BE READY. BE SAFE. BE SECURE.

BSIDES LISBON 2015 SECURITY DATA METRICS AND MEASUREMENTS AT SCALE

WHO

TIAGO HENRIQUES

CEO and Founder @ BinaryEdge

- BSc Software Engineering / University of Brighton
- MSc Computer Security and Forensics / University of Bedfordshire
- 8 Years experience in Information Security consultancy, leadership and research

TIAGO MARTINS

CTO and Co-Founder @ BinaryEdge

- BSc and MSc Computer Science / University of Lisbon
- 7 Years experience developing real-time systems and high-volume data processing

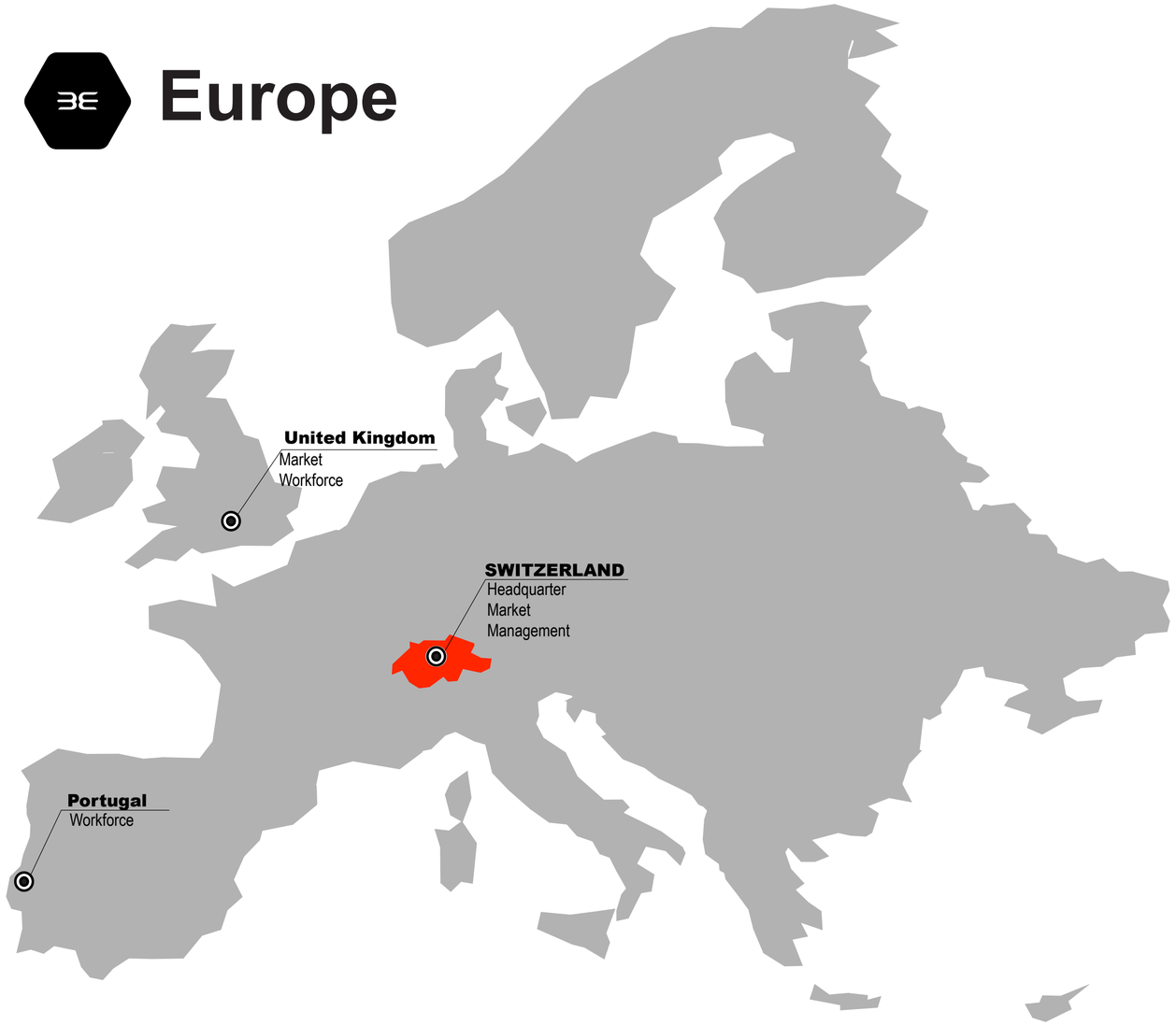
ROBERTO BARBOSA

COO and Head of DataScience at BinaryEdge

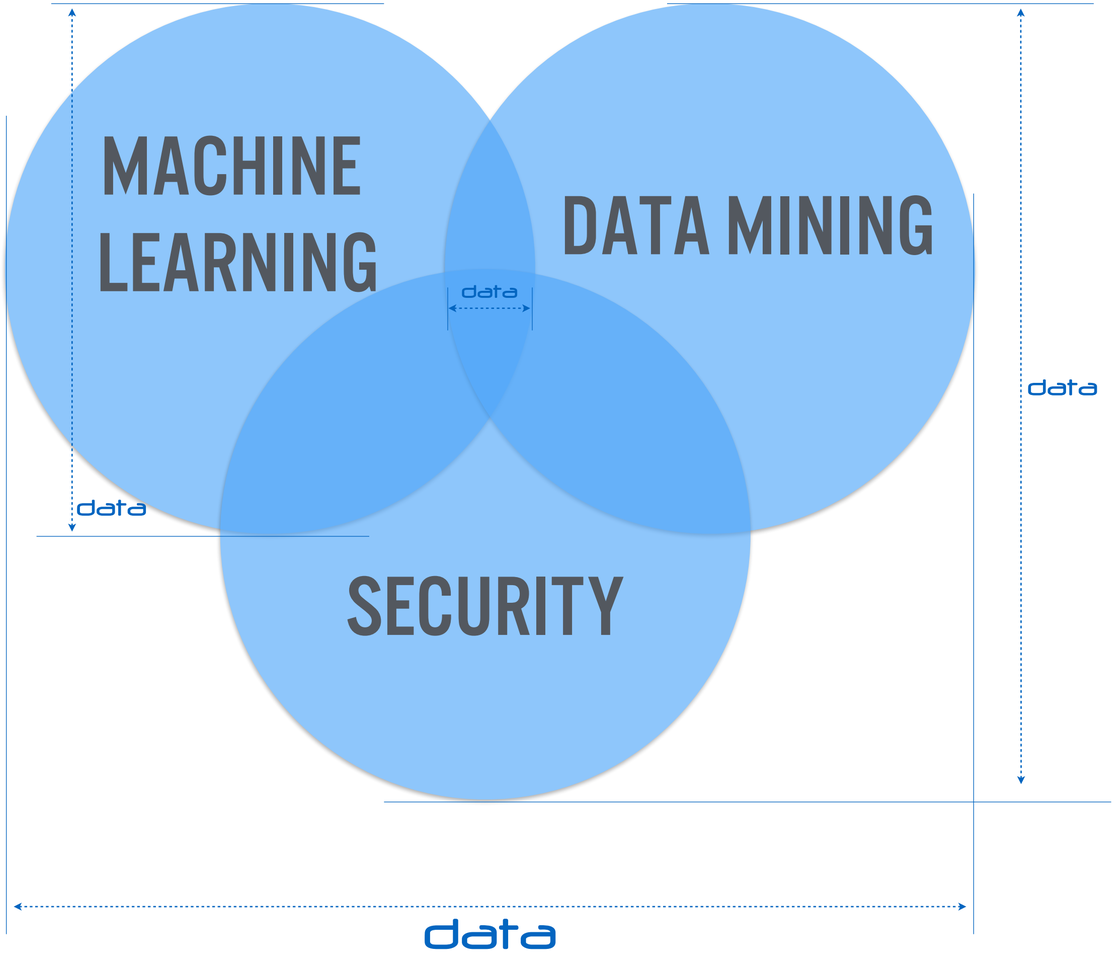
- More than 20 years on the IT sector
- ex-Engineer at Sun Microsystems
- Former Philip Morris corporate Auditor
- expert on High Scalability and Availability on the Finance Sector (UBS, Citigroup and Leonteq) and mobile startup.

