:00:56
NA1 (1)	success	at 7/9/2015, 4:03:08 PM	Will run again at approximately: **7/9/2015, 10:03:08 PM**	Elapsed: 0:00:07 Median: 0:00:08 Longest: 0:00:30

Timestamp	Status	Message	Payload	Interval
7/9/2015, 4:03:07 PM	success	Transferred 2203 rows of data	2203	No Interval
7/9/2015, 4:02:59 PM	running	None	None	No Interval



Webapp

View
- backbone.js
- Bootstrap CSS

Task / Helper / Controllers

Scheduler Process

Worker Process

Command Line Tool

Core Libraries

Task Service

Environment Service

Helper Service

Tasks

Helpers

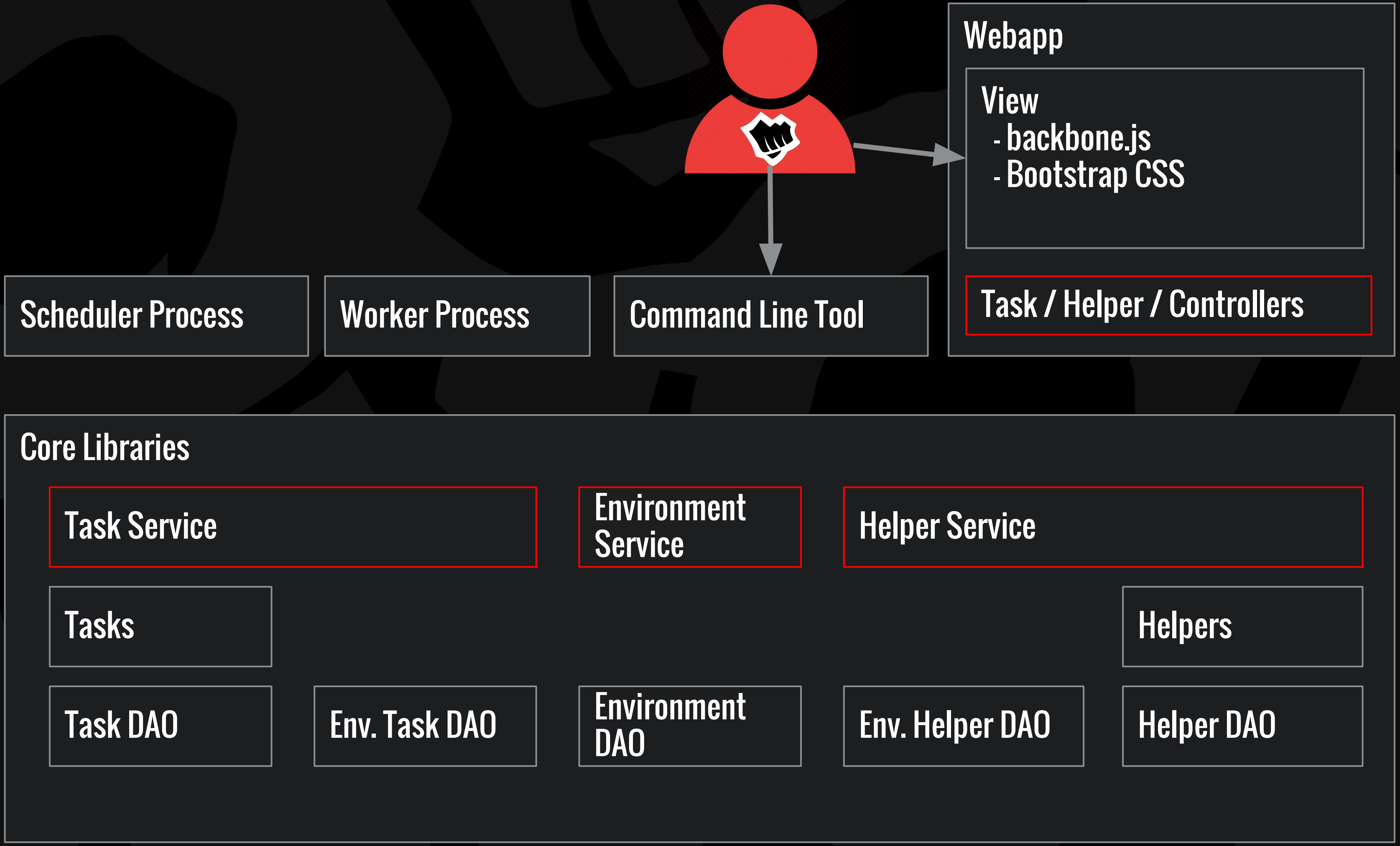
Task DAO

Env. Task DAO

Environment DAO

Env. Helper DAO

Helper DAO



Webapp

View
- backbone.js
- Bootstrap CSS

Task / Helper / Controllers

