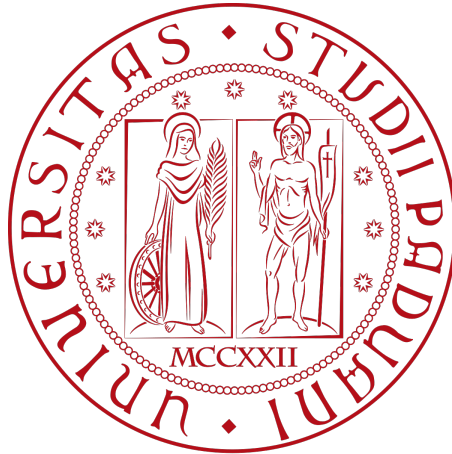


Università degli Studi di Padova

Dipartimento di Matematica «Tullio Levi-Civita»

Corso di Laurea in Informatica



**Archeologia Digitale e Rinascimento del Codice:
Modernizzazione dei Sistemi Legacy attraverso la
Migrazione Automatizzata COBOL - Java**

Tesi di Laurea

Relatore

Prof. Tullio Vardanega

Laureando

Annalisa Egidi

Matricola: 1216745

Anno Accademico 2024/2025

Ringraziamenti

Sommario

L'elaborato descrive i processi, gli strumenti e le metodologie coinvolte nello sviluppo di un sistema di migrazione automatizzata per la modernizzazione di sistemi *legacy*, in particolare sulla conversione di applicazioni COBOL verso Java.

Nel dominio applicativo di interesse dell'elaborato:

- **Migrazione automatizzata:** è il processo di conversione di sistemi informatici da tecnologie obsolete a moderne architetture, preservando la logica di *business* originale;
- **Legacy Systems:** sistemi informatici datati ma ancora operativi, spesso critici per le organizzazioni, difficili da mantenere e integrare con tecnologie moderne (ingl. *legacy systems*).

Il progetto, sviluppato nel corso del tirocinio presso l'azienda Miriade Srl (d'ora in avanti **Miriade**), ha la peculiarità di aver esplorato inizialmente un approccio tradizionale basato su *parsing* deterministico per poi scegliere una soluzione innovativa basata su intelligenza artificiale generativa, dimostrando come l'AI possa cambiare drasticamente i tempi e la qualità dei risultati nel campo della modernizzazione *software*.

Struttura del testo

Il corpo principale della relazione è suddiviso in 4 capitoli:

Il **primo capitolo** descrive il contesto aziendale in cui sono state svolte le attività di tirocinio curricolare, presentando Miriade come ecosistema di innovazione tecnologica e analizzando le metodologie e tecnologie all'avanguardia adottate dall'azienda;

Il **secondo capitolo** approfondisce il progetto di migrazione COBOL - Java, delineando il contesto di attualità dei sistemi *legacy*, gli obiettivi del progetto e le sfide tecniche identificate nella modernizzazione di applicazioni COBOL verso architetture Java moderne;

Il **terzo capitolo** descrive lo sviluppo del progetto seguendo un approccio cronologico, dal *parser* tradizionale iniziale al *pivot* verso l'intelligenza artificiale, documentando le metodologie di lavoro, i risultati raggiunti e l'impatto trasformativo dell'AI sui tempi di sviluppo;

Il **quarto capitolo** sviluppa una retrospettiva sul progetto, analizzando le *lessons learned*, il valore dell'AI come *game changer* nella modernizzazione *software*, la crescita professionale acquisita e le prospettive future di evoluzione della soluzione sviluppata.

Le **appendici** completano l'elaborato con:

- **Acronimi:** elenco alfabetico degli acronimi utilizzati nel testo con le relative espansioni;
- **Glossario:** definizioni dei termini tecnici e specialistici impiegati nella trattazione;
- **Sitografia:** fonti consultate e riferimenti sitografici utilizzati per la redazione dell'elaborato.

Indice

1	Miriade: un ecosistema di innovazione tecnologica	1
1.1	L'azienda nel panorama informatico e sociale	1
1.2	Metodologie e tecnologie all'avanguardia	1
1.3	Architettura organizzativa	3
1.4	Investimento nel capitale umano e nella ricerca	5
2	Il progetto di migrazione <i>COBOL-Java</i>	8
2.1	Contesto di attualità	8
2.2	Obiettivi dello stage	10
2.2.1	Obiettivi principali	10
2.2.2	Obiettivi operativi	10
2.2.3	Metriche di successo	11
2.3	Vincoli	12
2.4	Pianificazione concordata	12
2.5	Valore strategico per l'azienda	13
2.6	Aspettative personali	14
3	Sviluppo del progetto: dal parser tradizionale all'AI	16
3.1	Setup iniziale e metodologia di lavoro	16
3.2	Primo periodo: immersione nel mondo COBOL	16
3.2.1	Studio del linguaggio e creazione progetti test	16
3.2.2	Mappatura dei pattern e analisi di traducibilità	18
3.2.3	Valutazione delle soluzioni esistenti	19
3.3	Secondo periodo: sviluppo del parser tradizionale	21
3.3.1	Implementazione del parser Java	21
3.3.2	Analisi critica e limiti dell'approccio	24
3.4	Terzo periodo: pivot verso l'intelligenza artificiale	25
3.4.1	Valutazione delle API di AI generativa	25
3.4.2	<i>Design</i> del sistema <i>AI-powered</i>	26
3.5	Quarto periodo: implementazione della soluzione AI-driven	28
3.5.1	Sviluppo del prompt engineering	28
3.5.2	Implementazione del translator completo	29
3.5.3	Generazione automatica di progetti Maven	30
3.6	Risultati raggiunti	32

3.6.1	Impatto dell'AI sui tempi di sviluppo	32
3.6.2	Analisi qualitativa dei risultati	32
3.6.3	Risultati quantitativi	32
4	Valutazioni retrospettive e prospettive future	33
4.1	Analisi retrospettiva del percorso	33
4.2	L'AI come <i>game changer</i> nella modernizzazione <i>software</i>	33
4.3	Crescita professionale e competenze acquisite	33
4.4	Valore della formazione universitaria nell'era dell'AI	33
4.5	<i>Roadmap</i> evolutiva e opportunità di sviluppo	33
5	Lista degli acronimi	34
6	Glossario	35
7	Sitografia	36

Elenco delle figure

Figura 1	Ecosistema Atlassian - dashboard Jira	2
Figura 2	Ecosistema Atlassian - dashboard Confluence	3
Figura 3	Struttura organizzativa delle divisioni Miriade	4
Figura 4	Funzioni nella sezione Analytics	5
Figura 5	Impegni etici e morali aziendali di Miriade	6
Figura 6	Interfaccia utente e codice COBOL tipici dei sistemi legacy	8
Figura 7	Confronto tra architettura monolitica dei mainframe e architettura moderna a microservizi	9
Figura 8	Diagramma di Gantt della pianificazione del progetto	13
Figura 9	Rappresentazione della metodologia Agile applicata al progetto	15

1 Miriade: un ecosistema di innovazione tecnologica

Miriade, come realtà nel panorama Information Technology (IT) italiano, si distingue per il suo approccio innovativo rispetto all'ecosistema completo del dato e alle soluzioni informatiche correlate. L'azienda, che ho avuto l'opportunità di conoscere durante il mio percorso di *stage*, si caratterizza per una filosofia aziendale orientata all'innovazione continua e all'investimento nel capitale umano, elementi che la rendono un ambiente particolarmente stimolante per la crescita professionale di figure *junior*.

1.1 L'azienda nel panorama informatico e sociale

Miriade si posiziona strategicamente nel settore dell'analisi dati e delle soluzioni informatiche, operando con quattro aree funzionali principali: *Analytics*, *Data*, *System Application* e *Operation*. L'azienda ha costruito nel tempo una solida reputazione nel mercato attraverso la capacità di fornire soluzioni innovative che rispondono non solo alle esigenze tecniche dei clienti, ma che prestano particolare attenzione alle relazioni umane e alle realtà del territorio.

Ciò che distingue *Miriade* nel contesto competitivo è la sua *vision* aziendale, che integra le competenze tecnologiche con una forte responsabilità sociale. L'azienda implementa attivamente azioni a supporto di società e cooperative del territorio, dimostrando come l'innovazione tecnologica possa essere un veicolo di sviluppo sociale ed economico locale. Questa attenzione alla dimensione sociale si riflette anche nell'approccio alle risorse umane, con una particolare propensione a individuare e coltivare giovani energie fin dalle scuole e università attraverso tirocini curriculari che permettono una crescita personale durante il percorso di studi.

La clientela di Miriade spazia tra i medi e grandi clienti, includendo sia realtà del settore privato che pubblico. Questa diversificazione del *portfolio* clienti permette all'azienda di confrontarsi con problematiche tecnologiche variegate, mantenendo una costante spinta all'innovazione e all'adattamento delle soluzioni proposte.

1.2 Metodologie e tecnologie all'avanguardia

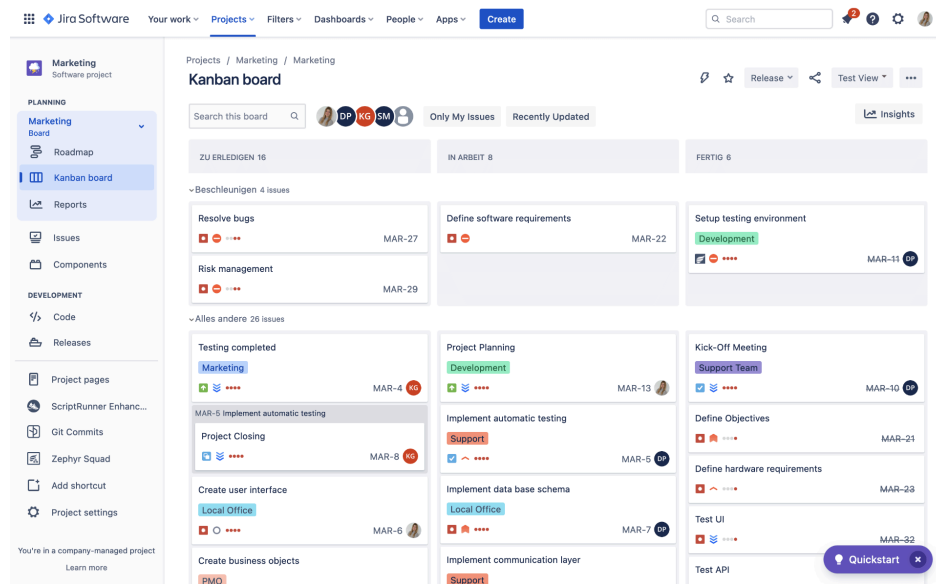
L'approccio metodologico di Miriade si fonda sull'adozione dell'*Agile* come filosofia operativa pervasiva, che permea tutti i processi aziendali e guida l'organizzazione del lavoro quotidiano. Durante il mio *stage*, ho potuto osservare direttamente come questa metodologia venga implementata attraverso *stand-up* giornalieri e *sprint* settimanali, creando un ambiente di lavoro dinamico e orientato agli obiettivi.

