

UNIVERSITÀ TELEMATICA INTERNAZIONALE
UNINETTUNO

INGEGNERIA INFORMATICA
Corsi di Laurea Magistrale - Big Data

Introduzione ai Big Data



Progetto su R, MapReduce e Hadoop "Contabilità ditta"

Docente:
Prof. Pirrone Daniele

Studente:
Emanuele Coltro
mat: 671HHHINGINFOR

Anno Accademico 2022-2023

Abstract

Il presente paper ha l'obiettivo di esaminare le potenzialità e le performance degli strumenti utilizzati per la gestione e l'analisi dei Big Data. Viene approfondito il processo che porta alla produzione di report grafici, tabelle e schemi a partire dai dati di un problema o di una esigenza. La relazione fornisce informazioni dettagliate su come impostare un progetto, dalla scelta e impostazione dell'ambiente di sviluppo alla contestualizzazione delle decisioni prese. Inoltre, vengono allegati i codici per la generazione dei risultati e dei grafici.

Il progetto in esame mira a soddisfare l'esigenza dell'azienda (fittizia) STEC-CAPARAPETUTTI S.R.L. di risolvere un problema di contabilità mediante l'estrazione di alcune feature dai dati. A tal fine, sono stati impiegati strumenti avanzati per l'elaborazione dei dati, tra cui algoritmi di MapReduce e tecniche di visualizzazione avanzata.

L'analisi dei dati è stata eseguita utilizzando il linguaggio R per una prima analisi, seguito dall'utilizzo di un tool scritto in JavaScript per prototipare il sistema di MapReduce e risolvere il problema.

Indice

| | | |
|----------|--------------------------------|----------|
| 1 | Introduzione | 3 |
| 1.1 | Che cos'è LCA | 3 |
| 1.2 | Da dove nasce | 4 |
| 1.3 | Utilizzo e obiettivi | 5 |
| | Allegato 1 | 7 |

1 Introduzione

L'analisi del ciclo di vita (LCA) è una tecnica utilizzata per valutare l'impatto ambientale di un prodotto, processo o attività per l'intero ciclo di vita, dall'estrazione delle materie prime alla sua eliminazione. È uno strumento completo che considera tutte le fasi del ciclo di vita di un prodotto, compresa la produzione, l'uso e lo smaltimento.

È particolarmente rilevante dal punto di vista della sostenibilità, poiché permette di valutare l'impatto ambientale di un prodotto o servizio lungo tutto il suo ciclo di vita, evitando che i miglioramenti locali comportino solo lo spostamento dell'impatto ambientale altrove. In questo modo, si può affrontare la causa del problema alla radice e non solo i sintomi.

Con l'aumento della popolazione mondiale e dello sviluppo economico, l'impatto ambientale delle nostre attività sta diventando sempre più evidente.

Questo strumento può essere utilizzato da aziende, organizzazioni e dal settore pubblico per prendere decisioni informate sulla gestione ambientale dei loro prodotti e servizi: emissioni di gas a effetto serra, l'inquinamento dell'aria, dell'acqua e del suolo, e l'uso di risorse naturali. L'analisi del ciclo di vita può anche essere utilizzata per identificare i punti critici del ciclo di vita di un prodotto o servizio, in modo da poterli migliorare.

1.1 Che cos'è LCA

Le questioni ambientali stanno giocando un ruolo sempre più importante nel processo decisionale a ogni livello: politico, economico, industriale e individuale. Più di una semplice tendenza passeggera, l'attenzione crescente dedicata ai problemi ambientali deriva da una semplice osservazione: a causa della sua limitata capacità di assorbire gli effetti delle attività umane, l'ambiente stabilisce un limite allo sviluppo della società. Questo limite è già stato raggiunto in molte regioni del pianeta[2].

