BIG DATA ANALYTICS & THREAT PREDICTION

MITIGATE

CIP Workshop

September 19-20, 2017

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
www.maggioli.it

IN THIS PRESENTATION…

MITIGATE

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

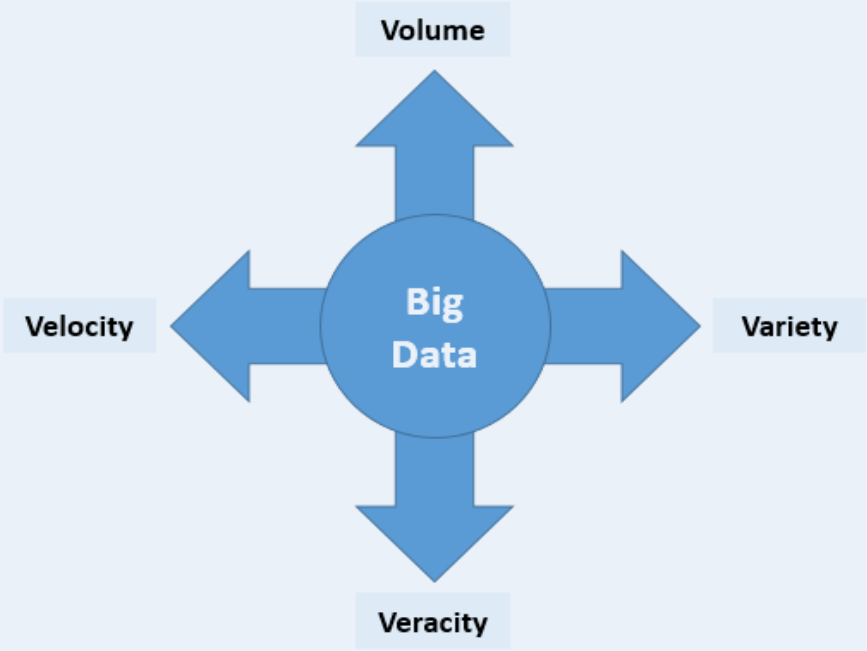
WHAT IS BIG DATA?

MITIGATE

- 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

MITIGATE

- 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

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

MITIGATE

- 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

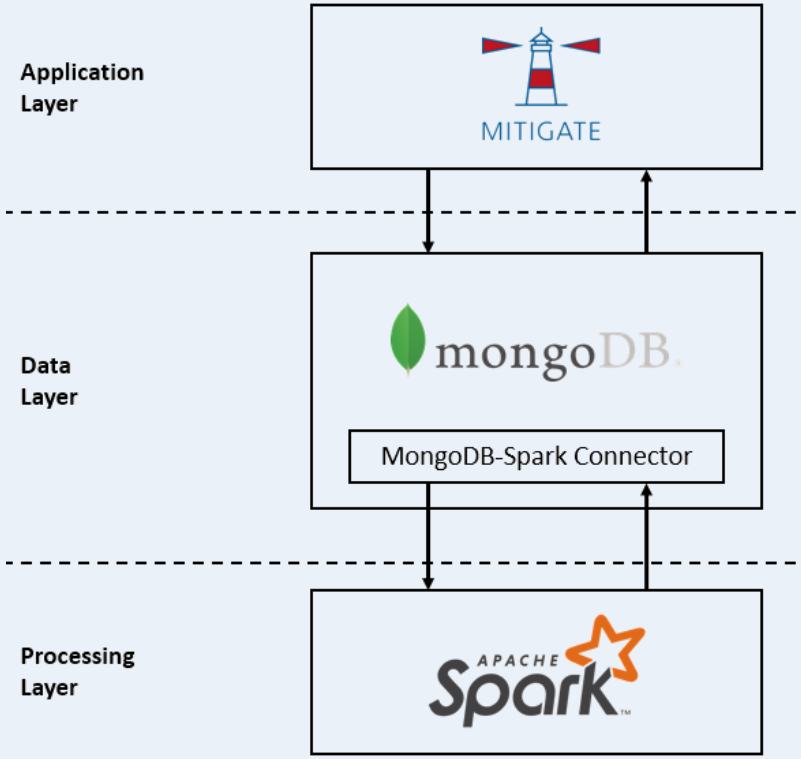
MITIGATE

ARCHITECTURE

MITIGATE

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

MITIGATE

- 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.

SPARK

MITIGATE

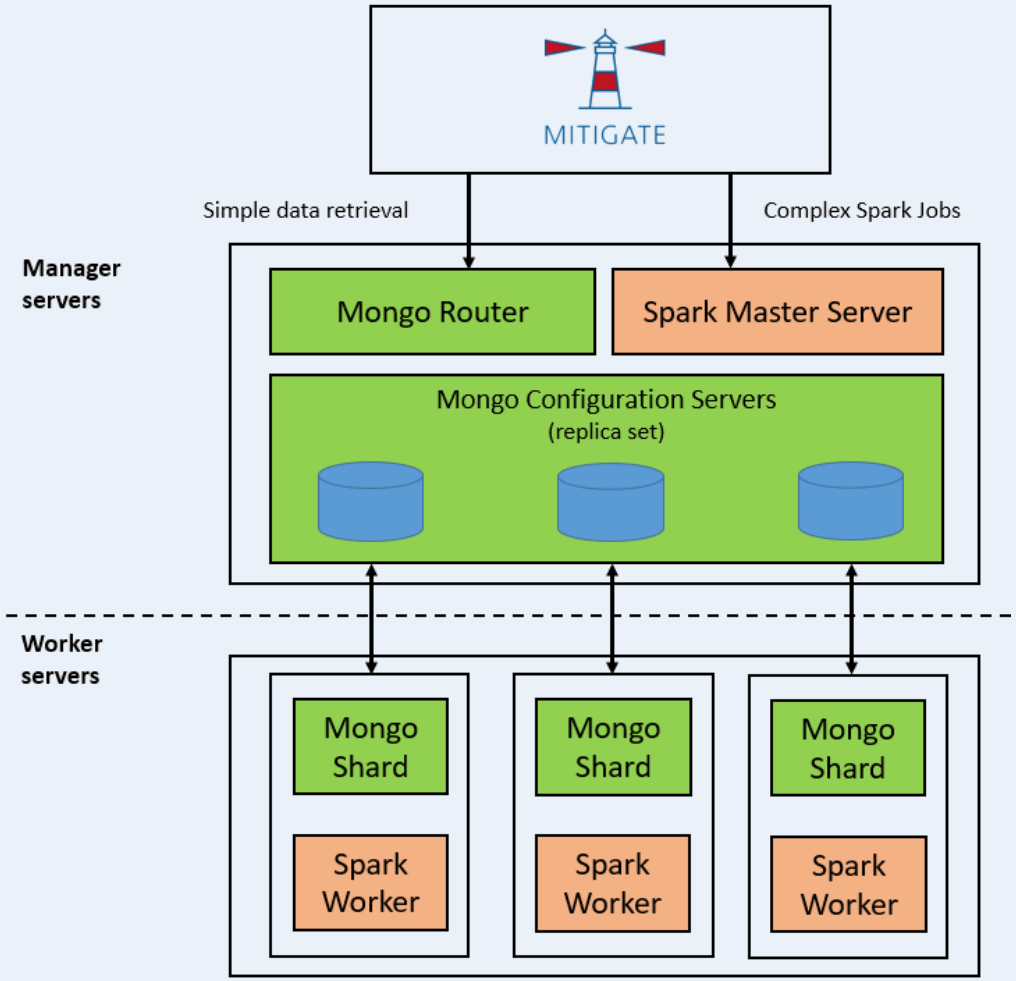
- Apache Spark is an open source engine for big data processing designed to be:
  - **Fast**, 100x faster than Apache Hadoop by exploiting in-memory parallel computing
  - **General purpose**, covers a wide range of workloads previously required separate systems, (ETL, queries, machine learning, streaming)
- 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