Scheduler Process

Worker Process

Command Line Tool

Core Libraries

Task Service

Environment Service

Helper Service

Tasks

Helpers

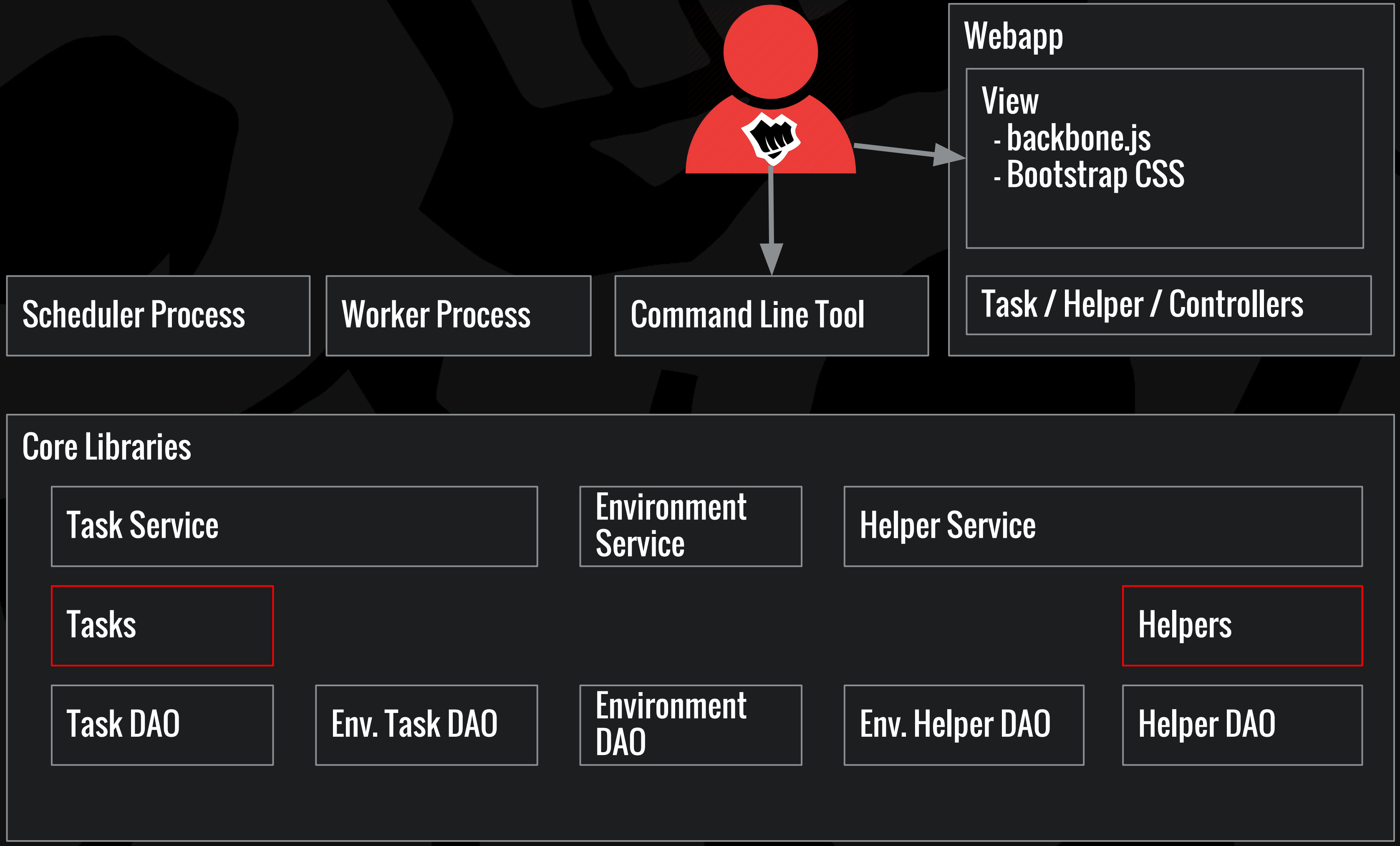
Task DAO

Env. Task DAO

Environment DAO

Env. Helper DAO

Helper DAO



Webapp

View
- backbone.js
- Bootstrap CSS

Task / Helper / Controllers

Scheduler Process

Worker Process

Command Line Tool

Core Libraries

Task Service

Environment Service

Helper Service

Tasks

Helpers

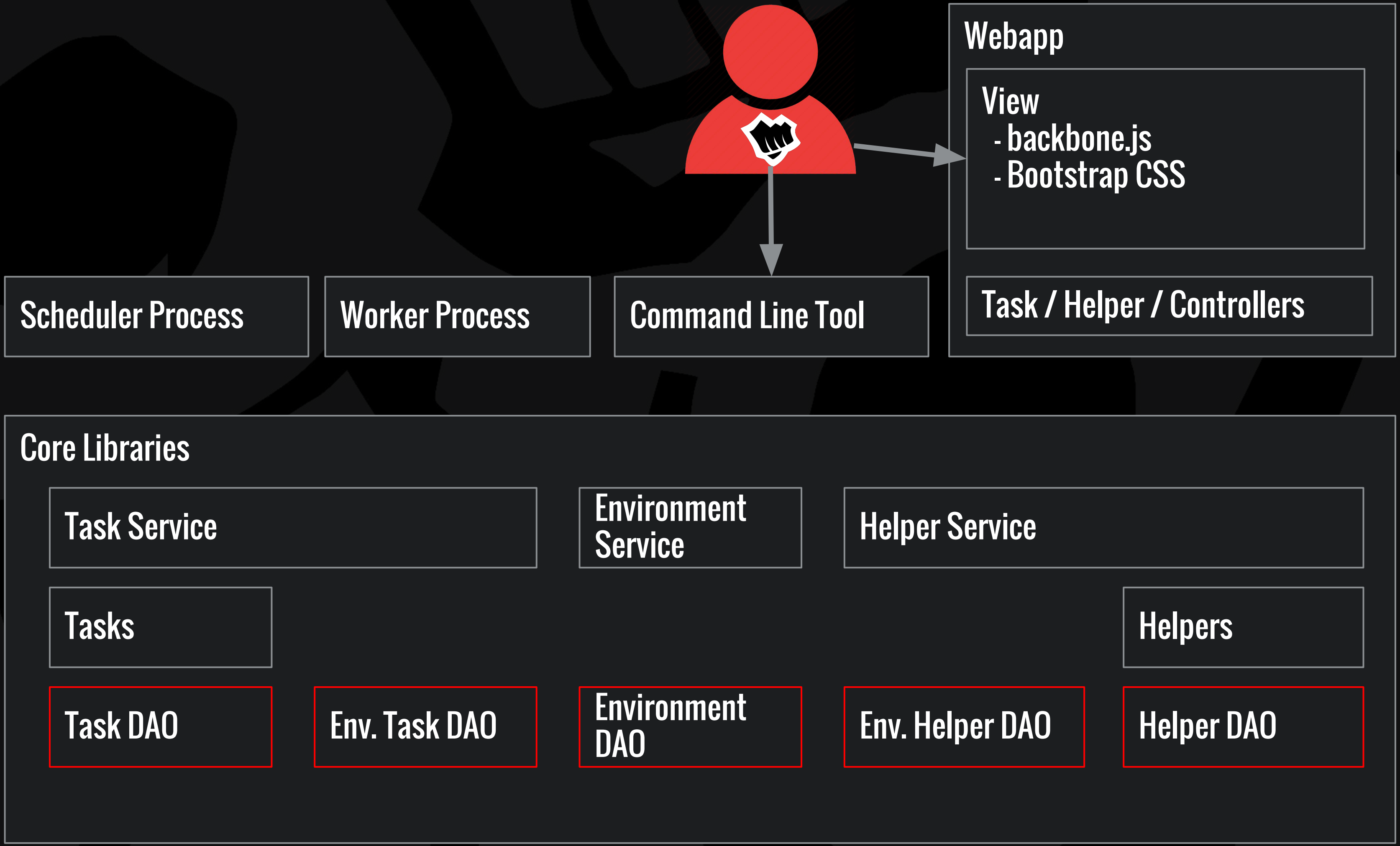
Task DAO

Env. Task DAO

Environment DAO

Env. Helper DAO

Helper DAO



Webapp

View
- backbone.js
- Bootstrap CSS

Task / Helper / Controllers

Scheduler Process

Worker Process

Command Line Tool

Core Libraries