I concetti di sostenibilità sono costantemente discussi nei titoli, ma diventa sempre più difficile agire. Per garantire un futuro sostenibile, le dichiarazioni e gli studi devono essere seguiti da azioni significative che riducano efficacemente l'impatto ambientale e che possano persino migliorare la situazione. Perché un'azione sia efficiente, devono essere soddisfatte tre condizioni:

- Le soluzioni tecnologiche devono essere disponibili.
- Diverse soluzioni devono essere priorizzate e selezionate le migliori pratiche, tenendo conto dell'efficienza ambientale, dei costi e dei vincoli economici risultanti.

- Le azioni dovrebbero essere ottimizzate per ridurre ulteriormente gli impatti.

L'analisi del ciclo di vita (LCA) è uno strumento decisionale che affronta specificamente questa esigenza di selezionare e ottimizzare le soluzioni tecnologiche disponibili. Ciò è fondamentale quando le risorse finanziarie sono limitate, poiché Barlow ha affermato in modo leggermente provocatorio: "Il problema non è come affrontare il singolo problema: l'ingegneria è disponibile o può essere sviluppata per affrontare quello. Piuttosto, il problema è come decidere le priorità. Il mondo semplicemente non può permettersi di fare tutto". L'analisi del ciclo di vita è un complemento agli sviluppi tecnologici, poiché evidenzia quali processi dovrebbero essere migliorati in ordine di priorità.

Rilevante inoltre dal punto di vista della sostenibilità, poiché copre l'intero ciclo di vita di un prodotto o servizio, evitando che i miglioramenti locali comportino solo lo spostamento dell'impatto ambientale altrove. Differisce poi da altri metodi ambientali collegando le prestazioni ambientali alla funzionalità, quantificando le emissioni inquinanti e l'uso di materie prime in base alla funzione del prodotto o del sistema.

Nonostante l'analisi del ciclo di vita presenti molti vantaggi, non è priva di limiti. Analogamente alla contabilità economica per la stima del costo effettivo di un prodotto, la contabilità ecologica richiede un certo numero di ipotesi che devono essere logiche e coerenti. Alcune applicazioni dell'LCA sono state duramente criticate, suggerendo che un determinato metodo di analisi del ciclo di vita è stato selezionato per ottenere i risultati attesi dal finanziatore dello studio.

Per coprire un insieme più ampio di impatti ambientali, è emersa la necessità di contabilizzare le emissioni inquinanti nell'aria, nell'acqua e nel suolo. Ciò ha portato a sviluppi metodologici, inizialmente nell'industria del packaging, che alla fine sono stati applicati a tutti i settori economici, poiché è emerso che il prodotto spesso aveva un impatto molto più grande della sua confezione.

1.2 Da dove nasce

Tre organizzazioni sono state coinvolte nello sviluppo e nella standardizzazione dell'analisi del ciclo di vita: la Società di tossicologia ambientale e chimica (SETAC), il Programma delle Nazioni Unite per l'ambiente (UNEP) e la ISO. A partire dai primi anni '90, SETAC ha offerto una piattaforma di scambio scientifico per lo sviluppo dell'analisi del ciclo di vita, ancora in corso oggi attraverso presentazioni di conferenze e gruppi di lavoro.

Dal 2002, l'iniziativa del ciclo di vita è stata un importante quadro istituzionale per lo sviluppo dei metodi dell'analisi del ciclo di vita e il loro utilizzo nell'industria. Lanciata da SETAC e UNEP, questa iniziativa mira a sviluppare

e diffondere strumenti pratici per valutare soluzioni, rischi, vantaggi e svantaggi associati a prodotti e servizi durante tutto il loro ciclo di vita. La prima fase si è svolta nel periodo 2002-2007, sviluppando un consenso sulle modalità di approccio del ciclo di vita. Ciò è stato seguito da una seconda fase dal 2007 al 2011, con l'obiettivo di diffondere la consapevolezza e l'utilizzo degli approcci del ciclo di vita in tutto il mondo. La terza fase (dal 2012 ad oggi) sta sviluppando un consenso sugli indicatori di impatto e fornendo orientamenti per l'analisi del ciclo di vita organizzativa, che considera gli impatti del ciclo di vita di una determinata azienda o organizzazione, compresa la fornitura, l'uso e lo smaltimento dei suoi prodotti e servizi.