WHERE

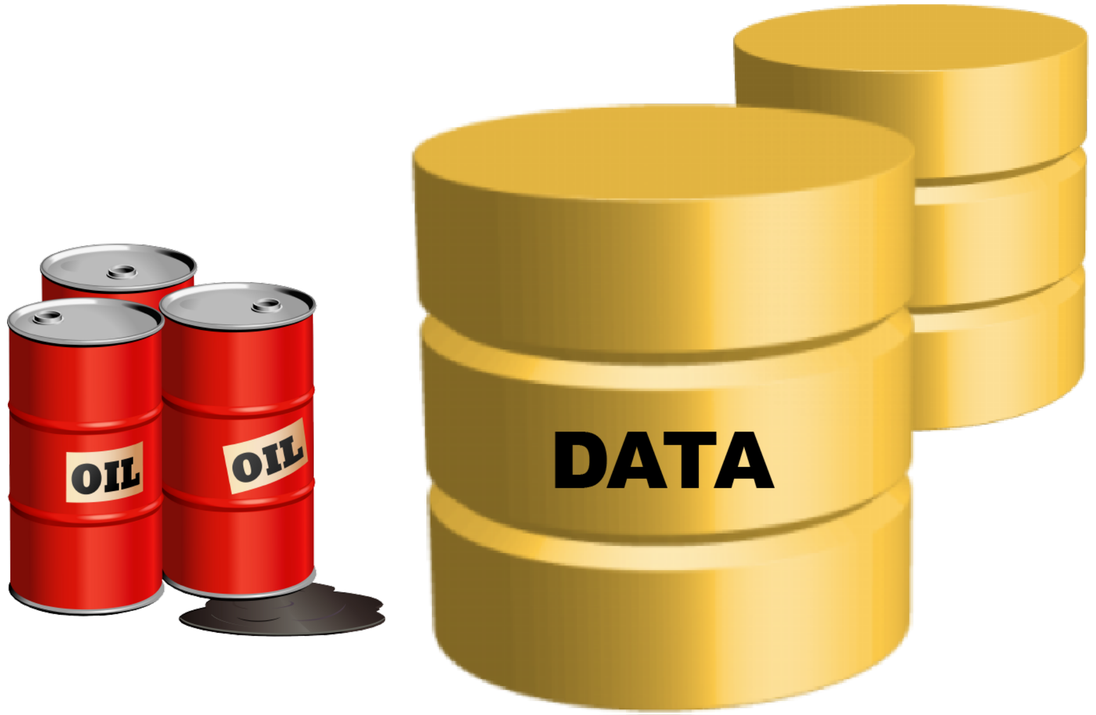
Europe



WHAT



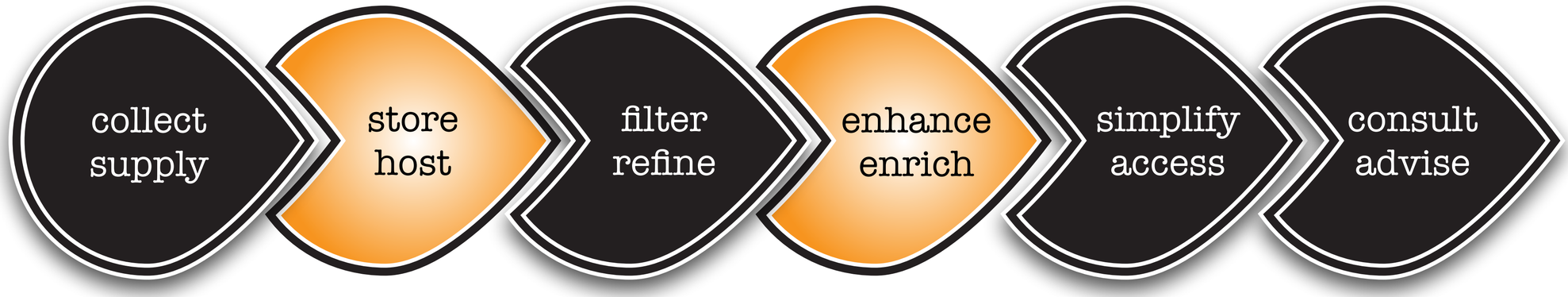
NEW COMMODITY



DATA IS THE NEW OIL

NEW CURRENCY

DATA BUSINESS MODEL



Gather and Sell raw Data

Hold onto someone else's data for them

Strip out problematic records or data fields or release interesting data subsets

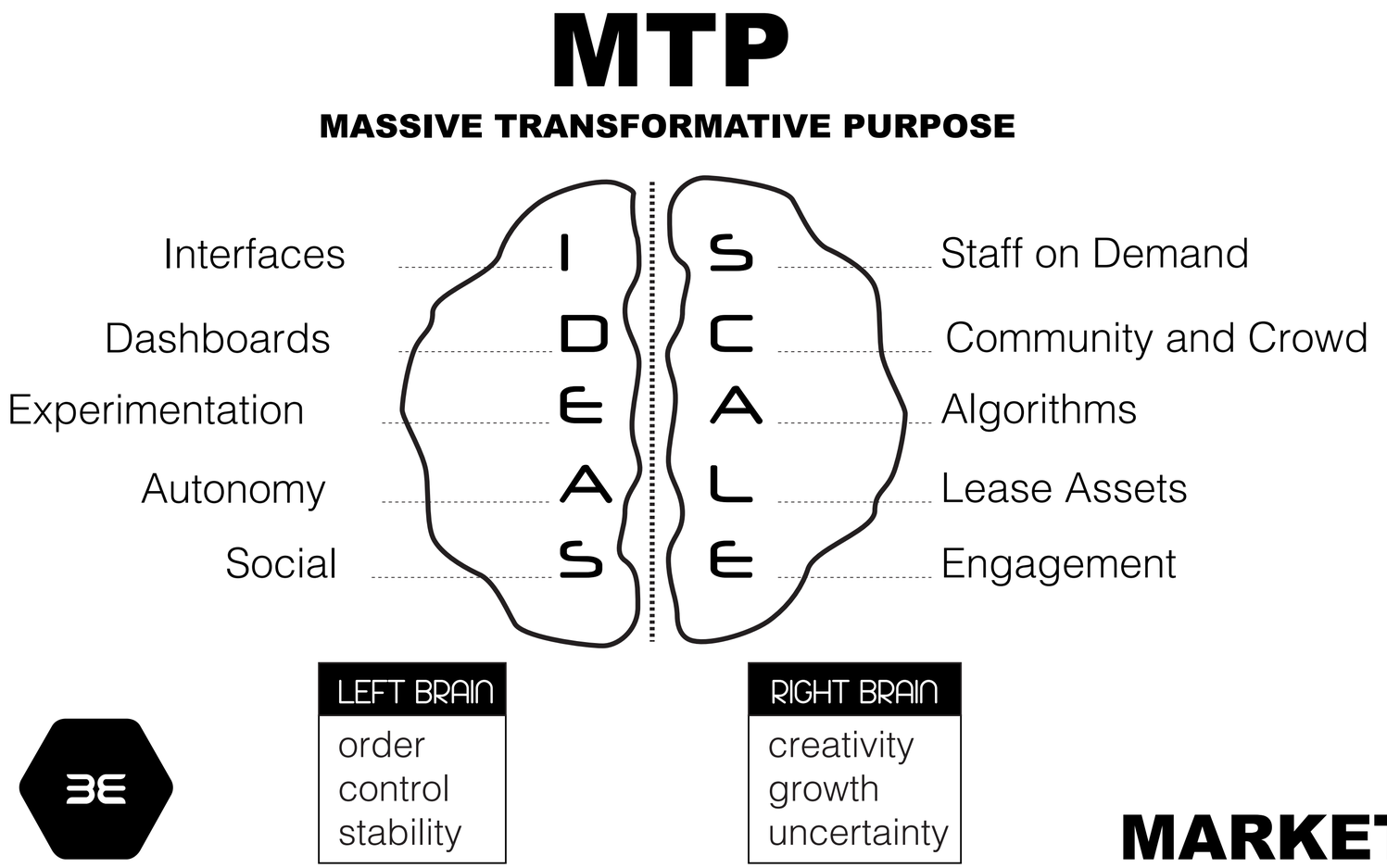
Blend in other datasets to create a new and interesting picture

Help people cherry-pick the data they want in the format they prefer

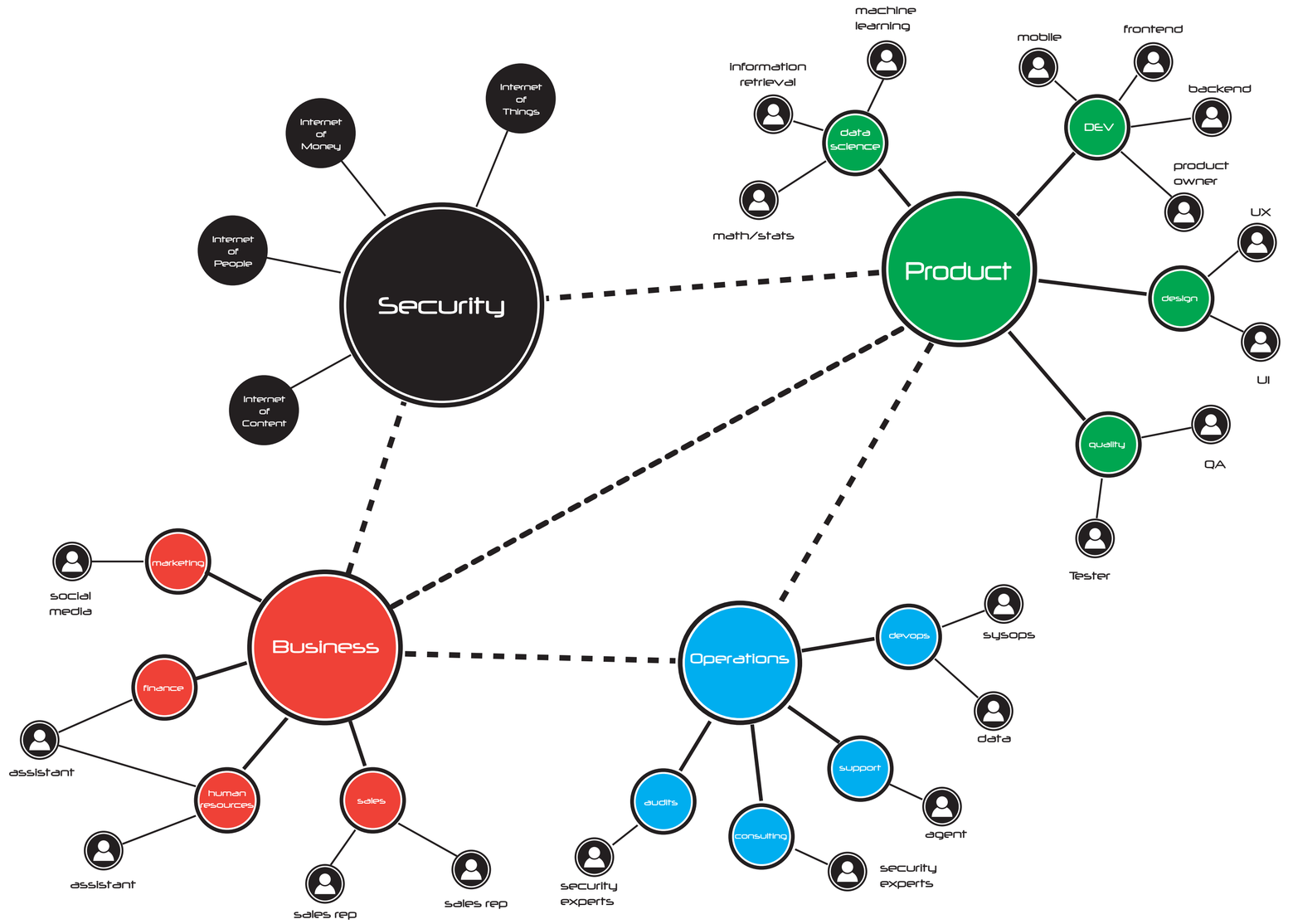
Provide guidance on others' data efforts

