

Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA "

CORSO DI LAUREA IN INFORMATICA



Introduzione al big data computing ed una sua applicazione

Tesi di laurea triennale

Relatore

Prof. Tullio Vardanega

Laureando

Stefano Panozzo

ANNO ACCADEMICO 2017-2018

Perché sognare è importante, ma impegnarsi concretamente lo è ancora di più.

— Pan di Stelle - Mulino Bianco

Sommario

Il presente documento descrive il lavoro svolto durante il periodo di stage del laureando Stefano Panozzo presso l'azienda I.T. Euro Consulting S.r.l. di Padova. Lo stage è stato svolto alla conclusione del percorso di studi della Laurea Triennale ed è durato in totale 320 ore. Gli obiettivi da raggiungere erano molteplici.

La prima richiesta dell'azienda era analizzare la struttura del [cluster](#) in cui risiedevano i dati utilizzati in seguito. Successivamente era richiesta l'analisi e la trasformazione del dataset di interesse per estrarre ed ottenere nuove informazioni utili per creare un modello che prevedesse il target desiderato. Infine, era richiesta la progettazione e lo sviluppo di una [web app](#) per la rappresentazione dei risultati ottenuti in precedenza. I primi due capitoli del presente documento hanno lo scopo di presentare il contesto aziendale in cui è stato sostenuto lo stage e di spiegare come il progetto di stage si renda utile all'interno della strategia aziendale. Il terzo capitolo documenta lo svolgimento dello stage descrivendo le attività che sono state portate a termine, i punti salienti del progetto stesso e le principali scelte progettuali. Il quarto ed ultimo capitolo presenta infine una valutazione dello svolgimento dello stage rispetto agli obiettivi aziendali e alle conoscenze acquisite dallo studente.

Ringraziamenti

Innanzitutto, vorrei esprimere la mia gratitudine al Prof. Tullio Vardanega, relatore della mia tesi, per l'aiuto e il sostegno fornitomi durante la stesura del lavoro.

Desidero ringraziare con affetto i miei genitori per il sostegno, il grande aiuto e per essermi stati vicini in ogni momento durante gli anni di studio.

Ho desiderio di ringraziare poi i miei amici per tutti i bellissimi anni passati insieme e le mille avventure vissute, che mi hanno fatto crescere e diventare la persona che sono ora.

Padova, Dicembre 2018

Stefano Panozzo

“Perché sognare è importante, ma impegnarsi concretamente lo è ancora di più.”

— Pan di Stelle

Indice

1	Il contesto aziendale	1
1.1	Il profilo aziendale	1
1.2	Dominio applicativo	2
1.3	Tecnologie utilizzate	2
1.4	Processi aziendali	2
1.5	Tipo di clientela	2
	Glossario	3
	Acronimi	5
	Bibliografia	7

Elenco delle figure

Elenco delle tabelle

Capitolo 1

Il contesto aziendale

Esempio di utilizzo di un termine nel glossario
[Application Program Interface \(API\)](#).

Esempio di citazione in linea
site:agile-manifesto.

1.1 Il profilo aziendale

I.T. Euro Consulting S.r.l.¹ è un'azienda di medie dimensioni con sede legale a Padova, nata nel 2007 e facente parte del gruppo SCAI, presente su tutto il territorio italiano. Dalla sua nascita si è sempre occupata prevalentemente di consulenza, *System Integration* ed *Application Management*, in ambito ICT, operando in tutti i principali settori di mercato: bancario ed assicurativo, industria, pubblica amministrazione e servizi. Nel corso degli anni l'azienda ha consolidato le proprie conoscenze soprattutto nei seguenti ambiti, offrendo svariati servizi:

- * **Big Data:** supporto alle aziende nel loro processo di crescita e cambiamento, tramite moderne soluzioni di *Business Intelligence* e la possibilità di prevedere scenari ed eventi futuri e prendere le più opportune decisioni operative o di business grazie all'analisi della gran mole di dati che ogni giorno vengono creati. Vengono quindi offerti servizi di big data engineer, big data scientist, big data architect e big data administrator;
- * **Internet of Things:** soluzioni end to end, basate su tecnologie leader di mercato che consentono di indirizzare in modo efficace la realizzazione di sistemi Iot accelerando la realizzazione di componenti web e mobile per la raccolta, la visualizzazione e l'analisi dei dati;
- * **Reference Architecture:** intesa come best practice e struttura di base per un insieme di domini applicativi all'interno di un'organizzazione, la quale agevola il continuo allineamento dei processi e delle strategie con le giuste soluzioni tecnologiche. Vengono quindi offerti servizi di assessment, design e consulenza;

¹<https://www.itecons.it>.

- * **DevOps:** automatizzazione delle attività manuali nelle diverse fasi del [Software Development Lifecycle](#). Il modello DevOps non si concentra esclusivamente sull' introduzione di nuovi tool, ma è inteso come una combinazione di cultura, processi unita agli strumenti di automazione. Vengono quindi offerti servizi di assessment e consulenza;
- * **System Integration:** servizi di consulenza o interventi progettuali per aiutare le aziende a gestire al meglio le proprie strutture tecnologiche complesse e soluzioni applicative per semplificare la coesione fra i vari sottosistemi che compongono la struttura;
- * **Application Management:** servizi di manutenzione correttiva, adattativa ed evolutiva di soluzioni applicative durante il loro intero ciclo di vita;
- * **Customer Relationship Management:** con l'obiettivo di ottenere una visione completa per perseguire uno scenario di *Single Customer View*, abilitante al dialogo one-to-one tra l'organizzazione ed il proprio cliente indipendentemente dalle canalità attraverso le quali avviene l'interazione;
- * **System & Data Administration:** servizio consultivo svolto avvalendosi di un insieme di strategie, processi e regole che consentono di gestire i sistemi e trattare i dati fondamentali per lo sviluppo aziendale. Vengono quindi offerti servizi di database administration, database security, data governance, data analysis e scheduling management.

1.2 Dominio applicativo

Introduzione all'idea dello stage.

1.3 Tecnologie utilizzate

1.4 Processi aziendali

1.5 Tipo di clientela

Glossario

cluster (indicato anche come computer cluster) insieme di macchine connesse tra loro che lavorano in parallelo. L'utilizzo di questi sistemi permette di distribuire un'elaborazione molto complessa tra le varie macchine, aumentando la potenza di calcolo del sistema e/o garantendo una maggiore disponibilità di servizio, a prezzo di un maggior costo e complessità di gestione dell'infrastruttura: per essere risolto, il problema che richiede molte elaborazioni, viene infatti scomposto in sottoproblemi separati i quali vengono risolti ciascuno in parallelo su tutti i nodi che compongono il cluster.. [v](#)

dataset . [v](#)

web app sistema di tipo client-server in cui l'interfaccia utente e la logica client-side viene eseguita in un browser web.. [v](#)

Acronimi

API [Application Program Interface](#). 1

Bibliografia