



**DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI
DEL TERRITORIO E DELLE COSTRUZIONI**

**RELAZIONE PER IL CONSEGUIMENTO DELLA
LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE**

***Studio e dimensionamento di un'architettura di tipo
Edge Computing per la nuova giga factory di Enel
Green Power***

SINTESI

RELATORI

Prof. Antonella Martini
*Dipartimento di Ingegneria dell'Energia, dei Sistemi,
del Territorio e delle Costruzioni*

Dott. Lorenzo Ricciardi Celsi, Ph.D.
Senior Consultant, ELIS Consulting & Labs

CANDIDATO

Giuseppe Intagliata
g.intagliata@studenti.unipi.it

Studio e dimensionamento di un'architettura di tipo Edge Computing per la nuova giga factory di Enel Green Power

Giuseppe Intagliata

Sommario

Questo lavoro di tesi nasce da un progetto formativo, della durata di 4 mesi, svolto all'interno del programma Junior Consulting, presso ELIS Consulting & Labs a Roma. Tale progetto, commissionato e svolto in collaborazione con Enel Green Power (EGP), ha come obiettivo principale il design di un'architettura di tipo Edge Computing per l'acquisizione e l'elaborazione dei dati, che verranno generati nella nuova Giga Factory 3SUN per la produzione dei moduli fotovoltaici. Data l'ingente quantità di dati provenienti dai vari dispositivi installati in fabbrica, si è reso necessario introdurre un'architettura Edge, capace di collezionare tale mole di dati, elaborarli "on board", e fornire un feedback al sistema con bassissimi tempi di latenza, consentendo, quindi, un'immediata risoluzione dei problemi che si verificano durante il processo di produzione. A tale scopo, il progetto ha riguardato (1) la realizzazione della topologia TO-BE del sistema di controllo, (2) la selezione degli hardware più adatti al sistema stesso.

Abstract

This thesis work stems from a training project, lasting 4 months, carried out within the Junior Consulting program, at Elis Consulting & Labs in Rome. This project, commissioned and carried out in collaboration with Enel Green Power (EGP), has as its main objective the design of an Edge Computing architecture for the acquisition and processing of data, which will be generated in the new 3SUN Giga Factory for the production of photovoltaic modules. Given the large amount of data coming from the various devices installed in the factory, it was necessary to introduce an Edge architecture, capable of collecting this amount of data, processing it "on board", and providing feedback to the system with very low latency times and thus enabling immediate troubleshooting of problems occurring during the manufacturing process. To this end, the project involved (1) the realization of the TO-BE topology of the control system, (2) the selection of hardware best suited for the system itself.

1. Contesto

Questo lavoro di tesi è frutto di un progetto formativo, della durata di quattro mesi, svolto all'interno del programma Junior Consulting, presso ELIS Consulting & Labs a Roma. Il progetto è stato commissionato da Enel Green Power (EGP), che è un'azienda presente all'interno del Gruppo Enel e che gestisce e sviluppa attività di generazione dell'energia da fonti rinnovabili a livello globale. L'azienda opera con oltre 1200 impianti in tutti e 5 i continenti e la sua capacità installata, proveniente da fonti rinnovabili, è di 54GW.

EGP sta realizzando una nuova giga factory per la produzione di moduli fotovoltaici, e il progetto ha come obiettivo principale il design di un'architettura di tipo Edge Computing per l'acquisizione e l'elaborazione dei dati che verranno generati dai vari dispositivi installati lungo la linea produttiva di questa fabbrica. Grazie all'introduzione dell'architettura Edge si vogliono migliorare i meccanismi di controllo del processo produttivo, riducendo così i tempi di intervento del gruppo di manutenzione in caso di insorgenza di problemi di non conformità o di deriva del processo e migliorando, allo stesso tempo, l'efficienza delle celle realizzate. Da parte di EGP sono stati presentati quattro casi d'uso da risolvere, ma al fine di non violare il segreto industriale all'interno di questo lavoro, non sono stati riportati molti dati sensibili e non sono stati trattati due casi d'uso su quattro.

