



PRESENTED BY



# Hadoop as a Service

How to build & operate a **multi-tenant** enterprise data lake supporting operational & streaming analytics

11:15–11:55 Capital Suite 10/11

**Phill Radley**

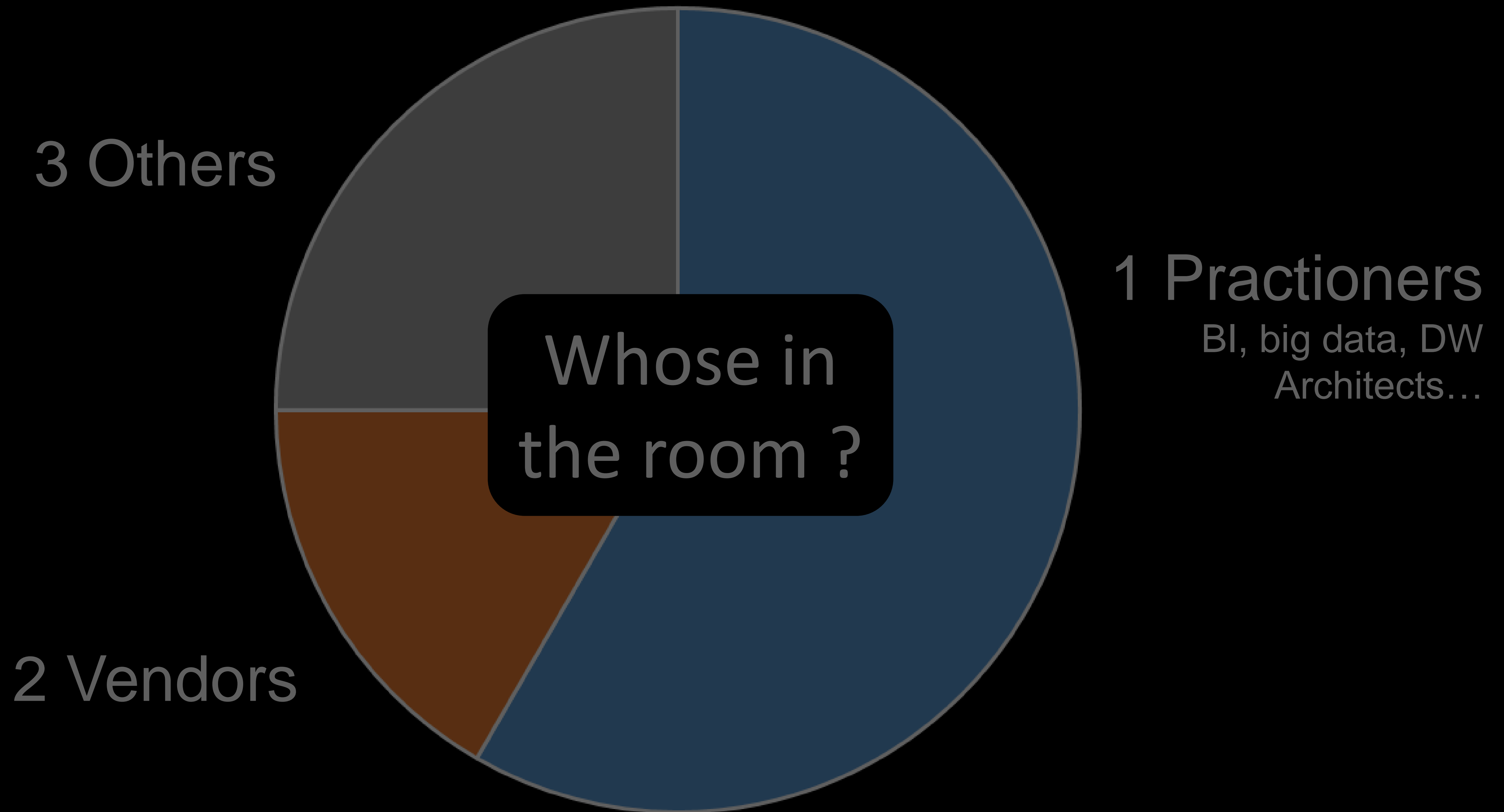
Chief Data Architect - BT

These slides @

[conferences.oreilly.com/strata/strata-eu/public/schedule/detail/57589](https://conferences.oreilly.com/strata/strata-eu/public/schedule/detail/57589)

[strataconf.com](https://strataconf.com)