L'azienda utilizza sia *Kanban* che *Scrum*, adattando la metodologia alle specifiche esigenze progettuali e alle preferenze del cliente. Questa flessibilità metodologica dimostra la maturità organizzativa di Miriade e la sua capacità di adattare i processi alle diverse situazioni operative. Ho potuto constatare personalmente come gli *stand-up* mattutini fossero momenti fondamentali per l'allineamento del *team*, permettendo una comunicazione trasparente sullo stato di avanzamento delle attività e una rapida identificazione di eventuali impedimenti.

Lo *stack tecnologico* adottato riflette l'attenzione dell'azienda per gli strumenti di collaborazione e versionamento. L'*Atlassian Suite* costituisce la spina dorsale dell'infrastruttura collaborativa aziendale, utilizzata in modo strutturato e pervasivo per diverse finalità:

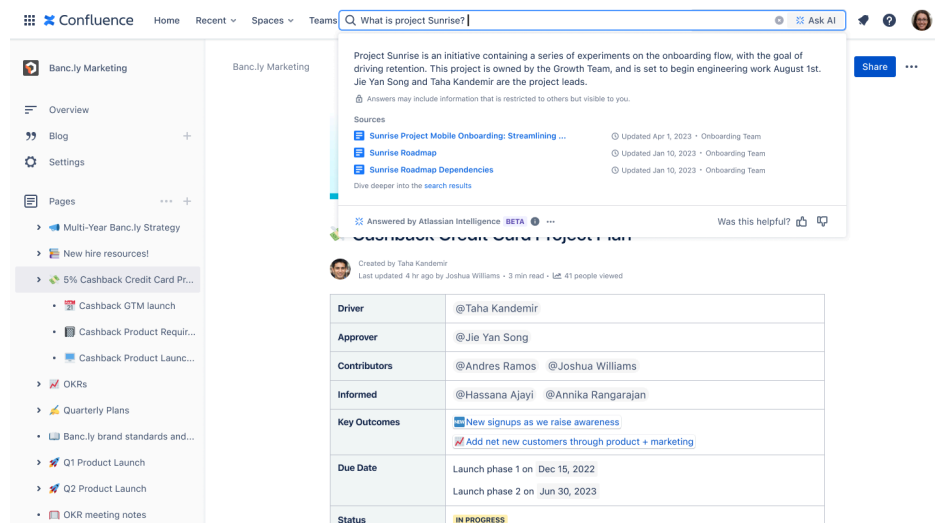
- **Confluence** per la gestione della *knowledge base* aziendale e la documentazione tecnica
- **Jira** per il *tracking* delle attività e la gestione dei progetti
- **Bitbucket** per il versionamento del codice e la collaborazione nello sviluppo

Durante il mio percorso, ho potuto apprezzare l'importanza che l'azienda attribuisce alla cultura del versionamento e della documentazione. Le Figura 1) e Figura 2 mostrano parte dell'ecosistema Atlassian integrato utilizzato quotidianamente in azienda, che ha rappresentato per me un elemento fondamentale nell'apprendimento delle pratiche professionali di sviluppo software.



Fonte: <https://www.peakforce.dev>

Figura 1: Ecosistema Atlassian - dashboard Jira



Fonte: <https://support.atlassian.com>

Figura 2: Ecosistema Atlassian - dashboard Confluence

La formazione continua sulle tecnologie emergenti è parte integrante della cultura aziendale. L'area *Analytics*, in particolare, mantiene un *focus* costante sull'esplorazione e implementazione di soluzioni basate su Artificial Intelligence (AI) e Large Language Models (LLM), che rappresentano il naturale proseguimento di quello che precedentemente veniva incasellato come «*big data*» ed è parte integrante della strategia aziendale.

1.3 Architettura organizzativa

L'architettura organizzativa di Miriade si distingue per la sua struttura «piatta». L'azienda ha adottato un modello organizzativo che prevede solo due livelli gerarchici: l'amministratore delegato e i responsabili di area. Questa scelta strutturale facilita la comunicazione diretta e riduce le barriere comunicative, creando un ambiente di lavoro agile e responsabilizzante.

Le quattro aree funzionali principali - *Analytics*, *Data*, *System Application* e *Operation* - operano con un alto grado di autonomia, pur mantenendo una forte interconnessione attraverso aree trasversali. Queste aree trasversali, composte da persone provenienti dalle diverse divisioni, si occupano di attività di innovazione a vari livelli, come *DevOps*, *Account Management* e *Research & Development*. Questa struttura matriciale permette una *cross-fertilizzazione* delle competenze e favorisce l'innovazione continua. La rappresentazione visuale in Figura 3 illustra chiaramente questa struttura organizzativa interconnessa.



Figura 3: Struttura organizzativa delle divisioni Miriade

La divisione *Analytics*, nella quale ho avuto il piacere di lavorare, guidata da Arianna Bellino, conta attualmente 17 persone ed è in veloce crescita. Rappresenta il motore di innovazione dell'azienda, specializzandosi nella gestione del dato, dal dato grezzo all'analisi avanzata, tramite approcci e tecnologie AI e LLM *based*, con *focus* sull'automazione dei processi e alla riduzione delle attività routinarie. I membri del *team* non hanno ruoli rigidamente definiti, ma piuttosto funzioni che possono evolversi in base alle esigenze progettuali e alle competenze individuali. Ho osservato dipendenti che svolgevano funzioni diverse quali:

- Pianificazione e gestione progetti
- Attività di prevendita e consulenza
- Ricerca e sviluppo di nuove soluzioni
- Sviluppo *software* e *data analysis*

Questa fluidità organizzativa crea un ambiente stimolante dove ogni persona può contribuire in modi diversi, favorendo la crescita professionale multidisciplinare. Durante lo stage, ho potuto interagire con colleghi che ricoprivano diverse funzioni, beneficiando della loro esperienza e prospettive diverse. La varietà di funzioni all'interno della divisione *Analytics* è rappresentata in Figura 4, che evidenzia la natura dinamica e multifunzionale del team.



Figura 4: Funzioni nella sezione Analytics

Il ruolo dello stagista in questo ecosistema aziendale è particolarmente valorizzato. Non viene visto come una risorsa marginale, ma come parte integrante del team, con la possibilità di contribuire attivamente ai progetti e di proporre soluzioni innovative. Il sistema di tutoraggio è strutturato con l'assegnazione di un *tutor* dell'area specifica e di un *mentor* che può provenire anche da altre aree. Il *tutor* segue il percorso tecnico dello stagista, mentre il *mentor* fornisce supporto a livello emotivo e di inserimento aziendale.

Particolarmente apprezzabili sono gli incontri settimanali chiamati «tiramisù», dedicati ai nuovi entrati in azienda. Durante questi momenti, vengono analizzate le possibili difficoltà relazionali o comunicative riscontrate durante la settimana, con il supporto di una figura dedicata. Questo approccio dimostra l'attenzione dell'azienda non solo alla crescita tecnica, ma anche al benessere e all'integrazione dei propri collaboratori.

1.4 Investimento nel capitale umano e nella ricerca

L'investimento nel capitale umano rappresenta uno dei pilastri fondamentali della strategia aziendale di Miriade. Durante il mio stage, ho potuto constatare come l'azienda non si limiti a dichiarare l'importanza delle risorse umane, ma implementi concretamente politiche e programmi volti alla valorizzazione e crescita delle persone, come ad esempio incontri, riflessioni e azioni sulla Parità di Genere, sulla quale sono certificati come azienda.



Fonte: <https://www.miriade.it>

Figura 5: Impegni etici e morali aziendali di Miriade

Come si può osservare in Figura 5, l'azienda si esprime esplicitamente riguardo i propri valori.

Il processo di selezione riflette questa filosofia: l'azienda ricerca persone sensibili, elastiche, proattive e autonome, ponendo l'enfasi sulle caratteristiche personali piuttosto che esclusivamente sulle competenze tecniche pregresse, un approccio che permette di costruire *team* coesi e motivati, capaci di affrontare sfide tecnologiche in continua evoluzione.

I programmi di formazione continua sono strutturati e costanti. L'azienda investe significativamente nella crescita professionale dei propri dipendenti attraverso:

- Corsi di formazione tecnica su nuove tecnologie
- Certificazioni professionali
- Partecipazione a conferenze e *workshop*
- Sessioni di *knowledge sharing* interno
- Progetti di ricerca e sviluppo che permettono sperimentazione

Il rapporto consolidato con le università rappresenta un altro aspetto distintivo dell'approccio di Miriade al capitale umano. Gli *stage* non sono visti come semplici adempimenti formativi, ma come veri e propri laboratori di sperimentazione tecnologica. Nel mio caso specifico, il progetto di migrazione *COBOL-Java* è stato scelto appositamente per valutare le capacità di *problem solving* e apprendimento, con maggiore attenzione al processo seguito piuttosto che al solo risultato finale.

