

**<Irpel>**

**<Servizio di creazione di una procedura informatica in grado di elaborare i dati in possesso di IRPET al fine di ricostruire le storie lavorative dei toscani che operano presso altre regioni, con l'adeguata formazione di personale IRPET relativa al programma informatico sopra richiesto>**

Proposta Tecnica

Danilo Maurizio

20 Ottobre 2022

(CIG): ZB63826C6E

# Indice

<b>Indice</b>	<b>2</b>
<b>Versione del documento</b>	<b>3</b>
<b>Scenario e Obiettivi principali</b>	<b>4</b>
<b>Attività di adeguamento previste</b>	<b>4</b>
Punti di attenzione	6
Versionamento del codice	7
<b>Takeover sulle procedure e piano formativo</b>	<b>7</b>
La procedura	7
Bitbucket	8
Prerequisiti	8

## Versione del documento

Versione	Data	Note	Autori
v 1.0	20 Ottobre 2022	Offerta Tecnica	Danilo Maurizio

## Scenario e Obiettivi principali

La collaborazione tra Ehm Lab (EHM) e Irpet si pone come obiettivo quello di adeguare le procedure di valorizzazione della Base Dati Amministrativa IDOL sviluppate nel corso del tempo. La fonte dati ha subito infatti modifiche che impongono la riscrittura di porzioni di codice. Allo stesso tempo risulta necessario adeguare la *codebase* per la correzione di un bug, per gestire una problematica dovuta all'obsolescenza dello stack (RDBMS e *temporary table*), per l'ottimizzazione dei tempi di esecuzione di alcune parti di procedura e per recepire una nuova esigenza in tema di analisi.

Contestualmente le modifiche verranno sviluppate e concordate per aumentare la presa sul progetto da parte degli specialisti interni Irpet (formazione e *training on the job*).

## Attività di adeguamento previste

Il progetto prevede attività di manutenzione, ordinaria e straordinaria delle procedure attualmente in produzione, in queste specifiche direzioni:

1. Manutenzione straordinaria - imposta per via delle modifiche alle logiche e ai tracciati di base processati dalla attuale *codebase* con particolare riferimento a:
  - a. revisione tracciati - alcune colonne non sono più presenti nelle tabelle in export e allo stesso tempo se ne sono aggiunte di nuove. Alcune strutture in export possono invece dirsi superate;
  - b. revisione date - il formato tecnico dei campi data risulta ora diverso per alcune delle tabelle in export;
  - c. superamento *flag\_master* - il campo specifico era pivot per la gestione delle attività di deduplica e validazione delle CO. Una procedura a monte dei processi di analisi Irpet mette in coerenza i dati e rende superflua la

gestione del campo stesso.

2. Evolutive - ci riferiamo allo sviluppo di specifico codice reso necessario da nuovi fabbisogni informativi da soddisfare; il riferimento è a strutture dati per l'analisi dei lavoratori con domanda di lavoro espressa da aziende o UL extra Toscana.
3. Bug Fixing - la normale attività di analisi dei dati ha portato in emersione una misclassificazione dei tempi di durata dei rapporti di lavoro quando presenti specifiche comunicazioni di trasformazione oraria (TOR).
4. Ottimizzazione - oltre alle normali attività di ottimizzazione del codice e di riorganizzazione degli step di processing della procedura imposti da quanto ai punti precedenti si rende necessaria anche una specifica attività di *tuning* che va a colmare una lacuna/bug dello stack tecnologico in uso.

**In riferimento al punto 1** si rende necessario rivedere il gruppo di script che si occupa di fare l'ingestion dei dati csv originari, che definisce i metadata delle strutture RDBMS ospitanti (dal DDL all'indexing - dove necessario) e il lineage per tutte quelle istruzioni SQL che usano i dati modificati. Allo stesso tempo il codice va depurato di tutte le istruzioni e strutture intermedie che gestiscono il *flag\_master*.

**In riferimento al punto 2** si rende necessario un fork del template ETL *01\_rapporti\_lavoro\_unica.tpl* che con una diversa batteria di condizioni di inclusione ed esclusione permetta di materializzare una nuova struttura (la *CO\_UNICA\_ET* nello schema %stg1) capace di isolare le comunicazioni obbligatorie (e quindi i rapporti) di lavoratori toscani in aziende/UL localizzate fuori dalla Regione Toscana. La richiesta prevede infatti di abilitare l'analisi valutativa di attività di formazione e necessita di seguire i lavoratori anche fuori dai confini regionali.