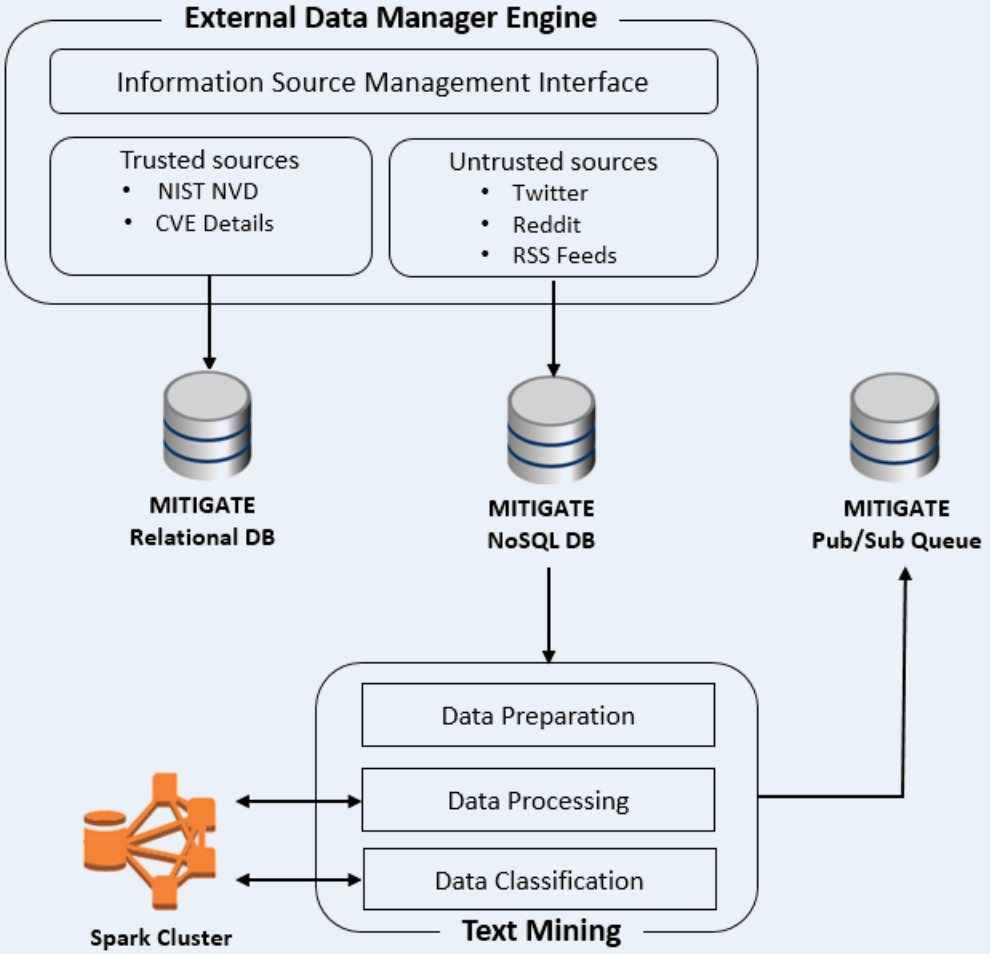
HORIZONTAL SCALING

MITIGATE



A CLOSER LOOK

MITIGATE

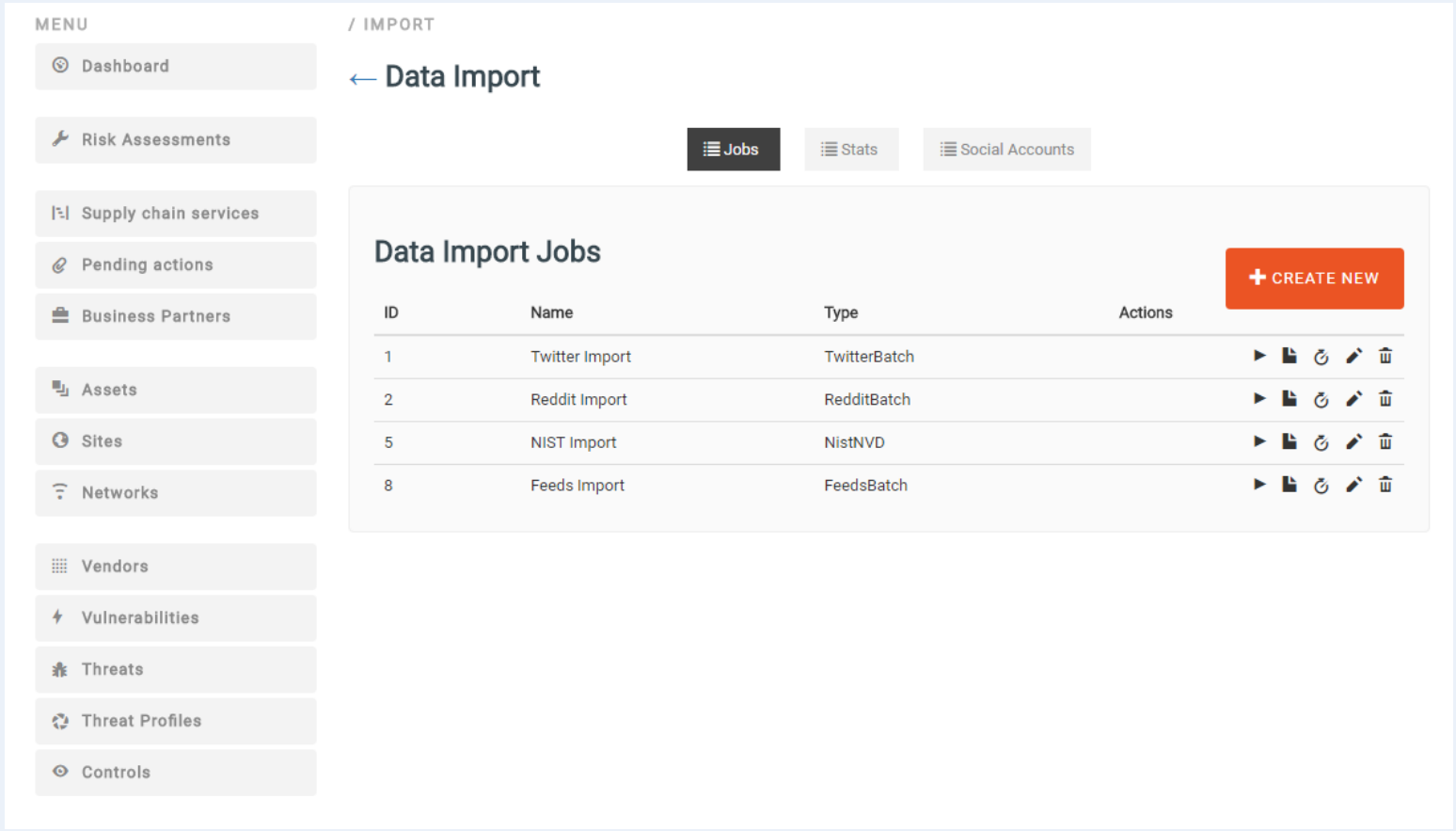


AN EXAMPLE

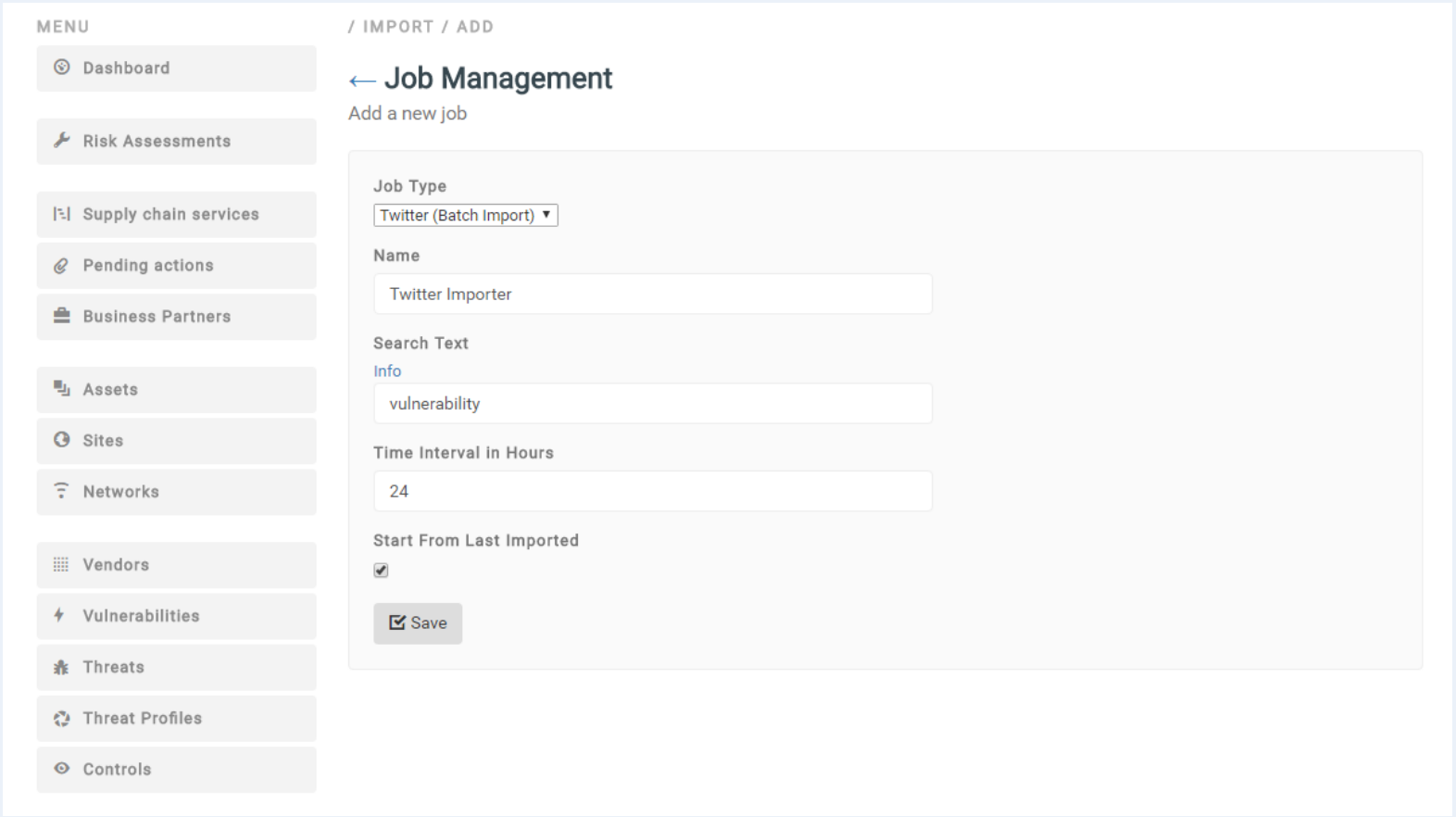