L'equilibrio tra formazione e produttività negli *stage* è gestito con attenzione. Inizialmente, lo *stage* è orientato totalmente sulla formazione, per poi evolvere gradualmente verso un bilanciamento equilibrato tra formazione e contributo produttivo quando lo stagista diventa sufficientemente autonomo. Nel mio caso però, trattandosi di *stage* curricolare per tesi,

l'intero percorso è stato focalizzato sulla formazione, permettendomi di esplorare in profondità tecnologie e metodologie senza la pressione di *deadline* produttive immediate.

L'investimento in risorse *junior* è visivamente significativo, questo approccio permette all'azienda di formare professionisti allineati con la propria cultura e metodologie.

In conclusione, Miriade si presenta come un ecosistema aziendale dove l'innovazione tecnologica e la valorizzazione del capitale umano si integrano sinergicamente. L'esperienza di stage in questo contesto ha rappresentato un'opportunità unica di crescita professionale, permettendomi di osservare e partecipare a dinamiche aziendali mature e orientate al futuro. La combinazione di una struttura organizzativa agile, metodologie all'avanguardia, forte investimento nelle persone e attenzione alla responsabilità sociale crea un ambiente ideale per affrontare le sfide tecnologiche contemporanee.

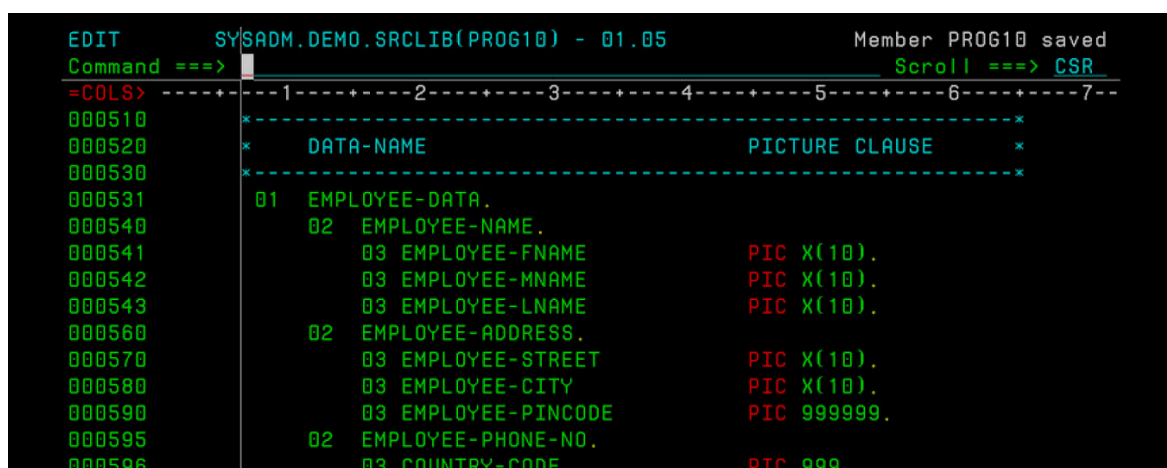
2 Il progetto di migrazione *COBOL-Java*

Il progetto di *stage* proposto da Miriade si inserisce in un contesto tecnologico di particolare rilevanza per il settore IT contemporaneo: la modernizzazione dei sistemi *legacy*. Durante il mio percorso, ho avuto l'opportunità di confrontarmi con una problematica comune a molte organizzazioni, in particolare nel settore bancario e assicurativo, dove i sistemi COBOL continuano a costituire l'impalcatura portante di infrastrutture critiche per il *business*.

2.1 Contesto di attualità

I sistemi legacy basati su Common Business-Oriented Language (COBOL) rappresentano ancora oggi una parte significativa dell'infrastruttura informatica di molte organizzazioni, specialmente nel settore bancario, finanziario e assicurativo. Nonostante COBOL sia stato sviluppato negli anni '60, ha una presenza significativa nelle moderne architetture.

Figura 6 mostra un esempio tipico di interfaccia utente e codice COBOL, che evidenzia il contrasto netto con le moderne interfacce grafiche e paradigmi di programmazione attuali. Questa differenza visuale è solo la punta dell'iceberg delle sfide che comporta il mantenimento di questi sistemi in un ecosistema tecnologico in rapida evoluzione.



```
EDIT      SYSDM.DEMO.SRCLIB(PROG10) - 01.05      Member PROG10 saved
Command ==>                                     Scroll ==> CSR
=COLS>  ---+---1---+---2---+---3---+---4---+---5---+---6---+---7---
000510  *-----*
000520  *   DATA-NAME                               PICTURE CLAUSE   *
000530  *-----*
000531  01  EMPLOYEE-DATA.
000540      02  EMPLOYEE-NAME.
000541          03  EMPLOYEE-FNAME                      PIC X(10).
000542          03  EMPLOYEE-MNAME                      PIC X(10).
000543          03  EMPLOYEE-LNAME                      PIC X(10).
000560      02  EMPLOYEE-ADDRESS.
000570          03  EMPLOYEE-STREET                    PIC X(10).
000580          03  EMPLOYEE-CITY                      PIC X(10).
000590          03  EMPLOYEE-PINCODE                    PIC 999999.
000595      02  EMPLOYEE-PHONE-NO.
000596          03  COUNTRY-CODE                      PIC 999.
```

Fonte: <https://overcast.blog>

Figura 6: Interfaccia utente e codice COBOL tipici dei sistemi legacy

La problematica della *legacy modernization* va ben oltre la semplice obsolescenza tecnologica. Durante il mio *stage*, attraverso l'analisi della letteratura e il confronto con i professionisti del settore, in particolare ho avuto modo di confrontarmi con la sig.ra Luisa Biagi, analista COBOL, ho potuto identificare come i costi nascosti del mantenimento di questi sistemi includano:

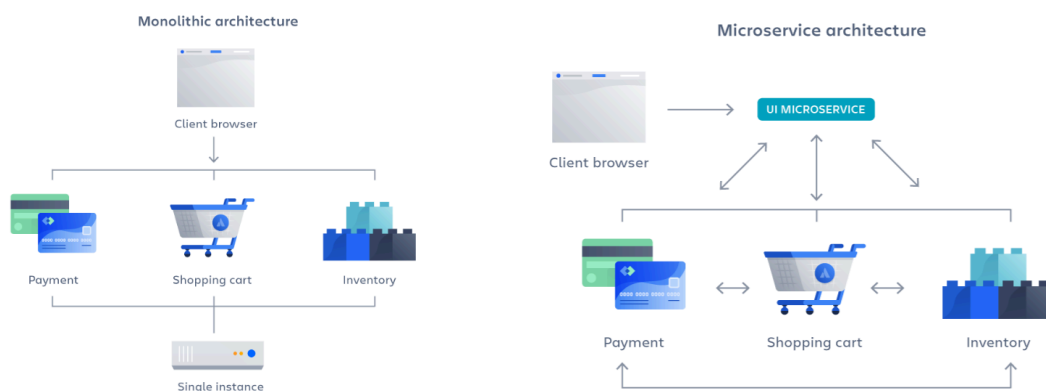
- La crescente difficoltà nel reperire sviluppatori COBOL qualificati [1]
- L'integrazione sempre più complessa con tecnologie moderne [2]
- I rischi operativi derivanti dall'utilizzo di piattaforme *hardware* e *software* che i *vendor* non supportano più attivamente [3]

Questi fattori si traducono in costi di manutenzione esponenzialmente crescenti e in una ridotta agilità nel rispondere alle esigenze di *business* in continua evoluzione.

I rischi associati al mantenimento di sistemi COBOL *legacy* nelle infrastrutture IT moderne sono molteplici e interconnessi:

- **Carenza di competenze:** La carenza di competenze specializzate crea una forte dipendenza da un *pool* sempre più ristretto di esperti, spesso prossimi al pensionamento [1].
- **Documentazione inadeguata:** La documentazione inadeguata o assente di molti di questi sistemi, sviluppati decenni fa, rende ogni intervento di manutenzione un'operazione ad alto rischio [4].
- **Incompatibilità tecnologica:** L'incompatibilità con le moderne pratiche di sviluppo come *DevOps*, *continuous integration* e *microservizi* limita in modo significativo la capacità delle organizzazioni di innovare e competere efficacemente nel mercato digitale [5].

Come illustrato in Figura 7, il contrasto tra l'architettura monolitica tipica dei sistemi *mainframe* e l'architettura moderna a microservizi evidenzia le sfide architetturali della migrazione. Questa differenza strutturale comporta non solo una riprogettazione tecnica, ma anche un ripensamento completo dei processi operativi e delle modalità di sviluppo.



Fonte: <https://www.atlassian.com>

Figura 7: Confronto tra architettura monolitica dei mainframe e architettura moderna a microservizi

La migrazione di questi sistemi verso tecnologie più moderne come *Java* rappresenta una sfida tecnica e una necessità strategica per garantire la continuità operativa e la competitività delle organizzazioni. *Java*, con il suo ecosistema maturo, la vasta *community* di sviluppatori e il supporto per paradigmi di programmazione moderni, si presenta come una delle destinazioni privilegiate per questi progetti di modernizzazione.

2.2 Obiettivi dello stage

Il macro-obiettivo era sviluppare un sistema prototipale di migrazione automatica da COBOL a *Java* che potesse dimostrare la fattibilità di automatizzare il processo di conversione, preservando la *business logic* originale e producendo codice *Java* idiomatrico e manutenibile.

2.2.1 Obiettivi principali

- **Esplorazione tecnologica:** Investigare e valutare diverse strategie di migrazione, dalla conversione sintattica diretta basata su regole deterministiche fino all'utilizzo di tecnologie di intelligenza artificiale generativa, identificando vantaggi e limitazioni di ciascun approccio.
- **Automazione del processo:** Sviluppare strumenti e metodologie che potessero automatizzare il più possibile il processo di conversione, riducendo l'intervento manuale e i conseguenti rischi di errore umano nella traduzione.
- **Qualità del risultato:** Garantire che il codice Java prodotto rispettasse standard di qualità professionale, con particolare attenzione alla leggibilità, manutenibilità e conformità alle convenzioni Java moderne.
- **Accessibilità della soluzione:** Fornire un'interfaccia utente (grafica o da linea di comando) che rendesse il sistema utilizzabile anche da personale non specializzato nella migrazione di codice.