2. Fasi del progetto

Il progetto si suddivide nelle seguenti fasi, mostrate in *Tabella 1*:

Tabella 1 - Fasi del progetto

| Fasi | Sottofasi | Step Metodologici | Risultati | § |
|--|--|--|---|---|
| Analisi della letteratura scientifica | | -Design della query -Ricerca articoli su Scopus -Analisi semantica -Selezione paper coerenti con lo scopo -Realizzazione di un riassunto delle informazioni trovate. | -Riassunto informazioni riguardanti l'Edge Computing | 3 |
| Studio della linea produttiva allo stato attuale | | | -Informazioni riguardanti lo stato dell'arte della linea produttiva | 4 |
| Analisi casi d'uso | -Studio dei casi d'uso -Realizzazione delle topologie AS-IS con spunti per la topologia TO-BE | | -Topologia AS-IS - Domande da porre a EDG rispetto ai punti critici emersi in fase di analisi dei casi d'uso | 5 |
| Realizzazione della topologia dell'infrastruttura TO-BE | -Acquisizione informazioni da EGP attraverso meeting periodici -Definizione delle feature tecniche -Realizzazione topologia TO-BE | | -Topologia TO-BE | 6 |
| Dimensionamento Edge Point | -Analisi dello storage -Analisi dei modelli che richiedono GPU e quelli che richiedono CPU -Analisi dell'ETL -Analisi delle risorse | | -Documento finale, riassuntivo delle analisi effettuate, da consegnare all'azienda cliente | 7 |

3. Analisi della letteratura scientifica

Quello dell'Edge computing è un dominio molto giovane, vasto e molteplici sono i campi di applicazione. Al fine di comprendere che cosa sia l'Edge Computing e come si possa successivamente applicare ai nostri casi d'uso, è stata svolta un'analisi delle pubblicazioni scientifiche tramite Scopus. Le attività svolte per estrapolare conoscenza dall'analisi degli articoli sono elencate nella seguente *Tabella 2*.

Tabella 2 - Tabella di sintesi dell'analisi della letteratura scientifica

| Attività | Metodologia e risultati |
|--|--|
| Creazione della chiave di ricerca | Criteria di ricerca: -Database: Scopus -Query: (edge AND computing) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (EXACTKEYWORD , "Deep Learning") OR LIMIT-TO (EXACTKEYWORD , "Deep Neural Networks") OR LIMIT-TO (EXACTKEYWORD , "Convolutional Neural Networks") OR EXCLUDE (EXACTKEYWORD , "Mobile Edge Computing") OR EXCLUDE (EXACTKEYWORD , "5G Mobile Communication Systems")) -Argomenti: Ingegneria e Computer science -Luogo: Titles, Keywords, Abstracts -Orizzonte temporale: from 2013 to 2023 Questa query ha portato a 1315 risultati. Tra questi, 15 sono stati raccolti come relativi all'attività successiva. |
| Analisi semantica | Metodologia: Worldcloud, Network analysis, Cooccurrence e topic modelling, per estrarre informazioni dai 15 paper precedentemente selezionati. |
| Selezione dei paper e realizzazione della sintesi dell'argomento | In base allo scopo del progetto, sono stati selezionati i migliori documenti di sintesi dell'argomento e sono stati estratti i concetti chiave. |

3.1 Risultati

Come già suggeritoci dall'analisi semantica, si è evidenziato nei documenti il fatto che l'architettura Edge non va a sostituirsi alla già utilizzata architettura Cloud, ma va ad affiancarla. Le ragioni, evidenziate in "Deep Learning With Edge Computing: A Review" (Chen, Jiasi & Ran, Xukan, 2019), sono le tre criticità del Cloud Computing, ovvero *latenza*, *scalabilità* e *privacy*. Queste criticità vengono affrontate durante lo studio d'introduzione di un sistema di Edge Computing, costituito principalmente da edge device, edge server, e cloud. La particolarità dell'edge computing risiede nel fatto che è possibile effettuare il computing vicino a dove i dati emergono.

Sono stati sintetizzati i punti di forza e di debolezza del modello in Figura 1.