Task Service

Environment Service

Helper Service

Tasks

Helpers

Task DAO

Env. Task DAO

Environment DAO

Env. Helper DAO

Helper DAO

FuETL STATISTICS

5213

ACTIVE REGIONAL ETLS

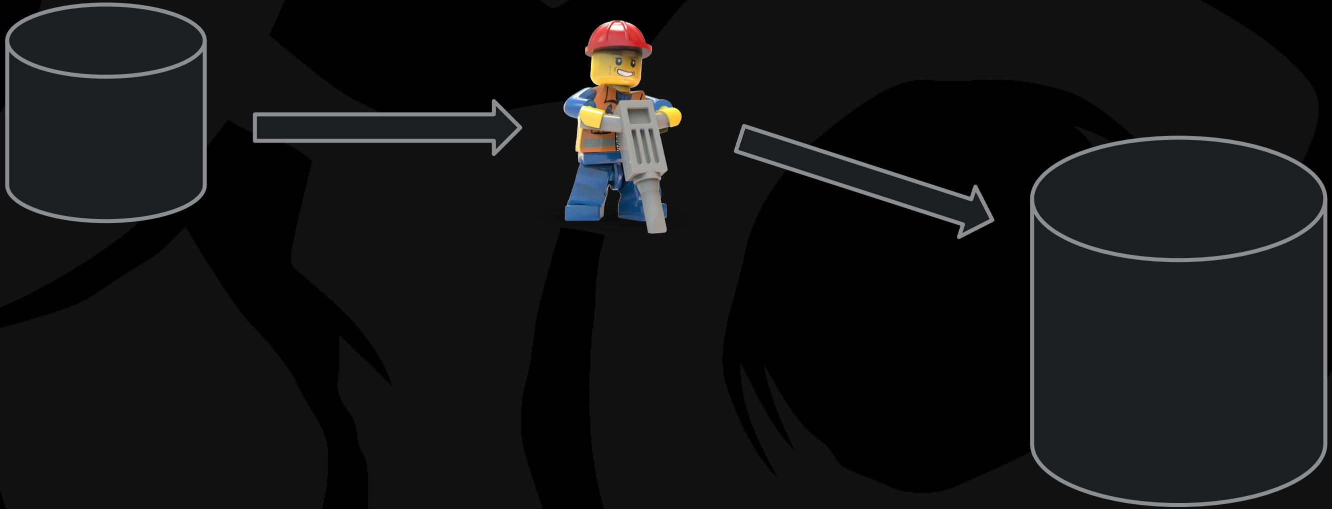
23125

DAILY ETL RUNS

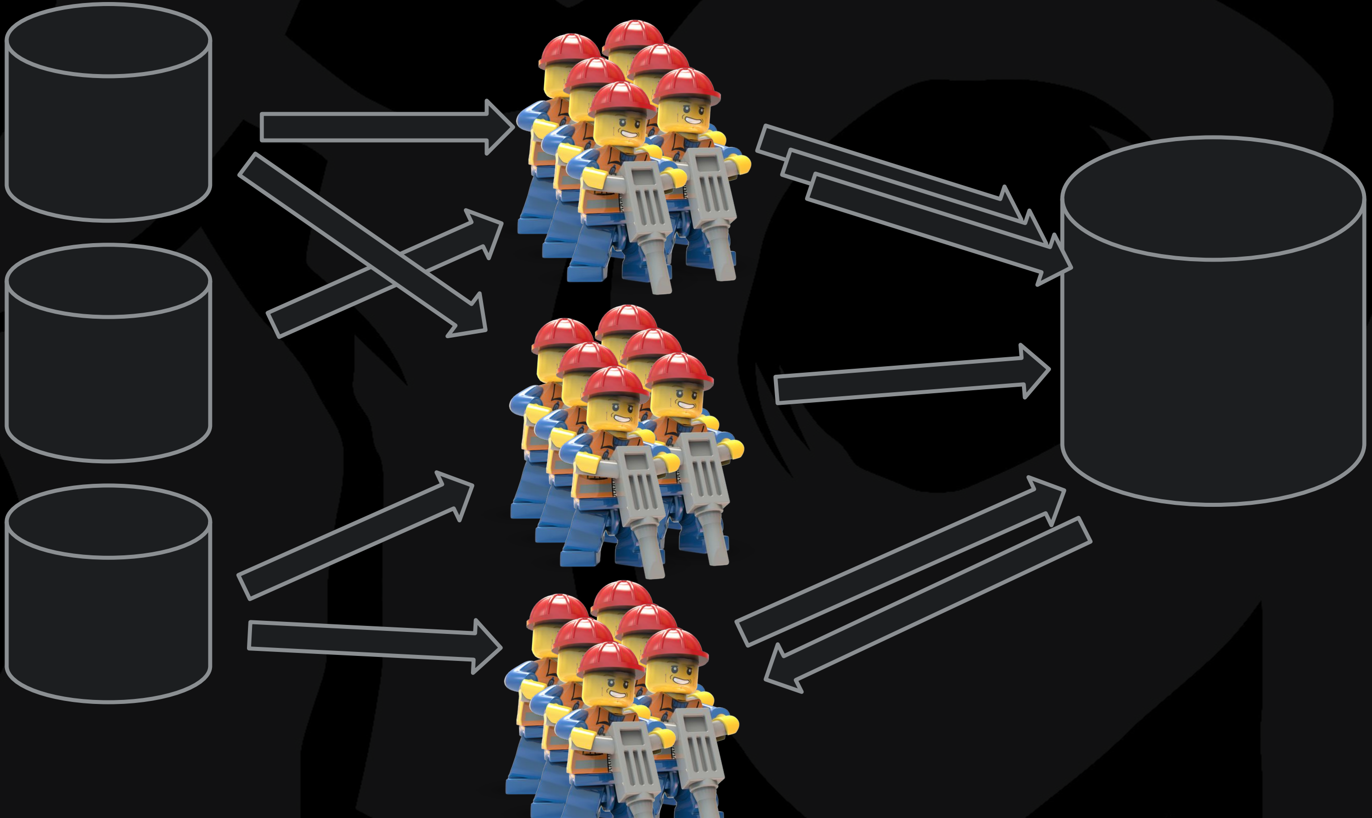
14 TB

DATA MOVED DAILY

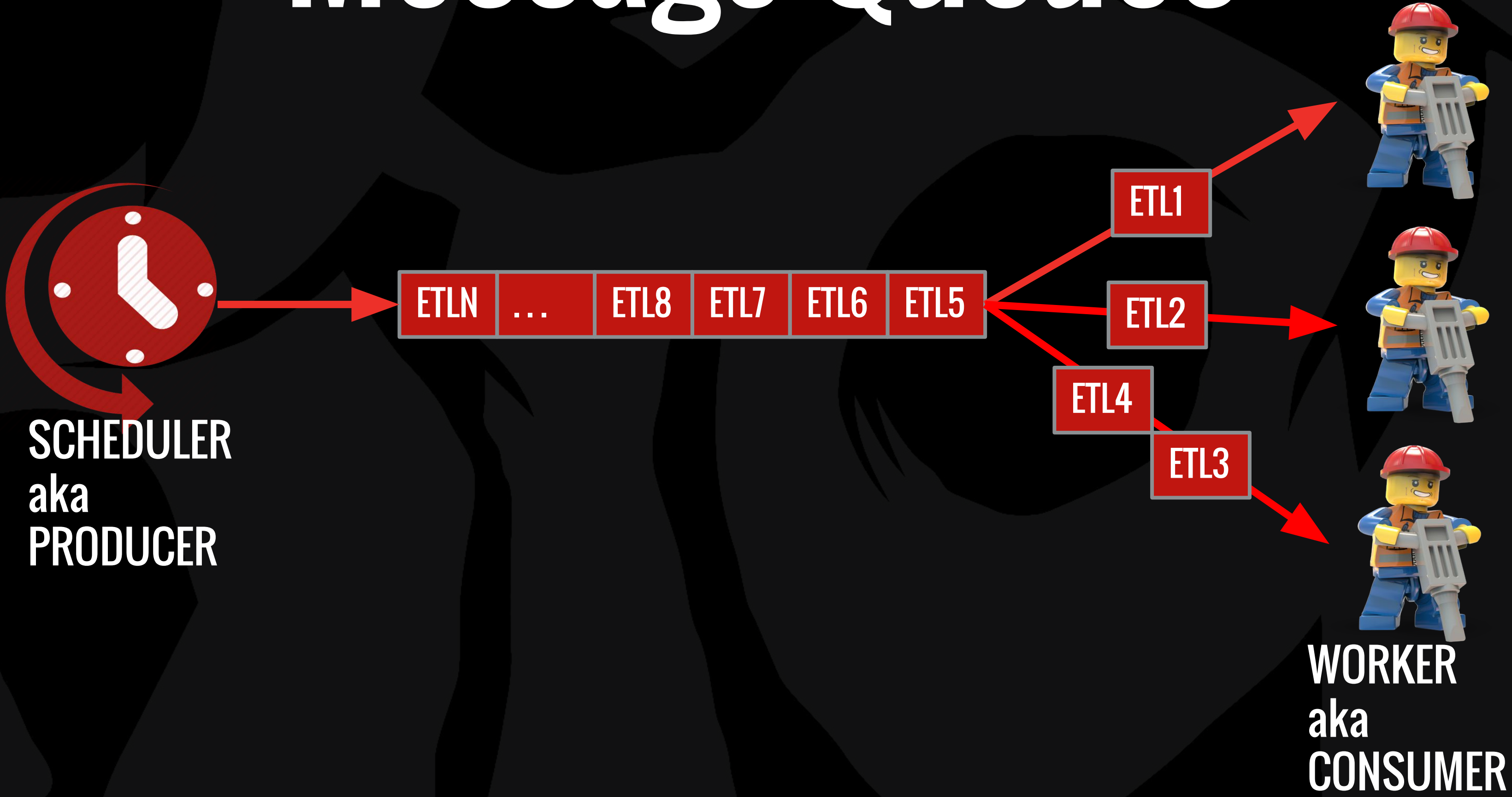
FuETL SCALING



FuETL SCALING



Message Queues



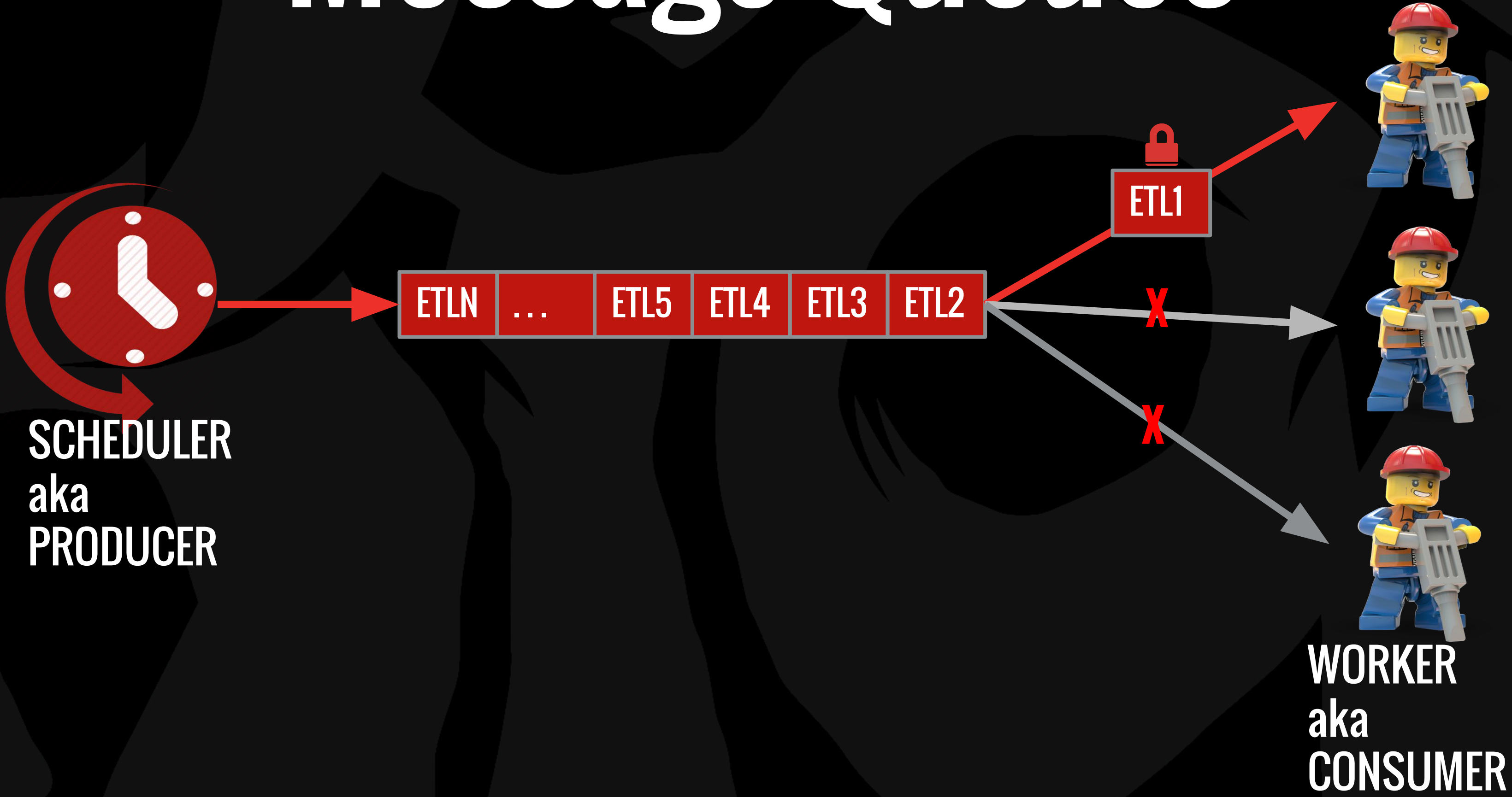
Message Queues

- Redundancy
- Delivery Guarantees
- Easy to Scale
- Asynchronous Communication
- Abstraction / Decoupling

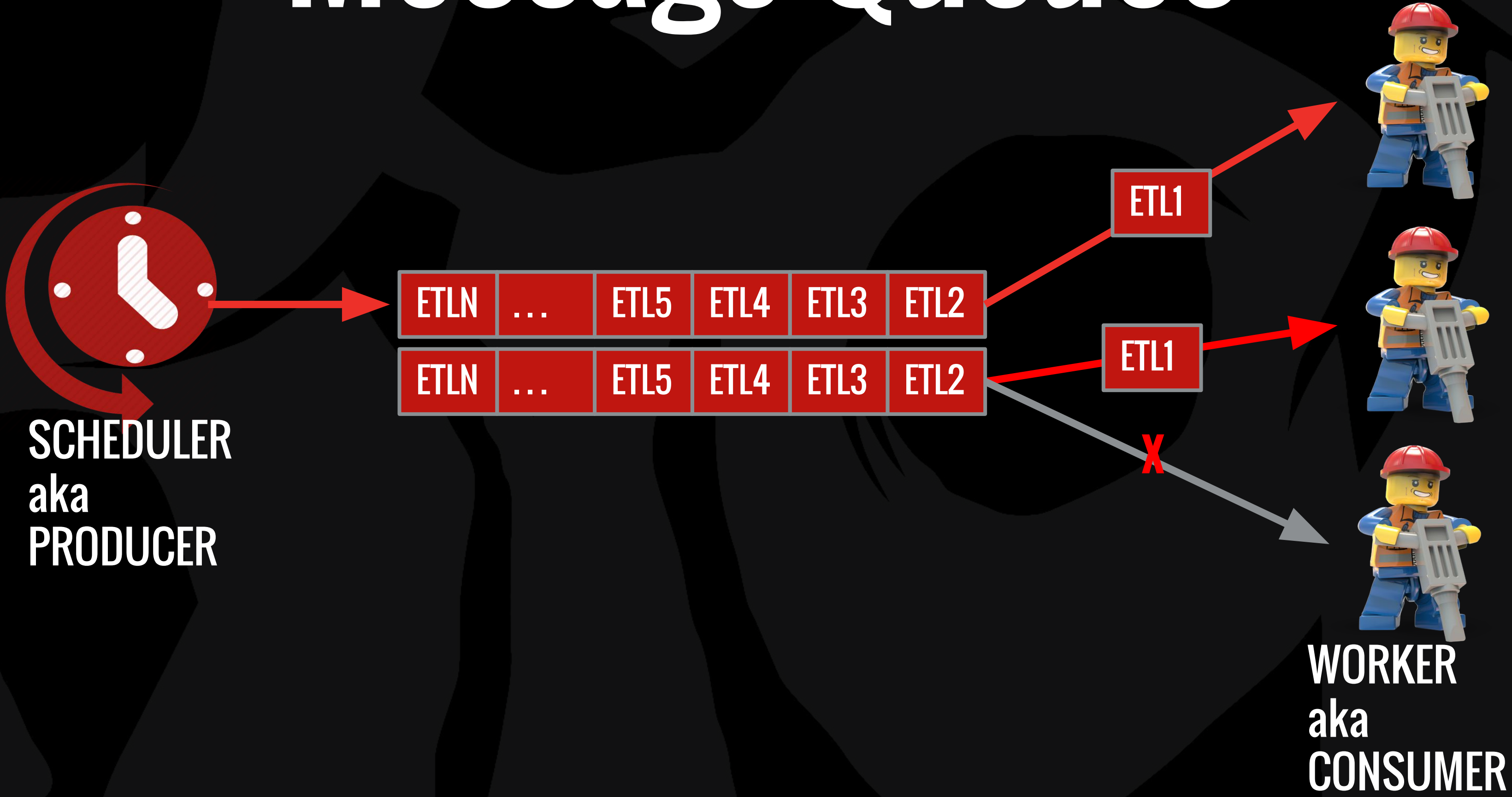
Message Queues

- Amazon Simple Queue Service
- Apache ActiveMQ
- RabbitMQ
- HornetQ
- Microsoft MQ (MSMQ)