2.2.2 Obiettivi operativi

Per rendere concreti e misurabili gli obiettivi principali, sono stati definiti obiettivi operativi specifici, classificati secondo tre livelli di priorità:

Obbligatori

- **OO01:** Sviluppare competenza nel linguaggio COBOL attraverso la produzione di almeno un progetto completo che includesse le quattro divisioni fondamentali (*Identification, Environment, Data e Procedure*)
- **OO02:** Esplorazione approfondita di diverse strategie di migrazione, dalla conversione sintattica diretta all'utilizzo di tecnologie di intelligenza artificiale generativa

- **OO03:** Implementare un sistema di conversione automatica che raggiungesse almeno il 75% di copertura delle divisioni
- **OO04:** Esplorare e documentare approcci distinti alla migrazione
- **OO05:** Completare la migrazione funzionante di almeno uno dei progetti COBOL sviluppati, validando l'equivalenza funzionale tra codice sorgente e risultato
- **OO06:** Produrre codice *Java* che rispettasse le convenzioni del linguaggio, includendo struttura dei *package*, nomenclatura standard e documentazione *JavaDoc*
- **OO07:** Fornire un'interfaccia utilizzabile (grafica o Command Line Interface (CLI)) per l'esecuzione del sistema di conversione
- **OO08:** Creare documentazione utente completa, includendo un *README* dettagliato con istruzioni di installazione, configurazione e utilizzo

Desiderabili

- **OD01:** Raggiungimento di una copertura del 100% nella conversione automatica del codice prodotto autonomamente
- **OD02:** Gestione efficace di costrutti COBOL complessi o non direttamente traducibili
- **OD03:** Implementazione di meccanismi di ottimizzazione del codice Java generato

Facoltativi

- **OF01:** Integrazione con sistemi di analisi statica per la verifica della qualità del codice generato
- **OF02:** Implementare un sistema di *reporting* dettagliato che producesse metriche sulla conversione, incluse statistiche di copertura, costrutti non convertiti e interventi manuali necessari
- **OF03:** Implementazione di funzionalità avanzate di *refactoring* del codice Java prodotto

2.2.3 Metriche di successo

Per valutare oggettivamente il raggiungimento degli obiettivi, erano state definite le seguenti metriche:

- **Copertura di conversione:** Percentuale di linee di codice COBOL convertite automaticamente senza intervento manuale
- **Equivalenza funzionale:** Corrispondenza *interfaccia utente* COBOL originale e *Java* convertito
- **Qualità del codice:** Conformità agli standard *Java* verificata tramite strumenti di analisi statica
- **Tempo di conversione:** Riduzione del tempo necessario per la migrazione rispetto a un approccio completamente manuale

- **Usabilità:** Capacità di utilizzo del sistema da parte di utenti con conoscenze base di programmazione

2.3 Vincoli

Il progetto si focalizzava sullo sviluppo di un sistema di migrazione automatica e questo aspetto caratterizzava le condizioni imposte per lo svolgimento del lavoro.

Vincoli temporali

- Durata complessiva dello *stage*: 320 ore
- Periodo: dal 05 maggio al 27 giugno 2025
- Modalità di lavoro ibrida: 2 giorni a settimana in sede, 3 giorni in modalità telematica
- Orario lavorativo: 9:00 - 18:00

Vincoli tecnologici

- Il sistema doveva essere sviluppato utilizzando tecnologie moderne e supportate
- Necessità di preservare integralmente la *business logic* contenuta nei programmi COBOL originali
- La soluzione doveva essere scalabile, capace di gestire progetti di diverse dimensioni
- Utilizzo strumenti di versionamento (*Git*) e di documentazione continua.

Vincoli metodologici

- Adozione dei principi *Agile* con *sprint* settimanali e *stand-up* giornalieri per allineamento costante
- Revisioni settimanali degli obiettivi con adattamento del piano di lavoro

2.4 Pianificazione concordata

La pianificazione del progetto seguiva un approccio flessibile, con revisioni settimanali che permettevano di adattare il percorso in base ai progressi ottenuti. La distribuzione delle attività era inizialmente stata organizzata come segue:

Prima fase - analisi e apprendimento COBOL (2 settimane - 80 ore)

- Studio approfondito del linguaggio COBOL e delle sue peculiarità
- Analisi di sistemi COBOL
- Creazione di programmi COBOL di test con complessità crescente
- Implementazione dell'interfacciamento con *database* relazionali

Seconda fase - sviluppo del sistema di migrazione (4 settimane - 160 ore)

- Analisi dei *pattern* di traduzione COBOL-Java del codice prodotto in fase precedente
- Sviluppo di uno *script* o utilizzo di *tool* esistenti per automatizzare la traduzione del codice COBOL in Java equivalente

- Gestione della traduzione dei costrutti sintattici, logica di controllo e interazioni con il *database*
- Definizione della percentuale di automazione raggiungibile e la gestione di costrutti COBOL complessi o non direttamente traducibili

Terza fase - *testing* e validazione (1 settimana - 40 ore)

- *Test* funzionali sul codice Java generato
- Confronto comportamentale con le applicazioni COBOL originali

Quarta fase - documentazione e consegna (1 settimana - 40 ore)

- Documentazione completa del sistema sviluppato
- Preparazione del materiale di consegna
- Presentazione finale dei risultati

La rappresentazione temporale dettagliata della pianificazione è visualizzata in Figura 8, che mostra la distribuzione delle attività lungo l'arco temporale dello stage.



Figura 8: Diagramma di Gantt della pianificazione del progetto

2.5 Valore strategico per l'azienda

In base a quanto ho potuto osservare e comprendere durante il periodo di *stage*, la strategia di gestione del progetto di migrazione *COBOL-Java* dell'azienda ospitante persegue i seguenti obiettivi:

- **Innovazione tecnologica:** l'interesse dell'azienda non era limitato allo sviluppo di una soluzione tecnica specifica, ma si estendeva all'osservazione dell'approccio metodologico e del metodo di studio che una risorsa *junior* con formazione universitaria avrebbe applicato a un problema complesso di modernizzazione *IT*.

- **Creazione di competenze interne:** Il progetto permetteva di sviluppare *know-how* interno su una problematica di crescente rilevanza, preparando l'azienda a potenziali progetti futuri.
- **Esplorazione di tecnologie emergenti:** Il progetto era stato concepito per esplorare la possibile applicazione dell'intelligenza artificiale generativa a problemi di modernizzazione del *software*. Questo ambito, all'intersezione tra *AI* e *software engineering*, può rappresentare una frontiera tecnologica di forte attualità e di interesse per un'azienda che opera già attivamente nel campo dell'*AI* e dei *Large Language Models*.
- **Sviluppo di *asset* riutilizzabili:** Sebbene il progetto fosse autoconclusivo, permetteva di ottenere risultati tangibili nel breve termine dello *stage*, ma con il potenziale di evolversi in soluzioni più ampie e commercializzabili.

2.6 Aspettative personali

La scelta di intraprendere questo *stage* presso Miriade è stata guidata da una combinazione di motivazioni tecniche e personali che si allineavano con il mio percorso formativo universitario. Tra le diverse opportunità di *stage* che avevo valutato, questo progetto si distingueva per due elementi fondamentali:

- **Libertà tecnologica:** La libertà concessami nell'esplorazione delle tecnologie da utilizzare rappresentava un'opportunità unica di sperimentazione e apprendimento.
- **Interesse per COBOL:** Il mio forte interesse nel scoprire di più sul linguaggio COBOL, un affascinante paradosso tecnologico che, nonostante la sua longeva età, continua a essere cruciale nello scenario bancario e assicurativo internazionale.

Il mio percorso di *stage* mirava principalmente all'acquisizione di competenze pratiche nel campo della modernizzazione di sistemi *legacy* e gestione progetti:

Obiettivi tecnici

- Comprendere la struttura e la logica dei programmi COBOL attraverso lo sviluppo di applicazioni di test
- Esplorare approcci concreti alla migrazione del codice, sia deterministici che basati su *AI*
- Produrre un prototipo funzionante di sistema di conversione, anche se limitato

Competenze da sviluppare

- Familiarità di base con il linguaggio COBOL e le sue peculiarità sintattiche
- Comprensione pratica delle sfide nella traduzione tra paradigmi di programmazione diversi
- Esperienza nell'utilizzo di tecnologie emergenti come l'*AI* generativa applicata al codice

Crescita professionale attesa

- Sviluppare autonomia nella gestione di un progetto aziendale, dalla pianificazione all'implementazione
 - Acquisire capacità di *problem solving* in contesti reali, con vincoli temporali e tecnologici definiti
 - Migliorare le competenze comunicative attraverso l'interazione con il *team* e la presentazione dei progressi
 - Apprendere metodologie di lavoro *Agile* applicate a progetti di ricerca e sviluppo.
- Figura 9 rappresenta visivamente l'approccio metodologico Agile che ho appreso e applicato durante lo stage, evidenziando il ciclo iterativo di pianificazione, sviluppo, testing e revisione che ha caratterizzato il mio percorso formativo.
- Sviluppare pensiero critico nella valutazione di soluzioni tecnologiche alternative



Fonte: <https://indevlab.com>

Figura 9: Rappresentazione della metodologia Agile applicata al progetto

3 Sviluppo del progetto: dal parser tradizionale all'AI

Il percorso di sviluppo del progetto di migrazione COBOL-Java ha attraversato diverse fasi evolutive, caratterizzate da sfide tecnologiche e cambiamenti strategici che hanno profondamente trasformato l'approccio iniziale. Questo capitolo analizza cronologicamente le fasi del progetto, dall'immersione iniziale nel linguaggio COBOL fino all'implementazione di una soluzione basata sull'intelligenza artificiale generativa, evidenziando come la flessibilità metodologica e l'apertura all'innovazione abbiano costituito fattori determinanti per il successo del progetto.

3.1 Setup iniziale e metodologia di lavoro

In questa sezione descriverò l'implementazione della metodologia *Agile* con *sprint* settimanali e *stand-up* giornalieri, illustrerò gli strumenti di sviluppo e l'ambiente tecnologico utilizzato, analizzerò la gestione del progetto attraverso Jira e Confluence, l'uso di Git e BitBucket per il versionamento con documentazione progressiva.