#StrataHadoop



# BT's big data journey



# BT Research & Innovation



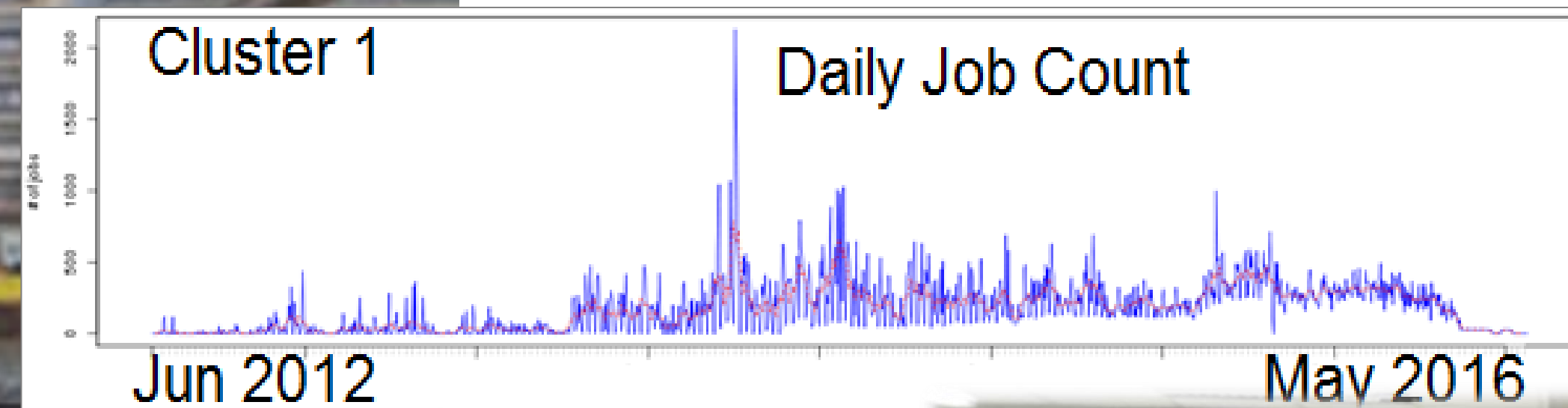
Experimental  
Cluster ~ 2008

2010 cloudera  
Community Edition

amazon  
web services



Jan 2014



Feb 2014 – map/reduce + HDFS



Oct 2016 – Spark + Impala + Kafka ~ 2.5x

#StrataData

## Production

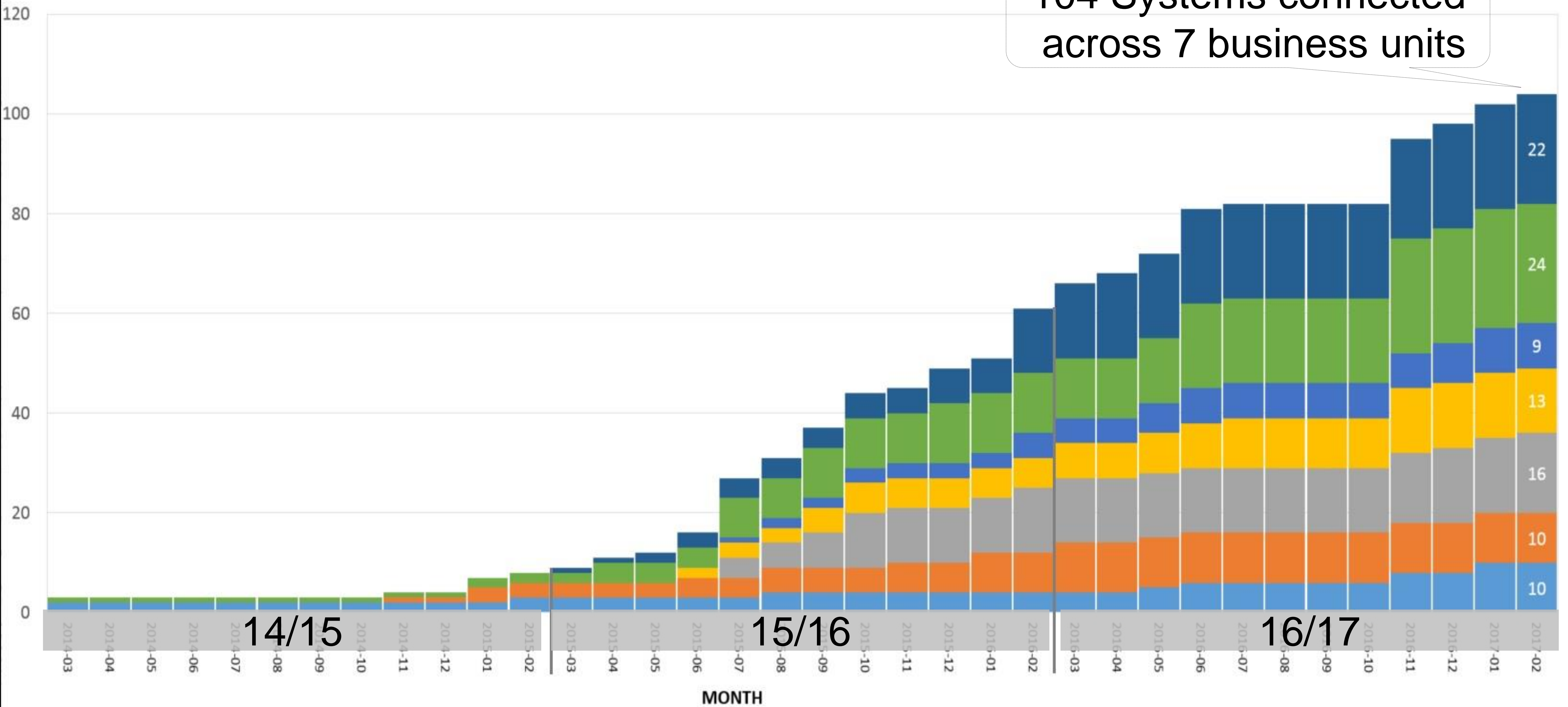
Strata  
DATA CONFERENCE



# Connecting legacy applications/data to HaaS

# APPS Connected

104 Systems connected  