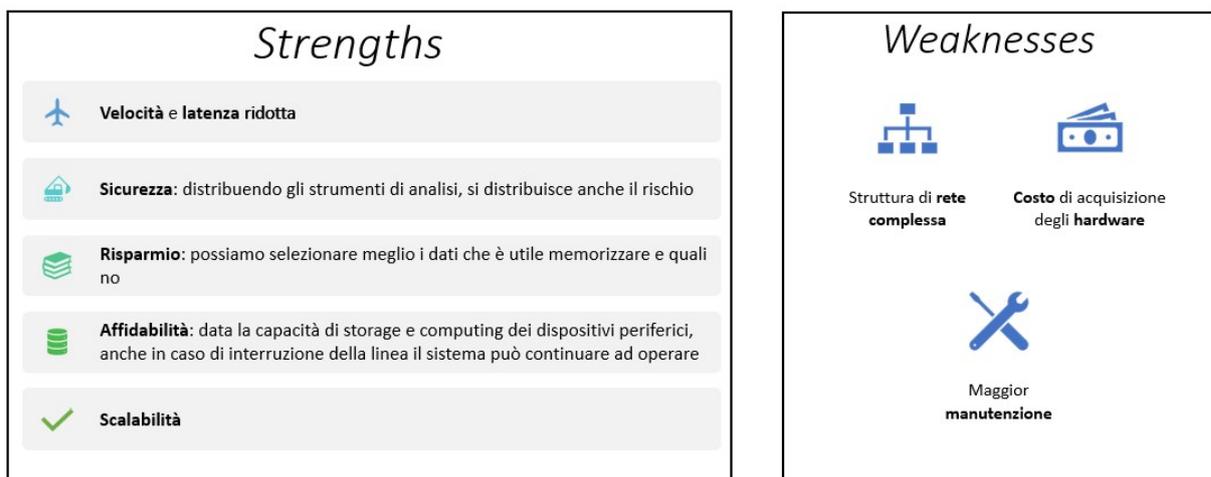


Figura 1 - Punti di Forza (Strengths) e di Debolezza (Weakness) dell'Edge Computing

Per poter eseguire un modello in un qualche punto della rete, prima esso deve essere addestrato e poi installato. La fase di addestramento e/o di installazione, possono seguire tre modalità diverse, che sono tre diverse topologie architetture, e queste sono:

1. *On-device computation*;
2. *Edge server computation*;
3. *Computing across edge device "misto"*.

Uno dei topic evidenziati dall'analisi effettuata riguardava l'offloading. A tal proposito, in letteratura viene discusso il processo di offloading del modello addestrato, del quale vengono evidenziate quattro modalità:

1. *Offloading binario* dei calcoli relativi al modello;
2. *Partizionamento del modello*, una parte del modello viene svolta sull'edge e una parte sul server (valido soprattutto per le reti neurali).
3. *Utilizzo di edge device e cloud*, una prima scrematura viene fatta sull'edge device con modelli leggeri, e in certe condizioni viene fatto un approfondimento del modello sull'edge server;
4. *Calcolo distribuito*, in cui le computazioni relative al modello vengono distribuite tra diversi edge device di supporto.

All'interno dei punti di forza è stata inserita la sicurezza: infatti, nel caso di architettura Edge, i dati provenienti dai vari dispositivi vengono elaborati nelle loro vicinanze e questo permette una distribuzione del rischio associato alla compromissione dei dati stessi.

4. Studio della linea produttiva allo stato attuale

Durante questa fase si è studiata la documentazione, riguardante la descrizione generale della linea produttiva, fornitaci da EGP.

Si è avuto così modo di comprendere il funzionamento della produzione dei moduli fotovoltaici a partire dai wafer grezzi fino al prodotto finito; quali macchine produttive siano necessarie al processo, quali quelle predisposte al controllo della linea stessa e come venga gestito l'handling.

Le macchine produttive sono:

- WIS (Wafer Inspection System) è una macchina di ispezione utilizzata per il controllo qualitativo dei wafer grezzi;
- PeCVD (Plasma Enhanced Chemical Vapor Deposition) dove vengono depositati gli strati di silicio amorfo drogato e non drogato per la generazione della giunzione, mentre nel PVD (Physical Vapor Deposition) viene eseguito il processo di sputterizzazione delle superfici del wafer;
- ci sono, inoltre, altre macchine produttive non menzionabili.

La maggior parte delle macchine di controllo opera in offline rispetto al flusso standard di produzione delle celle. Esse forniscono esclusivamente dati di misura quali, ad esempio, la riflettività, il lifecycle, il peso, etc....