3.2 Primo periodo: immersione nel mondo COBOL

Le prime due settimane del progetto sono state dedicate ad un'immersione completa nel linguaggio COBOL, dai paradigmi di programmazione moderni a cui ero abituata - con la loro enfasi su astrazione, modularità e riusabilità - a un approccio procedurale strutturato degli anni "60.

Questa fase di apprendimento intensivo si è rivelata fondamentale non solo per acquisire competenze tecniche, ma anche per comprendere la filosofia e il contesto storico che hanno plasmato COBOL e, di conseguenza, i sistemi legacy che ancora oggi sostengono infrastrutture critiche.

3.2.1 Studio del linguaggio e creazione progetti test

L'approccio allo studio del linguaggio COBOL ha combinato risorse storiche e moderne. I manuali tecnici IBM degli anni "80, sorprendentemente ancora attuali, sono stati integrati con tutorial online che tentavano di rendere COBOL accessibile ai programmatori moderni. Particolarmente preziosa si è rivelata la collaborazione con la programmatrice Luisa Biagi, analista COBOL con oltre trent'anni di esperienza, che ha fornito non solo conoscenze tecniche ma anche contesto pratico sull'utilizzo effettivo di COBOL in ambienti di produzione.

La struttura rigida del codice COBOL, articolata nelle quattro divisioni obbligatorie (Identification, Environment, Data e Procedure), rappresenta un paradigma radicalmente diverso rispetto ai linguaggi orientati agli oggetti. Tale struttura, lungi dall'essere un limite, riflette una filosofia progettuale volta a dare ordine e standardizzazione per una manutenibilità in progetti di grandi dimensioni sviluppati da team numerosi.

L'analisi delle singole divisioni ha rivelato aspetti significativi:

- La IDENTIFICATION DIVISION testimonia l'enfasi sulla documentazione incorporata nel codice, con campi dedicati per autore, data di installazione e osservazioni, riflettendo un'epoca in cui il codice sorgente costituiva spesso l'unica documentazione disponibile.
- La ENVIRONMENT DIVISION introduce esplicitamente considerazioni hardware e di sistema operativo nel codice sorgente. La necessità di specificare SOURCE-COMPUTER e OBJECT-COMPUTER evidenzia le sfide dell'era dei mainframe, dove la portabilità del software non poteva essere data per scontata.
- La sezione FILE-CONTROL, con la sua gestione esplicita dell'associazione tra file logici e fisici, richiede un cambio di mentalità significativo rispetto all'astrazione automatica fornita dai moderni sistemi operativi e strutture software.

Ho sviluppato, progressivamente, tre codici applicativi di complessità crescente:

- Il primo progetto, un sistema di gestione conti correnti bancari, implementa le operazioni fondamentali del banking: apertura conti, depositi, prelievi ed estratti conto. Questo sistema sfrutta la forza di COBOL nell'aritmetica decimale precisa per gestire saldi, fidi e transazioni finanziarie, interfacciandosi con un database PostgreSQL tramite SQL embedded per garantire l'integrità transazionale e la persistenza dei dati.
- Il secondo progetto, un sistema di gestione paghe e stipendi, aumenta significativamente la complessità introducendo calcoli multi-livello per IRPEF con scaglioni progressivi, trattenute previdenziali, detrazioni e addizionali. Il sistema gestisce presenze, straordinari e genera cedolini dettagliati, richiedendo la coordinazione tra molteplici tabelle correlate e l'implementazione di logiche di business complesse per il calcolo delle retribuzioni secondo la normativa fiscale italiana.
- Il terzo progetto, un sistema di gestione magazzino e inventario, rappresenta il culmine della complessità con funzionalità avanzate come la valorizzazione FIFO/LIFO/costo medio ponderato, gestione lotti, analisi ABC degli articoli, controllo scorte con punti di riordino automatici e gestione completa del ciclo ordini fornitori. Il sistema utilizza cursori SQL multipli, transazioni annidate e genera report sofisticati per l'inventario

fisico e l'analisi del valore di magazzino, dimostrando la capacità di COBOL di gestire processi aziendali articolati con elevati volumi di dati.

L'interfacciamento con database relazionali ha rappresentato una sfida particolare. Lavorando con PostgreSQL e DB2, ho approfondito le peculiarità dell'SQL embedded in COBOL. L'approccio differisce radicalmente dalle moderne Application Programming Interface (API) Java Database Connectivity (JDBC): il preprocessore COBOL-SQL analizza il codice sorgente, estrae le istruzioni SQL delimitate da EXEC SQL ... END-EXEC, e genera il codice COBOL appropriato per l'interazione con il database. La gestione delle variabili host e controllo esplicito degli errori attraverso SQLCODE e SQLSTATE implementa la richiesta del controllo esplicito del codice di ritorno dopo ogni operazione SQL. Inoltre, per ogni colonna del database è richiesta la corrispettiva variabile locale che funziona da ponte tra il programma e il database, queste variabili sono gestite attraverso la dichiarazione delle stesse nella WORKING-STORAGE SECTION e devono essere necessariamente compatibili per tipo e dimensione con la colonna del database corrispondente.

3.2.2 Mappatura dei pattern e analisi di traducibilità

Durante l'analisi dei pattern, ho classificato tre categorie principali:

- pattern con equivalenza diretta in Java
- pattern che richiedono trasformazioni con adattamento
- costrutti problematici senza equivalenti in java

3.2.2.1 Pattern di equivalenza diretta e costrutti base

- I tipi di dati primitivi seguono corrispondenze chiare
 - PIC 9(n) → int/long
 - PIC X(n) → String
 - PIC 9(n)V9(n) → BigDecimal
- Le strutture di controllo (IF-THEN-ELSE, EVALUATE) hanno equivalenti diretti in Java, rispettivamente if-else e switch
- Le operazioni aritmetiche base (ADD, SUBTRACT, MULTIPLY, DIVIDE) si mappano direttamente agli operatori java, con attenzione alla gestione della precisione
- Le dichiarazioni di variabili seguono pattern consolidati, con record gerarchici che diventano classi Java annidate
 - 01 level declaration diventano classi java
 - i record strutturati gerarchici del COBOL, con i loro level numbers, diventano classi Java annidate

- I PERFORM statements si convertono in metodi Java seguendo pattern specifici:
 - PERFORM UNTIL diventa while loop
 - PERFORM VARYING si trasforma in for loop
 - PERFORM THRU richiede metodi che incapsulano l'intera sequenza di paragrafi

3.2.2.2 Trasformazioni complesse e pattern di adattamento

Costrutti di complessità intermedia necessitano di strategie di conversione più sofisticate:

- GOTO, non supportato in java, richiede la ristrutturazione del flusso di controllo
- i COPY statements, che permettono l'inclusione di codice comune, si trasformano in classi java dedicate con struttura dati appropriata.
- la WORKING-STORAGE SECTION si converte in classi di storage con pattern specifici
- la SECTION e PARAGRAPH della PROCEDURE DIVISION diventano metodi Java, questo approccio mantiene la leggibilità del codice e facilita il debugging durante la fase di transizione
- l'SQL embedded (EXEC SQL) richiede la conversione in JDBC, con particolare attenzione alla gestione dei cursori e delle transazioni

3.2.2.3 Costrutti problematici e soluzioni architetturali

I costrutti senza equivalenti diretti rappresentano una parte critica della migrazione.

- L'ALTER statement, che modifica il comportamento dei GOTO a runtime, non ha alcun equivalente in Java e richiede una completa ristrutturazione utilizzando pattern Strategy
- Costrutti specifici del mainframe (UNSTRING, EXAMINE, NEXT SENTENCE) necessitano di implementazioni custom. Per esempio, UNSTRING si può convertire utilizzando String.split() con logica aggiuntiva per gestire i contatori e le condizioni di overflow

3.2.3 Valutazione delle soluzioni esistenti

Dopo la mappatura dei pattern comuni e l'analisi di pattern Java corrispondenti ho condotto un'analisi approfondita e sistematica delle soluzioni esistenti sul mercato, sia open-source che enterprise. Questa fase di ricerca e valutazione si è rivelata fondamentale non solo per comprendere lo stato dell'arte attuale, ma anche per identificare opportunità di innovazione e differenziazione.

Ho valutato ciascuna soluzione secondo criteri multipli:

- completezza della copertura COBOL
- qualità del codice generato
- facilità d'uso

- costo totale di proprietà
- supporto e manutenzione
- estensibilità
- personalizzazione

3.2.3.1 Soluzioni open-source

L'analisi del ProLeap COBOL parser, uno dei progetti open-source più maturi nello spazio di GitHub, ha rivelato un'architettura solida basata su ANother Tool for Language Recognition (ANTLR)⁴ con capacità complete di analisi sintattica. Tuttavia, il sistema si limitava alla generazione dell'Abstract Syntax Tree (AST), richiedendo l'implementazione separata della trasformazione AST COBOL → AST Java e della successiva generazione del codice. Ad ogni modo il ProLeap parser presentava una architettura modulare che permetteva l'estensione per nuovi costrutti.

3.2.3.2 Soluzioni enterprise

Il panorama presentava soluzioni commerciali sofisticate con prezzi corrispondentemente elevati. Essendo soluzioni chiuse al pubblico ho avuto modo di testare tali soluzioni solo in modo limitato, quando prove gratuite o soluzioni demo lo permettevano.

In particolare, grazie alle soluzioni di prova per sviluppatori, ho potuto mettere mano a IBM WatsonX Code Assistant for Z che rappresenta attualmente lo stato dell'arte nell'applicazione dell'intelligenza artificiale alla modernizzazione legacy. La soluzione IBM non si limita alla traduzione sintattica ma tenta di comprendere il contesto aziendale del codice tramite l'interazione e interconnessione con l'intelligenza artificiale. Durante la demo, la quale si presentava come chatbot integrato nell'IDE di riferimento (nel mio caso Visual Studio Code), ho osservato come il sistema potesse riconoscere schemi di dominio, desse suggerimenti di modernizzazione architetturale (conversione a microservizi dove appropriato), generazione di test automatici basati sulla comprensione del comportamento atteso, e documentazione automatica che catturava l'intento aziendale tramite domande di contesto oltre che puramente tecniche.