MITIGATE

DATA IMPORT – JOBS 1/6

MITIGATE

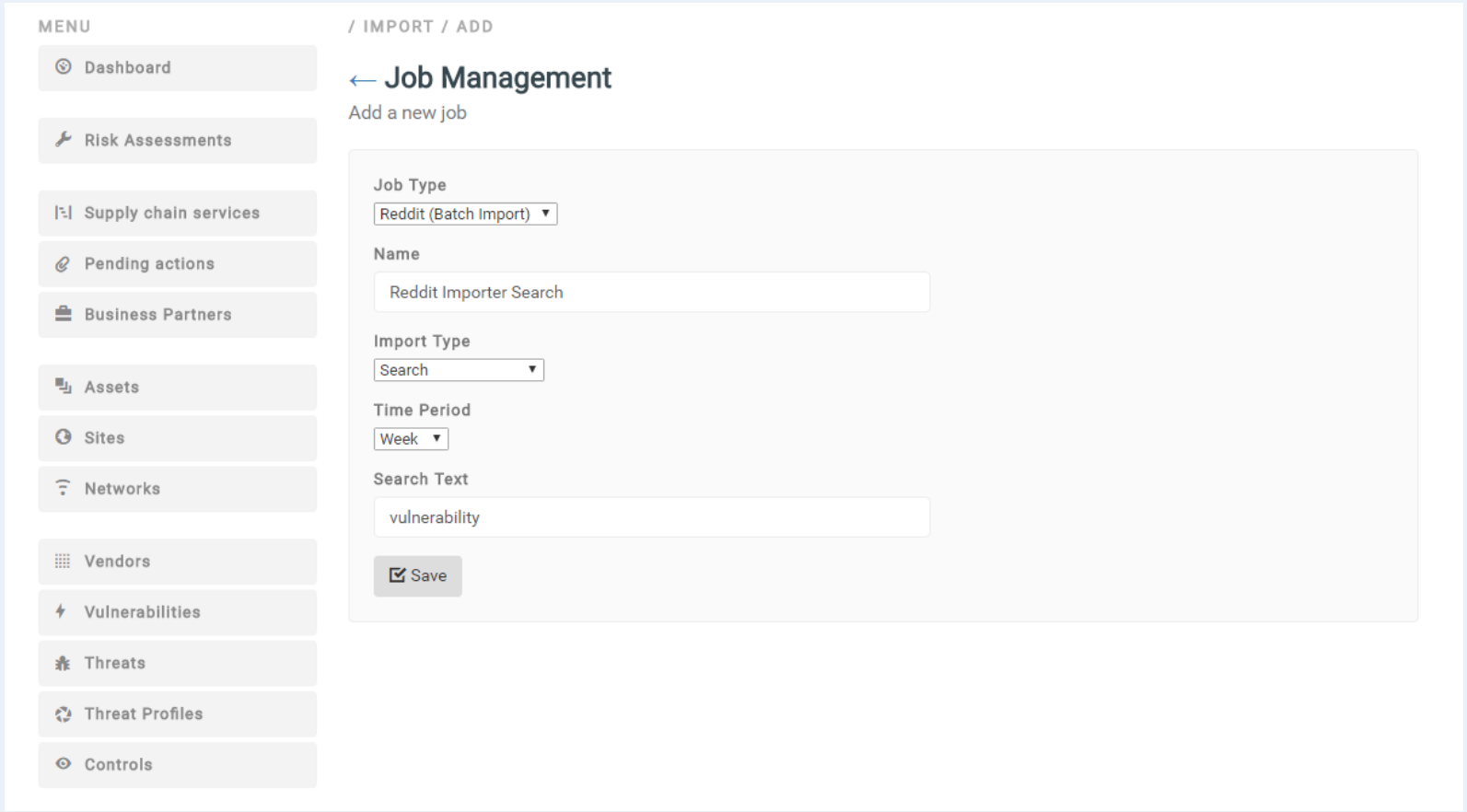


DATA IMPORT – JOBS 2/6



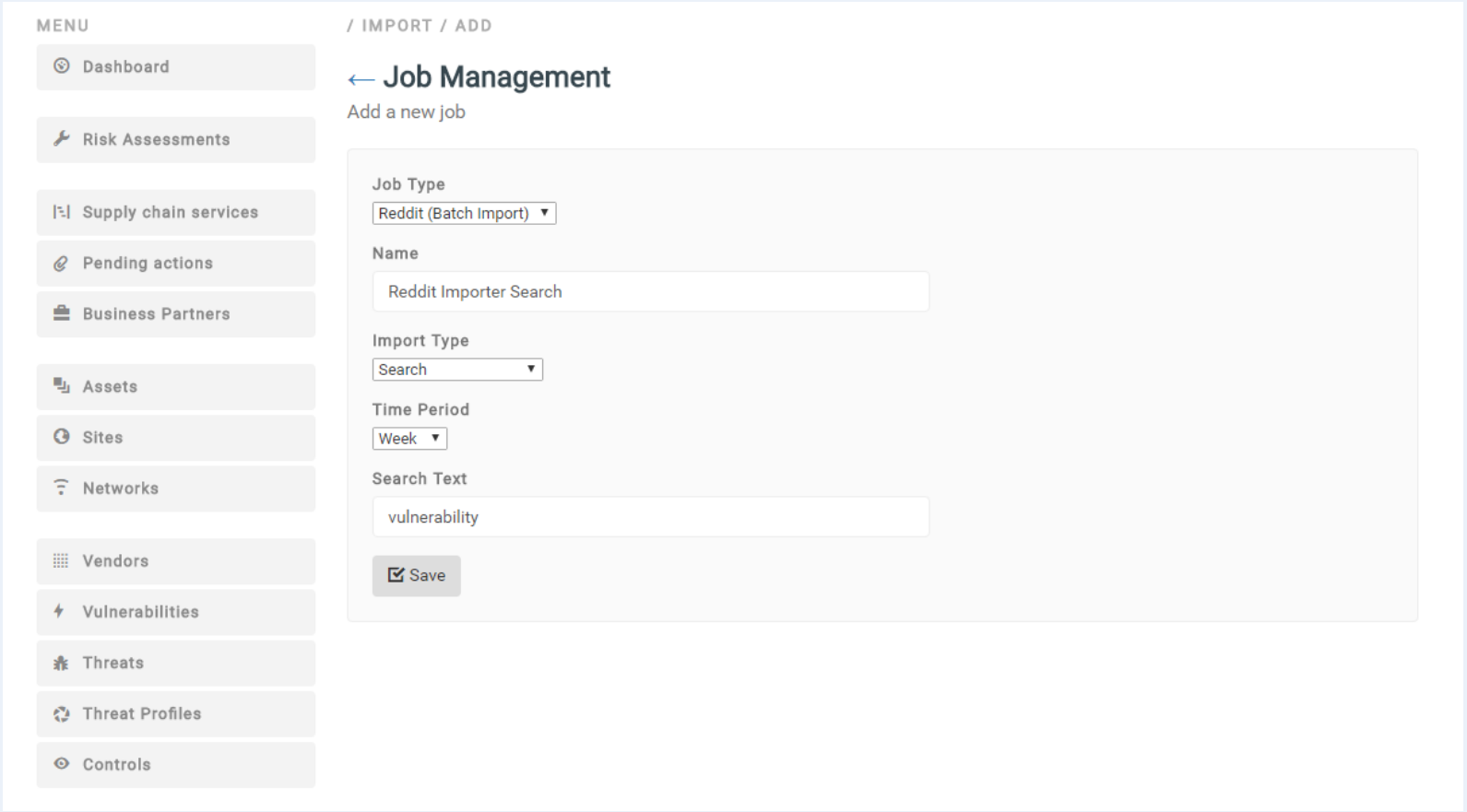