Tra le macchine di metrologia troviamo anche la macchina di fotoluminescenza (PL), la macchina di elettroluminescenza (EL) e la Automatic Optical Inspection (AOI). Le prime due operano online e generano immagini oltre che dati di misura. La AOI è in grado di scartare le celle qualora non soddisfino i requisiti di specifica.

Tutte le macchine (equipment) di processo della linea di produzione sono connesse ad un database centralizzato (DB) che si occupa di collezionare ed archiviare i dati prodotti dalle equipment di processo.

In particolare, essendo ogni cella fotovoltaica identificata univocamente con un codice e tracciata lungo tutta la linea di produzione, potremo estrarre tutti i dati di interesse e correlarli con le altre sorgenti dati.

È bene evidenziare che solo un sottogruppo dei dati di metrologia è acquisito dal database centralizzato. Per la restante parte ci saranno altre tipologie di interfacce per l'acquisizione dei dati all'interno dell'architettura Edge.

Non è possibile illustrare la linea produttiva con tutti i soggetti presenti per ragioni di segreto industriale.

5. Analisi casi d'uso

5.1 Descrizione dei casi d'uso

Enel Green Power ha richiesto la risoluzione di quattro casi d'uso, due dei quali sono discussi in questo lavoro di tesi, che chiameremo caso d'uso 1 e il caso d'uso 2. Si ricorda che per non violare il segreto industriale, non è stato possibile illustrare gli altri due casi d'uso.

Il caso d'uso 1 consiste nel processamento delle immagini di fotoluminescenza e nella loro correlazione con tutti gli altri dati collezionati dalle altre equipment di processo e di metrologia. Tutto ciò al fine di ottimizzare il processo in questo tratto di linea e migliorare, di conseguenza, l'efficienza finale della cella.

Il caso d'uso 2 riguarda il processamento delle immagini di elettroluminescenza e fotoluminescenza per individuare l'occorrenza di malfunzionamenti lungo la linea di produzione.

5.2 Realizzazione topologia AS-IS e prime ipotesi per la topologia TO-BE

Per prima cosa si è cercato di comprendere il tipo di informazioni generate dai vari attori, rispetto ai casi d'uso, e a chi vengono trasmesse. Come risultato di questo lavoro sono state realizzate delle topologie AS-IS, di natura logica, dei due casi d'uso. Sono state quindi elaborate delle domande da porre ad EGP sulle diverse problematiche emerse, riguardanti l'introduzione di un sistema Edge all'interno dell'impianto e i singoli casi d'uso.

6. Realizzazione topologia TO-BE

6.1 Acquisizione informazioni da EGP attraverso meeting periodici

Le topologie sono state rivisitate in modo continuativo, per un paio di mesi, in seguito a meeting periodici con EGP. Durante questa fase, il team di cui ho fatto parte ha presentato lo stato dell'arte del lavoro svolto fino a quel punto e i dubbi emersi.

6.2 Risultati

I feedback periodici ci hanno permesso di avere chiarimenti relativi agli aspetti generali del processo produttivo, e di approfondire la conoscenza utile ad elaborare le topologie TO-BE dei casi d'uso.

6.2.1 Aspetti generali che riguardano il processo produttivo

Le informazioni riguardanti questo paragrafo sono state ricavate ponendosi le seguenti domande:

- *Quali dati si desidera memorizzare? Per quanto tempo? Quanto spazio è richiesto?*

In generale, le informazioni provenienti dalle differenti equipment, quali metrologia, parametri elettrici, etc., sono già memorizzati e storicizzati nel DB centralizzato.

Nel caso specifico, risulta utile storicizzare:

- le immagini provenienti da PL ed EL,
- i file di log generati dai modelli di defect detection, che possono essere eseguiti direttamente dal PC industriale fornito di telecamere. Questi modelli possono realizzare sia alcune fasi di preelaborazione dei dati, sia analisi più dettagliate e definitive sui dati stessi;
- i file di log che provengono dalle diverse equipment.

L'analisi dello storage necessario sarà successivamente riportata.

- *Esiste un tema di latency all'interno dell'impianto di produzione?*

Gli edge point e l'infrastruttura IT concentrati nella stessa sede fisica e i dati che viaggiano alla velocità della luce, determinano che per distanze di 100m, con cavi in rame, la latenza è pari a 0,5 μ s.