across 7 business units





# Typical usecase 1... **nuisance calls**

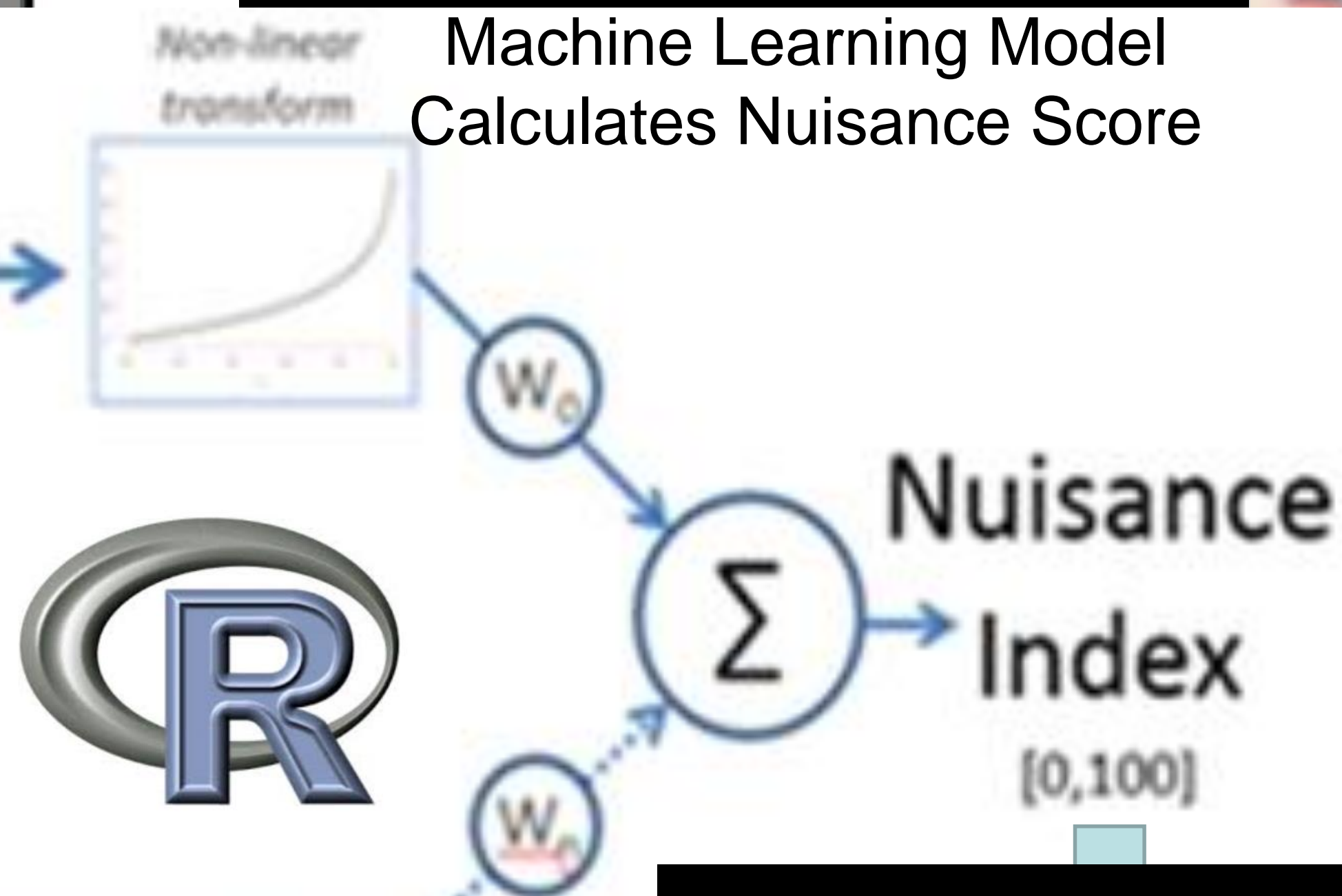
BT Call Protect

Fight nuisance calls, take control

Join Christine,  
fight nuisance  
calls

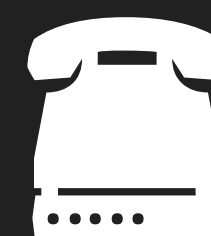
Call Features  
Duration  
Valid CLI  
Who hangs up  
...

Machine Learning Model  
Calculates Nuisance Score



Black List of Nuisance Call Centres  
(customer opt in)