ORGANISATION

BECOMING EXPONENTIAL ORGANIZATION



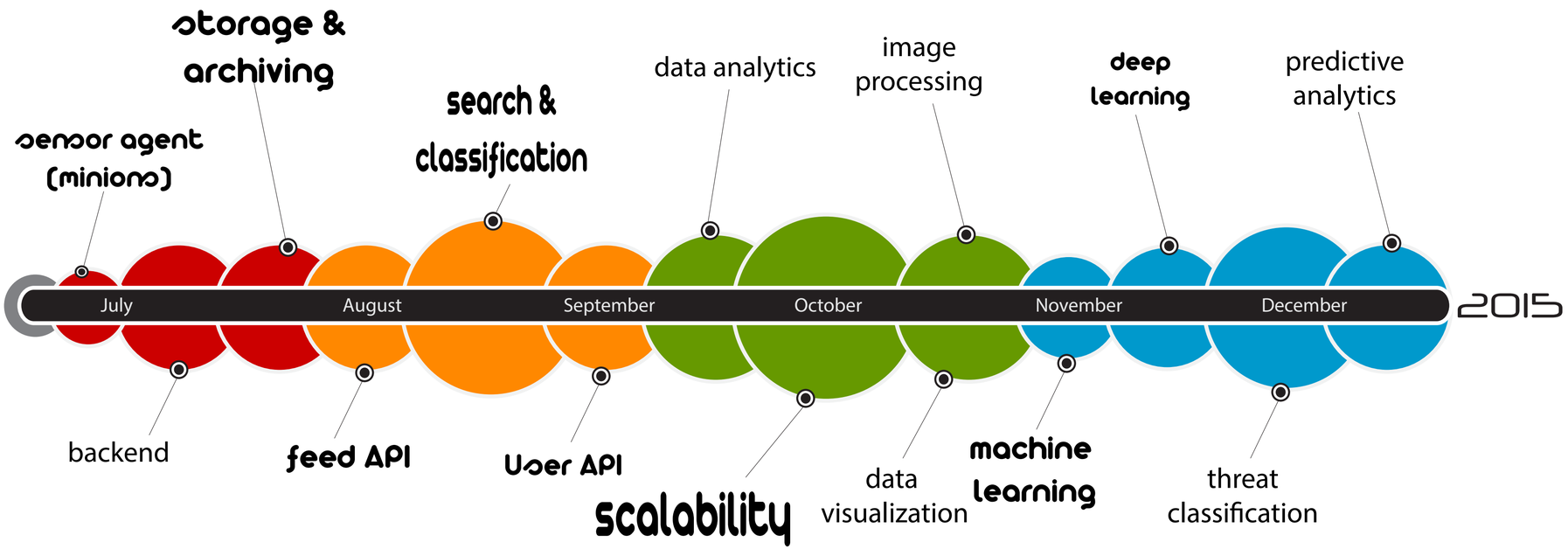
ORGANISATION RELATIONSHIP



DATA ARCHITECTURE DESIGN

Goals	Requirements	Results
• EASY TO UNDERSTAND	• UNDERSTANDABILITY	Simple Architecture
• EASY TO EXTEND	• EXTENSIBILITY	Loosely Coupled Services
• EASY TO CHANGE	• CHANGEABILITY	Built for replacement
• EASY TO REPLACE	• REPLACEABILITY	Self-dependency
• EASY TO DEPLOY	• DEPLOYABILITY	Immutability
• EASY TO SCALE	• SCALABILITY	Responsibility Segretation
• EASY TO RECOVER	• RESILIENCE	Decoupling and Isolation
• EASY TO CONNECT	• UNIFORM INTERFACE	API based
• EASY TO AFFORD	• COST EFFICIENT	On-demand computing

PRODUCT IMPROVEMENTS 2015



data - - - - - - - - - - → information - - - - - - → knowledge - - - - - - - → intelligence

MILESTONES

milestone

EFFORT

less

MORE

IMPORTANCE

average

GREATER

PHASE

POC | data | process | analysis | intel

ENGINEERING

METRICS COLLECTION AT LARGE SCALE

NO LEGACY TO MAINTAIN
LOTS OF EXPERIENCE IN THE TEAM
LOTS OF TECHNOLOGIES TO PICK FROM
MICRO SERVICE BASED APPROACH

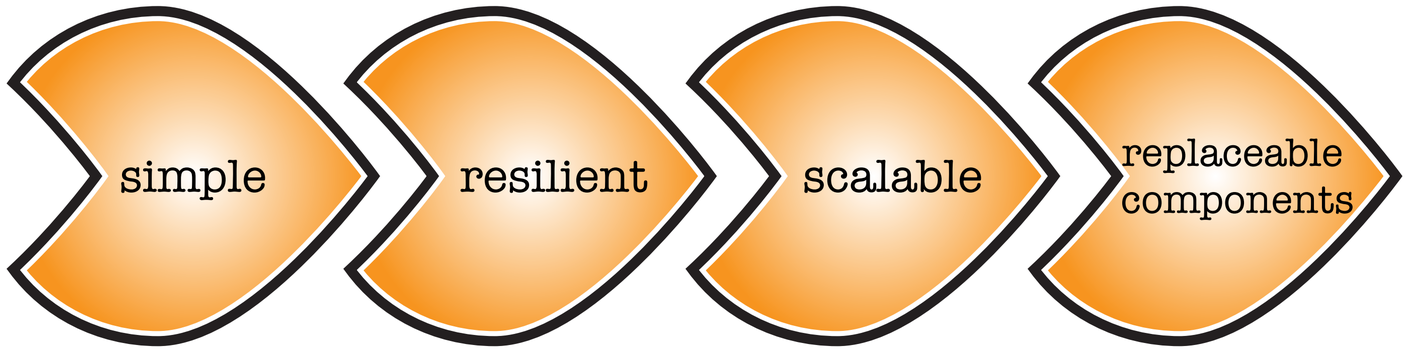
TECHNOLOGIES?
ARCHITECTURE?
PROTOTYPE?