- *Che banda è necessaria per il trasferimento dei dati?*

La banda della rete informatica dell'impianto è almeno 1Gb/s. Per trasferire i dati delle immagini, a partire dai dati forniti, si ottengono i seguenti risultati:

- le telecamere (i macchinari) di EL e PL acquisiscono (processano) migliaia di immagini all'ora. Si trascurano, per dimensione, i file di log;
- ogni immagine ha un peso di circa 500KB a cui corrispondono 4Mb;
- in media sono richiesti circa 44Mb/s per inviare le 10 immagini allo storage, che è nettamente inferiore alla banda disponibile;
- avendo la macchina fotografica un processo di tipo discontinuo, che produce un numero elevato di foto in pochi secondi, il numero massimo di immagini che possono essere inviate al secondo è pari a 250. Se il numero di foto supera questa soglia, si incorre nel problema della latenza della trasmissione di queste immagini.

È improbabile che ci possa essere un problema di banda.

- *Esiste in letteratura qualche riferimento specifico che ci possa fornire suggerimenti utili ad effettuare la computazione su edge point?*

È stata condotta una nuova analisi della letteratura di riferimento, sempre tramite Scopus. Si sono estrapolate informazioni sul numero di immagini utilizzate per eseguire un addestramento e un'inferenza di modelli dati. Inoltre, i paper ritrovati suggeriscono il tipo di hardware necessario per eseguire questi task.

- *Che impatto ha l'ETL (Extract, Transform, Loading) sugli edge point?*

L'ETL è il processo di estrazione, trasformazione e carico dei dati. A partire dal DB centralizzato è possibile estrarre i dati di produzione di interesse per poter essere poi analizzati. Quindi, quando si parla di ETL, ci si riferisce al processo di estrazione dati dal DB, di lavorazione di tali dati e del successivo salvataggio nello Storage.

6.2.2 Realizzazione topologia TO-BE dei casi d'uso

Se con l'avvento dell'edge computing i tempi di latenza si riducono notevolmente, bisogna prima chiedersi se tale riduzione sia strettamente necessaria per i casi d'uso in esame. I tempi di risposta cui si farà riferimento saranno sempre considerati come "tempi di risposta affinché l'operatore intervenga sul posto", e saranno categorizzati in "LENTO" (entro l'ora), "RAPIDO" (pochi minuti), "ISTANTANEO" (pochi secondi). In questo lavoro non vi sono casi d'uso con tempi di risposta istantanei. Le rappresentazioni architettoniche sono di natura logica e non fisica; quindi, alcuni edge point non periferici potrebbero poi essere accorpati, a livello fisico, in un unico edge point periferico. Nei casi d'uso, il colore nero dei box e delle frecce sta ad indicare l'attrezzatura già presente nell'impianto, invece il colore rosso indica l'attrezzatura non ancora presente. I colori blu e giallo servono solo a distinguere tra loro le equipment.

Ai fini della descrizione delle topologie, nella *Tabella 3* viene riportato l'elenco delle risorse logiche (Edge Point -EP), il tipo di risorsa, l'equipment collegato e l'uso prevalente.

Tabella 3 - Elenco delle risorse logiche con la specificazione del tipo, dell'equipment di riferimento e dell'utilizzo

| Risorsa logica | Tipo Risorsa | Equipment | Uso prevalente |
|----------------|----------------|-----------|-----------------------------|
| EP 1 | PC industriale | WIS | |
| EP 2 | PC industriale | PeCVD | |
| EP 3 | PC industriale | PL | |
| EP 4 | PC industriale | EL | |
| EP 5 | Edge Server | PL | ML e Storage PL |
| EP 6 | Edge Server | EL | ML e Storage EL |
| EP 7 | Edge Server | | Raccordo EP5, EP6, EP8 e DB |
| EP 8 | Edge Server | | Raccolta log |
| EP 9 | PC industriale | AWS | |

