

# BIG DATA ANALYTICS & THREAT PREDICTION

Armend Duzha  
EU Project Manager  
Maggioli Group | Santarcangelo di Romagna, Italy  
Tel.: +39 0541 628329  
Fax: +39 0541 621153  
E-mail: [armend.duzha@maggioli.it](mailto:armend.duzha@maggioli.it)  
[www.maggioli.it](http://www.maggioli.it)

06/10/2017

[www.mitigateproject.eu](http://www.mitigateproject.eu)



MITIGATE

CIP Workshop

September 19-20,  
2017



# IN THIS PRESENTATION...



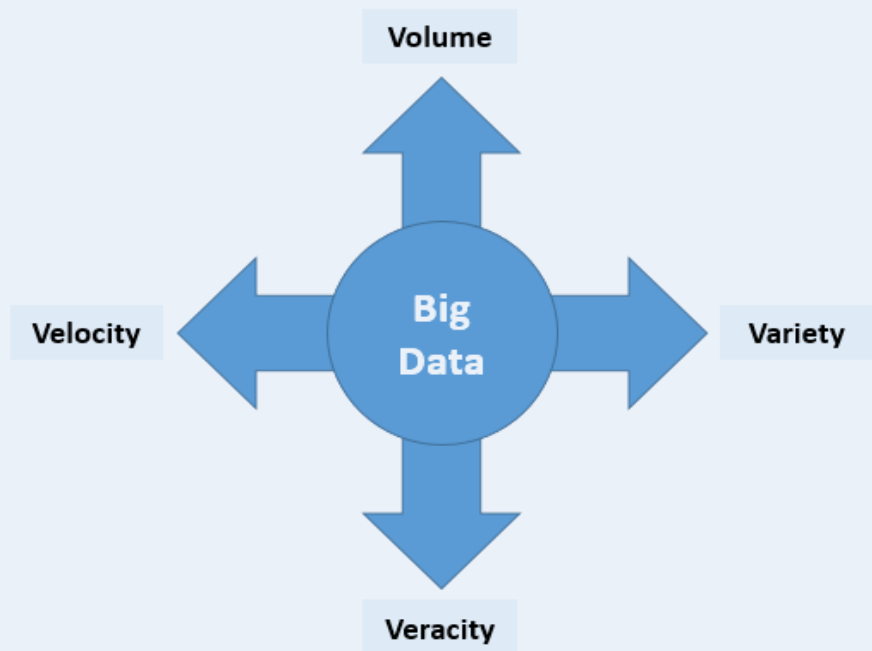
MITIGATE

- Quick Intro
  - Understanding Big Data
  - Proactive Threat Identification
- An Overview of MITIGATE Open Intelligence
- Q&A

# WHAT IS BIG DATA?



- A collection of large and complex data sets which are difficult to process using on-hand database management tools or traditional data processing applications
- Big Data refers to the tools, processes and procedures allowing an organization to create, manipulate, and manage a variety of data sets and storage facilities
- The challenges include capturing, storing, searching, sharing and analyzing



The four dimension (V's) of Big Data

# BIG DATA CHALLENGES IN SECURITY



- Efficiently handling large volumes of data and extracting potentially useful information.
- Inferring knowledge from complex heterogeneous data sources (*trusted* and *untrusted*).
- Understanding unstructured data in the right context.

# OVERALL GOALS OF BIG DATA ANALYTICS IN MITIGATE



- Take advantage of the massive amounts of data and provide right intervention to the right user at the right time.
  - Forecast potential future **(maritime) cyber threats**
  - Prediction of **zero-day exploits**
- Potentially benefit all the components of a port security system i.e., providers, users, and management.

# A SMART DATA SECURITY STRATEGY



- What are the sources of potential threats?
- Which assets are most vulnerable and likely to be targeted?
- Which processes/sub-processes need improvement?
- Was our audit effective?