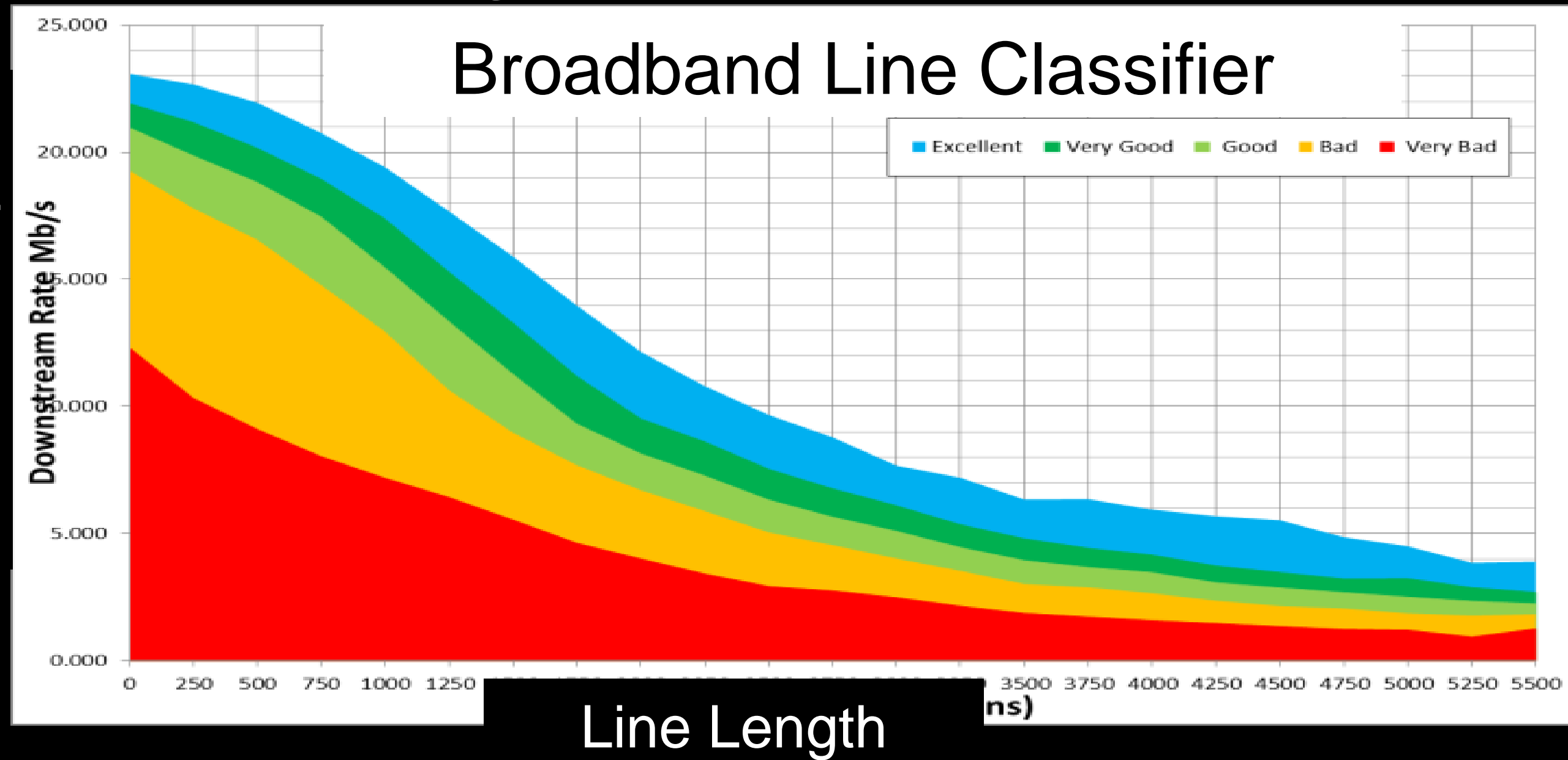
Divert call to Junk voicemail box





# Typical usecase 2... Broadband Speed

Broadband Speed



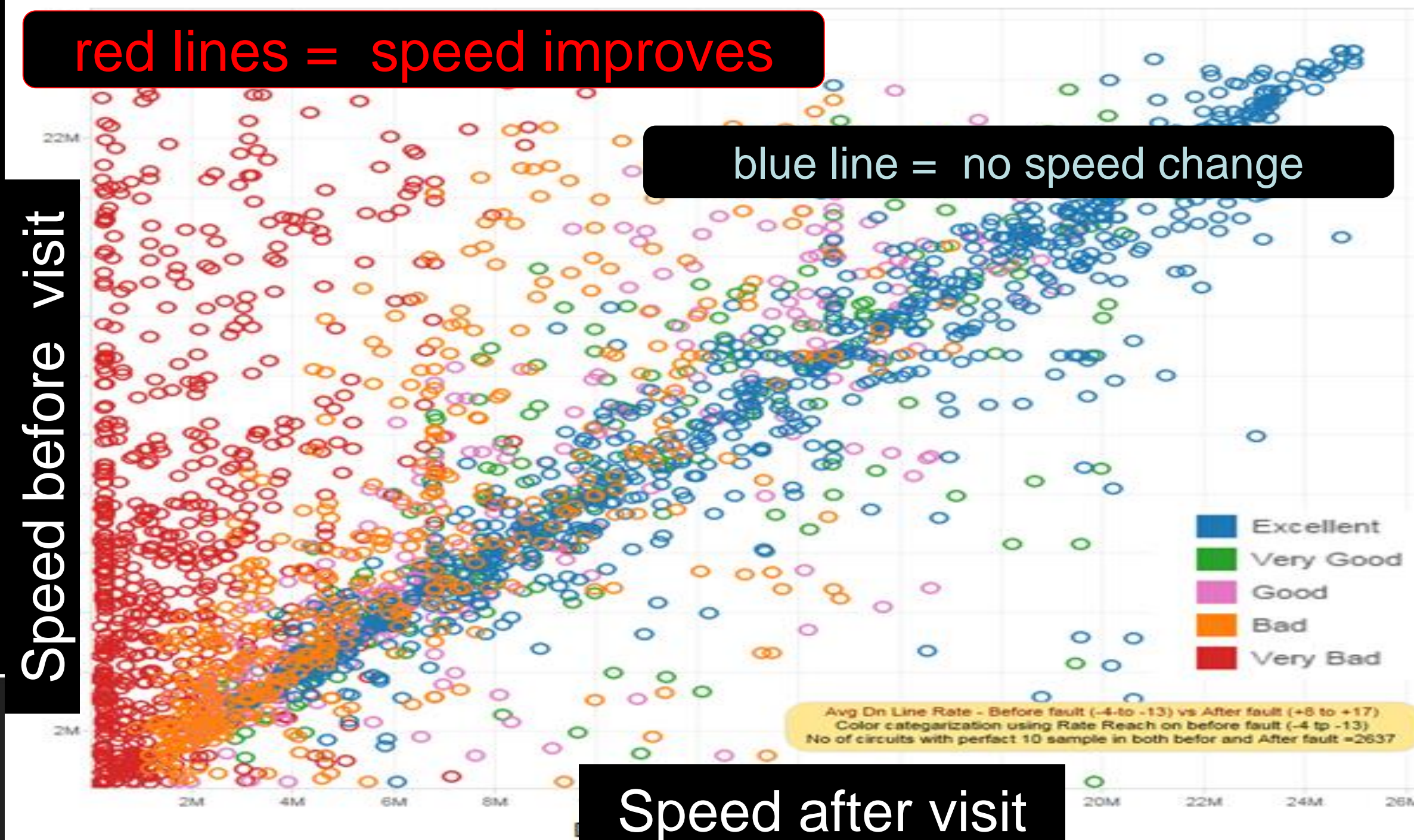
predict when engineer visits will improve speed  
(each dot is an engineer visit)