METRICS COLLECTION AT LARGE SCALE

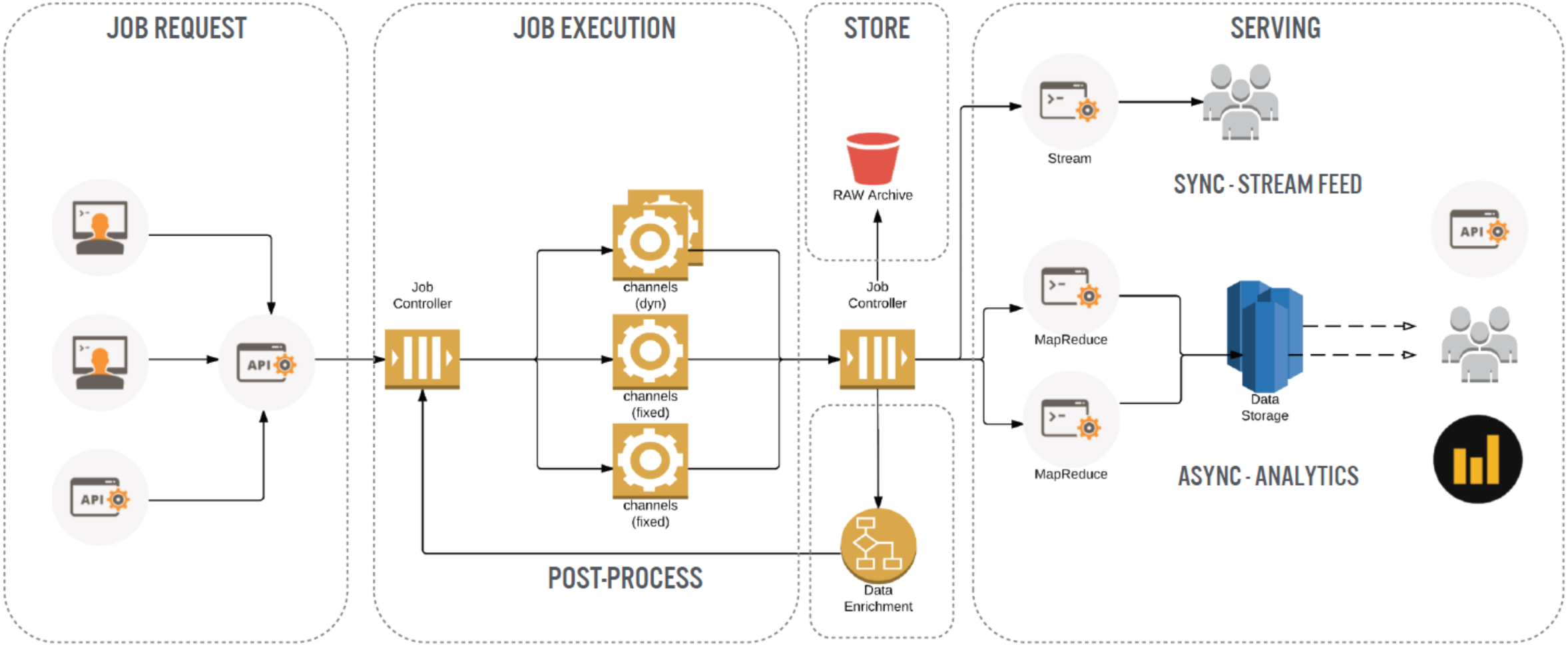
ARCHITECTURE

FOCUS ON ARCHITECTURE

TECHNOLOGY INDEPENDENT



ARCHITECTURE OVERVIEW



ARCHITECTURE - JOB REQUEST

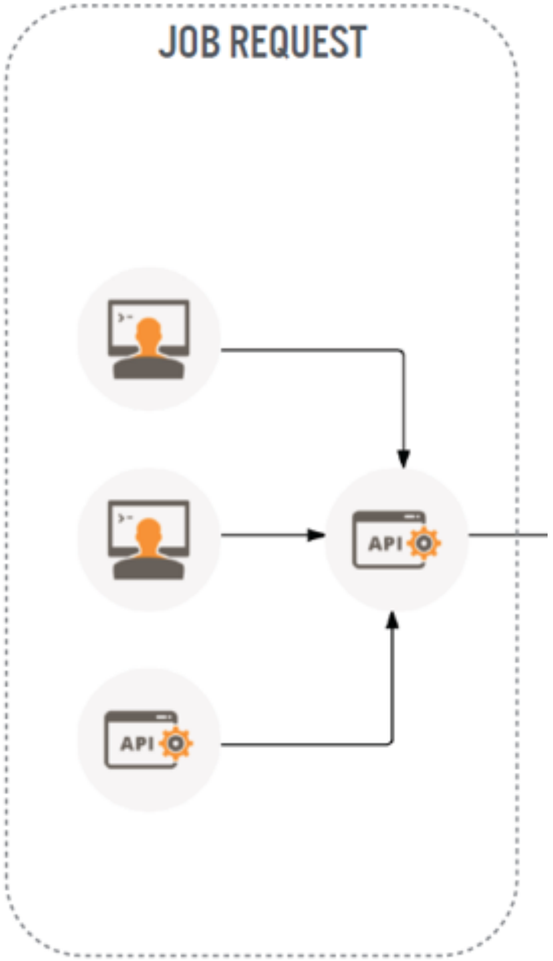
API ORIENTED

HTTP API
COMMAND LINE CLIENTS
MODULES
- PYTHON
- NODEJS
- GO

THIRD-PARTY APIS

JOB TYPES

DATA COLLECTION
DATA PROCESSING / ANALYTICS



ARCHITECTURE - JOB EXECUTION

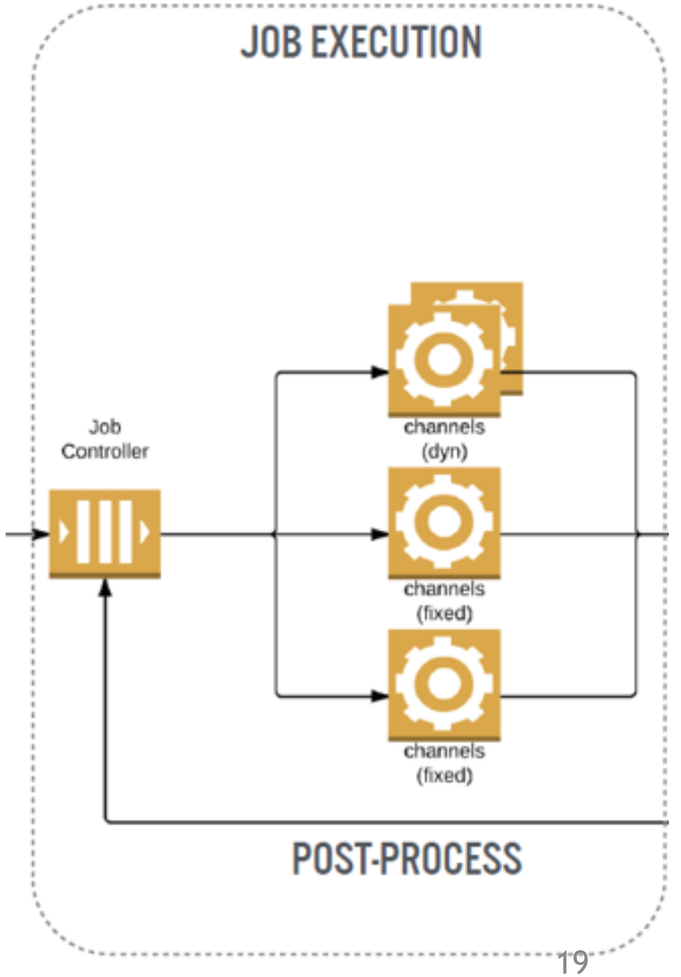
AGENTS LISTEN FOR WORK IN CHANNELS
TECHNOLOGIES
MULTIPLE TYPES OF AGENTS

AGENTS

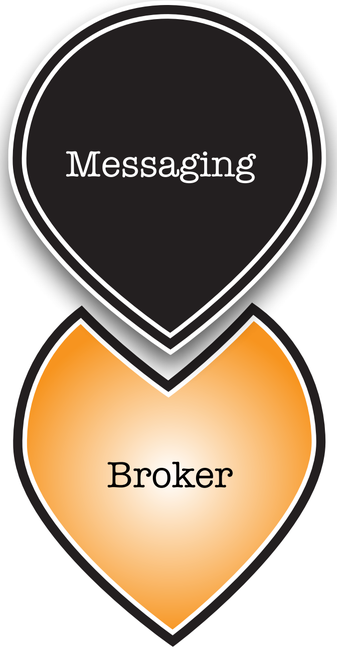
GO
PYTHON
NODEJS
SCALA
JAVA

JOB CONTROL

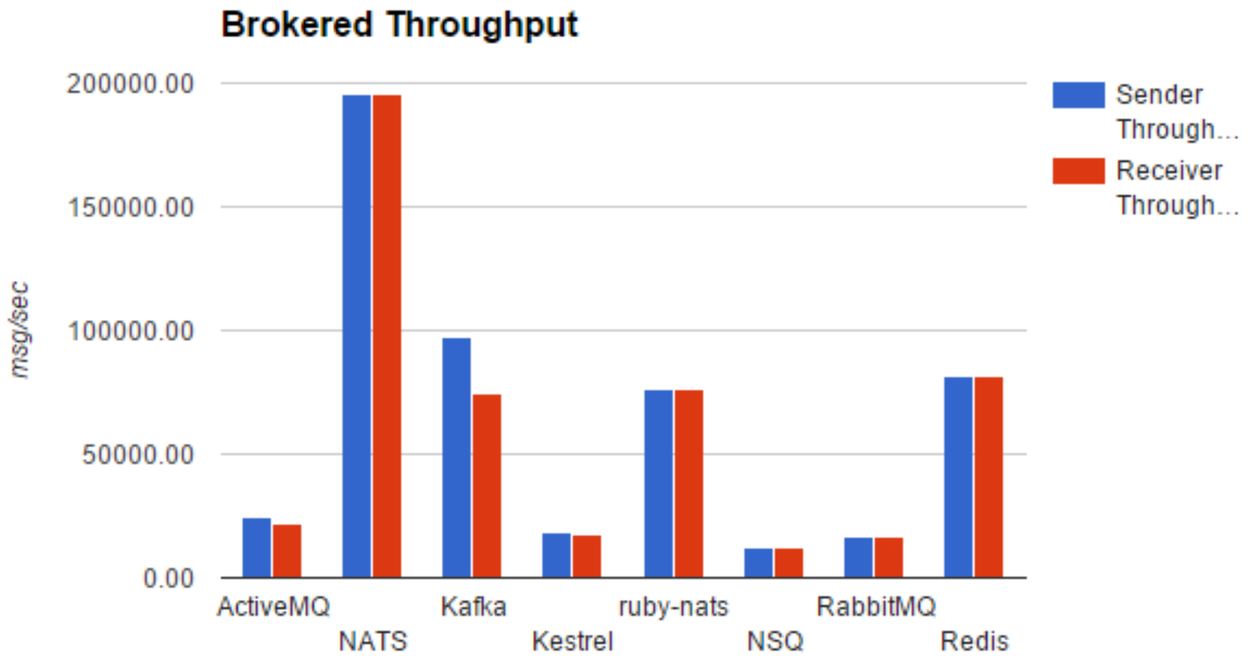
RABBITMQ
NSQ
REDIS
APOLLO



ARCHITECTURE - JOB EXECUTION



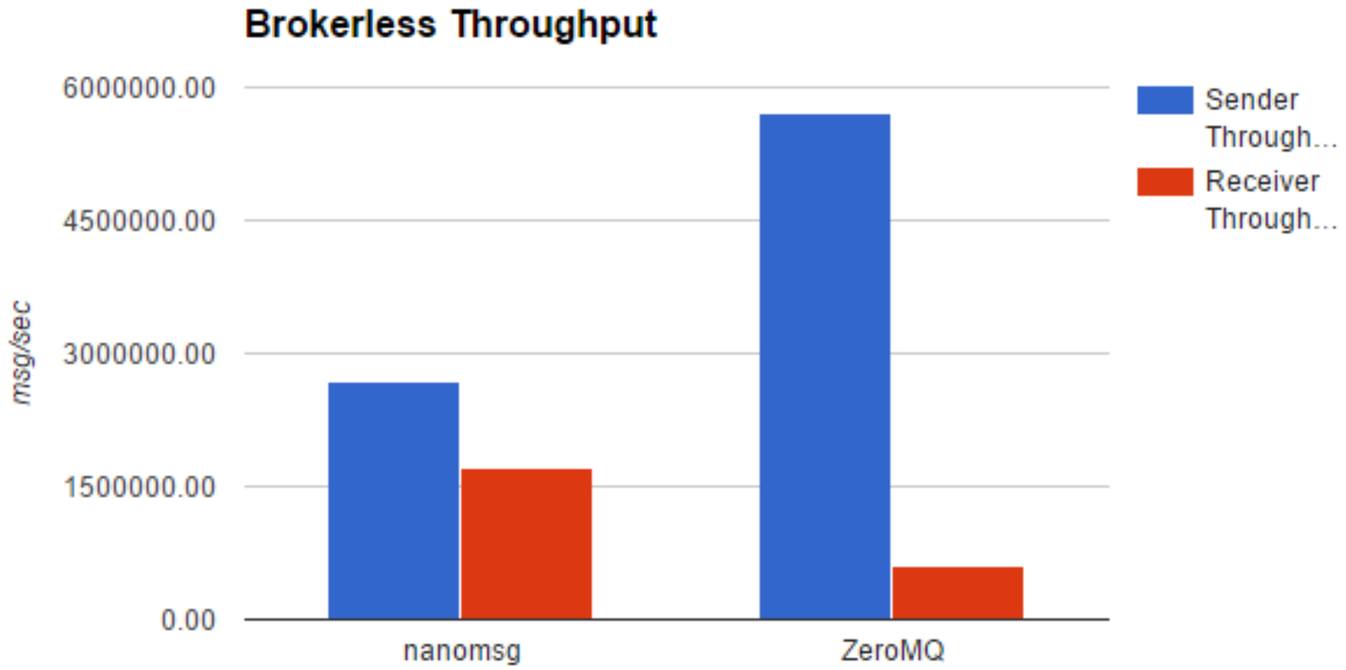
ACTIVEMQ
NATS
KAFKA
KESTREL
NSQ
RABBITMQ
REDIS
QPID
HORNETQ
APOLLO