Message Queues



Message Queues

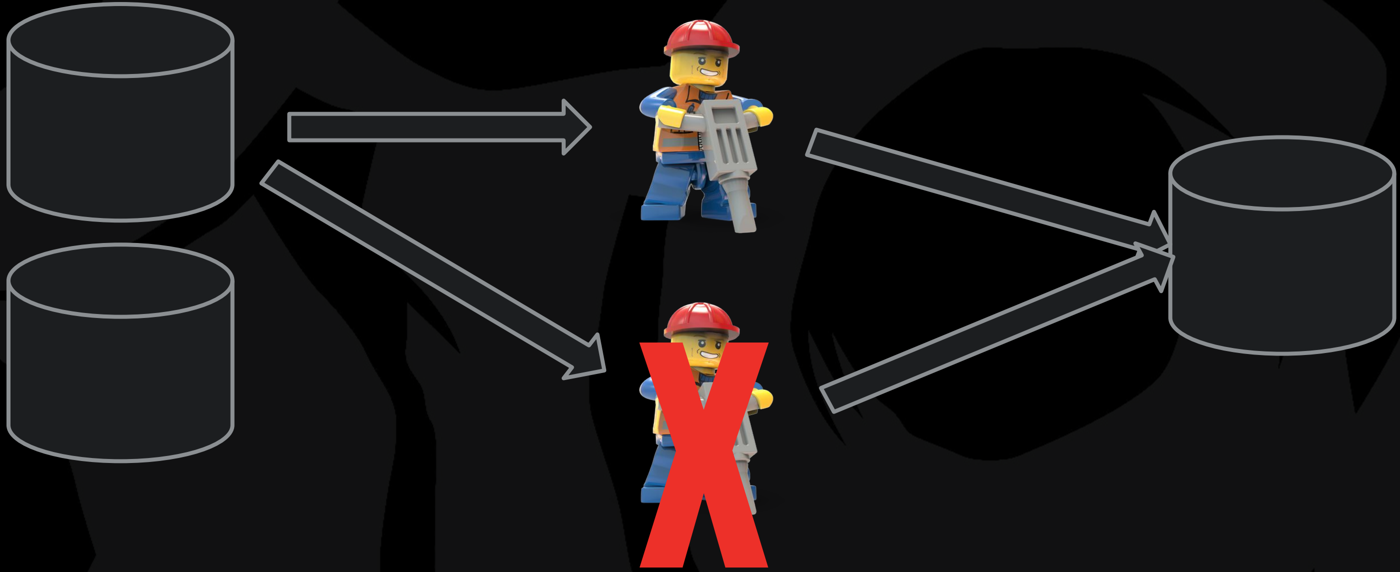


What will happen

In the big data / OLAP world….
(hint: no primary key validation)

```
INSERT INTO games_played
(SELECT * FROM games_played_na
 WHERE date >= ‘2015-10-25’)
```

KEEPING INTEGRITY



INGESTION

PULL-BASED / ETL

FU ETL

FuETL
- OLTP game data
- External Data Sources

PUSH-BASED

kafka

- Anything pushed to it
- Server logs

STORAGE

MASTER WAREHOUSE

amazon web services™
S3 Simple Storage Service

DATA AUDITING

Wasp

QUERY / VIEWS

AGGREGATE QUERIES

VERTICA
An HP Company

BATCH QUERIES

HIVE

SINGLE-ROW QUERIES

amazon DynamoDB

VIZ. TOOLS

tableau

platfora
You should know.

Kafka

The new hotness in big data

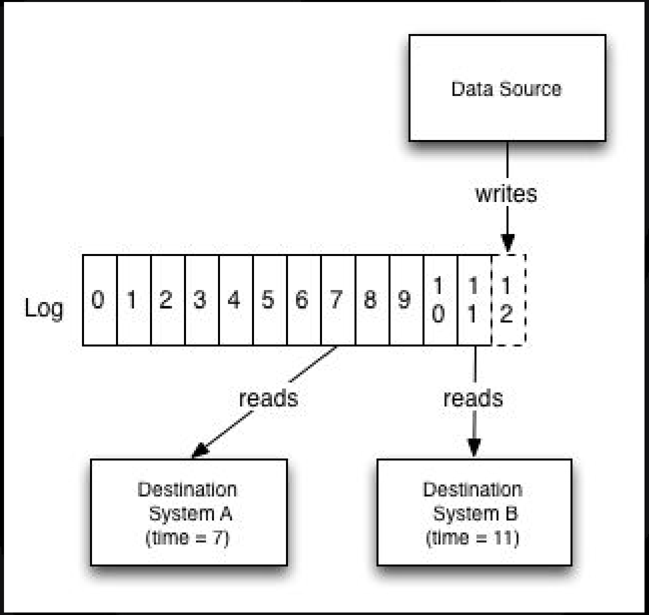
Open-source project maintained by Confluent

Very fast distributed message queue

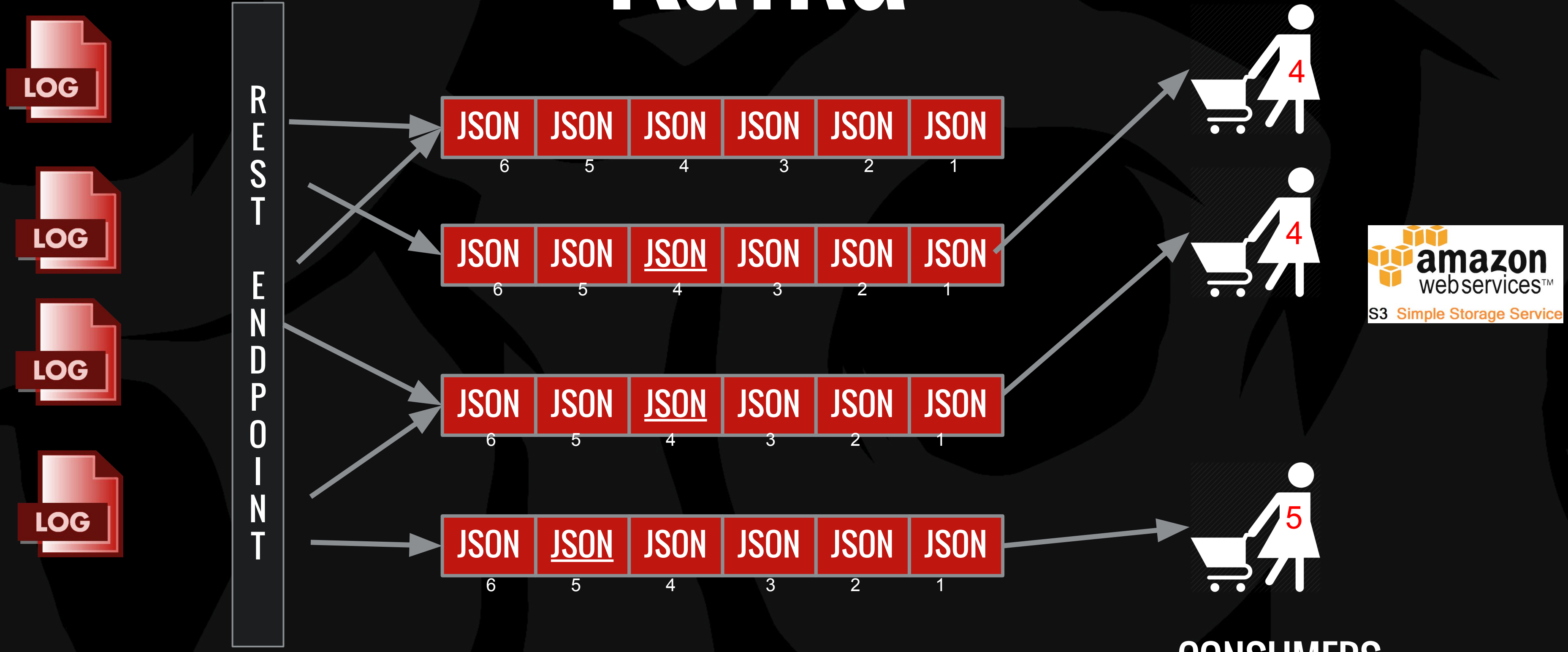
Data is replicated across “partitions” to ensure no loss

Kafka

Has a DB Commit Log
(ooh revolutionary)


Data Source
writes
Log
0
1
2
3
4
5
6
7
8
9
10
11
12
reads
reads
Destination System A (time = 7)
Destination System B (time = 11)