DATA IMPORT – JOBS 3/6

MITIGATE



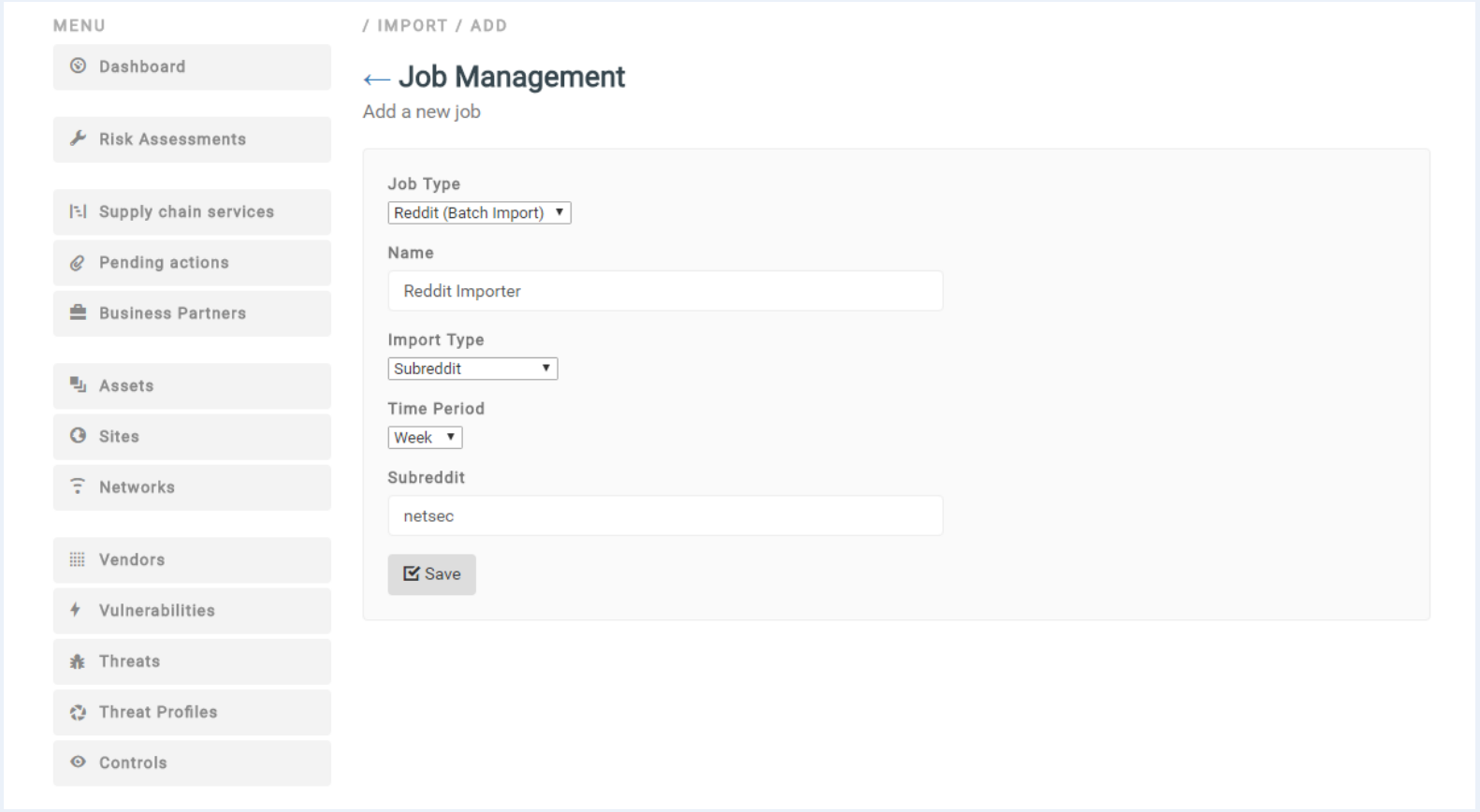
DATA IMPORT – JOBS 4/6

MITIGATE



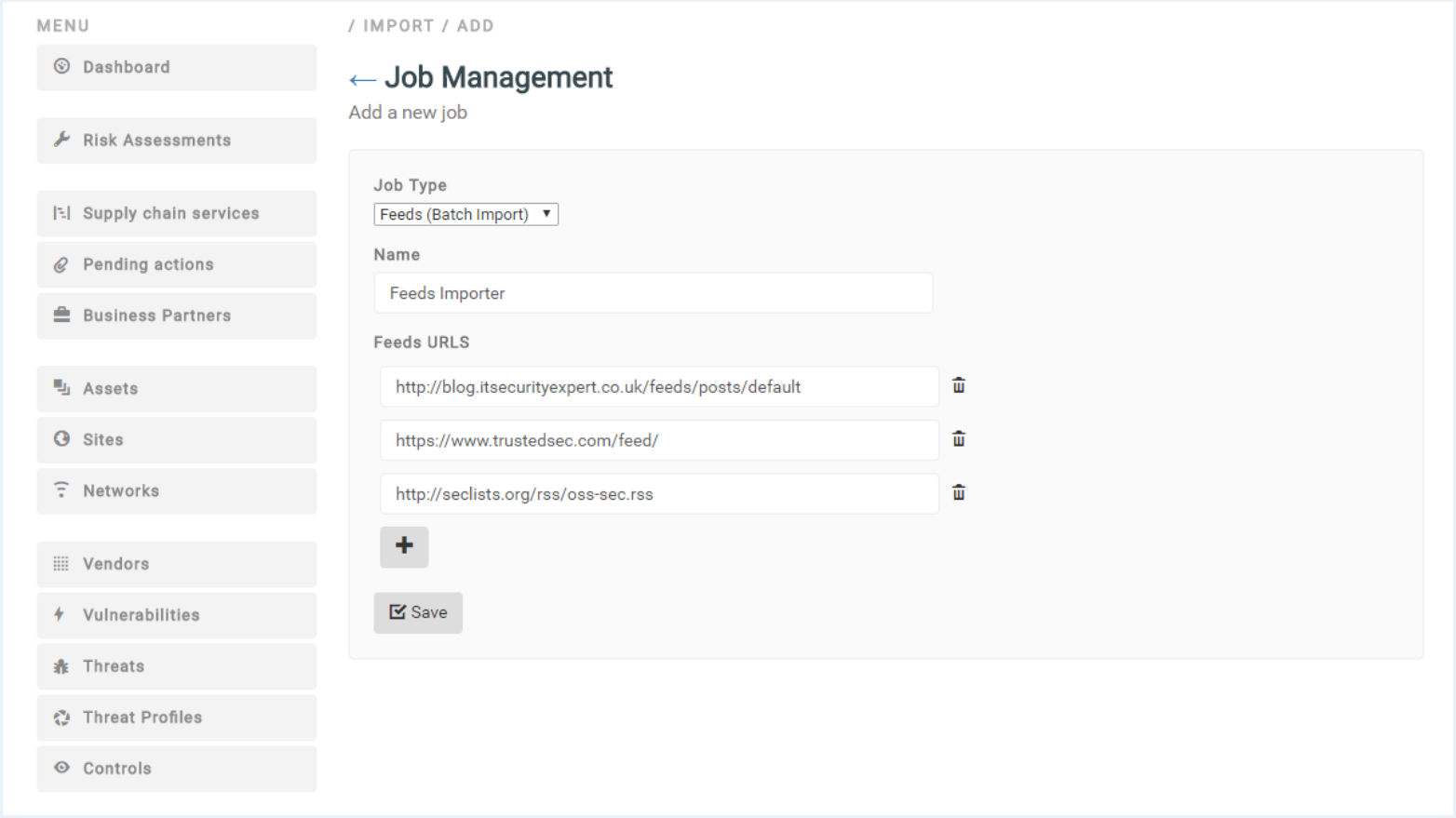
DATA IMPORT – JOBS 5/6