#StrataData

red lines = speed improves

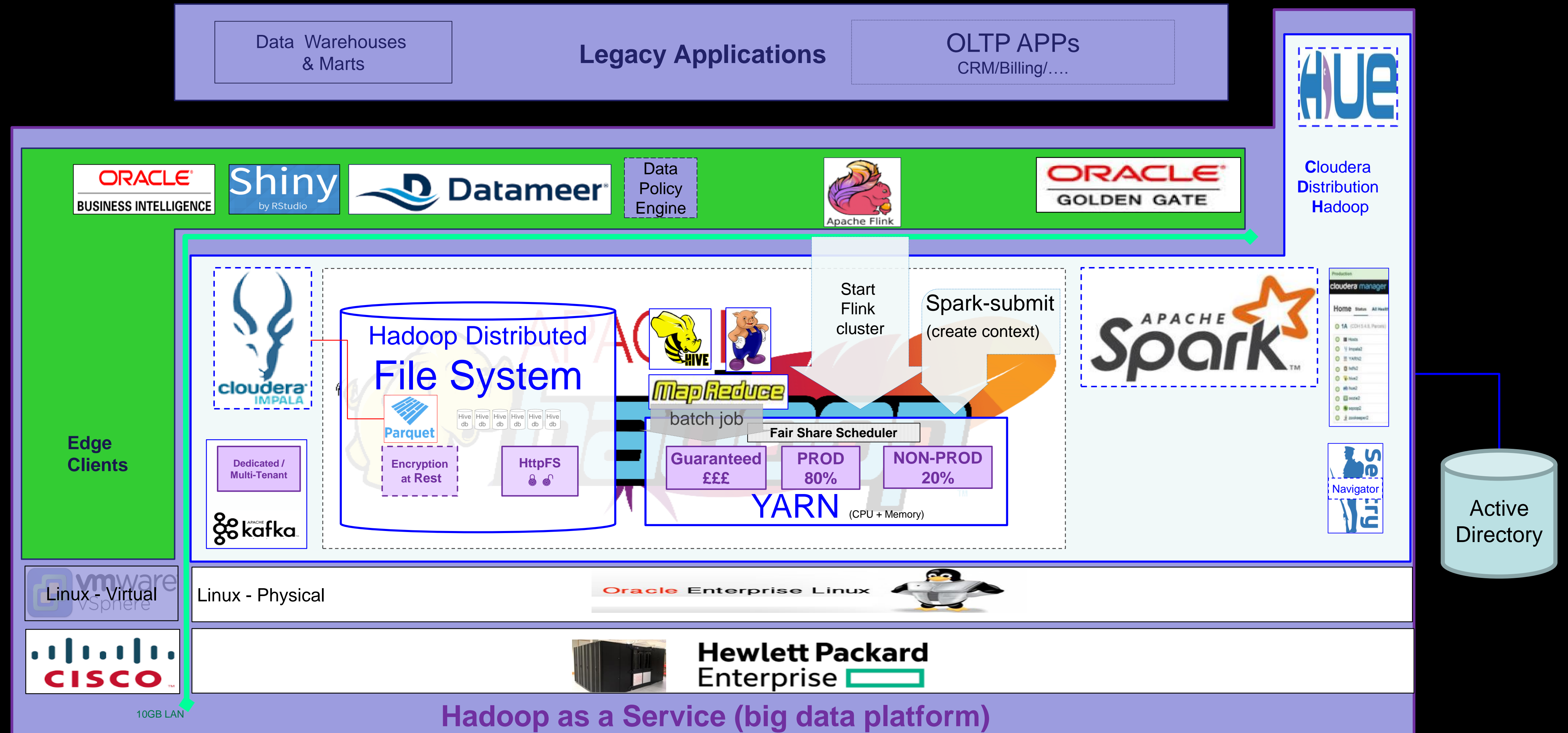
blue line = no speed change

Speed before visit





# Current Technology Stack





# Cluster Utilisation: Weekly

Total Containers Running