# MAKING SECURITY PROACTIVE



MITIGATE

Understand organization's risks, threats and vulnerabilities



Identify key metrics

- Measure adhere to policy



Measure risk

- Measure risk in real-time
- Measure risk based on people's actions/behavior



Use metrics to guide actions

- Target programs
- Spend efficiently

# AN OVERVIEW OF MITIGATE OPEN INTELLIGENCE



MITIGATE

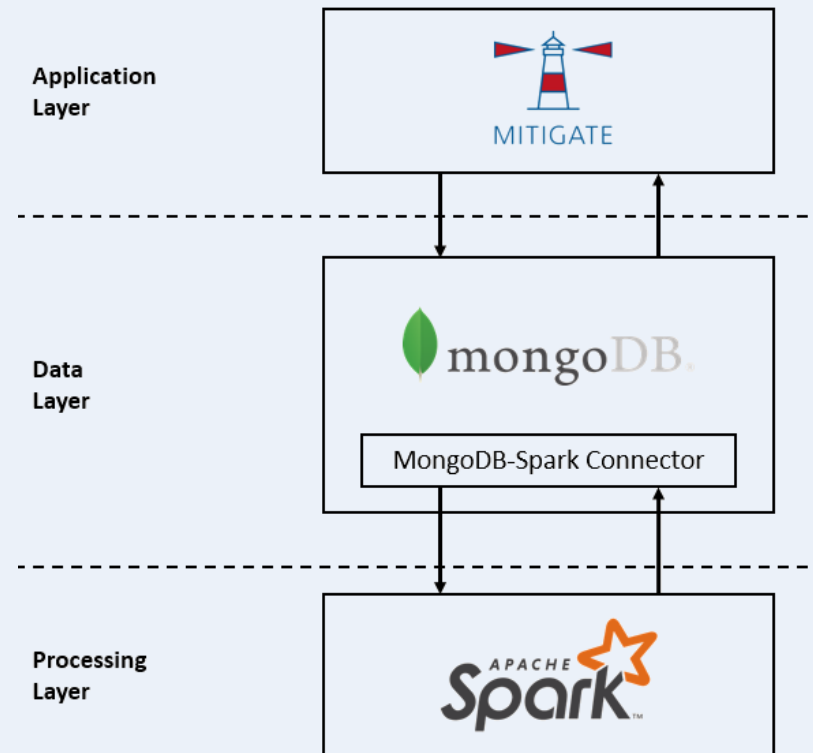


# ARCHITECTURE



Our goal is to provide MITIGATE users with a system for interactive Big Data Analytics, using an architecture composed by the following layers:

- Application layer: uses **MITIGATE** to provide an interactive analysis environment
- Data layer: uses **MongoDB** to efficiently handle huge amount of data
- Processing layer: uses **Spark** for efficient data analysis



- MongoDB is a **NoSQL document oriented database**, designed to manage huge amount of data providing **high performance, high availability** and **consistency of data**.
- MongoDB uses a **flexible JSON-based document** data model, with a schema-less approach.
- MongoDB provides a lot of **useful features** like indexing, a rich query language, an aggregation framework, replicas and sharding.

- Apache Spark is an open source engine for big data processing designed to be:
  - **Fast**, 100x faster than Apache Hadoop by exploiting in-memory parallel computing
  - **General purpose**, covers a wide range of workloads previously required separate systems, (ETL, queries, machine learning, streaming)
- Spark allows to integrate **many data sources** (HDFS, SQL, MongoDB, etc.), **cluster managers** (Yarn, EC2, Mesos) and supports many languages including **R, Java, Python** and **Scala**.