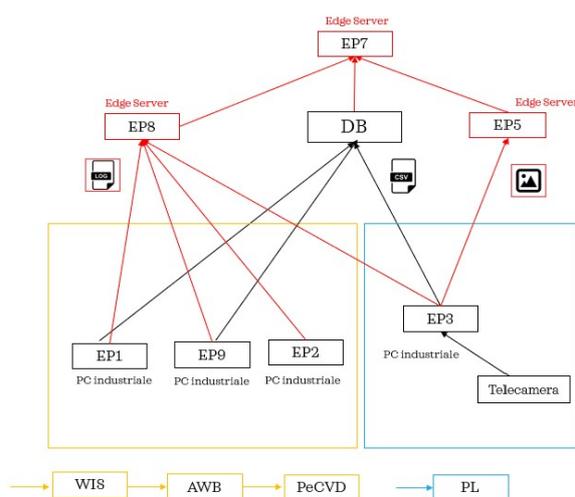


Figura 2 - Topologia del caso d'uso 1

In *Figura 2* viene mostrata la topologia del caso d'uso 1. Qui, per quanto riguarda i tempi di risposta, il processo di raccolta e analisi di dati così eterogenei e così numerosi, richiede molto tempo e la risposta non può che essere lenta.

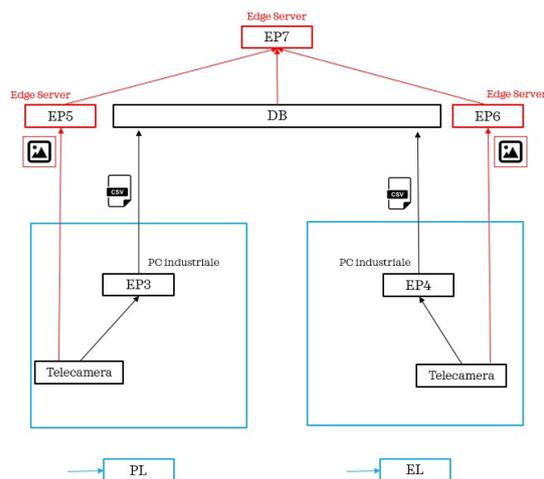


Figura 3 - Topologia del caso d'uso 2

In Figura 3 viene mostrata la topologia realizzata per il caso d'uso 2. Qui si vogliono individuare tempestivamente i problemi di handling che possono insorgere lungo la linea di produzione. È sufficiente un tempo di risposta rapido, perché poi l'alert deve essere confermato da un intervento umano.

I casi d'uso 3 e 4, come precedentemente detto, non possono essere riportati al fine di non violare il segreto industriale.

7. Dimensionamento degli Edge Point

Si sono dimensionati i vari edge point rispetto alla loro funzione (inferenza e/o addestramento del modello). Le topologie architetturelliche ritrovate in letterature sono, come detto, On-device computation, Edge server computation e Computing "misto". La scelta tra queste architetture dipende da:

- La necessità di bassa latenza, che comporta la scelta dell'architettura "on device computation";
- La necessità di una capacità di calcolo onerosa, che può comportare la scelta dell'architettura "Edge server computation" o "Computing misto";
- La necessità di integrare dati provenienti da dispositivi diversi, che costringe a ricorrere a Edge point non periferici e quindi al modello "Edge server computation" o "Computing misto".

Nei casi d'uso mostrati, si è in presenza del terzo fattore, in cui i dati provengono da edge point periferici diversi ed è necessario usare un server edge come punto di raccolta.

In particolare, la *Tabella 4* riassume e colloca questi concetti nei casi d'uso in esame:

Tabella 4 - Collocazione modelli

| COLLOCAZIONE MODELLI | TRAINING | PRODUZIONE | CASO D'USO |
|--|----------------------------|--------------------------------------|------------|
| Modello "leggero" per defect detection su telecamera di PL; Modello "robusto" più accurato per defect detection su EP5; Modello su EP7 per output finale del caso d'uso. | EP5, EP7 (+ CLOUD) | Telecamera PL, EP5, EP7 | 1 |
| Modello "leggero" per defect detection sulle telecamere di PL e EL; Modello "robusto" più accurato per defect detection su EP5 e EP6; Modello su EP7 per output finale del caso d'uso. | EP5, EP6, EP7 (+ CLOUD) | Telecamera PL e EL, EP5, EP6, EP7 | 2 |

7.1 Analisi dei modelli che richiedono la GPU

Prima di applicare un modello a livello di edge point, è necessario accertare la sua effettiva potenza di calcolo, ovvero la configurazione minima affinché esso sia in grado di realizzare analisi computazionali onerose. In tal senso si è deciso di fare una simulazione di tipo edge, con l'uso di un modello da applicare a numerosi campioni, monitorando ogni volta i tempi medi di analisi.