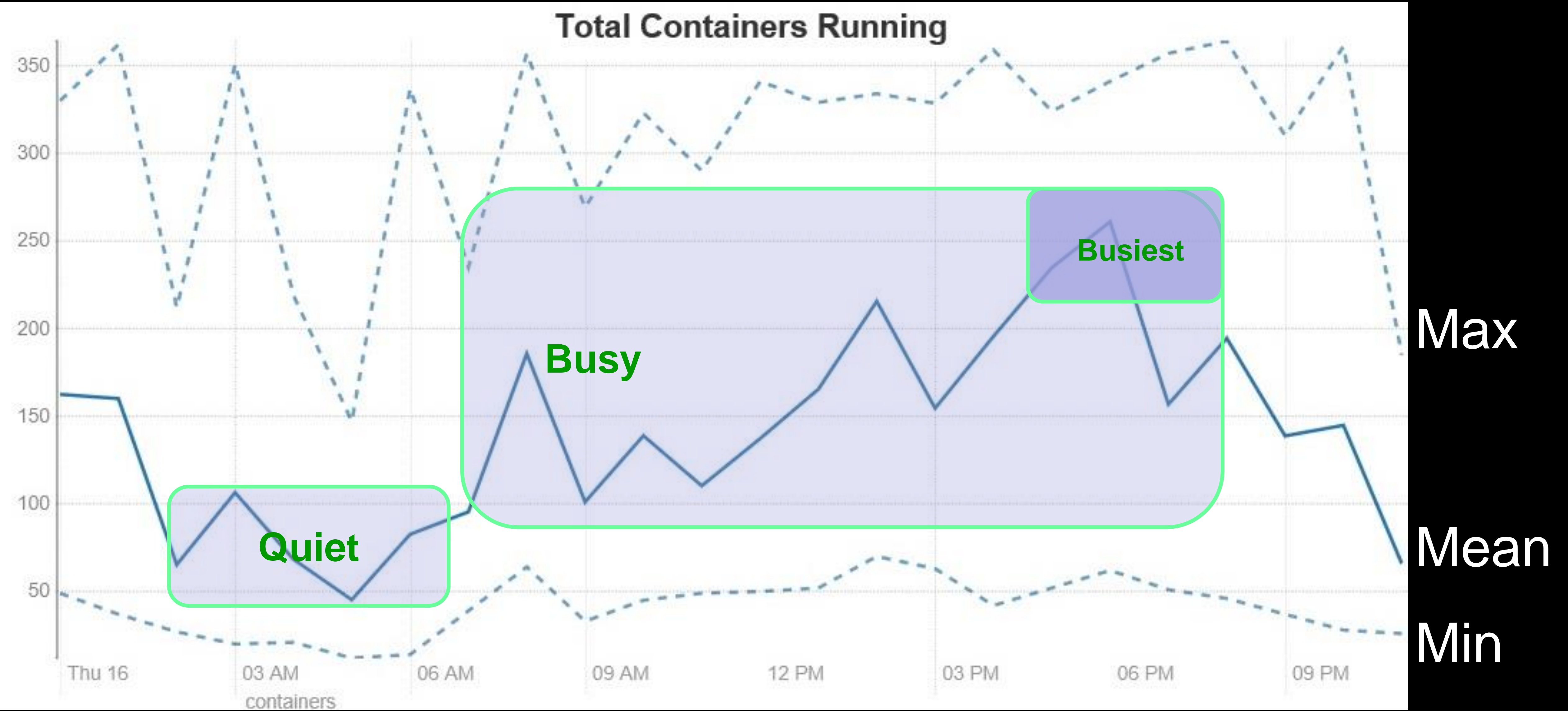
Max

Mean

Daily  
“Quiet” time  
3-6am

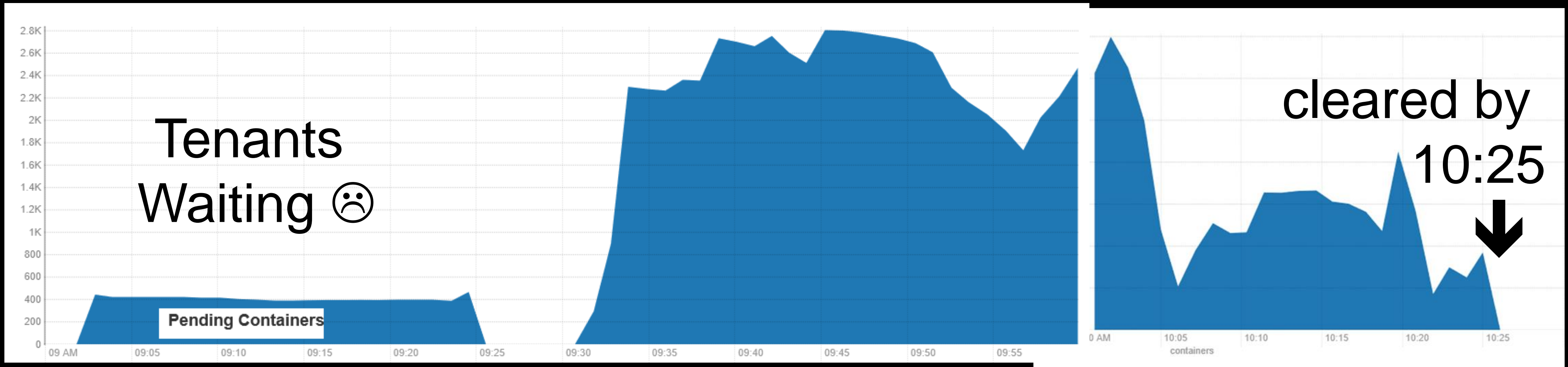
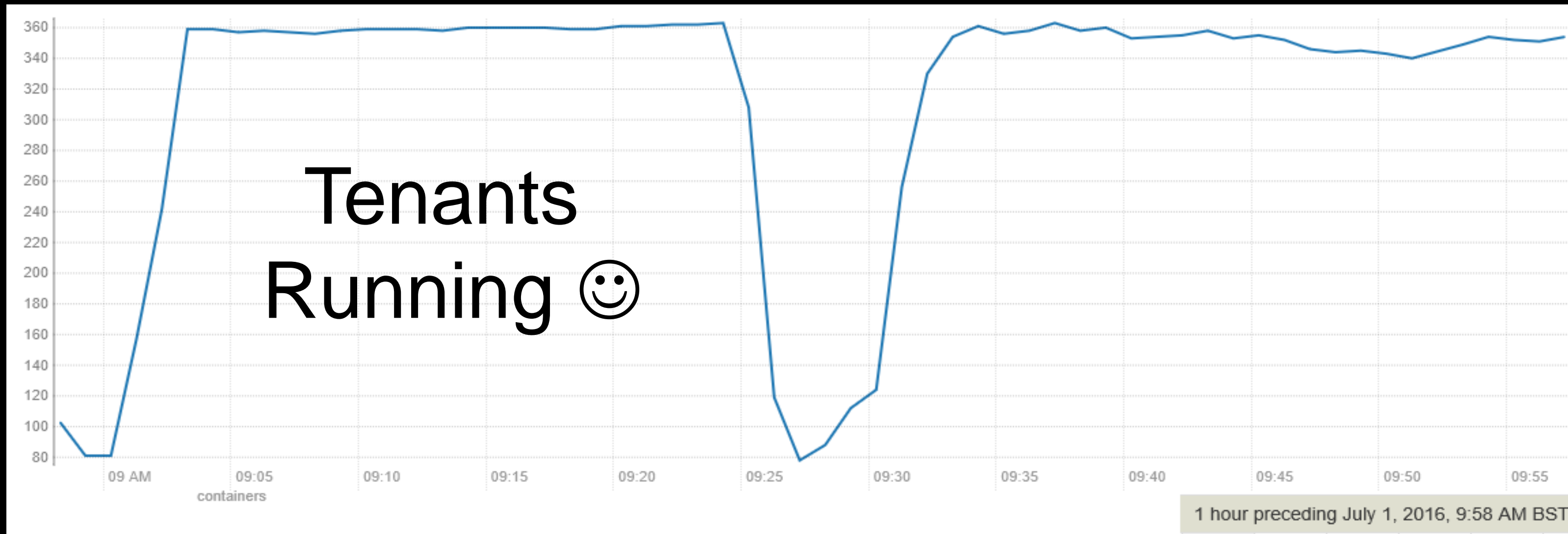
Min