http://bravenewgeek.com/dissecting-message-queues/

ARCHITECTURE - JOB EXECUTION

ZEROMQ
NANOMSG



http://bravenewgeek.com/dissecting-message-queues/

ARCHITECTURE - JOB EXECUTION

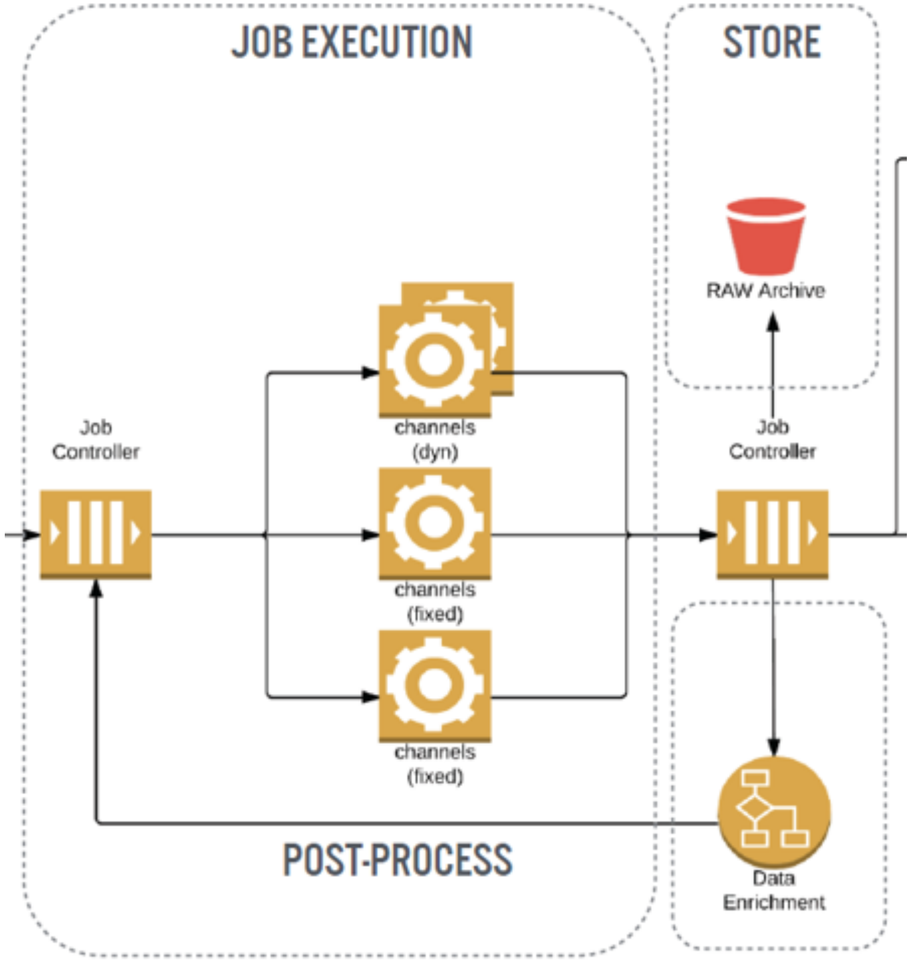
AMAZON
MICROSOFT
GOOGLE
REALTIME.CO
...

ARCHITECTURE - DATA ENRICHMENT

AGENTS CAN FEED OTHER AGENTS

DIFFERENT TYPES OF ENRICHMENT

- CLEAN DATA
- PROCESS DATA
- ALARMS



ARCHITECTURE - STORE

ALL INFORMATION IS STORED
- RAW DATA
- PROCESSED DATA

GEOLOCATE OF INFORMATION

ENCRYPTED DATA FOR EACH CLIENT

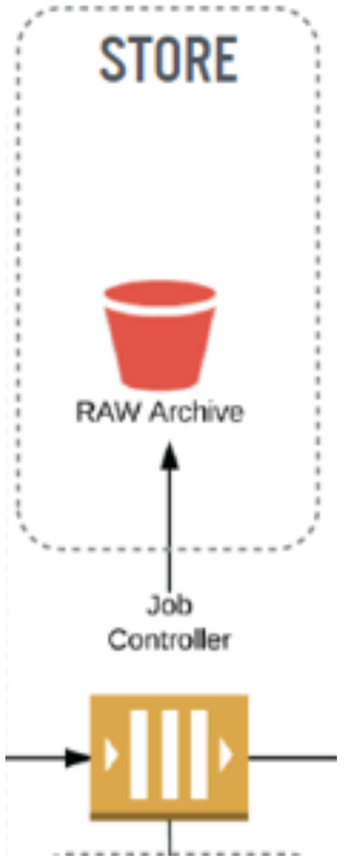
DATA STORAGE

DATABASE SOLUTIONS

- MONGODB
- ELASTICSEARCH
- CASSANDRA
- RIAK
- LUCENE

CLOUD SERVICES

- AMAZON S3
- AMAZON DYNAMODB
- AZUREDOCUMENTDB
- AZURE STORAGE
- GOOGLE CLOUD STORAGE
- GOOGLE BIGQUERY
- RACKSPACE CLOUD FILES
- CONSTANT CLOUD STORAGE
- SKYLABLE
- RUNABOVE