MITIGATE



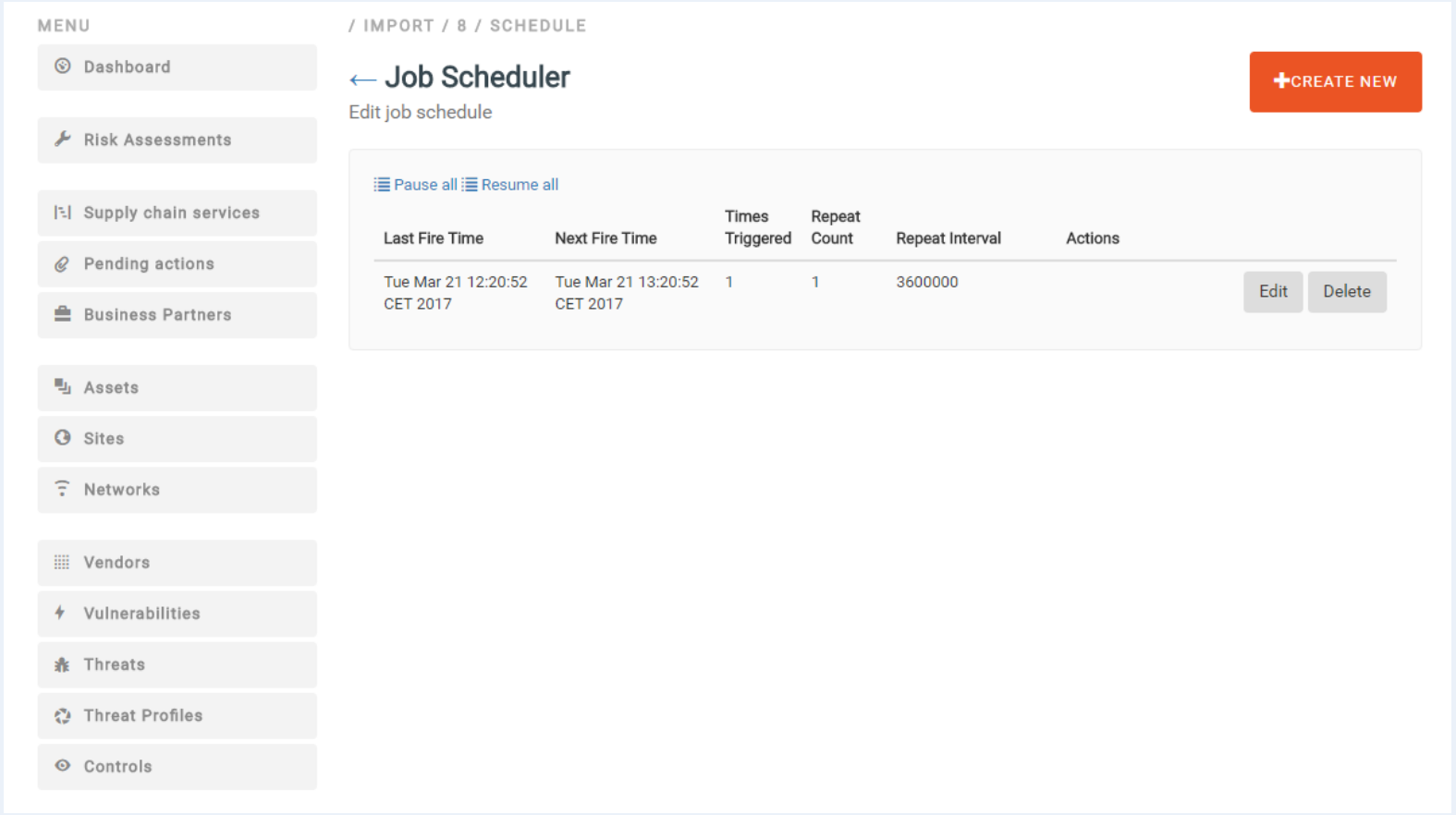
DATA IMPORT – JOBS 6/6

MITIGATE



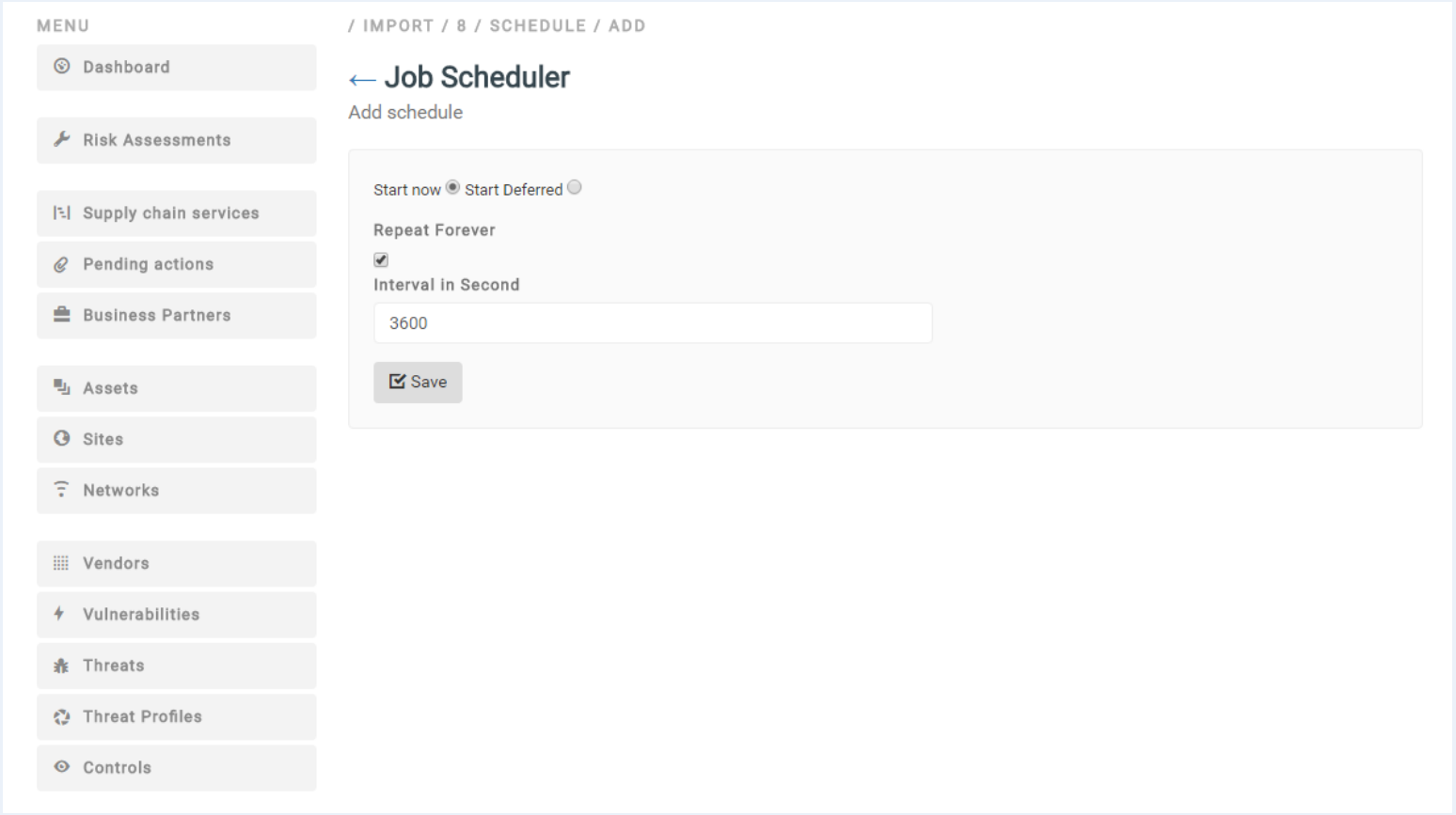
DATA IMPORT – JOBS SCHEDULER 1/3