Kafka



LOG

LOG

LOG

LOG

REST ENDPOINT

JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1

JSON JSON <u>JSON</u> JSON JSON JSON
6 5 4 3 2 1

JSON JSON <u>JSON</u> JSON JSON JSON
6 5 4 3 2 1

JSON <u>JSON</u> JSON JSON JSON JSON
6 5 4 3 2 1

4

4

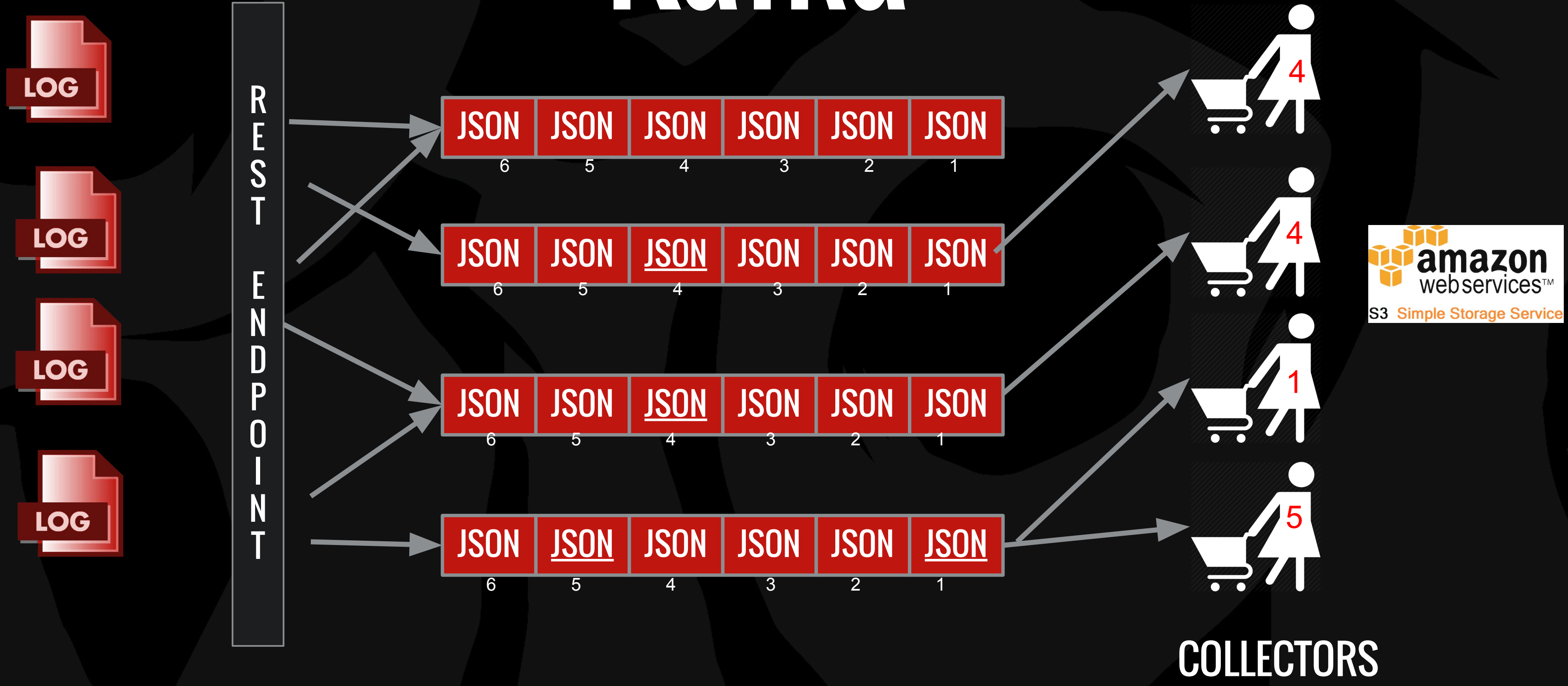
5

amazon webservices™

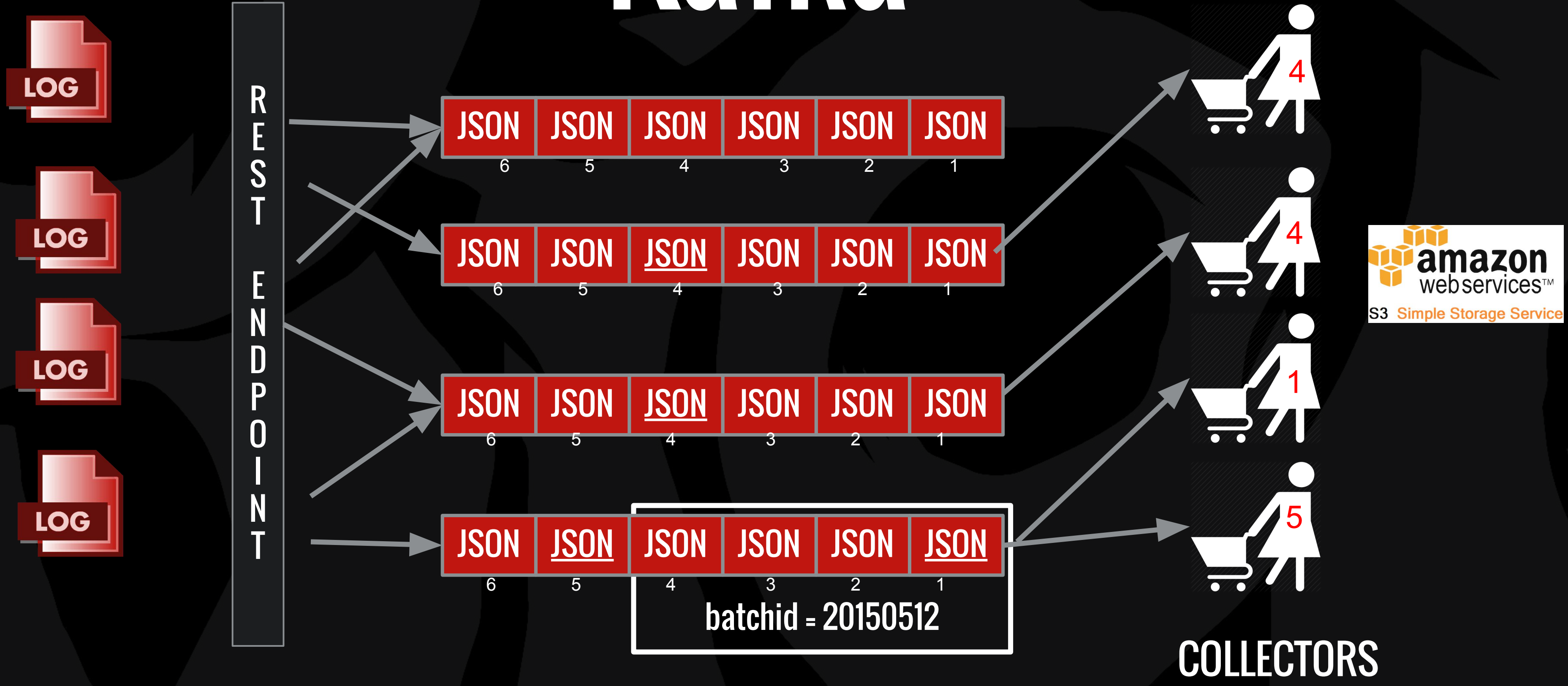
S3 Simple Storage Service

CONSUMERS

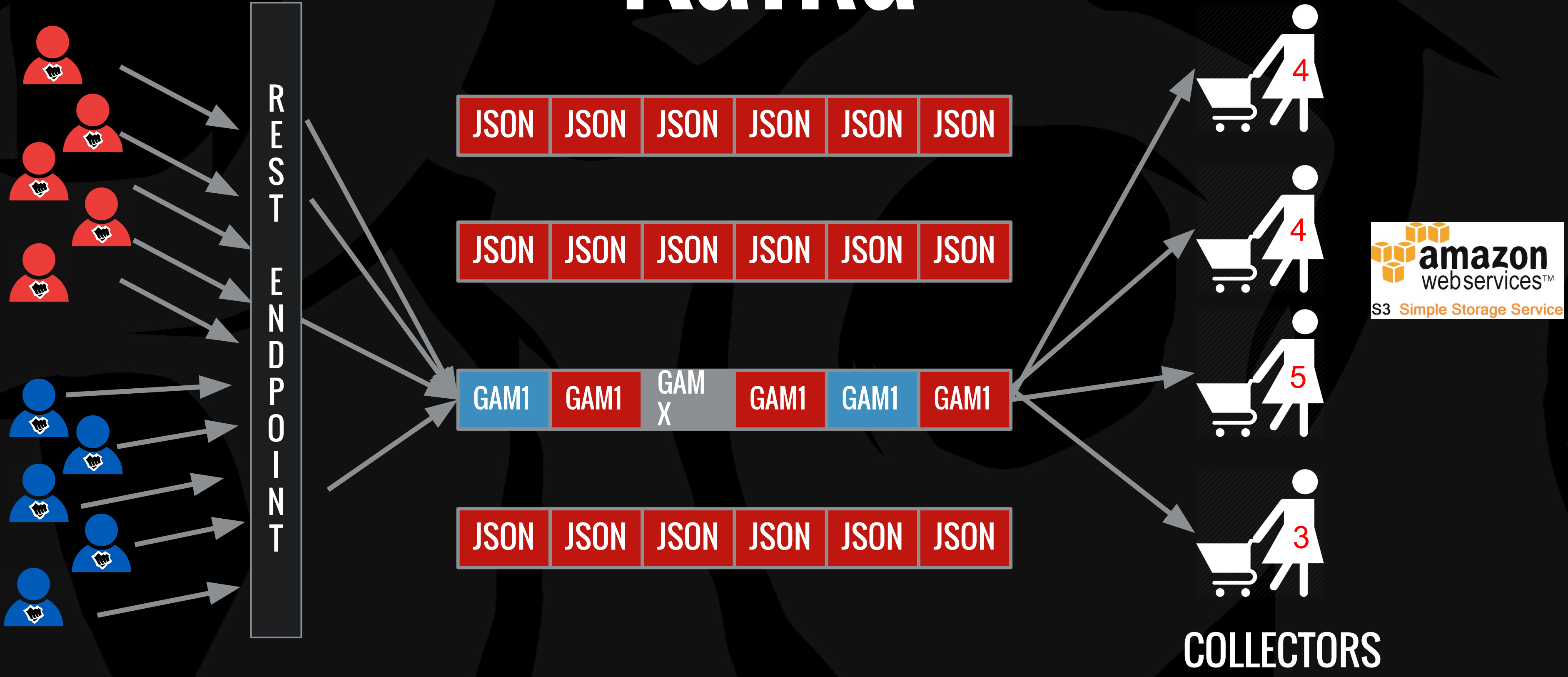
Kafka


LOG
LOG
LOG
LOG
REST ENDPOINT
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
4
4
1
5
amazon webservices™
S3 Simple Storage Service
COLLECTORS

Kafka


LOG
LOG
LOG
LOG
REST ENDPOINT
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
JSON JSON JSON JSON JSON JSON
6 5 4 3 2 1
batchid = 20150512
4
4
1
5
amazon webservices™
S3 Simple Storage Service
COLLECTORS

Kafka


REST ENDPOINT
JSON JSON JSON JSON JSON JSON
JSON JSON JSON JSON JSON JSON
GAM1 GAM1 GAM X GAM1 GAM1 GAM1
JSON JSON JSON JSON JSON JSON
4
4
5
3
amazon webservices™
S3 Simple Storage Service
COLLECTORS

Idempotency

Idempotent - an operation that will produce the same results if executed once or multiple times

EXAMPLE:

Non-Idempotent:

- `x = x * 5;`
- Submitting a purchase

Idempotent:

- `abs( abs(x) ) = abs(X)`
- Cancelling a purchase

Idempotent?

In the transactional OLTP world….

```
INSERT INTO games_played
(SELECT * FROM games_played_na
 WHERE date >= ‘2015-10-25’)
```

Idempotent?