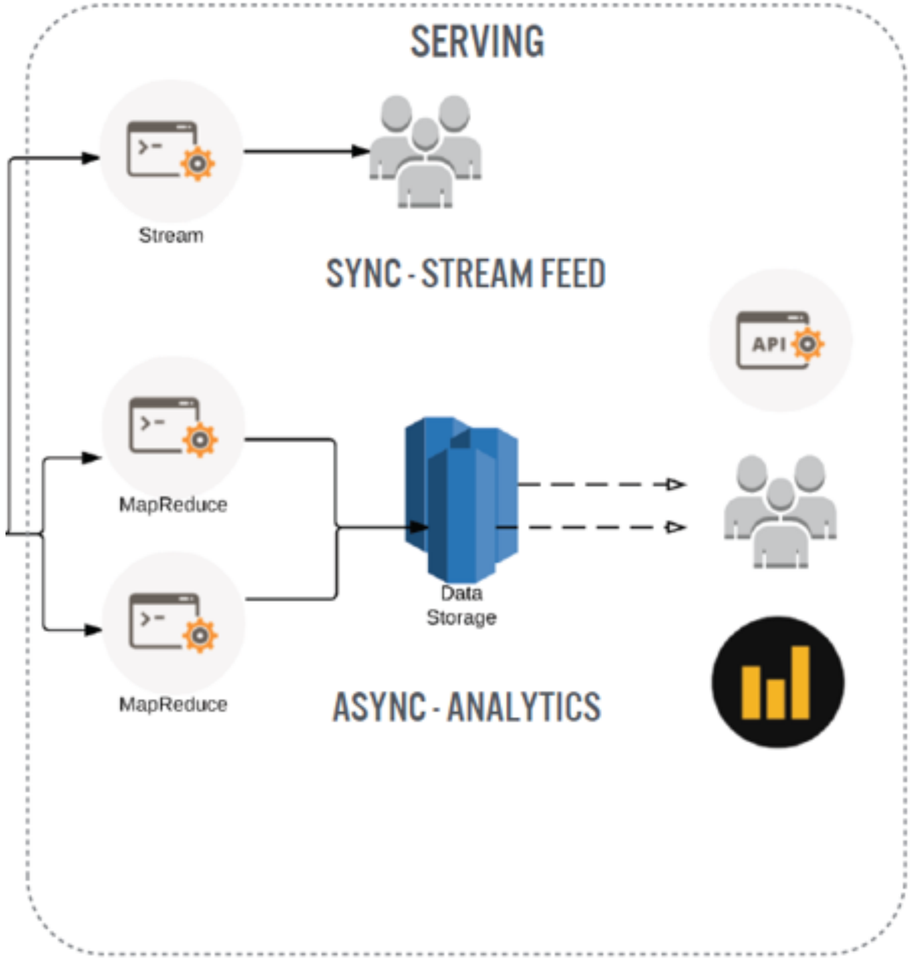
ARCHITECTURE - SERVING

DELIVERING DATA

- REALTIME - STREAMING
- STORAGE FOR ANALYTICS
- API
- RAW

DATA ANALYTICS

- KIBANA
- INFLUXDB
- DRUID



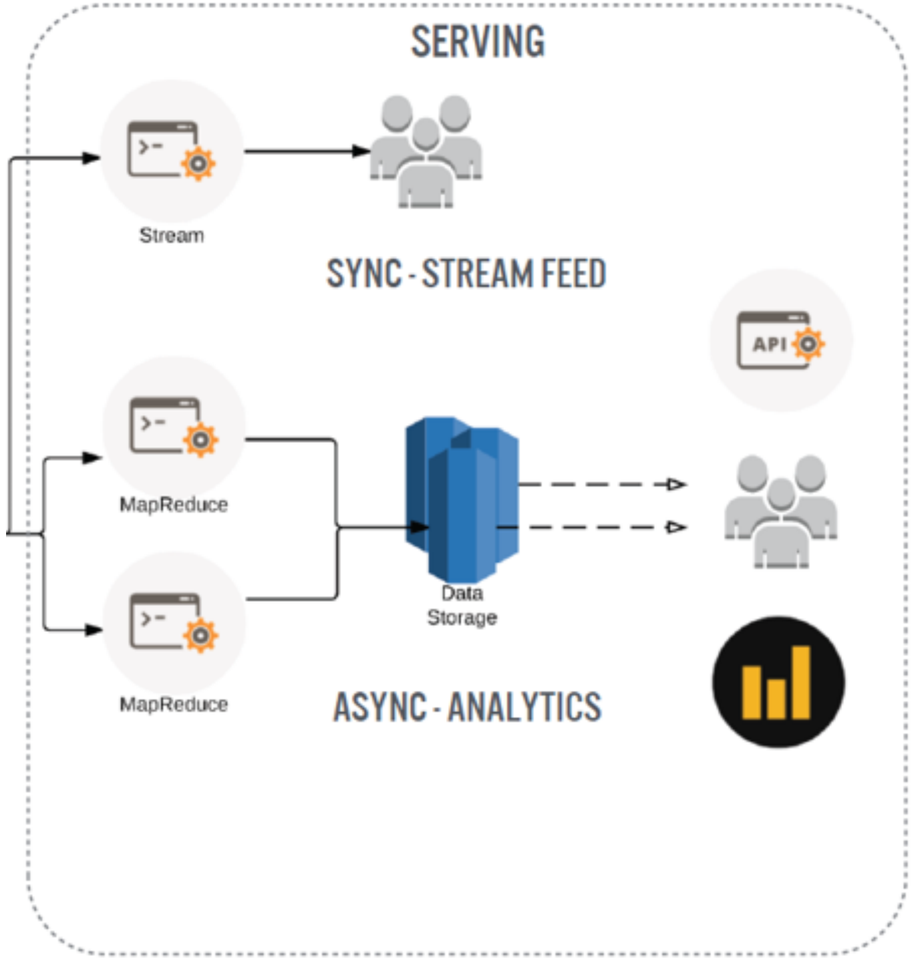
ARCHITECTURE - SERVING

DATA PROCESSING

- APACHE SPARK
- HADOOP
- AMAZON KINESIS

DATA INTELLIGENCE

- AMAZON MACHINE LEARNING/EMR
- GOOGLE PREDICTION API
- AZURE MACHINE LEARNING



AGENTS/ MINIONS

OUR AGENTS ARE VERY SIMPLE

- SIMPLE TASKS
- EASY TO MAINTAIN AND ADAPT

AGENTS CAN BE LOCATED/RUN ANYWHERE

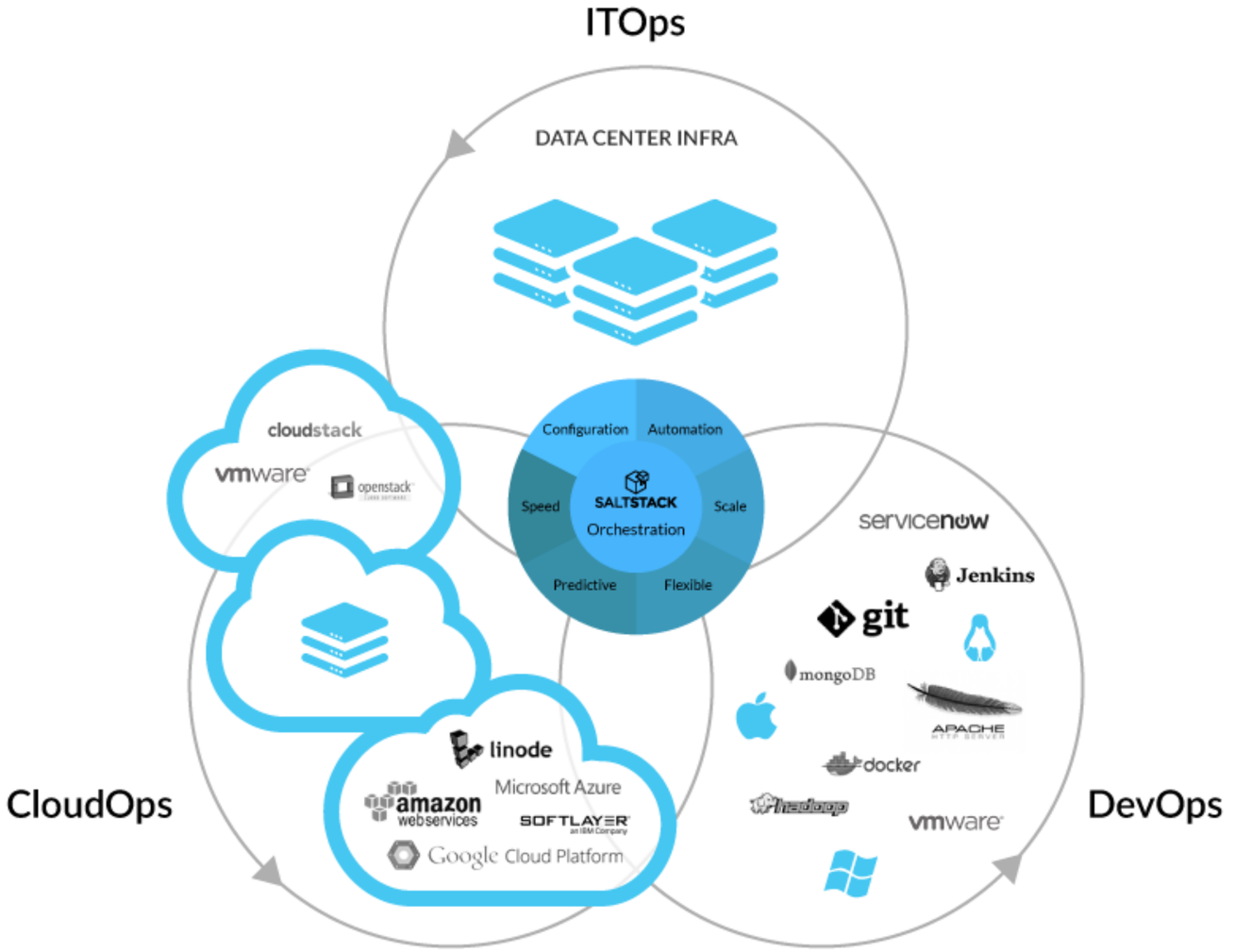
- GEO DISTRIBUTION
- CLOUDS
- DEDICATED SERVERS
- RASPBERRY PIS IN TIAGO HENRIQUES' DUAL GBIT CONNECTION