La soluzione presentava barriere significative per un progetto di ricerca:

- costo proibitivo per licenze, anche di sviluppo
- natura proprietaria che impediva personalizzazione profonde
- requisiti di infrastruttura aziendale
- vincolo con l'ecosistema IBM

Altre soluzioni enterprise che ho preso in analisi erano Micro Focus, Modern Systems e TSRI che risultavano però meno di valore per qualità-costo dei contenuti producibili, pertanto ho approfondito meno queste tecnologie.

Alla luce di tutte le ricerche, sia open-source che enterprise, ho potuto constatare un vuoto significativo sul tema nel mercato. Le soluzioni open-source, pur essendo accessibili, mancavano di sofisticazione e producevano risultati incompleti che richiedevano estese implementazioni e modifiche manuali. Le soluzioni aziendali, pur essendo potenti, erano inaccessibili per la maggior parte delle organizzazioni a causa dei costi¹.

Questo vuoto mi suggeriva un'opportunità per una soluzione che combinasse l'accessibilità dell'open-source con capacità più sofisticate di comprensione e trasformazione. La recente democratizzazione dell'AI generativa attraverso API accessibili apriva possibilità precedentemente riservate solo a fornitori e clienti con risorse massive.

Inoltre, l'analisi delle soluzioni esistenti mi ha fatto notare schemi comuni:

- le migrazioni che preservavano la logica di business erano le più richieste
- l'importanza della documentazione e della tracciabilità nel processo di migrazione
- la necessità di un approccio interattivo col personale umano per una validazione iterativa del codice prodotto
- il valore di preservare la conoscenza di dominio incorporata nel codice legacy

Queste riflessioni hanno formato significativamente gli approcci che avrei adottato nelle fasi successive del progetto, suggerendo che una soluzione efficace avrebbe dovuto combinare «automatizzazione intelligente» con comprensione semantica profonda, preservazione della logica aziendale con modernizzazione dell'implementazione e accessibilità.

3.3 Secondo periodo: sviluppo del parser tradizionale

Forte delle conoscenze acquisite sul linguaggio COBOL e dell'analisi delle soluzioni esistenti, ho intrapreso lo sviluppo dell'analizzatore sintattico basato su un approccio deterministico ibrido tradizionale che conciliasse l'utilizzo di tecnologie open-source, ProLeap, e l'implementazione autonomo di un sistema di traduzione dell'AST COBOL in un AST Java corrispondente e successivo passaggio a codice sorgente.

3.3.1 Implementazione del parser Java

L'obiettivo era implementare un sistema modulare ed estensibile che potesse crescere incrementalmente man mano che nuovi costrutti COBOL venivano supportati.

¹ I costi delle soluzioni enterprise variano da €100.000 a €1.000.000 per licenza annuale, a seconda della dimensione del deployment

L'architettura seguiva il classico modello di compilatore a pipeline:

1. Il **Lexer di ProLeap** gestiva la tokenizzazione del codice COBOL, affrontando le peculiarità del linguaggio come la sensibilità alla colonna (le colonne 1-6 riservate per numeri di linea, colonna 7 per indicatori speciali, colonne 8-72 per il codice effettivo), la gestione dei commenti e delle linee di continuazione
2. Il **Parser di ProLeap** costruiva l'albero sintattico astratto (AST) basandosi sulla grammatica COBOL definita in ANTLR4
3. Il **Traduttore custom** (da implementare) trasformava l'AST COBOL nell'AST Java corrispondente, gestendo le differenze semantiche tra i due linguaggi
4. Il **Generatore di codice** (da implementare) convertiva l'AST Java annotato in codice sorgente Java idiomatico

L'implementazione del traduttore custom e generatore di codice è proceduta in parallelo, per divisioni COBOL:

- **IDENTIFICATION DIVISION**, punto di partenza naturale per la sua semplicità strutturale e prevedibilità. Questa divisione, contenendo principalmente metadati senza impatto diretto sulla logica del programma, forniva contesto essenziale per la comprensione del sistema. Ho sviluppato un analizzatore basato su pattern matching che estraeva sistematicamente le informazioni standard: PROGRAM-ID, AUTHOR, INSTALLATION, DATE-WRITTEN e REMARKS. La strategia di conversione per questa divisione prevedeva una trasformazione semanticamente ricca dei metadati:
 - Il PROGRAM-ID veniva convertito nel nome della classe Java principale, applicando le convenzioni di denominazione Java (trasformazione da KEBAB-CASE o UNDERSCORE_CASE a CamelCase)
 - Le informazioni di contesto (AUTHOR, INSTALLATION, DATE-WRITTEN) venivano preservate in un blocco JavaDoc strutturato in testa alla classe, mantenendo la completa tracciabilità con il programma originale e rispettando gli standard di documentazione Java
- La **ENVIRONMENT DIVISION** ha presentato le prime sfide sostanziali. Questa divisione, che specifica l'ambiente di esecuzione del programma COBOL includendo informazioni su hardware, sistema operativo e mappatura dei file, riflette un'epoca in cui tali dettagli erano critici per l'esecuzione. Nel contesto moderno, molte di queste informazioni risultano obsolete o vengono gestite attraverso meccanismi completamente diversi. Ho sviluppato una strategia di conversione che preservava le informazioni semanticamente rilevanti mentre scartava quelle puramente storiche:

- La CONFIGURATION SECTION, contenente SOURCE-COMPUTER e OBJECT-COMPUTER, veniva convertita in commenti documentativi strutturati, preservando l'informazione per riferimento storico senza impatto sul codice generato
- Nella INPUT-OUTPUT SECTION, particolarmente la FILE-CONTROL, ogni dichiarazione di file in COBOL include non solo il nome logico del file ma anche dettagli critici sulla sua organizzazione (sequenziale, indicizzata, relativa), modalità di accesso (sequenziale, random, dinamica), e strategie di gestione degli errori.

Ho implementato un sistema di mapping che traduceva le dichiarazioni COBOL in configurazioni Java moderne. Ad esempio, una dichiarazione COBOL come:

```
SELECT CUSTOMER-FILE ASSIGN TO "CUSTMAST.DAT"
      ORGANIZATION IS INDEXED
      ACCESS MODE IS RANDOM
      RECORD KEY IS CUSTOMER-ID
```

veniva trasformata in una configurazione che utilizzava classi di accesso ai file Java con appropriati livelli di astrazione, preservando la semantica COBOL (accesso indicizzato, chiave di record) mentre si integrava con le API Java moderne per I/O.

- La **DATA DIVISION** ha rappresentato la sfida tecnica più significativa di questa fase. La complessità derivava dal sistema gerarchico di definizione dei dati in COBOL, che utilizza numeri di livello (01-49, 66, 77, 88) per definire strutture dati annidate con semantiche specifiche. Ho implementato un analizzatore ricorsivo che costruiva una rappresentazione interna completa della gerarchia dei dati, gestendo:
 - I livelli 01 che definiscono record di primo livello
 - I livelli 02-49 che creano strutture gerarchiche
 - Il livello 66 per la ridefinizione di gruppi di campi
 - Il livello 77 per variabili indipendenti
 - Il livello 88 per valori condizionali (condition names)

La conversione in strutture Java appropriate si è rivelata particolarmente complessa. Un approccio naïve di mappare ogni elemento di gruppo su una classe Java e ogni elemento elementare su un campo produceva codice eccessivamente verboso e non idiomático. Ho quindi sperimentato due strategie complementari:

- **Appiattimento selettivo:** per gerarchie semplici, i campi venivano appiattiti in una singola classe con nomi composti (es. CUSTOMER-NAME-FIRST diventava customerNameFirst)

- **Preservazione gerarchica:** per strutture complesse o quando la gerarchia aveva significato semantico, utilizzavo classi Java annidate che riflettevano la struttura originale

La gestione delle **PICTURE clauses** ha richiesto un'attenzione particolare. Queste clausole non definiscono solo tipo e dimensione dei dati, ma anche formattazione, gestione del segno, allineamento decimale e altre caratteristiche di presentazione. Ho sviluppato un mini-parser specializzato per le PICTURE clauses che le decompone in attributi gestibili:

- PIC 9(5) → int con validazione del range
- PIC X(30) → String con lunghezza massima
- PIC 9(7)V99 → BigDecimal con precisione specificata
- PIC S9(5) COMP-3 → gestione speciale per packed decimal

3.3.2 Analisi critica e limiti dell'approccio

Dopo tre settimane di sviluppo intensivo, dedicando la maggior parte del tempo all'implementazione e affinamento dell'analizzatore, i limiti intrinseci dell'approccio tradizionale sono diventati evidenti. Quello che era iniziato come un esercizio accademico ambizioso ma fattibile si era trasformato in un progetto di complessità esponenzialmente crescente.

La PROCEDURE DIVISION, che contiene la logica applicativa vera e propria, ha rappresentato il punto di rottura: la complessità dei costrutti e delle loro interazioni suggeriva mesi aggiuntivi di sviluppo solo per una copertura parziale.

Il sistema che avevo sviluppato copriva circa il 25% dei costrutti presenti nei codici prodotti nel primo periodo ed una percentuale ancora meno soddisfacente rispetto ai costrutti necessari per gestire programmi del mondo reale.

Continuare l'esplorazione lineare suggeriva almeno altri 2-3 mesi di sviluppo solo per completare la copertura sintattica, ma la complessità non era lineare - ogni nuovo costrutto interagiva con quelli esistenti in modi che richiedevano ristrutturazione del codice esistente.

In una sessione di retrospettiva con la tutor aziendale, Arianna Bellino, abbiamo analizzato criticamente i risultati ottenuti e le prospettive future arrivando alla necessità di esplorazione di alternative più innovative.