# Cluster Utilisation: Daily



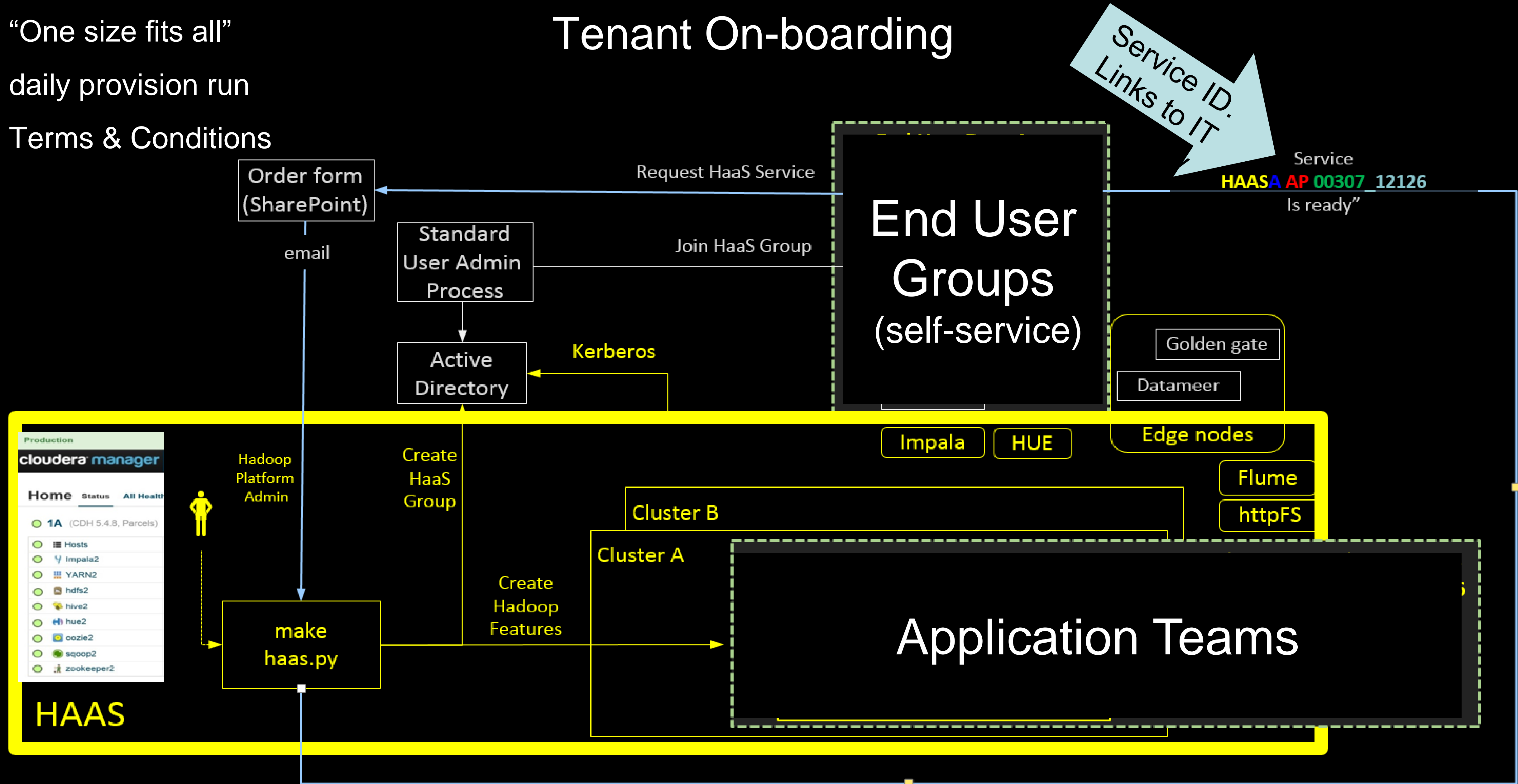


# Cluster Utilisation: “Busy Hour”



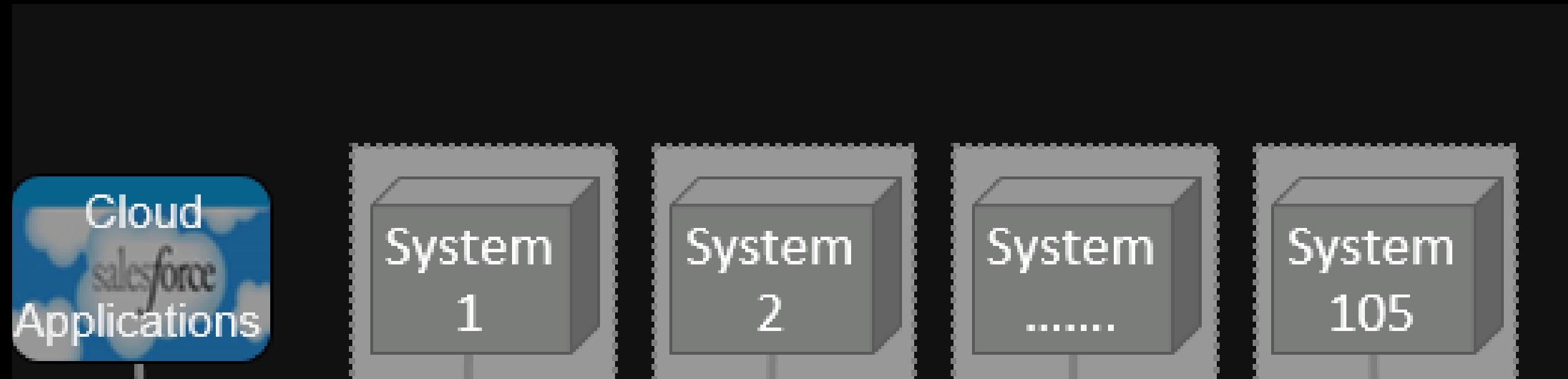
- “One size fits all”
- daily provision run
- Terms & Conditions