MITIGATE



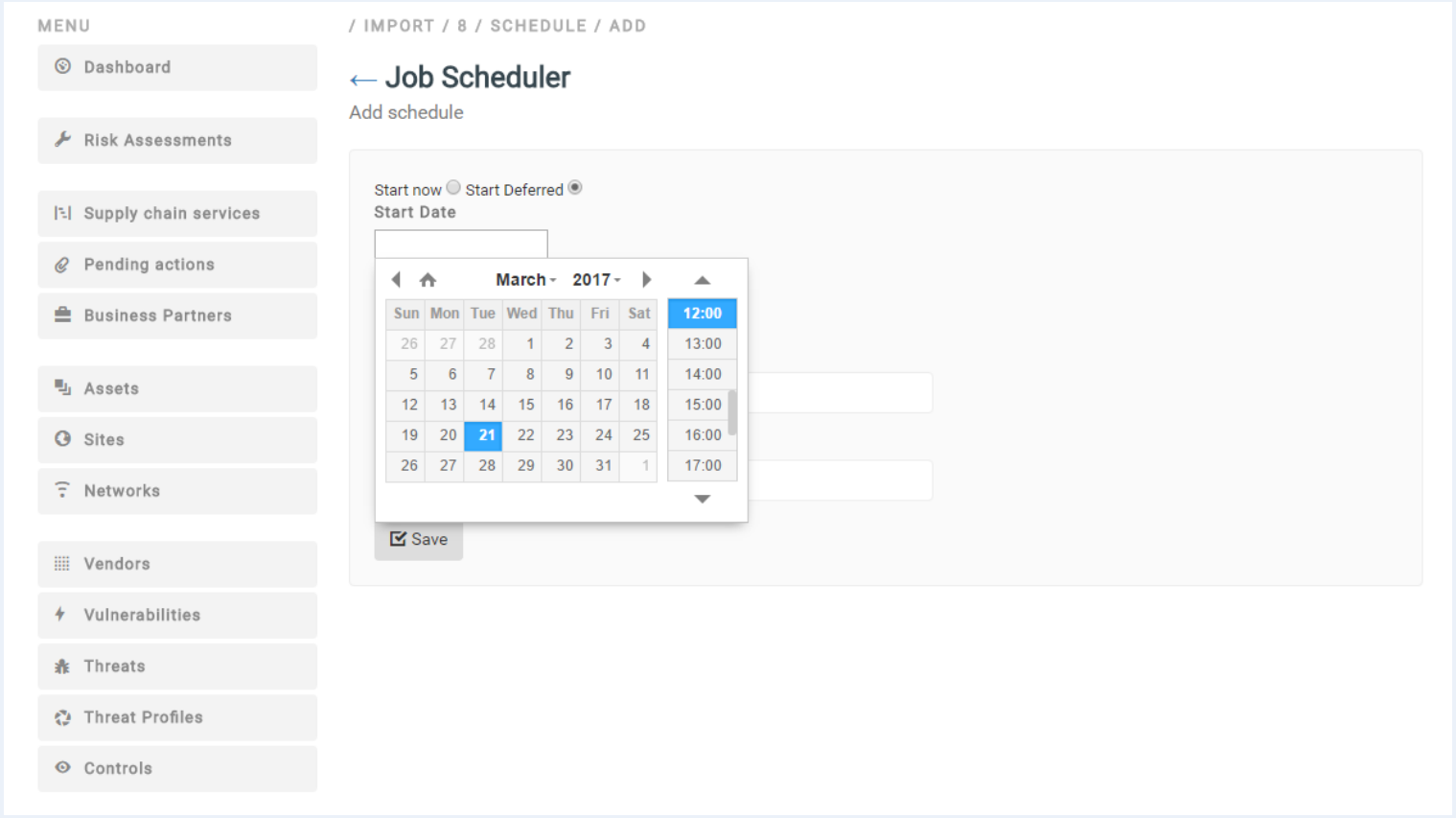
DATA IMPORT – JOBS SCHEDULER 2/3

MITIGATE



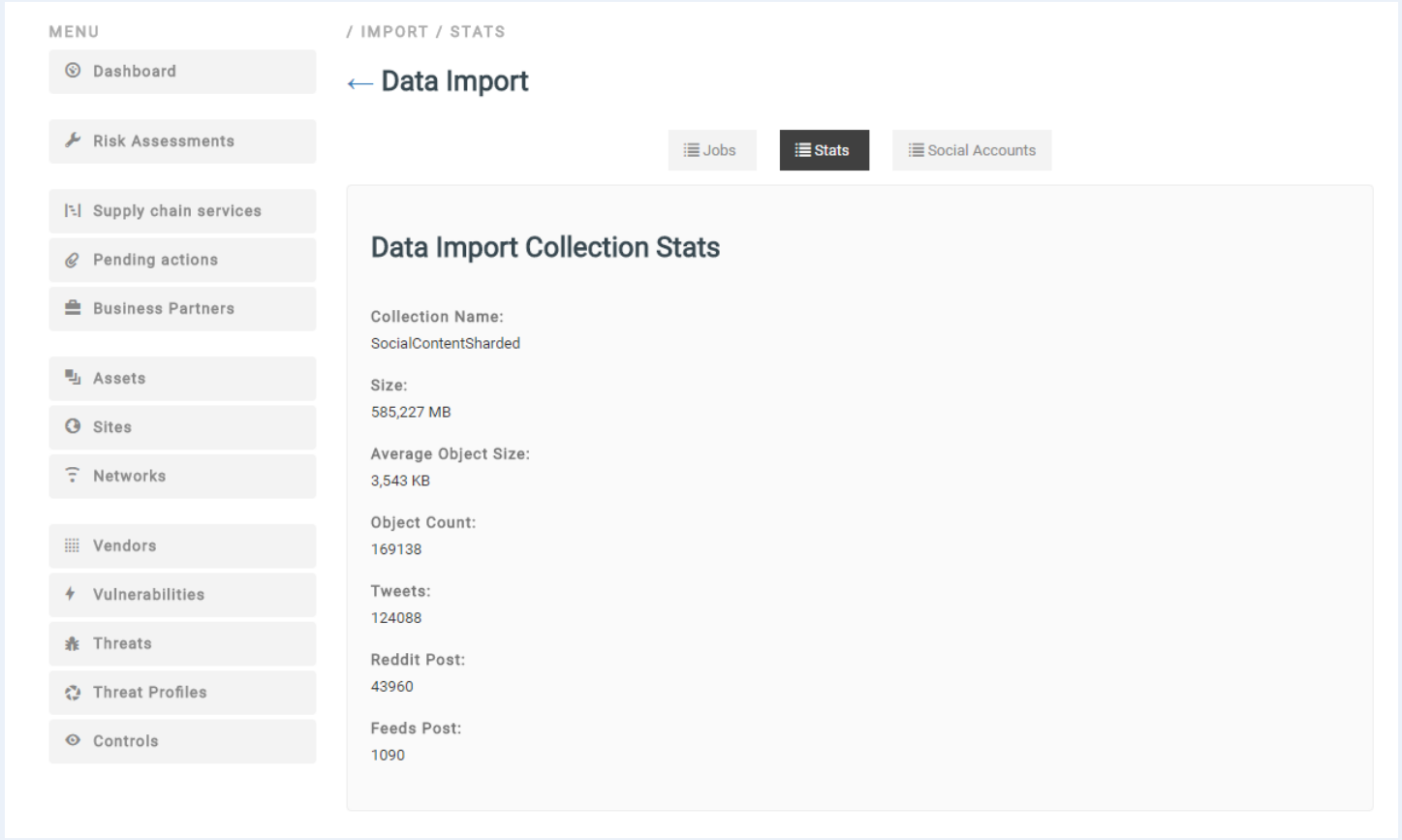
DATA IMPORT – JOBS SCHEDULER 3/3