Durante gli anni '80 e '90, ISO¹ ha pubblicato oltre 350 standard relativi alle questioni ambientali, in particolare, la serie ISO 14000 sui sistemi di gestione ambientale, aggiornando e fornendo un quadro per le aziende per gestire l'impatto ambientale delle loro attività e misurare le loro prestazioni ambientali. La serie ISO 14040 (14040-14049) è dedicata all'analisi del ciclo di vita (Tabella 1.2). Il primo standard (ISO 14040) stabilisce le linee guida per l'esecuzione di un'analisi del ciclo di vita. ISO 14044 ha sostituito ISO 14041, 14042 e 14043 nel 2006 per descrivere le fasi di inventario, valutazione dell'impatto e interpretazione. Esempi della sua applicazione sono presentati in ISO 14047 e 14049, e ISO 14048 descrive il formato di documentazione dei dati.

1.3 Utilizzo e obiettivi

Molte associazioni di imprese e aziende industriali già utilizzano l'approccio LCA nel quadro della sostenibilità. Questo metodo è sempre più utilizzato dall'industria per aiutare a ridurre il carico ambientale complessivo lungo l'intero ciclo di vita dei beni e dei servizi, migliorando la competitività dei prodotti dell'azienda e nella comunicazione con gli organi governativi.

Oltre a supportare la presa di decisioni, l'LCA viene ampiamente utilizzato per migliorare la progettazione del prodotto, ad esempio nella scelta dei materiali, nella selezione delle tecnologie, nei criteri di progettazione specifici e nella considerazione del riciclaggio.

Un'altra possibilità interessante fornita dall'LCA è il benchmarking delle opzioni del sistema di prodotto, che può essere utilizzato anche nella decisione di acquisto e investimenti tecnologici, nei sistemi di innovazione, ecc. È uno strumento unico in quanto fornisce una panoramica dei compromessi a monte e a valle associati alle pressioni ambientali, alla salute umana e al consumo di ri-

¹ISO produce standard internazionali per la maggior parte dei settori tecnologici. Gli standard ISO sono adattati alle applicazioni industriali e derivano da un consenso tra esperti di vari background, tra cui industria, tecnologia, economia e accademia.

sorse. Queste informazioni a macro scala completano altre valutazioni sociali, economiche ed ambientali.

Anche il settore pubblico fa uso del pensiero del ciclo di vita nelle consultazioni con gli stakeholder e nell'attuazione delle politiche. Ciò garantisce che si tenga in considerazione il quadro generale nelle valutazioni ambientali orientate alle politiche, considerando i compromessi a monte e a valle. L'LCA contribuisce infatti a una politica di prodotto efficiente fornendo informazioni preziose aggiuntive sulle prestazioni ambientali di beni e servizi.

Diversi sono i livelli di contribuzione nelle strategie politiche come:

- supporto all'attuazione della strategia tematica dell'UE sull'uso sostenibile delle risorse naturali[1];
- strategia tematica sulla prevenzione e il riciclaggio dei rifiuti
- definizione di criteri di eco-design, contribuendo alla definizione di obiettivi di prestazione nell'ambito del Piano d'azione per la tecnologia ambientale (ETAP) e per i prodotti ad uso energetico nell'ambito della Direttiva EuP, negli appalti pubblici verdi (GPP) e nelle dichiarazioni ambientali di prodotto (EPD).

E' infine importante tenere presente che l'utilizzo dell'analisi del ciclo di vita è un semplice strumento di supporto alla decisione, piuttosto che uno strumento decisionale, poiché ha uno specifico focus. In particolare, tende a escludere gli impatti economici e sociali, nonché la considerazione di questioni ambientali più locali. Pertanto, è necessario utilizzarlo in combinazione con altri strumenti per identificare le aree di miglioramento potenziale.