# Tenant On-boarding





# Hybrid Architecture

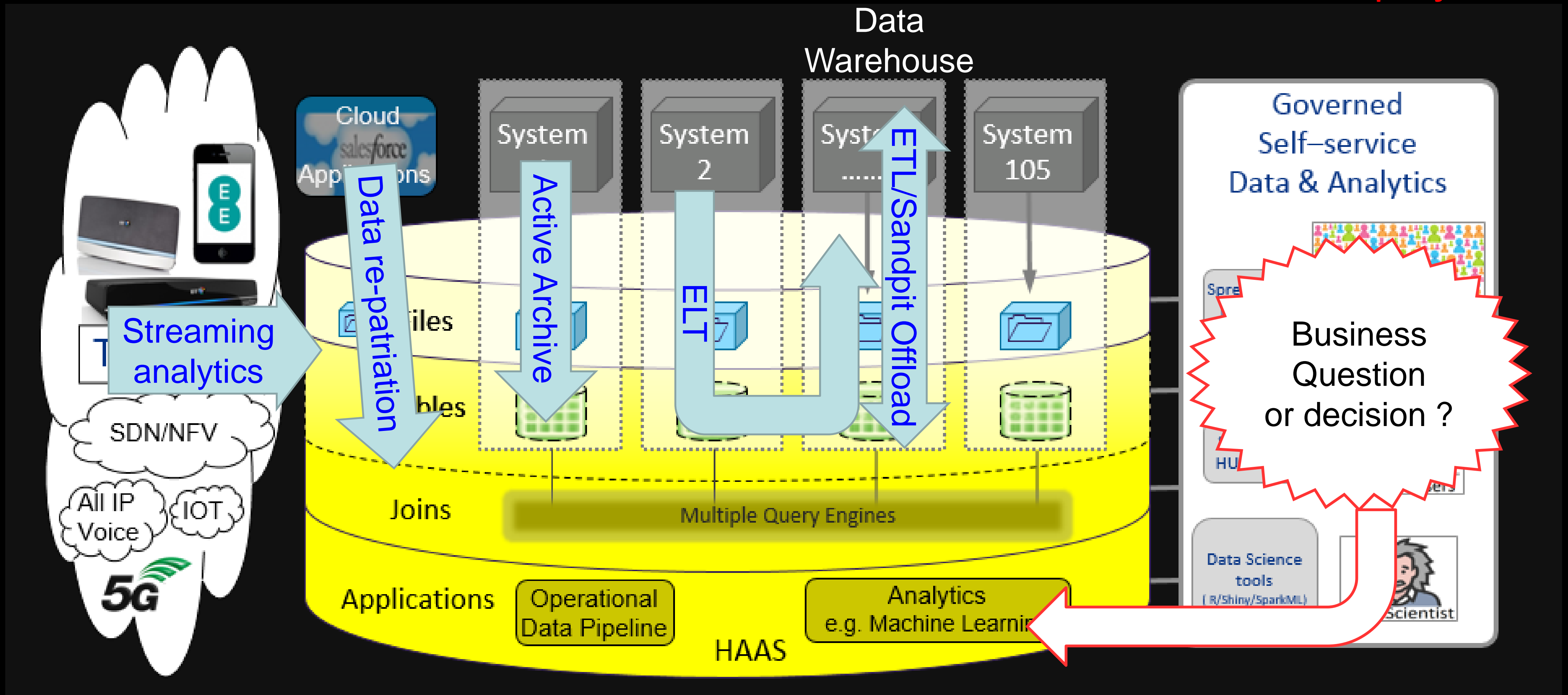


Legacy systems

# Adoption / On-boarding Patterns

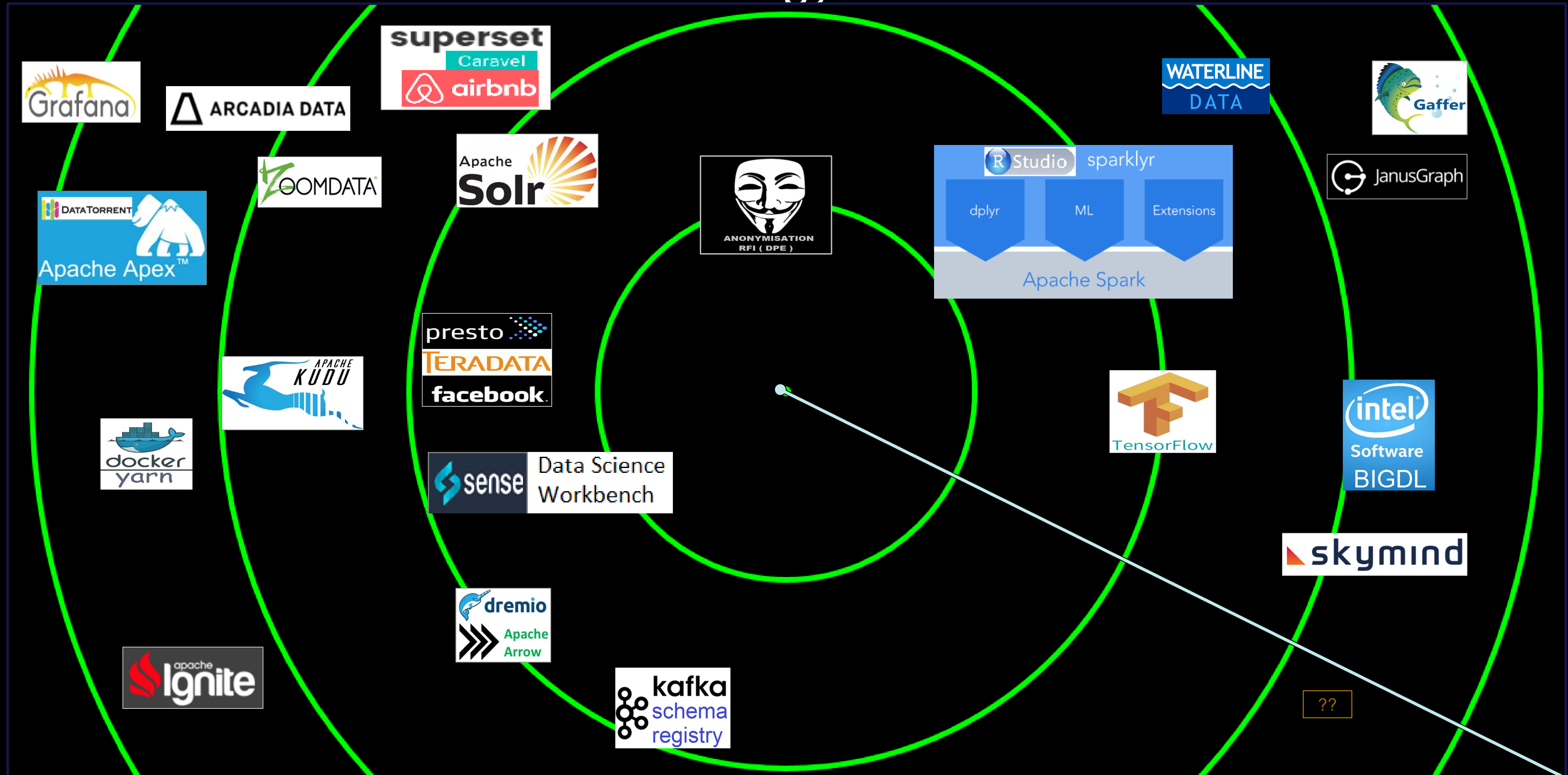
IT led projects

Business led projects





# Technology Radar



# Hadoop as a Service

# Questions ?

**Meet the Expert with Phillip Radley (BT)**

**12:05–12:45 Wednesday**

**Capital Hall (N24) - O'Reilly booth - Table B**

Rating & slides @

[conferences.oreilly.com/strata/strata-eu/public/schedule/detail/57589](https://conferences.oreilly.com/strata/strata-eu/public/schedule/detail/57589)

[phillip.radley@bt.com](mailto:phillip.radley@bt.com)



# Hadoop as a Service, how to build & operate an enterprise data lake supporting operational & streaming analytics

- Migrating from MapReduce to YARN, Spark, and Impala
- Migrating a live cluster to a new hardware stack
- Adding HDFS encryption
- Adding in second site for disaster recovery (adopting Airbnb's gold/silver cluster approach)
- Evolving the architecture to support streaming applications (using Kafka, Spark, and Flink)