MITIGATE



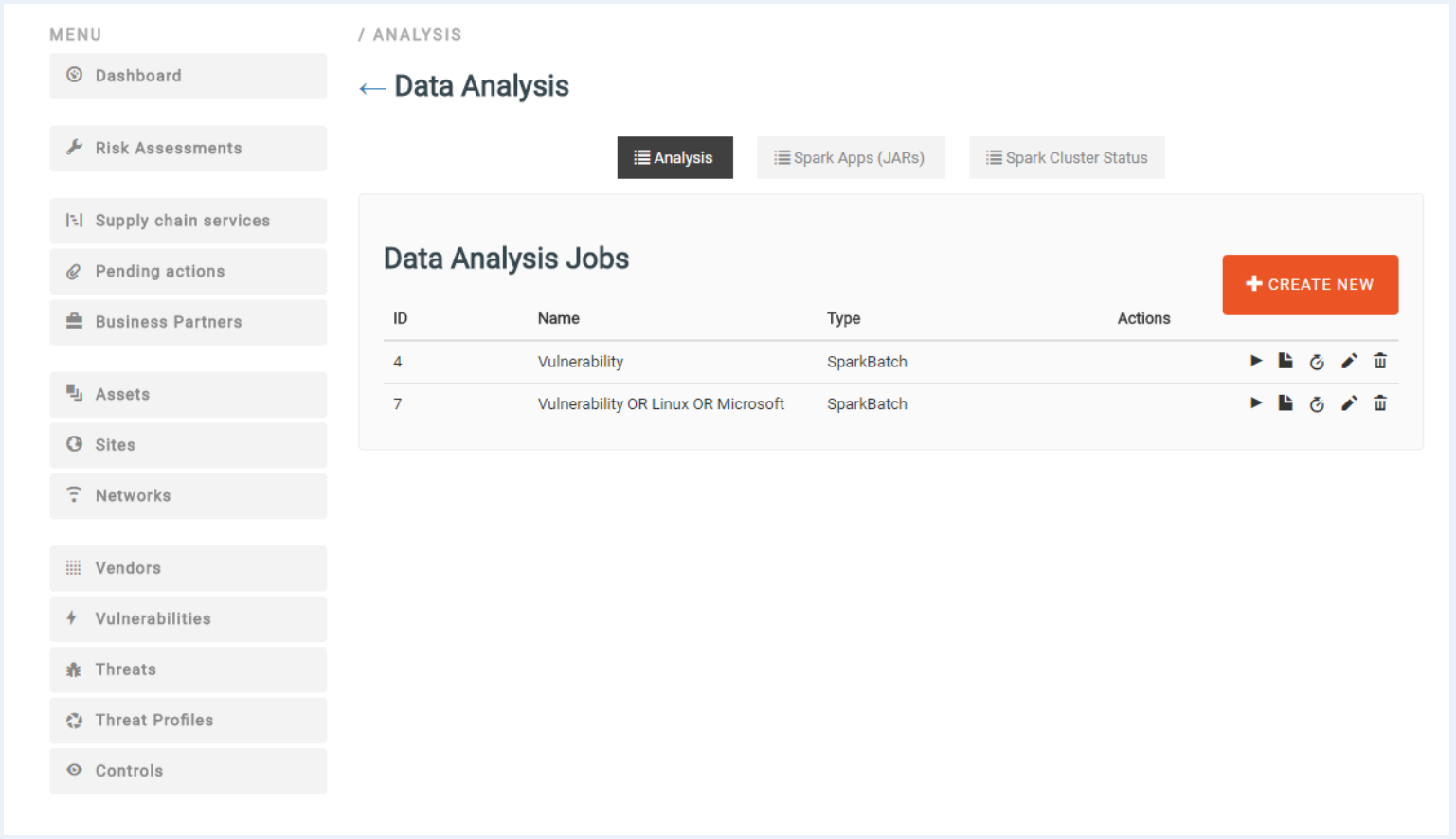
DATA IMPORT – STATS

MITIGATE



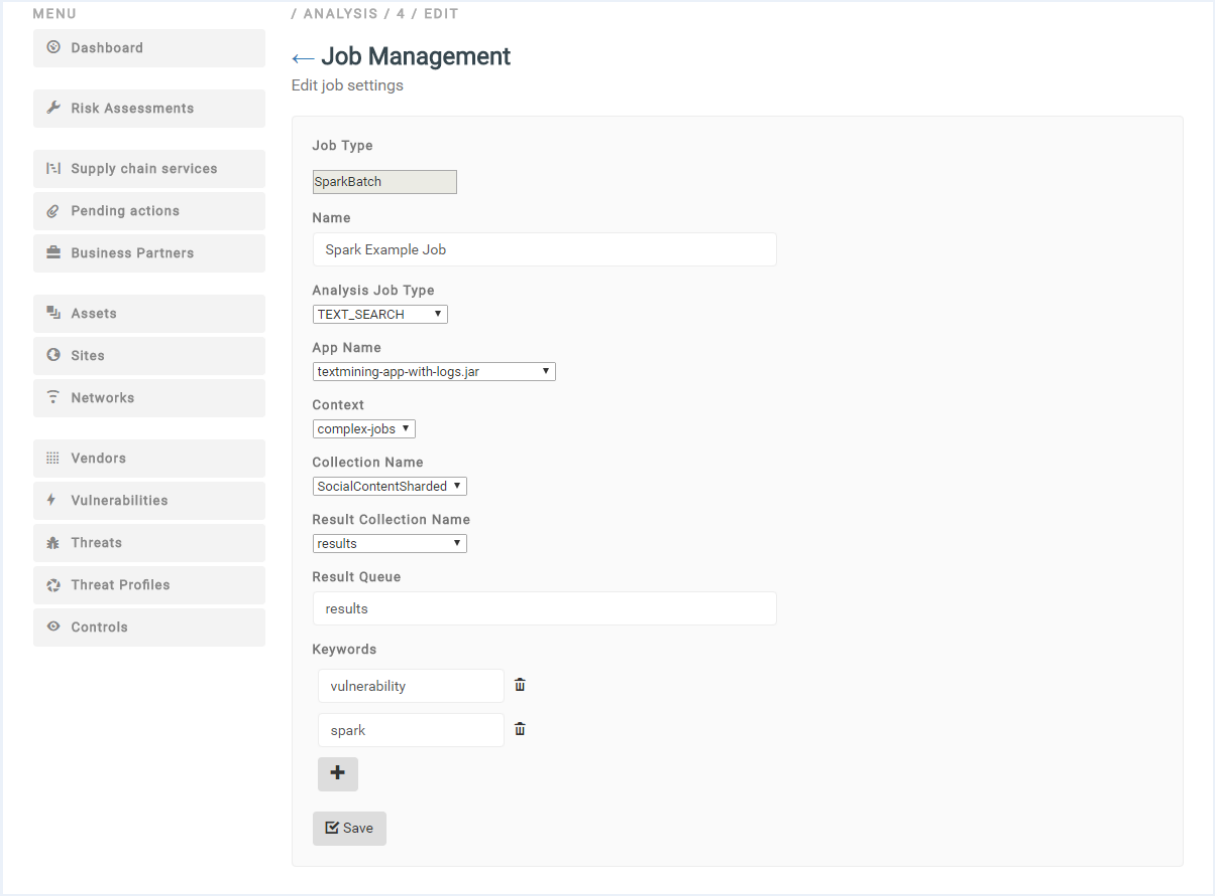
DATA ANALYSIS 1/4

MITIGATE



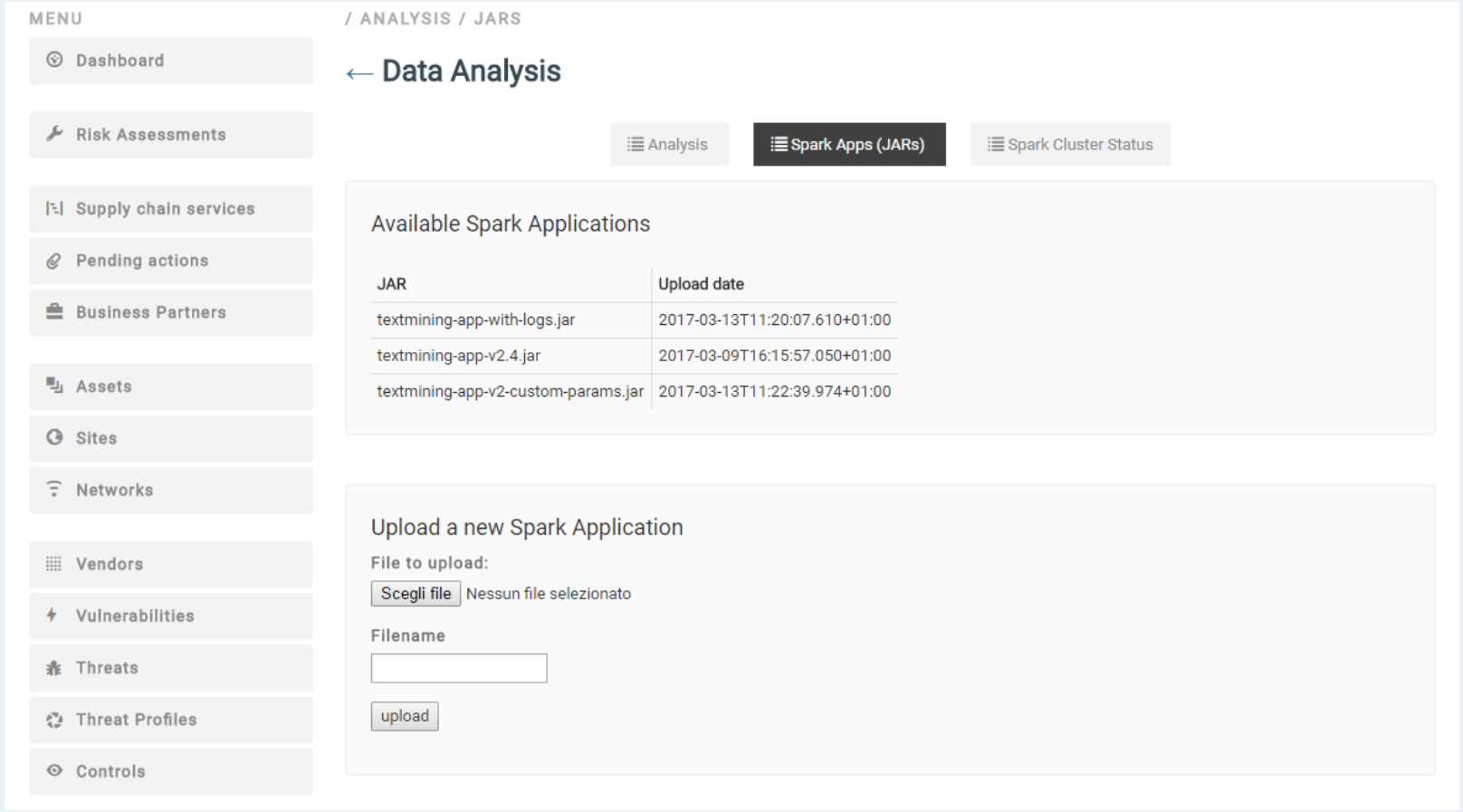