Allegato 1

Fattori di standardizzazione in base alle tematiche

Energia Primaria

```
1 # Esempio di codice Python
2 import numpy as np
3
4 def incmatrix(genl1,genl2):
5     m = len(genl1)
6     n = len(genl2)
7     M = None #to become the incidence matrix
8     VT = np.zeros((n*m,1), int) #dummy variable
9
10    #compute the bitwise xor matrix
11    M1 = bitxormatrix(genl1)
12    M2 = np.triu(bitxormatrix(genl2),1)
13
14    for i in range(m-1):
15        for j in range(i+1, m):
16            [r,c] = np.where(M2 == M1[i,j])
17            for k in range(len(r)):
18                VT[(i)*n + r[k]] = 1;
19                VT[(i)*n + c[k]] = 1;
20                VT[(j)*n + r[k]] = 1;
21                VT[(j)*n + c[k]] = 1;
22
23            if M is None:
24                M = np.copy(VT)
25            else:
26                M = np.concatenate((M, VT), 1)
27
28            VT = np.zeros((n*m,1), int)
29
30    return M
```

Listing 1: Python example

```
1
2 #JOB 1
3 media_mesi <- aggregate(documenti_vendita$costo, by = list(
4     mesi), FUN = mean)
5
6
7
8 colnames(media_mesi) <- c("data","vendite_medie")
9
10
11
12 nomi_mesi <- c("Gen", "Feb", "Mar", "Apr", "Mag", "Giu",
```



```

9           "Lug", "Ago", "Set", "Ott", "Nov", "Dic")
10
11 media_mesi <- media_mesi %>%
12   mutate(mese = substring(media_mesi$data, 5, 6))
13 media_mesi$anno <- substr(media_mesi$data, 1, 4)
14
15 media_mesi <- media_mesi %>%
16   mutate(mese_testuale = nomi_mesi[as.numeric(mese)])
17
18 media_mesi <- media_mesi %>%
19   mutate(data_formattata = paste(anno, mese_testuale, sep = "
    -"))
20
21 media_mesi <- media_mesi %>%
22   select(data, vendite_medie, data_formattata)
23
24
25
26 #JOB 2
27 varianza_mesi = aggregate(documenti_vendita$costo, by = list(
    mesi), FUN = var)
28 colnames(varianza_mesi) <- c("data", "varianza")
29 varianza_mesi$anno <- substr(varianza_mesi$data, 1, 4)
30 varianza_mesi$mese <- substr(varianza_mesi$data, 5, 6)

```

Questo indicatore considera la richiesta di energia primaria per l'intero ciclo di vita del prodotto considerato, tenendo conto, ad esempio, della trasformazione dei materiali combustibili in energia elettrica.

A questo indicatore contribuiscono quindi i materiali combustibili con il loro contenuto di energia primaria.

Il fattore di caratterizzazione è in questo caso il potere calorifico del materiale considerato.

Effetto Serra

L'indicatore effetto serra viene calcolato considerando, tra le sostanze emesse in aria, quelle che contribuiscono al potenziale riscaldamento globale del pianeta Terra.

La quantità in massa di ciascuna sostanza, calcolata sull'intero ciclo di vita del prodotto, viene moltiplicata per un coefficiente di peso chiamato potenziale di riscaldamento globale (GWP, Global Warming Potential). Sommando poi i contributi delle varie sostanze, si ottiene il valore aggregato dell'indicatore.

Le sostanze che contribuiscono all'effetto serra sono principalmente: CO₂, CH₄, N₂O, CFC, gli HCFC e gli HFC.