# MONGO SPARK CONNECTOR

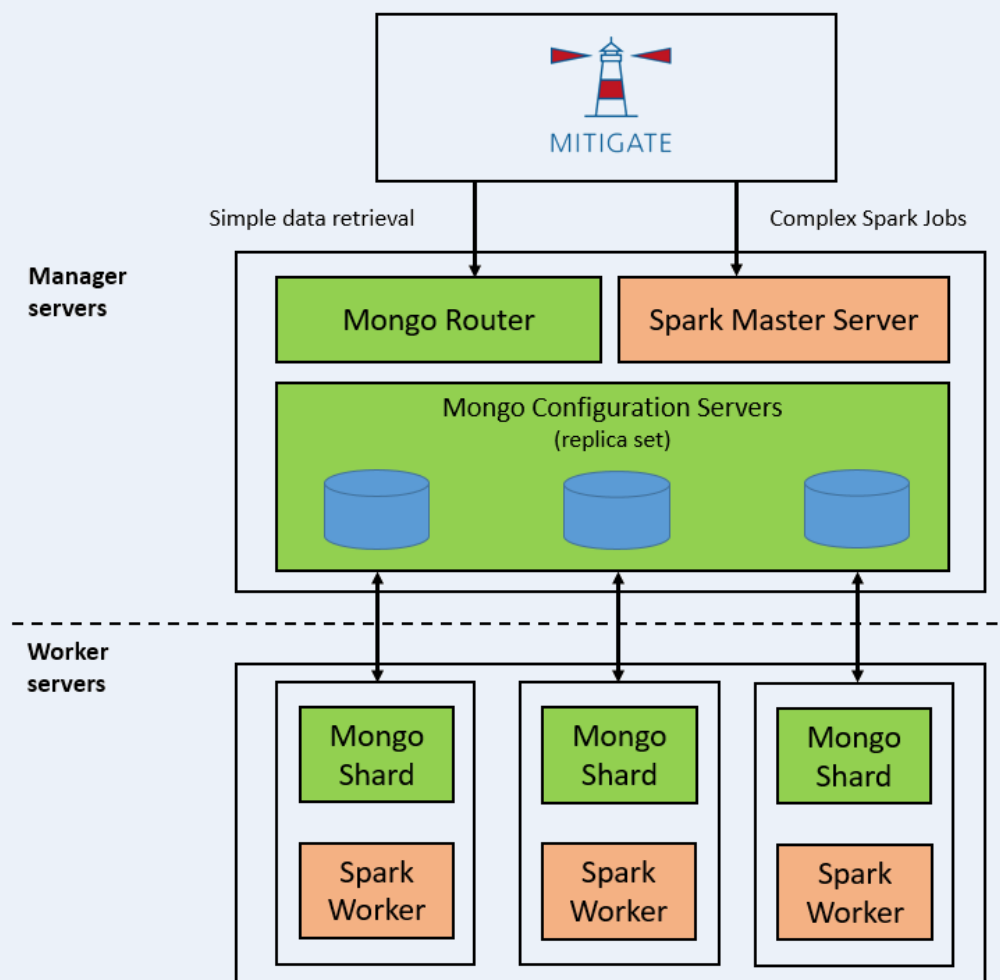


- Officially developed from Mongo team
- Under active development
- Supports data locality
- Lightweight
- Developer-friendly: we just need to import a library