MONITORING

NEW RELIC
LOGENTRIES
SERVER DENSITY
CLOUD WATCH
GRAFANA
LOGSTASH

DEPLOYMENT

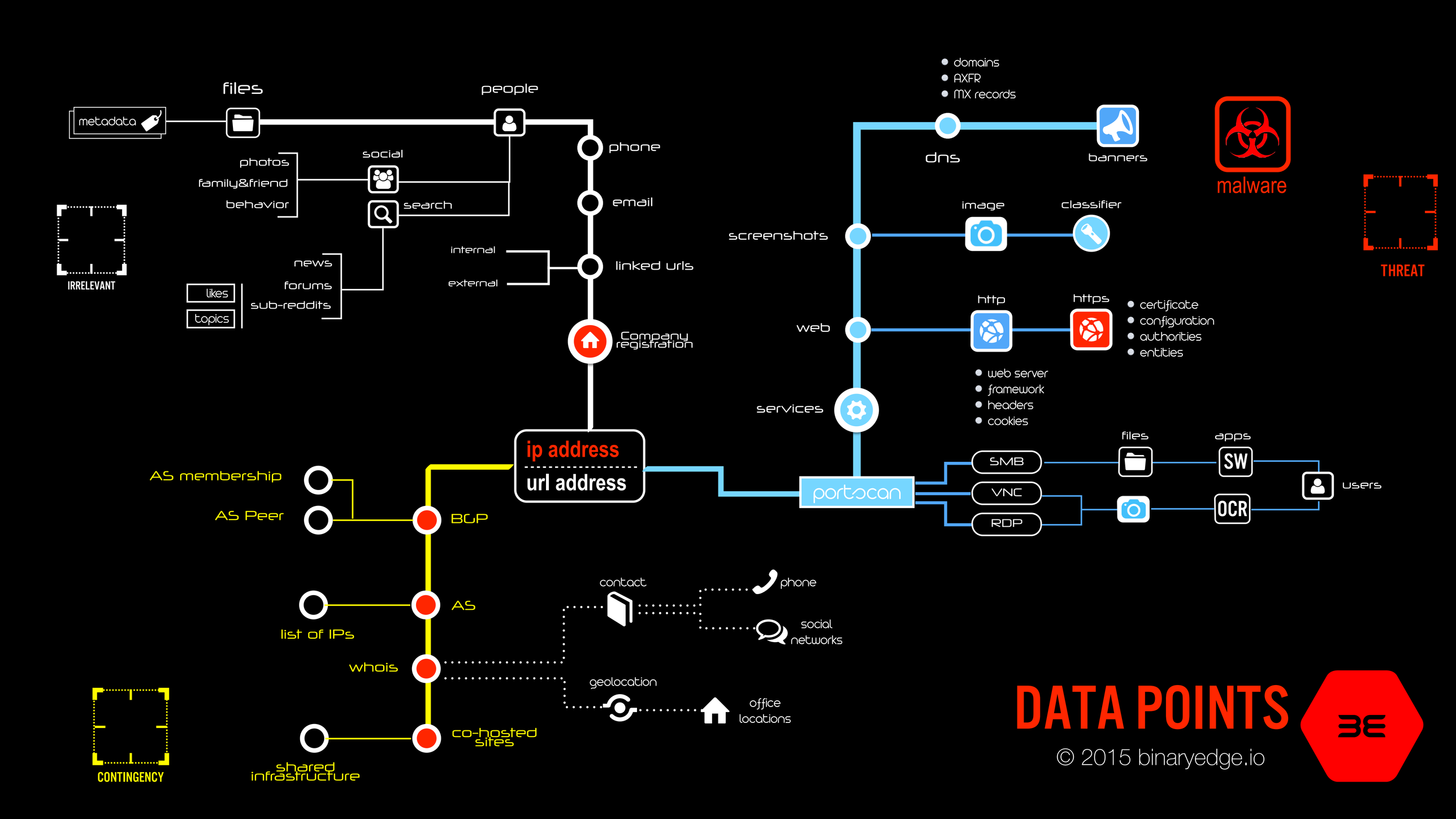
ANSIBLE
PUPPET
DOCKER
SALTSTACK
ETCD



SALTSTACK.COM

MACHINE LEARNING

CHALLENGES IN DATA MINING


files
people
metadata
photos
family&friend
behavior
social
search
phone
email
internal
external
linked urls
news
forums
sub-reddits
likes
topics
Company registration
IRRELEVANT
ip address
url address
AS membership
AS Peer
BGP
AS
list of IPs
whois
co-hosted sites
shared infrastructure
CONTINGENCY
contact
phone
social networks
geolocation
office locations
domains
AXFR
MX records
dns
banners
malware
THREAT
screenshots
image
classifier
web
http
https
certificate
configuration
authorities
entities
services
web server
framework
headers
cookies
portscan
SMB
VNC
RDP
files
apps
SW
OCR
users
DATA POINTS
© 2015 binaryedge.io

MACHINE LEARNING TECHNIQUES

- ARTIFICIAL NEURAL NETWORK (ANN)
- SUPPORT VECTOR MACHINE (SVM)
- DECISION TREES
- BAYESIAN NETWORKS (BNS)
- K-NEAREST NEIGHBOUR (KNN)
- HIDDEN MARKOV MODEL (HMM)

MACHINE LEARNING - WHY?

CLASSIFICATION

DETECTION

CLUSTERING

AUTOMATION

CORRELATION

PREDICTION

ANALYSIS

MEASUREMENTS ON OUR OWN DATA

SUPPORT - INDICATES WHICH PERCENTAGE OF DATA ON STORAGE SHOWS CORRELATION

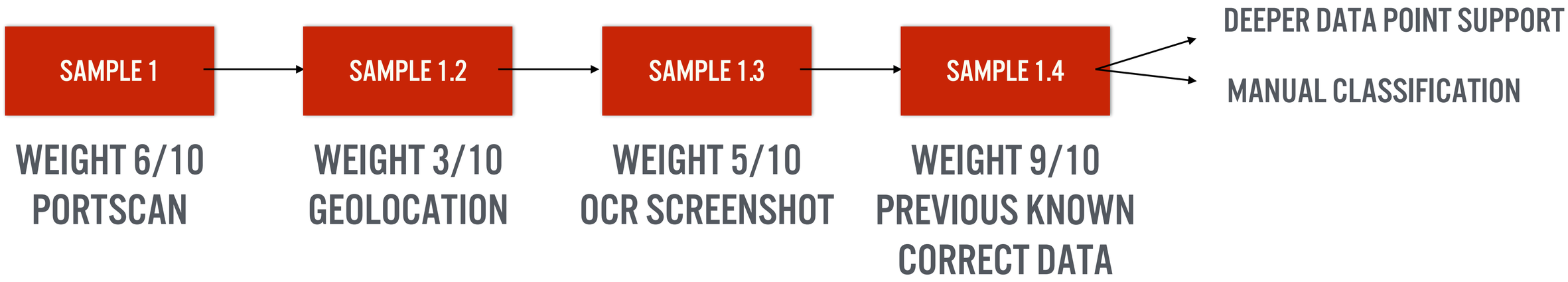
$$Support(A \Rightarrow B) = \frac{|\#T_i \mid A, B \in T_i|}{N},$$

CONFIDENCE - INDICATES PROBABILITY OF OUR ASSUMPTION BEING CORRECT

$$Confidence(A \Rightarrow B) = \frac{|\#T_i \mid A, B \in T_i|}{|\#T_i \mid A \in T_i|}.$$

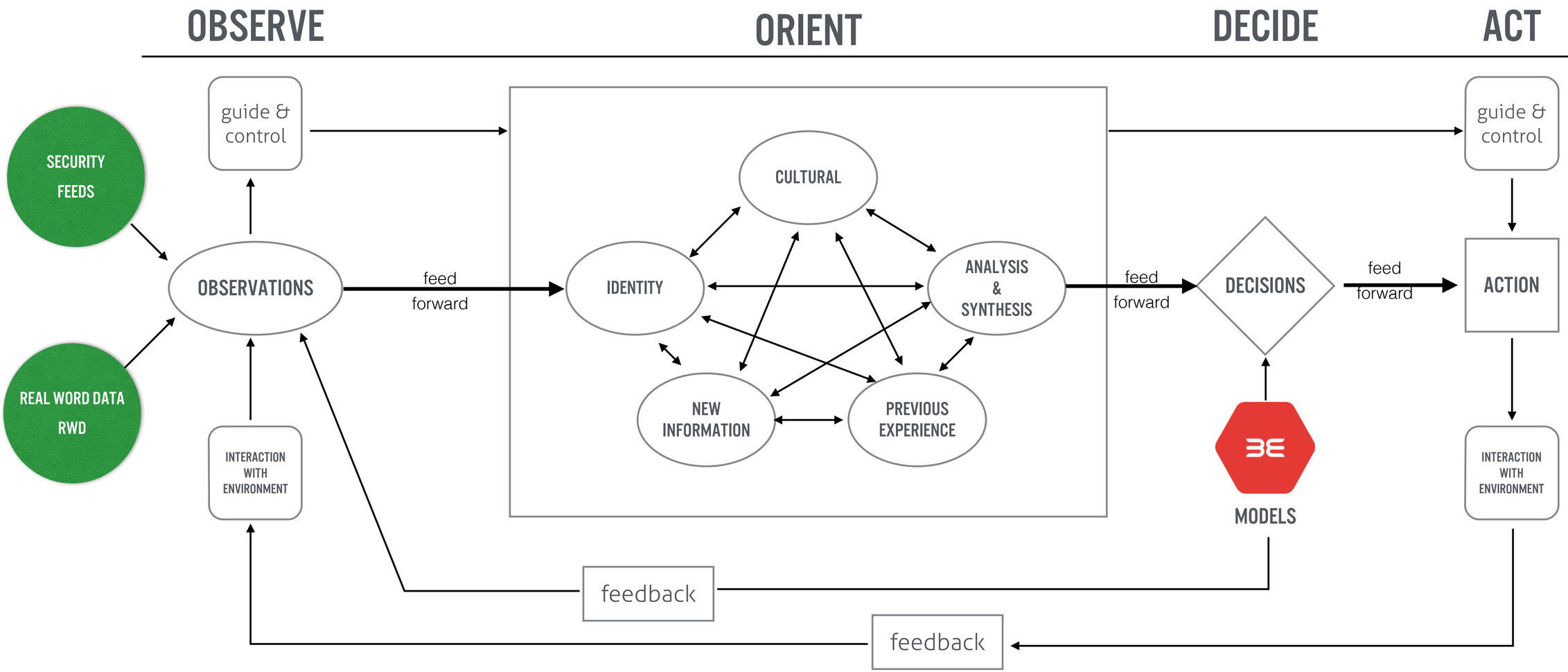
IMPROVING OUR OWN DATA

- KALMAN FILTER
- ADABOOST (ADAPTIVE BOOST)