DATA ANALYSIS 2/4

MITIGATE



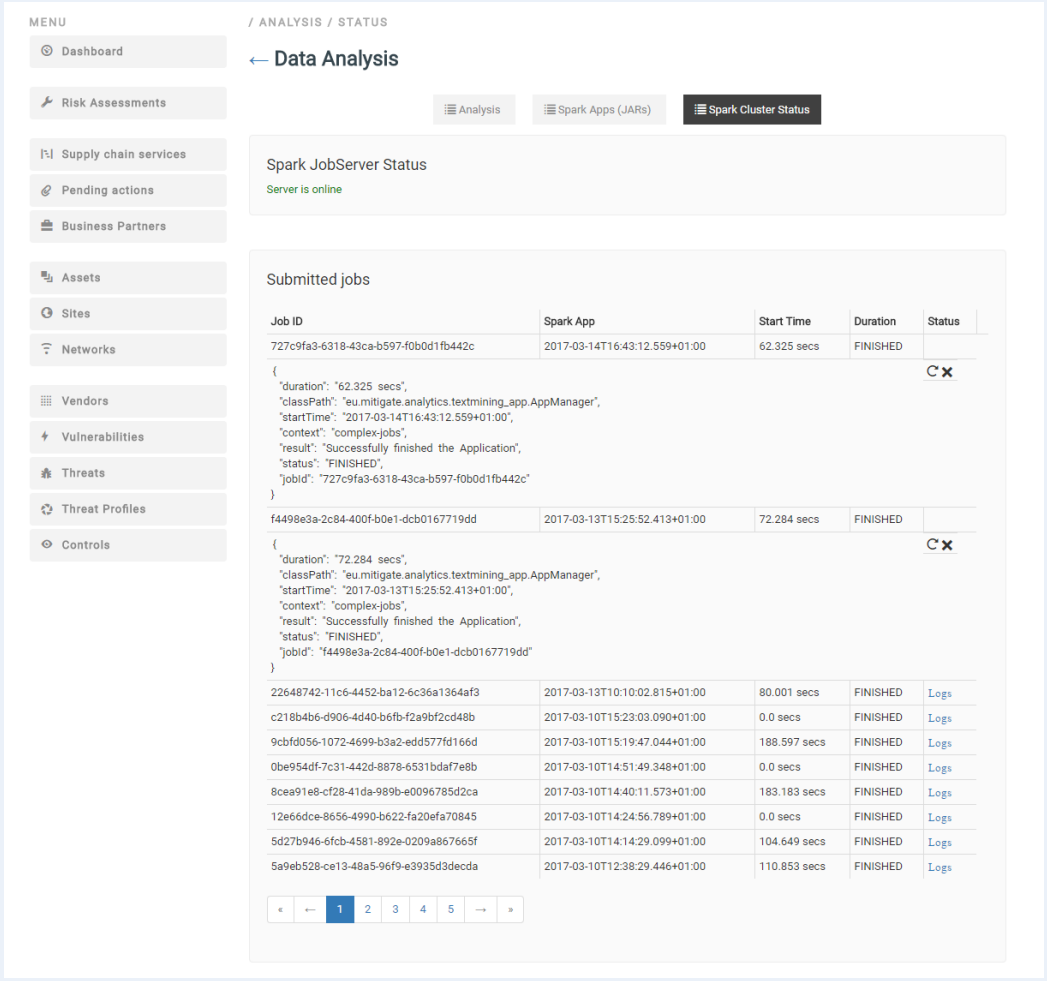
DATA ANALYSIS 3/4

MITIGATE



DATA ANALYSIS 4/4

MITIGATE




QUESTIONS?
COMMENTS?