In riferimento al punto 3 si rende necessario rivedere la condizione riferibile all'istruzione `WHEN tipo_mov = 'TOR'` nel blocco ELT del template `01_rapporti_lavoro_unica.tpl`.

In riferimento al punto 4 e in virtù di un naturale processo di obsolescenza<sup>1</sup> dello stack HW/SW si rende necessario rivedere l'uso delle *temporary table* di MariaDB/MySQL nella versione attualmente in produzione (10.1.23-MariaDB-1~trusty based on MySQL 5.5.5). Al crescere della mole dei dati si presenta un eccessivo consumo di spazio sulle cartelle /tmp fino al determinarsi di una completa saturazione che blocca la procedura con specifico *error code*. Al fine di limitare tale eventualità la procedura sarà sottoposta ad *audit* e le tabelle temporanee più grosse sostituite dalla materializzazione e successiva cancellazione delle tabelle di passaggio/strutture intermedie.

## Punti di attenzione

Al fine di garantire al meglio l'esecuzione delle attività richieste segnaliamo questi punti di attenzione come opportunità:

- disponibilità di un ambiente di test - sarebbe opportuno che per le attività dettagliate nel presente documento fosse predisposto un ambiente di test. Diversamente provvederemo allo sviluppo utilizzando degli *schema* dedicati alle attività di test nella macchina di produzione;
- disponibilità di nuove versioni dei dati di input - per esaminare le modifiche ai file di input riferiti al punto (1) del paragrafo precedente e per riscrivere le parti di codice necessarie è necessario entrare in possesso di vecchie e nuove versioni dei file di estrazione da IDOL;

---

<sup>1</sup> il tema dell'obsolescenza dello stack HW/SW è stato sovente sottolineato. Eventuali ulteriori attività di revisione straordinaria della *codebase* potrebbero suggerire una revisione completa della stessa.

- verifiche puntuali - le modifiche richieste e documentate nel presente documento necessitano di verifiche puntuali (UAT - User Acceptance Test). Le attività di verifica devono essere condotte congiuntamente con questo schema di responsabilità: le verifiche di tipo tecnico/funzionale sono a guida EHM, con Irpet informato, mentre le verifiche di dominio (sul contenuto scientifico degli aggregati prodotti) sono a guida Irpet con EHM informata.

## Versionamento del codice

Il codice prodotto è già sottoposto a *versioning* su di un *repository* Bitbucket (Atlassian). Alla conclusione dello sviluppo la ownership del *repository* verrà attribuita a personale Irpet con l'obiettivo di facilitare il *takeover* sulle procedure da parte di Irpet stessa. Da quel momento in poi la gestione del master sarà in capo a Irpet.

## Takeover sulle procedure e piano formativo

Come richiesto il progetto prevede di aumentare il livello di autonomia del team Irpet con particolare riferimento al rilascio della procedura oggetto del presente documento tecnico. Nello specifico vanno previste attività formative, erogate in FAD, da calendarizzare opportunamente con questi contenuti:

1. l'anatomia della procedura;
2. bitbucket, git e gli ambienti per la gestione del codice.

## La procedura

La codebase oggetto di manutenzione ha accumulato quasi 10 anni di vita. Per comprendere appieno i vari processi di trasformazione del dato (full SQL) è necessario un'esplorazione congiunta del codice assieme al trasferimento di alcune logiche utili a facilitare il processo di ingegneria inversa. In questo modo personale con competenze sul dominio dati delle comunicazioni obbligatorie (CO) potrà interpretare i costrutti SQL che materializzano le strutture analitiche nei vari schema DB.

## Bitbucket

Bitbucket (Atlassian - <https://bitbucket.org/>) è una piattaforma web basata su Git che consente di salvare il proprio codice in repository ad accesso privato (o pubblico se utile/necessario), facilitando lo sviluppo collaborativo del software. Git è un software (open source e completamente free) per il controllo di versione distribuito, utilizzabile da interfaccia a riga di comando e creato da Linus Torvalds nel 2005 per facilitare lo sviluppo del kernel Linux. Nel frattempo è diventato lo strumento di controllo di versione più diffuso.

Il personale Irpet verrà formato al fine di apprendere i principali comandi necessari per governare lo sviluppo e la manutenzione della procedura ovvero la gestione e il versioning del codice.

## Prerequisiti

I prerequisiti necessari per l'attività formativa sono una conoscenza ottima del linguaggio SQL, buona del linguaggio python e di base della shell/bash del S.O. linux Ubuntu.