Si è scelto un modello di classificazione, adatto alle immagini, basato su rete neurale, appartenente alla famiglia EfficientNet e addestrato in ambiente TensorFlow. Si sono poi selezionate circa 100

immagini, ognuna di circa 500KB, simili a quelle previste dal nostro progetto, sui cui è stata addestrata la rete. Su altre 160 immagini è stato poi realizzato il test, salvando ogni volta il tempo complessivo per portare a termine il task. Il test è stato eseguito, sia con CPU che con GPU, prima su un PC locale e poi su Virtual Machine operanti su AWS.

I risultati sono riportati in *Tabella 5*.

Tabella 5 - Risultati della simulazione

| Macchina | Numero GPU | Numero CPU | RAM CPU | Risultato – CPU (sec) | Risultato - GPU (sec) | Immagini al secondo CPU | Immagini al secondo GPU |
|----------|------------|------------|---------|-----------------------|-----------------------|-------------------------|-------------------------|
| PC | | | 16 | 113 | 34 | 1,4 | 4,7 |
| g4dn | 1 | 4 | 16 | 72 | 18 | 2,2 | 8,9 |
| g4dn | 1 | 8 | 32 | 50 | 18 | 3,2 | 8,9 |
| g4dn | 1 | 16 | 64 | 38 | 18 | 4,2 | 8,9 |
| g4dn | 1 | 32 | 128 | 34 | 18 | 4,7 | 8,9 |
| g4dn | 4 | 48 | 192 | 31 | 18 | 5,2 | 8,9 |

Quindi, una macchina locale, dotata di GPU, ottiene risultati soddisfacenti. Se si ipotizzassero 9 immagini al secondo da analizzare, basterebbero macchine con 8 core e 16 GB di RAM.

7.2 Analisi dei modelli che si basano su CPU

Possiamo estrapolare le informazioni necessarie dal progetto “Digital Twin”, dove è stato realizzato un modello basato su una famiglia di modelli detti “Random Forest”.

Nel caso specifico, il PC locale aveva le seguenti caratteristiche:

- GPU: NVIDIA GeForce GTX 1650. RAM: 4GB
- CPU: I7-9750H CPU 2.60GHz, 6 Cores, 12 Logical Processor
- RAM: 16GB

7.3 Analisi dell’ETL

In generale non è facile dimensionare i server per l’ETL. Nel progetto “Digital Twin” si sono analizzati i dati provenienti da circa 200.000 wafer in due differenti periodi ed è stato realizzato un processo di ETL dei dati provenienti da tutte le equipment di processo con salvataggio dei dati intermedi e finali. Partendo da esso si sono potute recuperare le informazioni riguardanti la componentistica hardware utilizzata. Per fare ciò, è stato infatti utilizzato un PC locale con le seguenti caratteristiche:

- GPU: NVIDIA GeForce GTX 1650. RAM: 4GB
- CPU: I7-9750H CPU 2.60GHz, 6 Cores, 12 Logical Processor
- RAM: 16GB

Il processo aveva una durata ragionevole, ovvero pochi minuti per circa 200.000 wafer. Per questo motivo, si può ritenere che per realizzare l’ETL sia sufficiente un server di piccole dimensioni.

Relativamente alla parte storage del processo di ETL, l’analisi mostra che:

- per i file raw (originali), è necessario uno spazio di archiviazione di circa 600MB al giorno, quindi 219GB all’anno
- per i file semilavorati, è necessario uno spazio di archiviazione di circa 500MB al giorno, quindi 182GB all’anno

7.4 Analisi delle risorse

Nell’ambito Cloud esistono cinque tipi di risorse hardware: per uso generale; ottimizzate per il calcolo senza GPU; ottimizzate per il calcolo con GPU; ottimizzate per lo storage; ottimizzate per la memoria.

L'analisi delle risorse disponibili ci ha permesso l'elaborazione delle seguenti ipotesi:

- Gli EP5 e 6 possono essere entrambe composte da due macchine fisiche:
 - Una per lo storage con almeno 200TB di hard disk, per contenere un anno di dati. Essa può appartenere alla categoria "ottimizzate per lo storage"
 - Una per analisi dei modelli legati alle immagini, che può appartenere alla categoria "ottimizzate per il calcolo con GPU".