# HORIZONTAL SCALING



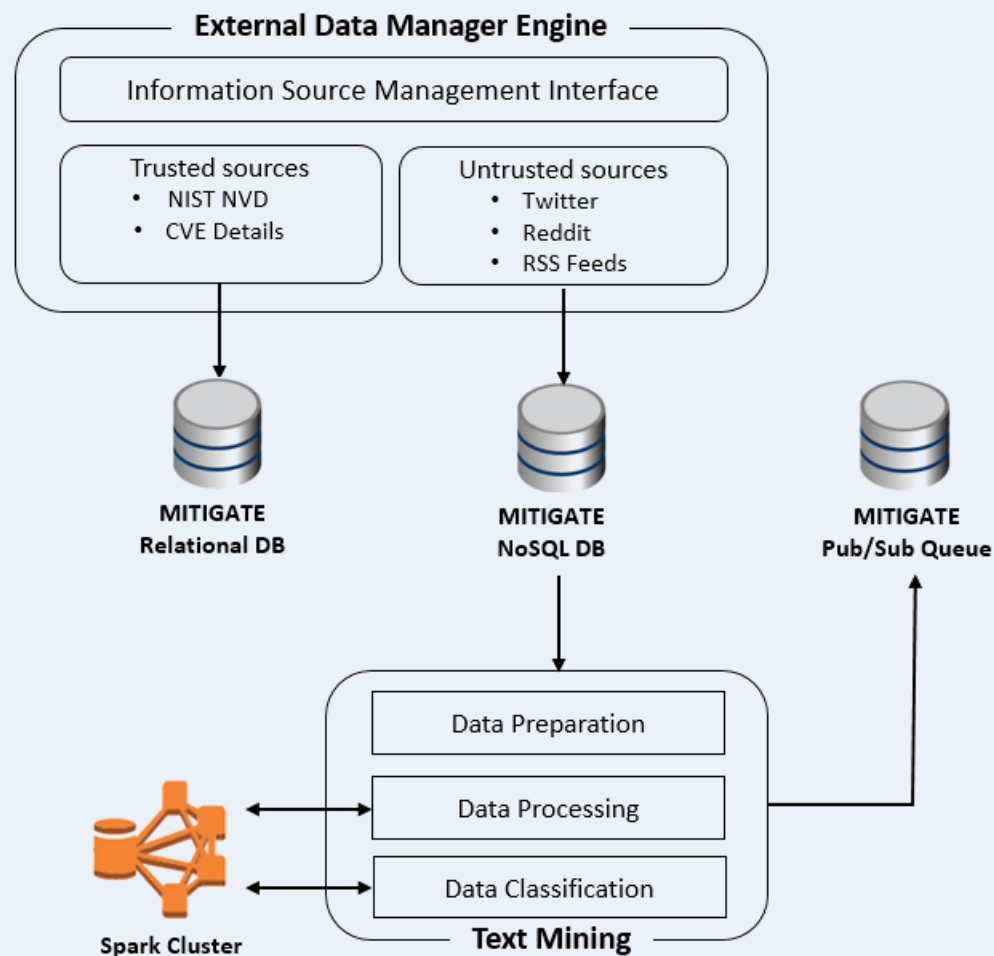
MITIGATE



# A CLOSER LOOK



MITIGATE



# AN EXAMPLE



MITIGATE

**EXAMPLE**

# DATA IMPORT – JOBS 1/6



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT

← Data Import

Jobs

Stats

Social Accounts

Data Import Jobs

+ CREATE NEW

ID	Name	Type	Actions
1	Twitter Import	TwitterBatch	<div>▶ 📄 ↺ ✎ 🗑</div>
2	Reddit Import	RedditBatch	<div>▶ 📄 ↺ ✎ 🗑</div>
5	NIST Import	NistNVD	<div>▶ 📄 ↺ ✎ 🗑</div>
8	Feeds Import	FeedsBatch	<div>▶ 📄 ↺ ✎ 🗑</div>



# DATA IMPORT – JOBS 2/6



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT / ADD

← Job Management

Add a new job

Job Type

Twitter (Batch Import) ▼

Name

Twitter Importer

Search Text

[Info](#)