3.4 Terzo periodo: pivot verso l'intelligenza artificiale

La necessità di identificare alternative metodologiche innovative, combinata con l'analisi delle soluzioni enterprise esistenti e le evidenze raccolte durante lo sviluppo del parser tradizionale, ha determinato l'adozione dell'intelligenza artificiale generativa come approccio risolutivo per il problema della migrazione COBOL-Java. L'evoluzione recente delle capacità dei modelli linguistici di grandi dimensioni (LLM) nel dominio della comprensione e generazione del codice ha aperto nuove prospettive per affrontare la complessità intrinseca della traduzione tra paradigmi di programmazione eterogenei.

3.4.1 Valutazione delle API di AI generativa

L'organizzazione ospitante, Miriade, disponeva di accesso enterprise all'API di Google Gemini Pro, fattore che ha determinato la scelta tecnologica e permesso di focalizzare l'analisi sulle capacità specifiche del modello per il task di traduzione del codice.

Google Gemini Pro presentava caratteristiche tecniche particolarmente adatte al progetto di migrazione. Il modello, già dall'interazione via chat dal sito <https://gemini.google.com/>, con upload dei file da tradurre, dimostrava una comprensione sofisticata non solo della sintassi dei linguaggi di programmazione, ma anche della semantica sottostante. Questa comprensione contestuale si rivelava essenziale per produrre traduzioni che preservassero l'intento originale del codice mentre lo modernizzavano per l'ecosistema Java.

La valutazione delle prestazioni del modello ha comportato test sistematici su tre dimensioni principali:

1. **Consistenza delle traduzioni:** verifiche ripetute hanno confermato che input identici producevano output funzionalmente equivalenti *across* multiple sessioni, con variazioni minime limitate a scelte stilistiche non impattanti sulla logica.
2. **Gestione della complessità:** il modello dimostrava capacità di tradurre costrutti COBOL avanzati preservando la logica di business anche in presenza di pattern procedurali complessi e interdipendenze tra moduli.
3. **Qualità del codice prodotto:** l'output generato rispettava consistentemente gli standard Java moderni, producendo codice idiomatrico che un developer Java considererebbe naturale e manutenibile.

Un aspetto cruciale nella valutazione riguardava la gestione dei limiti di token. I programmi COBOL enterprise possono essere estremamente verbosi, con sezioni di dichiarazione dati che occupano centinaia di righe. Gemini Pro offriva un limite di token sufficientemente elevato per gestire la maggior parte dei programmi senza necessità di segmentazione, carat-

teristica che semplificava notevolmente l'architettura del sistema eliminando la complessità della gestione di traduzioni parziali e successive riconciliazioni.

3.4.2 *Design del sistema AI-powered*

Il sistema è stato progettato seguendo tre principi architetturali fondamentali che ne guidano ogni aspetto implementativo:

- **Comprensione semantica olistica:** il sistema non si limita a mappare costrutti sintattici ma analizza il contesto aziendale del codice tramite richiesta di dettagli, interpreta l'intento oltre la forma superficiale, e preserva la logica di dominio durante la trasformazione. Questo approccio permette, ad esempio, di riconoscere che una serie di PERFORM statements in COBOL implementa un pattern di elaborazione batch e tradurlo in un design pattern Iterator in Java, mantenendo la semantica mentre si modernizza l'implementazione.
- **Integrazione contestuale:** COBOL raramente esiste in isolamento, essendo tipicamente integrato con sistemi di gestione dati attraverso SQL embedded o file system proprietari. Il design prevede l'analisi congiunta di codice e schema database, permettendo al sistema di comprendere le relazioni tra logica applicativa e struttura dei dati. Questa visione integrata produce codice Java che non solo traduce le operazioni COBOL ma ottimizza anche l'interazione con il database secondo pattern moderni come connection pooling e prepared statements.
- **Automazione end-to-end:** l'obiettivo di minimizzare l'intervento umano ha guidato la progettazione di un sistema che produce non solo codice tradotto ma progetti completi pronti per il deployment. Questo include la generazione automatica della struttura del progetto, la configurazione del build system, la risoluzione delle dipendenze, e la produzione di documentazione.

L'implementazione si articola in tre moduli principali, ciascuno con responsabilità ben definite ma interconnesse attraverso interfacce chiare:

- Il **modulo di traduzione** costituisce il cuore del sistema. La sua architettura interna gestisce la costruzione di prompt ottimizzati che codificano la conoscenza domain-specific necessaria per guidare il modello nella traduzione. Il modulo implementa meccanismi di gestione dell'interazione con l'API, includendo retry logic con *backoff esponenziale* per gestire eventuali limitazioni di rate o errori transitori. La validazione dell'output avviene attraverso parsing del codice Java generato per assicurare completezza sintattica e presenza di tutti gli elementi strutturali attesi. La configurazione dei

parametri generativi ha richiesto una fase di sperimentazione per identificare i valori di generazione ottimali:

- La temperatura è stata impostata a 0.1, valore estremamente basso che garantisce output deterministici e consistenti, essenziale per un processo di migrazione che richiede ripetibilità.
- Il parametro top-p, configurato a 0.9, e top-k, limitato a 20, sono stati calibrati per bilanciare la capacità del modello di esplorare soluzioni diverse mantenendo al contempo un controllo stretto sulla qualità e coerenza dell'output.
- Il limite di 20.000 token di output era sufficiente per i codici sviluppati nel primo periodo.
- Il **modulo di packaging** trasforma il codice Java raw in un progetto Maven completo. Il punto di forza del modulo risiede nella sua capacità di analizzare il codice generato per identificare automaticamente le dipendenze necessarie. Utilizzando nuovamente Gemini, il modulo esamina gli import statements e l'uso effettivo delle API per determinare non solo quali librerie sono necessarie ma anche le versioni appropriate basandosi su compatibilità e best practice correnti. La generazione del file pom.xml avviene dinamicamente, producendo configurazioni complete che includono sia le dipendenze che anche la configurazione appropriata dei plugin per compilazione, testing, e packaging.
- Il **modulo di orchestrazione** fornisce il layer di coordinamento che trasforma i componenti individuali in un sistema coeso. Implementa una pipeline di esecuzione che gestisce il flusso dei dati tra i moduli, monitora il progresso della conversione, e gestisce condizioni di errore con strategie di recovery appropriate. Il logging da terminale strutturato fornisce visibilità completa sul processo di conversione, essenziale per debugging e audit in contesti enterprise.

Il flusso di elaborazione segue una sequenza logica che massimizza le probabilità di successo della conversione. La fase iniziale di acquisizione e validazione assicura che gli input siano completi e ben formati, prevenendo errori downstream. Segue una fase di pre-processamento dove il codice viene normalizzato e preparato per l'analisi, rimuovendo elementi non significativi, preservando struttura e commenti per la comprensione del contesto.

La fase di generazione del prompt rappresenta il momento critico dove la conoscenza sulla migrazione viene codificata in forma processabile dal modello. Il prompt non è una semplice richiesta ma una specifica dettagliata che include il ruolo del modello, il contesto della traduzione, esempi di pattern di trasformazione, e requisiti specifici per l'output. La

costruzione del prompt sfrutta template parametrizzati che vengono istanziati con il codice specifico e lo schema database, assicurando consistenza mentre si adatta al contesto specifico.

L'elaborazione della risposta del modello richiede parsing e validazione. Il sistema deve estrarre il codice Java dalla risposta del modello, che può includere spiegazioni o metadati aggiuntivi, validare la completezza e correttezza sintattica del codice estratto, e prepararlo per le fasi successive di packaging assicurando che tutti gli elementi necessari siano presenti.

3.5 Quarto periodo: implementazione della soluzione AI-driven

L'implementazione operativa del sistema di migrazione basato su intelligenza artificiale ha trasformato il design concettuale in una soluzione funzionale sorprendente, capace di gestire non solo la complessità del codice COBOL autoprodotta ma anche di eventuali necessità enterprise. Questo periodo è stato caratterizzato da un approccio iterativo dove ogni componente veniva sviluppato, testato sui programmi COBOL di esempio creati nella prima fase, e raffinato basandosi sui risultati ottenuti.

3.5.1 Sviluppo del prompt engineering

Il *prompt engineering* era l'elemento più critico per il successo della traduzione *AI-driven*. Il processo di sviluppo del prompt ha seguito una metodologia empirica basata su cicli di sperimentazione e affinamento.

Il prompt doveva stabilire chiaramente il contesto operativo, definendo il modello come un compilatore capace di comprendere non solo sintassi ma semantica e intento aziendale. Questa definizione di ruolo si è dimostrata cruciale per orientare il comportamento del modello verso traduzioni che privilegiassero la preservazione della logica di business rispetto alla traduzione letterale.

Il corpo del prompt include sezioni strutturate che guidano il modello attraverso il processo di traduzione:

- La sezione di input fornisce il codice COBOL completo insieme allo schema SQL quando disponibile, permettendo al modello di comprendere il contesto completo dell'applicazione
- Le istruzioni di traduzione specificano come gestire costrutti specifici, fornendo mappature esplicite per tipi di dato, pattern di trasformazione per strutture di controllo, e linee guida per la gestione di costrutti senza equivalenti diretti in Java.

L'ottimizzazione iterativa del prompt ha richiesto analisi sistematica dei risultati di traduzione. Ogni fallimento o traduzione sub-ottimale forniva informazioni preziose su ambiguità o lacune nelle istruzioni. Pattern comuni di errore includevano:

- Gestione inadeguata delle transazioni database, inizialmente risolta aggiungendo istruzioni specifiche sulla struttura dei blocchi try-catch e la gestione del rollback
- Traduzione letterale di costrutti COBOL che produceva codice Java non idiomático, affrontata attraverso esempi di pattern di trasformazione preferiti
- Perdita di informazioni sui tipi di dato durante la conversione, corretta specificando mappature esplicite e regole di inferenza

La gestione dei casi speciali ha richiesto particolare attenzione nella formulazione del prompt. Costrutti COBOL come REDEFINES, che permettono interpretazioni multiple della stessa area di memoria, non hanno equivalenti diretti in Java. Il prompt è stato evoluto per includere strategie specifiche di gestione, suggerendo l'uso di classi wrapper con metodi di conversione espliciti. Similarmente, la gestione delle tabelle COBOL con OCCURS DEPENDING ON ha richiesto istruzioni per la creazione di strutture dati dinamiche appropriate in Java, tipicamente ArrayList o array ridimensionabili.