- L'EP7 ha quattro scopi e quindi potrebbe essere composta da quattro macchine fisiche:
 - Una per la realizzazione dell'ETL, che potrebbe appartenere alla categoria "Ottimizzate per la memoria";
 - Una per lo storage dei file grezzi e semilavorati dell'ETL, che può appartenere alla categoria "ottimizzate per lo storage". È difficile fare una stima dei dati, ma, come descritto nel lavoro di tesi, l'ammontare dei dati potrebbe crescere, raggiungerebbe difficilmente una quantità di dati pari a 1TB, che è la dimensione abituale di un hard disk;
 - Una per analisi dei modelli legati ai dati provenienti dal DB centralizzato. Essa potrebbe appartenere alla categoria di risorse "ottimizzate per il calcolo senza GPU". Qualora i modelli richiedessero anche un'analisi delle immagini, essa si può sostituire con una macchina appartenente alla categoria di risorse "ottimizzate per il calcolo con GPU";
 - Una per lo storage dei dati in un Data Warehouse (strutture tabellari colonnari adatte per la Business Intelligence). Questa macchina può appartenere alla categoria di risorse "ottimizzate per la memoria".

- L'EP8 è una risorsa logica pensata come edge point di storage dei log provenienti da tutte le equipment. La dimensione occupata dai file di log è piuttosto scarna, però il numero di macchinari usati nell'impianto è notevole, infatti è circa pari a 100. Quindi, considerando che ogni macchinario richiede 2MB per i file di log dell'ultimo mese e 2,4MB per i file di log di un anno, lo spazio totale richiesto è circa 240MB all'anno, e 200MB per l'ultimo mese di dati. Perciò è sufficiente un server che può appartenere alla categoria "ottimizzate per lo storage".

7.4.1 Prospetto conclusivo

Come risultato conclusivo dell'analisi svolta, è stato realizzato un prospetto riportato in *Tabella 6*.

Tabella 6 - Prospetto conclusivo dell'analisi svolta

| Risorsa fisica | Hard Disk | RAM | Numero | Risorsa logica | Locazione | Note |
|-----------------------------|-----------|------------|--------|----------------|-------------------------|--|
| Storage immagini PL | 180TB | | 1 | EP5 | Server room | Sistema RAID on premise. Un anno di dati. Server room per dimensione eccessiva |
| Storage immagini EL | 180TB | | 1 | EP6 | Server room | Sistema RAID on premise. Un anno di dati. Server room per dimensione eccessiva |
| Storage file ETL | 400GB | | 1 | EP7 | Server room | Sistema RAID on premise. Un anno di dati. |
| Storage log | 1GB | | 1 | EP8 | Server room | Un anno di dati |
| Storage file Data Warehouse | | | | EP7 | Server room | Da valutare caso per caso. Normalmente nettamente inferiori allo storage ETL |
| ML per PL | 64GB | 16GB (GPU) | 2 | EP5 | Edge point/ Server Room | Potrebbe aver bisogno delle immagini, che vengono inviate nella Server Room |
| ML per EL | 64GB | 16GB (GPU) | 2 | EP6 | Edge point/ Server Room | Potrebbe aver bisogno delle immagini, che vengono inviate nella Server Room |
| ML su dati di processo | 64GB | 8GB (CPU) | 1 | EP7 | Server room | |
| ETL | 64GB | 8GB (CPU) | 1 | EP7 | Server room | |

8 Conclusioni

L'obiettivo di questo progetto era quello di studiare l'architettura Edge computing e le sue applicazioni, allo scopo di inserirla all'interno di un contesto manifatturiero qual è la realizzazione di moduli fotovoltaici nella nuova Giga Factory di Enel Green Power.

A tale scopo sono state condotte le seguenti analisi:

- analisi della letteratura scientifica di riferimento;
- studio della linea produttiva;
- analisi dei casi d'uso;
- analisi delle risorse necessarie;
- dimensionamento degli edge point.

I risultati ottenuti dallo studio sono stati inseriti all'interno di un documento di sintesi presentato all'azienda cliente.

Sfruttando le informazioni contenute all'interno del documento, EGP può avere le idee chiare sul tipo di architettura e come essa possa essere implementata all'interno della sua linea produttiva.