La CO₂ è la sostanza di riferimento per questo indicatore, vale a dire che il suo coefficiente di peso è uguale a 1 e i valori dell'indicatore sono espressi in kg di CO₂ equivalente (kg CO₂ eq).

| Composto | Formula GWP100 | [kg CO ₂ /kg gas] |
|-------------------------------|----------------------------------|------------------------------|
| Diossido di carbonio | CO ₂ | 1 |
| Ossido di carbonio | CO | 2 |
| Metano | CH ₄ | 11 |
| Ossido di azoto | N ₂ O | 320 |
| CFC-11 | CFCI ₃ | 4.000 |
| CFC-12 | CF ₂ CI ₂ | 8.500 |
| Clorotrifluorometano (CFC-13) | CF ₃ CI | 11.700 |
| Tetrafluorometano (CFC-14) | CF ₄ | 9.300 |
| HCFC-22 | CHF ₂ CI | 1.700 |
| HCFC-125 | CHF ₂ CF ₃ | 3400 |
| Halon-1301 | CF ₃ Br | 5.600 |
| Diclorometano | CH ₂ CI ₂ | 25 |
| Cloroformio | CHCI ₃ | 15 |

Tabella 1: Fattori di standardizzazione per i principali responsabili dell'effetto serra, basati sul loro diretto contributo al riscaldamento globale con un tempo-orizzonte di 100 anni.

Assottigliamento della fascia di ozono stratosferico

La riduzione della fascia di ozono stratosferico viene calcolata come l'indicatore precedente, ma utilizzando diverse sostanze (CFC, HCFC) e un diverso coefficiente di peso, chiamato potenziale di riduzione dell'ozono (ODP, Ozone Depletion Potential).

La sostanza di riferimento in questo caso è un clorofluorocarburo, precisamente il CFC-11.

Eutrofizzazione

Questo indicatore valuta l'effetto dell'eutrofizzazione, ovvero l'aumento della concentrazione di sostanze nutritive negli ambienti acquatici. Le sostanze che contribuiscono al fenomeno dell'eutrofizzazione sono i composti a base di fosforo e azoto.

La sostanza di riferimento è il fosfato (PO₄) e il coefficiente di peso prende il nome di potenziale di nutrizione (NP, Nutrifcation Potential).

| Formula | NEP [kg NO ₃ -/kg compost] |
|-------------------|---------------------------------------|
| NO ₃ - | 1 |
| NO ₂ | 1.35 |
| NO _x | 1.35 |
| NO | 2.07 |
| N ₂ O | 2.82 |
| NH ₃ | 3.64 |
| HCN | 2.29 |
| N | 4.43 |
| PO ₄ — | 10.45 |
| P | 32.03 |

Tabella 2: Fattori di standardizzazione per i principali responsabili dell'effetto serra, basati sul loro diretto contributo al riscaldamento globale con un tempo-orizzonte di 100 anni.

Formazione di smog fotochimico (photo-smog)

Il termine "smog estivo" si riferisce a tutte le sostanze organiche volatili che portano alla formazione di ozono troposferico attraverso reazioni fotochimiche (in presenza di radiazione solare).

Il fattore di caratterizzazione utilizzato è chiamato "potenziale di formazione di ozono fotochimico" (POCP, Photochemical Ozone Creation Potential) e la sostanza di riferimento è l'etilene (C₂H₄).

| Composto | POCP [g C ₂ H ₄ /g di composto] |
|----------|---|
| metano | 0,007 |
| etano | 0,100 |
| propano | 0,500 |
| aldeidi | 0,3±0,2 |
| CO | 0,040 |
| metanolo | 0,123 |
| etanolo | 0,268 |

Tabella 3: Fattori di standardizzazione per i principali responsabili dello smog fotochimico.

Rifiuti Solidi

L'indicatore in questione raggruppa tutti i rifiuti di tipo solido generati in qualsiasi attività nel ciclo di vita di un prodotto, ad esempio durante la generazione di

energia elettrica necessaria per una lavorazione o durante la produzione delle lamiere di acciaio.

Non esistono fattori di caratterizzazione per questo indicatore, e ogni sostanza viene sommata alle altre tenendo semplicemente conto della quantità emessa in massa.

Riferimenti bibliografici

- [1] European Commission. Environment sustainable use of natural resources. url: <https://ec.europa.eu/environment/archives/natres/index.htm>. Accessed: 2023-04-10.
- [2] O. Jolliet, M. Saade, S. Shaked, and A. Jolliet. *Environmental Life Cycle Assessment*. Taylor & Francis, 2015.