In the big data / OLAP world….

```
INSERT INTO games_played
(SELECT * FROM games_played_na
 WHERE date >= ‘2015-10-25’)
```

Idempotency

Use application logic to make idempotent

```
msg = queue.pop;
if (processed_games.contains( msg.game_id )
{
    return; //do nothing
else {
    process_game(msg);
}
```

INGESTION

PULL-BASED / ETL

FU ETL

FuETL
- OLTP game data
- External Data Sources

PUSH-BASED

kafka

- Anything pushed to it
- Server logs

STORAGE

MASTER WAREHOUSE

amazon web services™
S3 Simple Storage Service

DATA AUDITING

Wasp

QUERY / VIEWS

AGGREGATE QUERIES

VERTICA
An HP Company

BATCH QUERIES

HIVE

SINGLE-ROW QUERIES

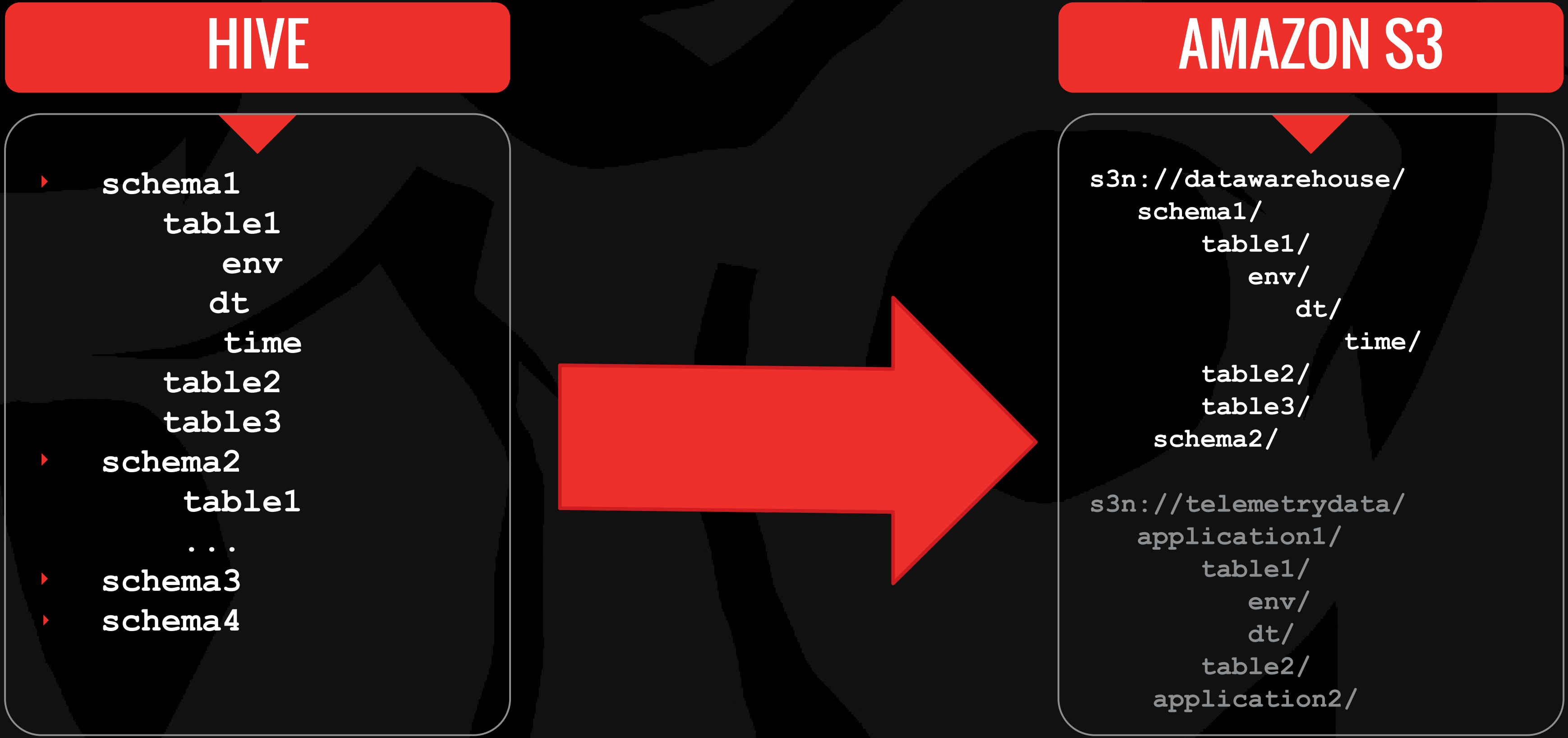
amazon DynamoDB

VIZ. TOOLS

+ableau

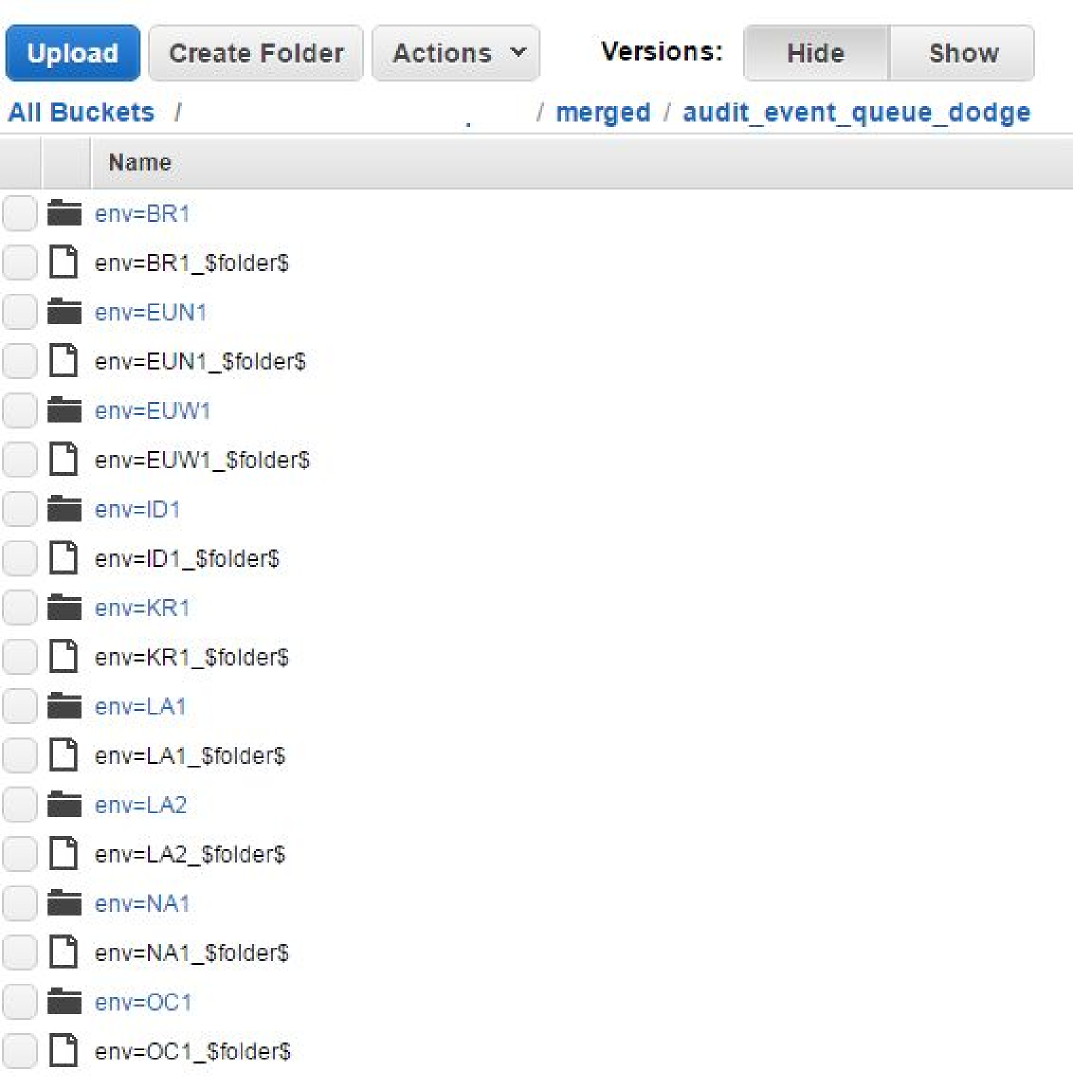
platfora
You should know.

AMAZON S3 STRUCTURE



```
- schema1
     table1
        env
       dt
        time
     table2
     table3
- schema2
      table1
      ...
- schema3
- schema4
```

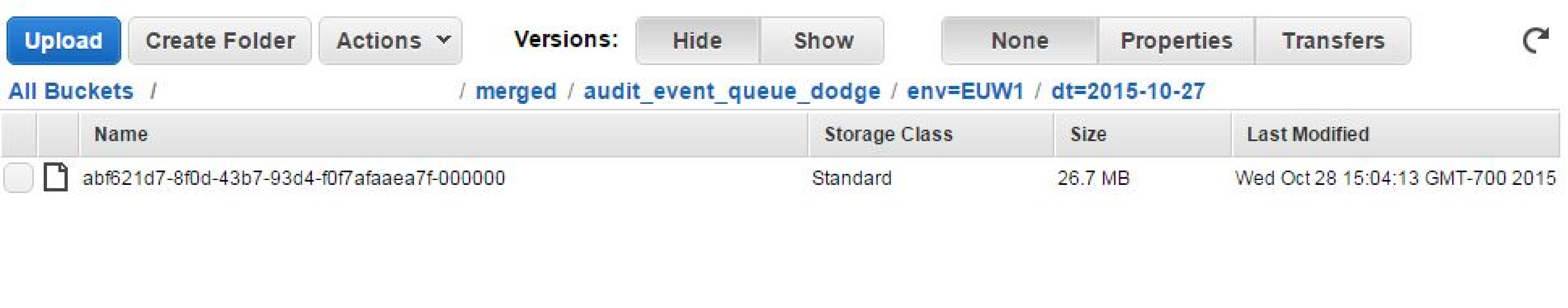
```
s3n://datawarehouse/
   schema1/
       table1/
          env/
             dt/
                time/
       table2/
       table3/
    schema2/

s3n://telemetrydata/
   application1/
       table1/
          env/
          dt/
       table2/
    application2/
```



All Buckets / . / merged / audit_event_queue_dodge

Name

env=BR1
env=BR1_$folder$
env=EUN1
env=EUN1_$folder$
env=EUW1
env=EUW1_$folder$
env=ID1
env=ID1_$folder$
env=KR1
env=KR1_$folder$
env=LA1
env=LA1_$folder$
env=LA2
env=LA2_$folder$
env=NA1
env=NA1_$folder$
env=OC1
env=OC1_$folder$