vulnerability

Time Interval in Hours

24

Start From Last Imported

☒

☒ Save

# DATA IMPORT – JOBS 3/6



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT / ADD

← Job Management

Add a new job

Job Type

Reddit (Batch Import) ▼

Name

Reddit Importer Search

Import Type

Search ▼

Time Period

Week ▼

Search Text

vulnerability

☒ Save

# DATA IMPORT – JOBS 4/6



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT / ADD

← Job Management

Add a new job

Job Type

Reddit (Batch Import) ▼

Name

Reddit Importer Search

Import Type

Search ▼

Time Period

Week ▼

Search Text

vulnerability

☒ Save

# DATA IMPORT – JOBS 5/6



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT / ADD

← Job Management

Add a new job

Job Type

Reddit (Batch Import) ▼

Name

Reddit Importer

Import Type

Subreddit ▼

Time Period

Week ▼

Subreddit

netsec

☒ Save

# DATA IMPORT – JOBS 6/6



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT / ADD

← Job Management

Add a new job

Job Type

Feeds (Batch Import) ▼

Name

Feeds Importer

Feeds URLs

http://blog.itsecurityexpert.co.uk/feeds/posts/default

https://www.trustedsec.com/feed/

http://seclists.org/rss/oss-sec.rss

+

☒ Save

# DATA IMPORT – JOBS SCHEDULER 1/3



MITIGATE

MENU

- Dashboard
- Risk Assessments
- Supply chain services
- Pending actions
- Business Partners
- Assets
- Sites
- Networks
- Vendors
- Vulnerabilities
- Threats
- Threat Profiles
- Controls

/ IMPORT / 8 / SCHEDULE

## ← Job Scheduler

Edit job schedule

+CREATE NEW

Pause all Resume all

Last Fire Time	Next Fire Time	Times Triggered	Repeat Count	Repeat Interval	Actions
Tue Mar 21 12:20:52 CET 2017	Tue Mar 21 13:20:52 CET 2017	1	1	3600000	<button>Edit</button> <button>Delete</button>

# DATA IMPORT – JOBS SCHEDULER 2/3



MITIGATE

## MENU

 Dashboard

 Risk Assessments

 Supply chain services

 Pending actions

 Business Partners

 Assets

 Sites

 Networks

 Vendors

 Vulnerabilities

 Threats

 Threat Profiles

 Controls

/ IMPORT / 8 / SCHEDULE / ADD

## ← Job Scheduler

Add schedule

Start now ☒ Start Deferred ☐

Repeat Forever



Interval in Second

3600

☒ Save

# DATA IMPORT – JOBS SCHEDULER 3/3



MITIGATE

MENU

- Dashboard
- Risk Assessments
- Supply chain services
- Pending actions
- Business Partners
- Assets
- Sites
- Networks
- Vendors
- Vulnerabilities
- Threats
- Threat Profiles
- Controls

/ IMPORT / 8 / SCHEDULE / ADD

## Job Scheduler

Add schedule

Start now ☐ Start Deferred ☒

Start Date

Sun	Mon	Tue	Wed	Thu	Fri	Sat	
26	27	28	1	2	3	4	12:00
5	6	7	8	9	10	11	13:00
12	13	14	15	16	17	18	14:00
19	20	21	22	23	24	25	15:00
26	27	28	29	30	31	1	16:00
							17:00

☒ Save



# DATA IMPORT – STATS



MITIGATE

MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ IMPORT / STATS

← Data Import

Jobs

Stats

Social Accounts

Data Import Collection Stats

Collection Name:  
SocialContentSharded

Size:  
585,227 MB

Average Object Size:  
3,543 KB

Object Count:  
169138

Tweets:  
124088

Reddit Post:  
43960

Feeds Post:  
1090

29/03/17

www.mitigateproject.eu

25

# DATA ANALYSIS 1/4



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ ANALYSIS

← Data Analysis

Analysis

Spark Apps (JARs)

Spark Cluster Status

Data Analysis Jobs

+ CREATE NEW

ID	Name	Type	Actions
4	Vulnerability	SparkBatch	<div><div></div><div></div><div></div><div></div><div></div></div>
7	Vulnerability OR Linux OR Microsoft	SparkBatch	<div><div></div><div></div><div></div><div></div><div></div></div>

