

DensityDesign @ Policloud

Daniele Ciminieri
daniele.ciminieri@polimi.it

Michele Mauri
michele.mauri@polimi.it

D E N -
S I T Y
G N +

Use case #1

FIRE project

OBIETTIVO

Scraping automatico di un sito aggregatore di news
([EMM newsbrief](#))

CRITICITÀ

download e parsing di tutti gli articoli riguardanti armi da fuoco pubblicati dal 1/1/2010 al 30/3/2015 nell' unione europea in 28 lingue diverse



raccolta dati **prolungata nel tempo** (circa 4 mesi)

SETUP

Scraper e Parser
Python



Entity Extraction
TextRazor (API)



Storage
MongoDB

RISULTATI

Attività di raccolta dati terminata in 6 mesi

Occasionali rallentamenti dovuti a periodi più o meno prolungati di downtime

Use case #2

Umbra

OBIETTIVO

Analisi testuale di circa **30 milioni** di tweet per filtrare post contenenti hashtag e parole sensibili

CRITICITÀ

mole di dati (>100GB) difficilmente gestibile con un database tradizionale

restituzione in *tempo reale* delle query

SETUP

Server di ricerca
ElasticSearch



Frontend del sito
HTML, CSS, JS

RISULTATI

La configurazione più performante delle VM Policloud (*medium*) si è rivelata **limitante** per Elastic



Il progetto è stato presentato come proof of concept su un insieme predefinito di keyword, cercate **offline** nel database

Use case #3

Sphaera

OBIETTIVO

Creazione di una web app di business intelligence per una catena di hotel, partendo dai dati delle prenotazioni

CRITICITÀ

sviluppo basato su tecnologie proprietarie ([Microsoft SQL server](#)), disponibili solo su piattaforma Windows

SETUP

Server DB
Windows



Middleware (API)
Python



Frontend del sito
HTML, CSS, JS

RISULTATI

Setup e sviluppo sono stati portati a termine senza problemi

A sviluppo terminato sono stati rilevati problemi di compatibilità tra *Windows Server 2012* e la piattaforma da parte di Policloud

VANTAGGI

Flessibilità

Sicurezza

Rapidità di setup

PROBLEMI

Downtime

Sono stati riscontrati frequenti periodi di downtime del servizio, di durata variabile



Servizio non ancora adeguato per ospitare siti di disseminazione e comunicazione, o applicazioni 24/7

Performance

Anche la configurazione più performante (*medium*) è limitante per applicazioni **compute-intensive**

PROSPETTIVE

Usare Policloud per **tutte le attività online**
del laboratorio (sito, blog, etc)

Usare una VM per **applicazioni cloud** di
project management, sincronizzazione
file e repository di codice

Sperimentare nuove tecnologie,
soprattutto in relazione a **Big Data**,
Internet of Things e **Machine Learning**.