All Buckets / / merged / audit_event_queue_dodge / env=EUW1 / dt=2015-10-27

Name	Storage Class	Size	Last Modified
abf621d7-8f0d-43b7-93d4-f0f7afaaea7f-000000	Standard	26.7 MB	Wed Oct 28 15:04:13 GMT-700 2015

INGESTION

PULL-BASED / ETL

FU ETL

FuETL
- OLTP game data
- External Data Sources

PUSH-BASED

kafka

- Anything pushed to it
- Server logs

STORAGE

MASTER WAREHOUSE

amazon web services™
S3 Simple Storage Service

DATA AUDITING

Wasp

QUERY / VIEWS

AGGREGATE QUERIES

VERTICA
An HP Company

BATCH QUERIES

HIVE

SINGLE-ROW QUERIES

amazon DynamoDB

VIZ. TOOLS

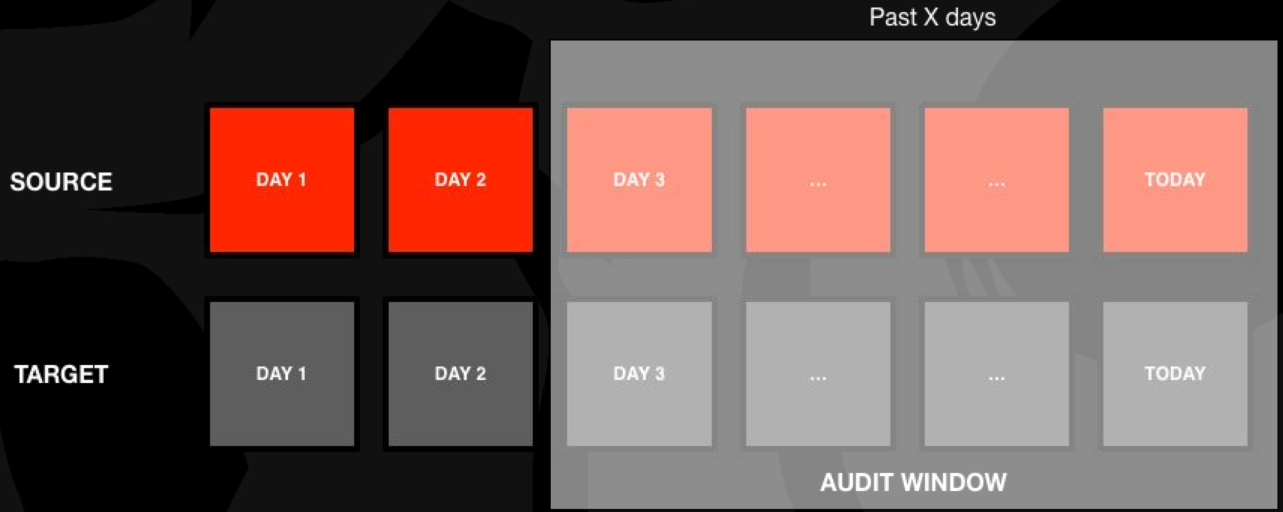
+ a b | e a u®

platfora
You should know.

Warehouse
Auditing
Service
Platform

- REST micro-service built with Java and docker.
- Source and target comparison.
- Reports and visualizations we can use to find problems.

HOW TO AUDIT



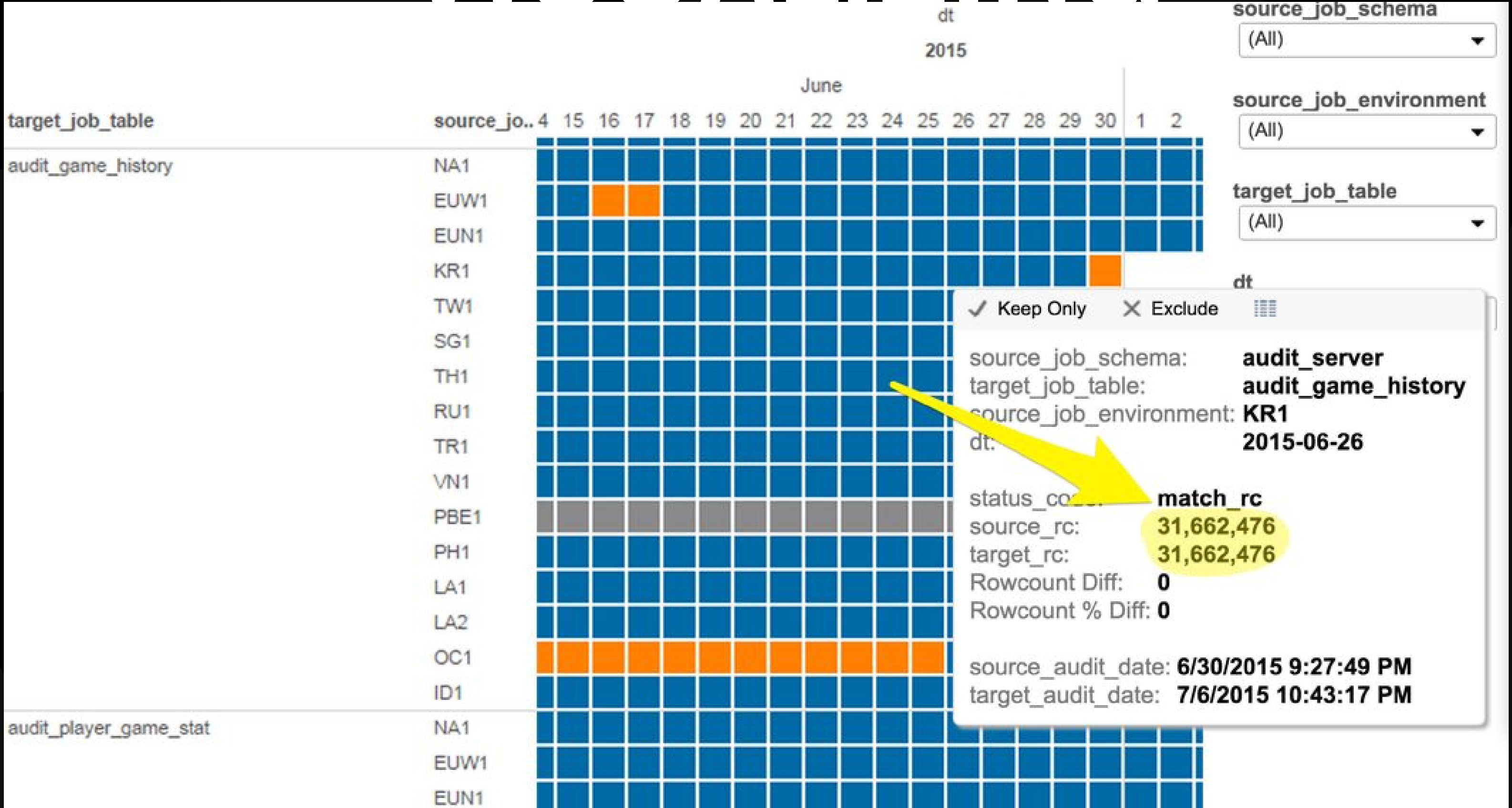
Past X days

SOURCE	DAY 1	DAY 2	DAY 3	...	...	TODAY
TARGET	DAY 1	DAY 2	DAY 3	...	...	TODAY

AUDIT WINDOW

HOW TO AUDIT

VISUALIZING



source_job_schema: (All)

source_job_environment: (All)

target_job_table: (All)

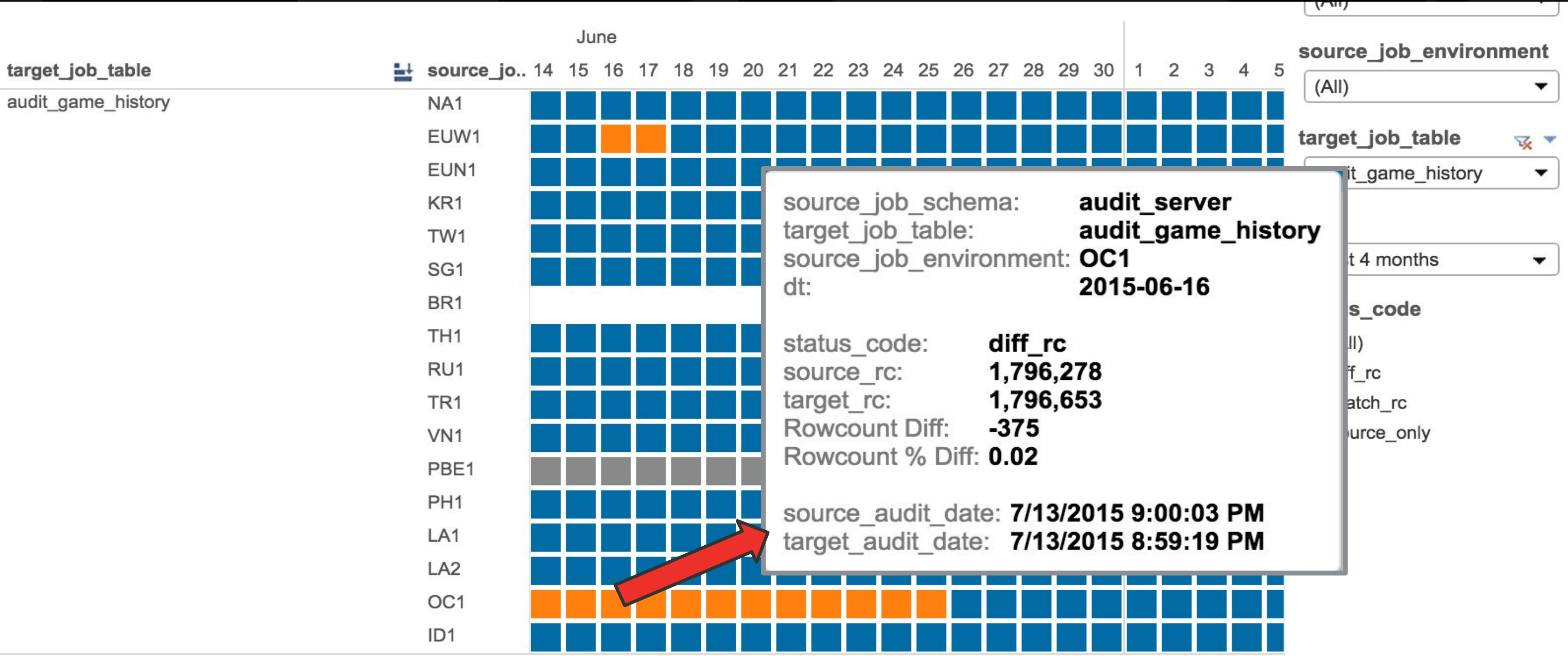
dt

source_job_schema: **audit_server**
target_job_table: **audit_game_history**
source_job_environment: **KR1**
dt: **2015-06-26**

status_code: **match_rc**
source_rc: **31,662,476**
target_rc: **31,662,476**
Rowcount Diff: **0**
Rowcount % Diff: **0**

source_audit_date: **6/30/2015 9:27:49 PM**
target_audit_date: **7/6/2015 10:43:17 PM**