# DATA ANALYSIS 2/4



MENU / ANALYSIS / 4 / EDIT

← Job Management

Edit job settings

**Job Type**  
SparkBatch

**Name**  
Spark Example Job

**Analysis Job Type**  
TEXT\_SEARCH

**App Name**  
textmining-app-with-logs.jar

**Context**  
complex-jobs

**Collection Name**  
SocialContentSharded

**Result Collection Name**  
results

**Result Queue**  
results

**Keywords**  
vulnerability  
spark  
+

☒ Save

# DATA ANALYSIS 3/4



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

/ ANALYSIS / JARS

← Data Analysis

Analysis

Spark Apps (JARs)

Spark Cluster Status

Available Spark Applications

JAR	Upload date
textmining-app-with-logs.jar	2017-03-13T11:20:07.610+01:00
textmining-app-v2.4.jar	2017-03-09T16:15:57.050+01:00
textmining-app-v2-custom-params.jar	2017-03-13T11:22:39.974+01:00

Upload a new Spark Application

File to upload:

Scegli file

Nessun file selezionato

Filename

upload

# DATA ANALYSIS 4/4



MENU

Dashboard

Risk Assessments

Supply chain services

Pending actions

Business Partners

Assets

Sites

Networks

Vendors

Vulnerabilities

Threats

Threat Profiles

Controls

ANALYSIS / STATUS

Data Analysis

Analysis

Spark Apps (JARS)

Spark Cluster Status

Spark JobServer Status

Server is online

Submitted jobs

Job ID	Spark App	Start Time	Duration	Status	
727c9fa3-6318-43ca-b597-f0b0d1fb442c	2017-03-14T16:43:12.559+01:00	62.325 secs	FINISHED		<div>Cx</div>
{ 'duration': '62.325 secs', 'classPath': 'eu.mitigate.analytics.textmining_app.AppManager', 'startTime': '2017-03-14T16:43:12.559+01:00', 'context': 'complex-jobs', 'result': 'Successfully finished the Application', 'status': 'FINISHED', 'jobid': '727c9fa3-6318-43ca-b597-f0b0d1fb442c' }					
f4498e3a-2c84-400f-b0e1-dcb0167719dd	2017-03-13T15:25:52.413+01:00	72.284 secs	FINISHED		<div>Cx</div>
{ 'duration': '72.284 secs', 'classPath': 'eu.mitigate.analytics.textmining_app.AppManager', 'startTime': '2017-03-13T15:25:52.413+01:00', 'context': 'complex-jobs', 'result': 'Successfully finished the Application', 'status': 'FINISHED', 'jobid': 'f4498e3a-2c84-400f-b0e1-dcb0167719dd' }					
22648742-11c6-4452-ba12-6c36a1364af3	2017-03-13T10:10:02.815+01:00	80.001 secs	FINISHED		<a href="#">Logs</a>
c218b4b6-d906-4d40-b6fb-f2a9bf2cd48b	2017-03-10T15:23:03.090+01:00	0.0 secs	FINISHED		<a href="#">Logs</a>
9cbfd056-1072-4699-b3a2-edd577fd166d	2017-03-10T15:19:47.044+01:00	188.597 secs	FINISHED		<a href="#">Logs</a>
0be954df-7c31-442d-8878-6531bdaf7e8b	2017-03-10T14:51:49.348+01:00	0.0 secs	FINISHED		<a href="#">Logs</a>
8cea91e8-cf28-41da-989b-e0096785d2ca	2017-03-10T14:40:11.573+01:00	183.183 secs	FINISHED		<a href="#">Logs</a>
12e66dce-8656-4990-b622-fa20efa70845	2017-03-10T14:24:56.789+01:00	0.0 secs	FINISHED		<a href="#">Logs</a>
5d27b946-6fcb-4581-892e-0209a867665f	2017-03-10T14:14:29.099+01:00	104.649 secs	FINISHED		<a href="#">Logs</a>
5a9eb528-ce13-48a5-96f9-e3993d3decda	2017-03-10T12:38:29.446+01:00	110.853 secs	FINISHED		<a href="#">Logs</a>

<

1

2

3

4

5

>