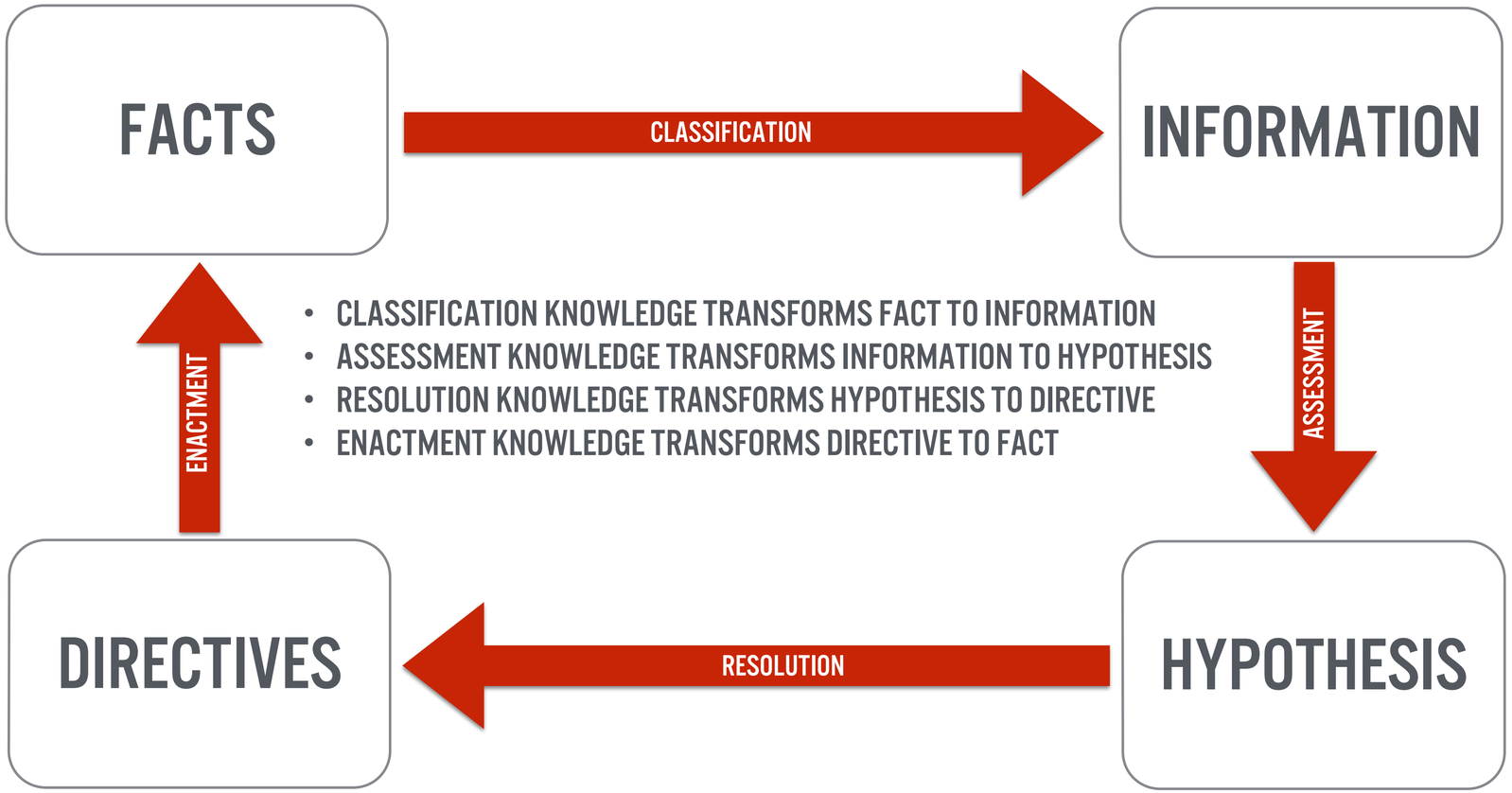


DATA CHAIN

CYBER INNOVATION LOOP



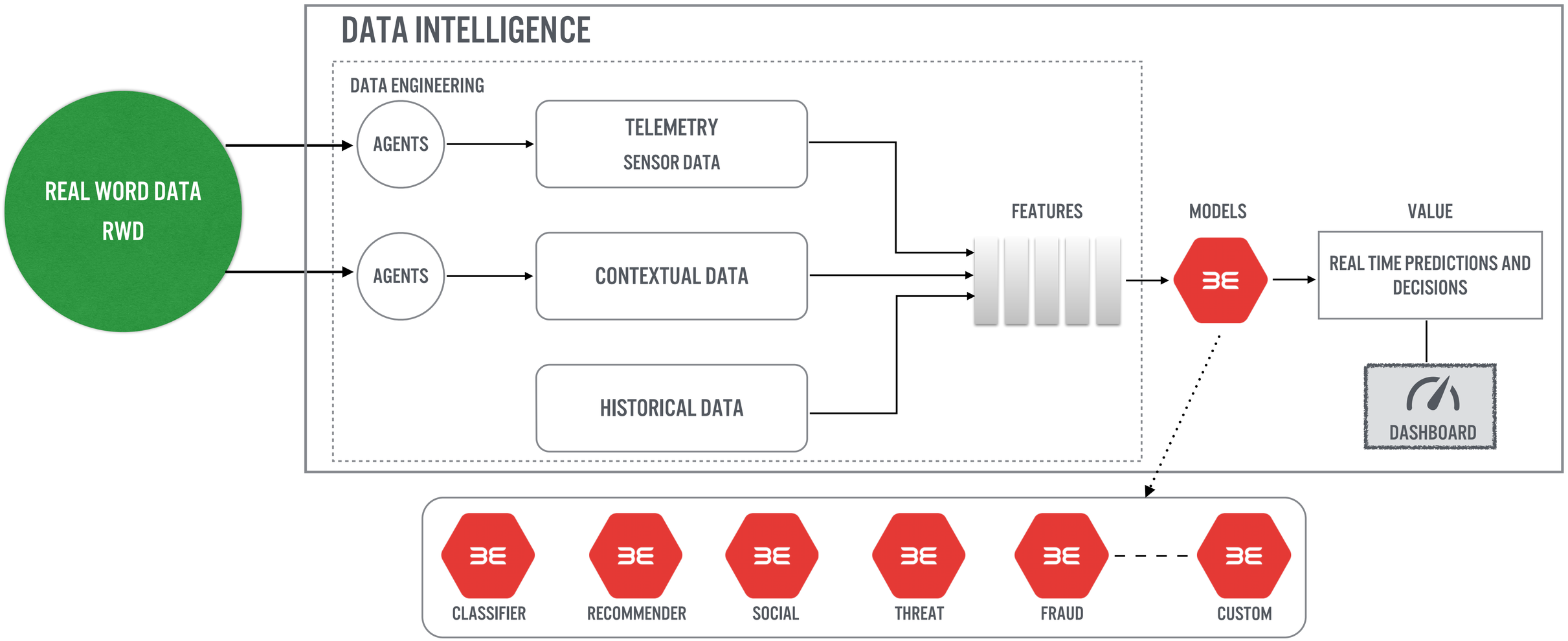
CYBER INNOVATION LOOP



KNOWLEDGE

DATA

CYBERSECURITY DATA SCIENCE



DEMO

DEMO

Port scanning from IP address [redacted]

The security team CSIRT-MU detected involvement of the IP address [redacted] llowing incident:

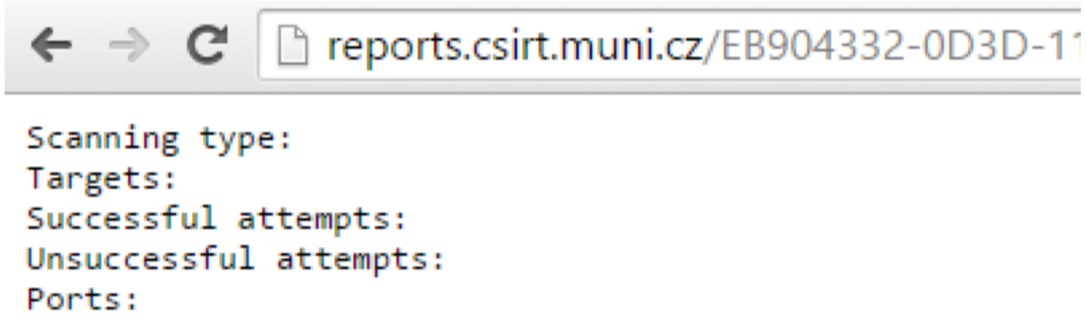
Incident type:	Port scanning
Time of detection:	2015-06-07 19:35:00 +0200
IP address:	[redacted]
Domain name:	---

Incident details can be found in the following files: details.txt

Best regards
CSIRT-MU, the security team of Masaryk University
http://csirt.muni.cz

Detailed information about the incident and the used detection method

DEMO


reports.csirt.muni.cz/EB904332-0D3D-11
Scanning type:
Targets:
Successful attempts:
Unsuccessful attempts:
Ports:

DEMO

IP Address 46.101.25.227 **is listed** in the CBL. It shows signs of being infected with a spam sending trojan, malicious link or some other form of botnet.

It was last detected at 2015-06-05 16:00 GMT (+/- 30 minutes), approximately 2 days, 23 hours, 29 minutes ago.

This IP is infected (or NATting for a computer that is infected) with the **Conficker** botnet.

More information about Conficker can be obtained from Wikipedia

Please **follow** these instructions.

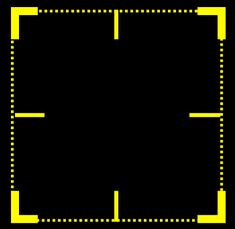
Dshield has a diary item containing many third party resources, especially removal tools such as Norton Power Eraser, Stinger, MSRT etc

One of the most critical items is to make sure that all of your computers have the MS08-067 patch installed. But even with the patch instal

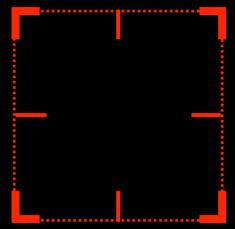
There are several ways to identify Conficker infections remotely. For a fairly complete approach, see Sophos.

If you have full firewall logs turned on at the time of detection, this may be sufficient to find the infection on a NAT:

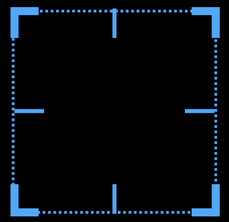
Your IP was observed making connections to TCP/IP IP address 38.229.183.123 (a conficker sinkhole) with a destination port 80, source
All of our detection systems use NTP for time synchronization, so the timestamp should be accurate within one second.



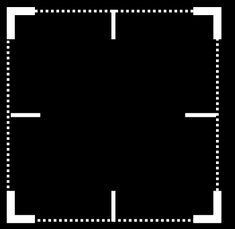
CONTINGENCY



THREAT



SAFE



IRRELEVANT

BE READY. BE SAFE. BE SECURE.

BINARYEDGE.IO

Finsterrütistrasse 4, 8134
Adliswil, ZURICH
Switzerland

+ 41 78 632 32 90

Email : th@binaryedge.io

www.binaryedge.io