VISUALIZING



source_job_environment: (All)

target_job_table: it_game_history

t 4 months

s_code

ll)

f_rc

atch_rc

urce_only

source_job_schema: **audit_server**
target_job_table: **audit_game_history**
source_job_environment: **OC1**
dt: **2015-06-16**

status_code: **diff_rc**
source_rc: **1,796,278**
target_rc: **1,796,653**
Rowcount Diff: **-375**
Rowcount % Diff: **0.02**

source_audit_date: **7/13/2015 9:00:03 PM**
target_audit_date: **7/13/2015 8:59:19 PM**

INGESTION

PULL-BASED / ETL

FuETL
- OLTP game data
- External Data Sources

PUSH-BASED

HONU
- Anything pushed to it
- Server logs

STORAGE

MASTER WAREHOUSE

amazon web services™
S3 Simple Storage Service

DATA AUDITING

Wasp

QUERY / VIEWS

AGGREGATE QUERIES

VERTICA
An HP Company

BATCH QUERIES

HIVE

SINGLE-ROW QUERIES

amazon DynamoDB

VIZ. TOOLS

+ableau

platfora
You should know.

HIVE

row_id	timestamp	col_a	col_b	col_c
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩

BATCH

VERTICA
An HP Company

row_id	timestamp	col_a	col_b	col_c
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	
		🟩	🟩	

OLAP

amazon
DynamoDB

row_id	timestamp	col_a	col_b	col_c
🟩	🟩	🟩	🟩	🟩

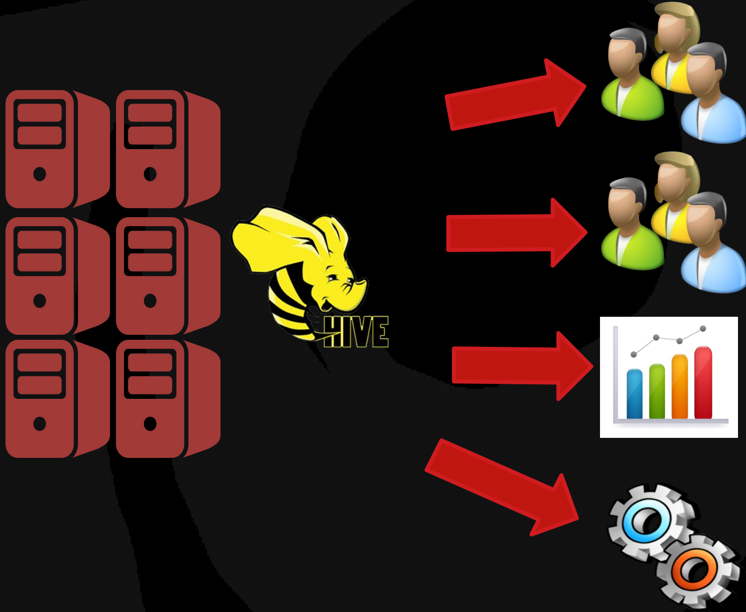
POINT

SCALING IN AWS

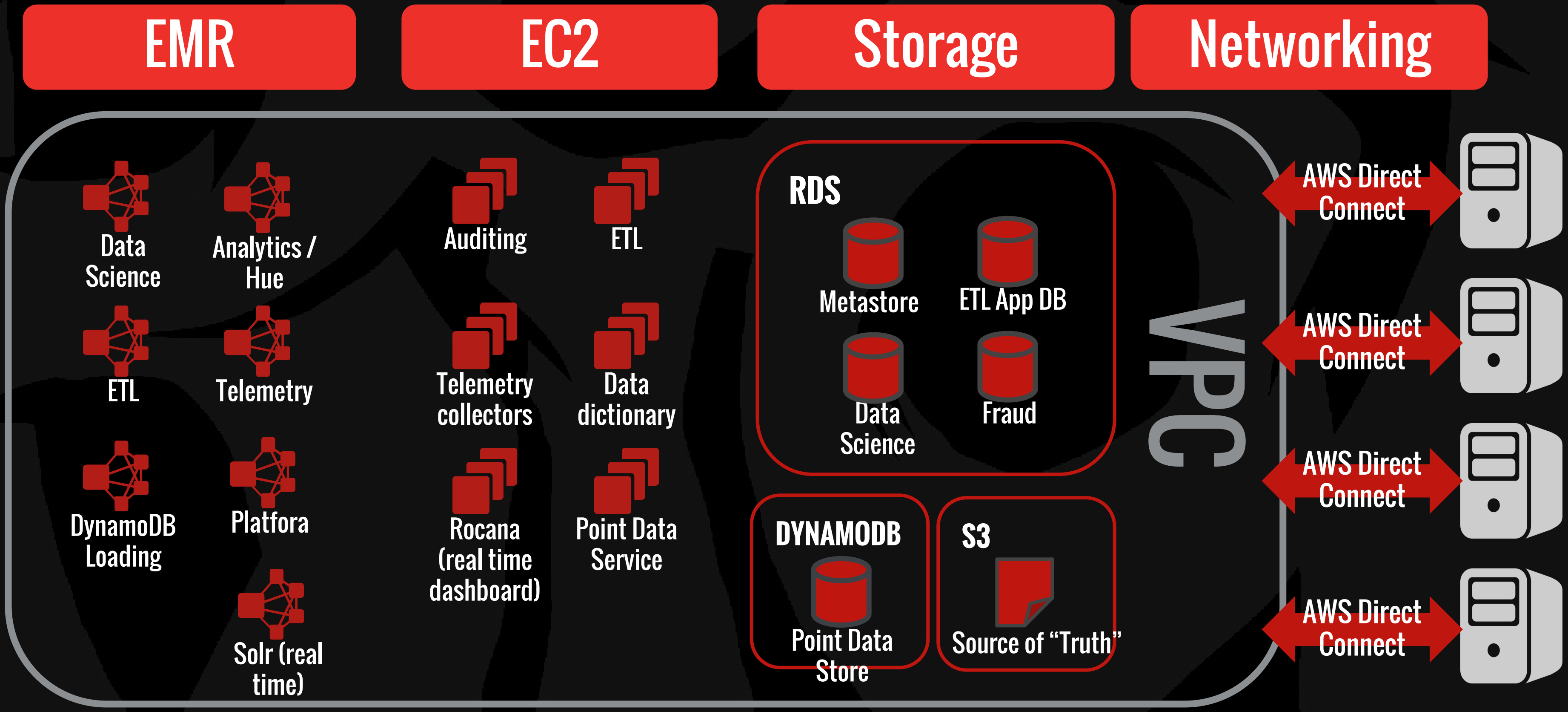
SCALING

RESOURCE CONTENTION

Hive .08 pre YARN, immature resource scheduling



AWS Infrastructure Today



EMR

EC2

Storage

Networking

Data Science

Analytics / Hue

ETL

Telemetry

DynamoDB Loading

Platfora

Solr (real time)

Auditing

ETL

Telemetry collectors

Data dictionary

Rocana (real time dashboard)

Point Data Service

RDS

Metastore

ETL App DB

Data Science

Fraud

DYNAMODB

Point Data Store

S3

Source of “Truth”

VPC

AWS Direct Connect

AWS Direct Connect

AWS Direct Connect

AWS Direct Connect

CONCLUSION

SEAN’S PRO TIPS OF THE DAY

DO

- Keep idempotency in mind and use MQ architecture
- Get an auditing solution for DW accuracy
- Prepare for multiple data access patterns
- Allocate time for tuning AWS infrastructure

DON’T

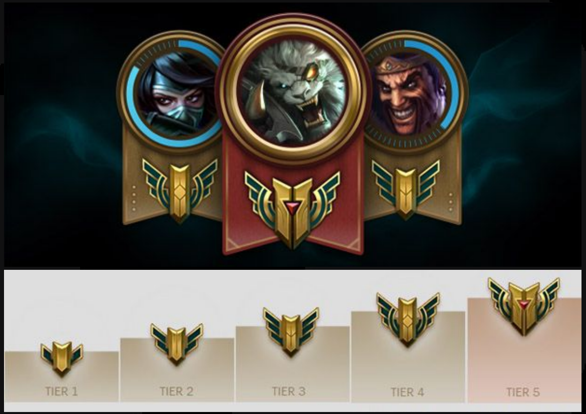
- Don’t underestimate simple problems in big data.
- Don’t forget to track cost. AWS bills can surprise you
- Don’t wait. Create S3 permissions and naming standards early
- Don’t stop. Believing

CHAMPION MASTERY

Custom rewards for mastering different champions

Intensive query that spans every game that every player has played

Improves player engagement

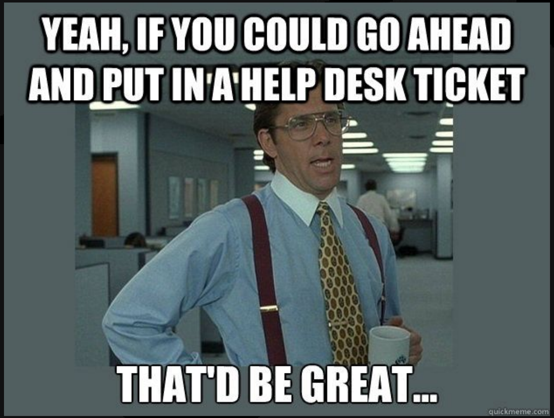


PLAYER SUPPORT

Full copy of our data warehouse in DynamoDB

Hive->DynamoDB Dynamic Partition

Support can answer questions faster than ever.



OFFENSIVE CHAT DETECTION

Data science team queries all chat messages in game

Sentiment analysis and classification

Identifies negative, offensive players and mutes them automatically.

QUESTIONS?

ENGINEERING BLOG

http://engineering.riotgames.com

CAREERS

http://www.riotgames.com/careers

SEAN MALONEY

✉ SMALONEY @ riotgames.com

🐦@SEAN_SEANNERY