Un aspetto innovativo dello sviluppo del prompt è stata l'inclusione di meta-istruzioni che guidano il processo di ragionamento del modello. Invece di fornire solo regole di traduzione meccaniche, il prompt incoraggia il modello a considerare, tramite analisi autonoma, l'intento del codice originale, e produrre soluzioni che un developer Java moderno considererebbe naturali. Questo approccio ha prodotto traduzioni significativamente migliori rispetto a prompt puramente prescrittivi.

3.5.2 Implementazione del translator completo

Lo sviluppo del sistema di conversione end-to-end ha richiesto l'integrazione di tutti i componenti in un flusso operativo coerente. L'implementazione si è concentrata sulla creazione di un sistema capace di gestire la varietà e complessità del codice COBOL reale.

Il sistema analizza la gerarchia dei level numbers per costruire una rappresentazione interna della struttura dati. Questa analisi identifica:

- Record di primo livello (level 01) che diventano classi Java principali
- Strutture subordinate che vengono mappate a inner classes o campi semplici basandosi sulla complessità
- Elementi ripetuti (OCCURS) che richiedono array o collezioni
- Ridefinizioni (REDEFINES) che necessitano di gestione speciale attraverso union-like patterns

Il processo di generazione delle classi Java corrispondenti applica convenzioni di naming standard, trasformando nomi COBOL in stile KEBAB-CASE in camelCase Java. La generazione include automaticamente metodi getter e setter appropriati, costruttori per inizializzazione, e metodi utility per conversioni quando ritenuto necessario dal modello.

La traduzione dell'SQL embedded ha richiesto particolare attenzione alla preservazione della semantica transazionale. Il sistema identifica i blocchi EXEC SQL attraverso pattern matching, estrae gli statement SQL e le variabili host coinvolte, e genera codice JDBC equivalente. La generazione utilizza sempre PreparedStatement per prevenire SQL injection, implementa gestione appropriata delle connessioni con pattern try-with-resources, preserva la logica di gestione errori COBOL attraverso mappature SQLCODE appropriate, e mantiene la semantica transazionale con commit e rollback espliciti.

Un aspetto critico dell'implementazione riguarda la preservazione della logica di business durante la trasformazione. Il translator riconosce pattern comuni nel codice COBOL procedurale e li trasforma in equivalenti object-oriented appropriati:

- I PERFORM statements vengono analizzati per determinare se rappresentano semplici chiamate di subroutine o pattern più complessi come iterazioni
- Le SECTION e PARAGRAPH della PROCEDURE DIVISION vengono trasformate in metodi Java, preservando la struttura logica mentre si adotta l'organizzazione object-oriented

Per quanto riguarda la gestione degli errori, COBOL spesso utilizza gestione degli errori implicita attraverso controlli di status code, mentre Java favorisce exception handling esplicito. Il translator aggiunge automaticamente blocchi try-catch appropriati dove necessario, preserva i codici di errore COBOL per compatibilità mentre aggiunge exception handling Java, implementa logging strutturato per facilitare debugging e manutenzione, e crea classi di eccezione custom quando pattern di errore specifici lo richiedono.

3.5.3 Generazione automatica di progetti Maven

La fase finale del processo di migrazione trasforma il codice Java generato in un progetto completo pronto per il deployment. Questa fase sfrutta nuovamente le capacità di Gemini per analizzare il codice e determinare tutti i requisiti di progetto.

L'analisi delle dipendenze inizia con l'esame degli import statements nel codice Java generato. Il sistema utilizza Gemini per comprendere non solo quali classi vengono importate ma come vengono utilizzate nel codice. Questa analisi contestuale permette di identificare le librerie necessarie con le versioni appropriate, determinare dipendenze transitive che

potrebbero essere richieste, escludere dipendenze non necessarie che potrebbero essere state importate ma non utilizzate, e risolvere potenziali conflitti di versione basandosi su best practice correnti.

La struttura del progetto Maven viene generata seguendo le convenzioni standard. Il sistema crea automaticamente la gerarchia di directory appropriata, posiziona il codice sorgente nelle location corrette secondo il package structure, prepara directory per risorse, configurazioni, e test, e predispone la struttura per facilitare future estensioni e manutenzione.

La generazione del file pom.xml rappresenta un elemento critico del processo. Il sistema produce configurazioni complete che specificano:

- Coordinate del progetto (groupId, artifactId, version) derivate dal nome del programma COBOL originale
- Proprietà del progetto includendo versioni Java, encoding, e altre configurazioni standard
- Dipendenze identificate attraverso l'analisi del codice con versioni appropriate
- Configurazione dei plugin Maven per compilazione, testing, e packaging
- Profili per diversi ambienti di deployment quando identificati dal contesto

La configurazione dei plugin Maven riceve particolare attenzione per assicurare che il progetto possa essere costruito e deployato senza modifiche manuali:

- Il maven-compiler-plugin viene configurato con la versione Java appropriata basata sulle feature utilizzate nel codice
- Il maven-jar-plugin include configurazione per generare JAR eseguibili con manifest appropriato
- Il maven-assembly-plugin viene configurato per creare «fat JARs» che includono tutte le dipendenze, semplificando il deployment

Il processo di build automatizzato verifica la correttezza della configurazione attraverso l'invocazione di Maven per compilare il codice, risolvere e scaricare tutte le dipendenze, eseguire eventuali test di base generati, e produrre gli artifact finali. Qualsiasi errore in questa fase viene catturato e reportato con suggerimenti per la risoluzione.

La generazione della documentazione completa il processo. Il sistema preserva i commenti COBOL originali e li arricchisce con informazioni sulla migrazione. I commenti di intestazione dei programmi COBOL vengono trasformati in JavaDoc comprensivi che includono informazioni sull'origine COBOL, data e versione della migrazione, note su trasformazioni significative applicate, e riferimenti alla documentazione originale quando disponibile.

Il risultato finale del processo è un progetto Java completo, moderno, e immediatamente utilizzabile.

3.6 Risultati raggiunti

3.6.1 Impatto dell'AI sui tempi di sviluppo

In questa sottosezione quantificherò la riduzione importante dei tempi rispetto all'approccio tradizionale, illustrerò il passaggio da mesi a giorni nel processo di conversione e analizzerò i risultati impossibili senza AI.

3.6.2 Analisi qualitativa dei risultati

In questa sottosezione descriverò il sistema completo di conversione COBOL-Java funzionante, analizzerò il codice Java idiomatrico e manutenibile prodotto e illustrerò la documentazione professionale automatizzata.

3.6.3 Risultati quantitativi

In questa sottosezione presenterò i dati concreti: tre progetti convertiti con successo, vasta copertura delle funzionalità e oltre 2000 linee di codice Java di qualità production-ready.

4 Valutazioni retrospettive e prospettive future

Qui introdurrò brevemente il contenuto delle sezioni sottostanti.

4.1 Analisi retrospettiva del percorso

In questa sezione analizzerò il soddisfacimento degli obiettivi al capitolo 2 grazie all'approccio AI, confronterò i risultati ottenuti con le stime iniziali basate sullo sviluppo tradizionale e identificherò le *lessons learned* e *best practices* emerse dal progetto.

4.2 L'AI come *game changer* nella modernizzazione *software*

In questa sezione descriverò come l'intelligenza artificiale abbia trasformato il progetto da «prototipo dimostrativo» a «soluzione potenzialmente completa», confronterò l'approccio sviluppato con soluzioni *enterprise* come IBM *WatsonX*, analizzerò il ruolo cruciale del *prompt engineering* e valuterò limiti e potenzialità dell'approccio AI-driven.

4.3 Crescita professionale e competenze acquisite

In questa sezione descriverò le *hard skills* acquisite in migrazione *legacy*, AI *engineering* e *prompt design*, analizzerò le *soft skills* sviluppate come *problem solving* e adattabilità, illustrerò la visione sistemica della modernizzazione IT maturata e la capacità di valutare e integrare pragmaticamente tecnologie emergenti.

4.4 Valore della formazione universitaria nell'era dell'AI

In questa sezione analizzerò come il percorso universitario mi abbia fornito le solide basi metodologiche essenziali per affrontare questa sfida tecnologica, valorizzando in particolare l'approccio al *problem solving* e il metodo di studio critico acquisiti. Descriverò come la formazione teorica ricevuta si sia rivelata fondamentale per comprendere e padroneggiare tecnologie emergenti come l'AI, evidenziando l'importanza dell'approccio universitario che insegna ad «imparare ad imparare».

4.5 Roadmap evolutiva e opportunità di sviluppo

In questa sezione descriverò le possibili evoluzioni della soluzione verso il supporto *multi-linguaggio* per altri sistemi *legacy*, analizzerò il potenziale di commercializzazione della soluzione e esplorerò l'uso di *multi-agent systems* per conversioni complesse.

5 Lista degli acronimi

AI: Artificial Intelligence

ANTLR: ANother Tool for Language Recognition

API: Application Programming Interface

AST: Abstract Syntax Tree

CLI: Command Line Interface

COBOL: Common Business-Oriented Language

IT: Information Technology

JDBC: Java Database Connectivity

JSON: JavaScript Object Notation

LLM: Large Language Model

6 Glossario

Agile: Metodologia di sviluppo software iterativa e incrementale

DevOps: Pratiche che combinano sviluppo software e operazioni IT

Kanban: Sistema di gestione del workflow visuale

Scrum: Framework Agile per la gestione di progetti complessi

backoff esponenziale: Strategia di retry con attese progressivamente più lunghe

legacy: Sistemi informatici datati ma ancora in uso

mainframe: Computer di grandi dimensioni per elaborazioni complesse

microservizi: Architettura software basata su servizi indipendenti

parser: Analizzatore sintattico che interpreta la struttura del codice

prompt engineering: Tecnica di formulazione di istruzioni per modelli AI

sprint: Periodo di tempo definito per completare un set di attività

stack tecnologico: Insieme di tecnologie software utilizzate per sviluppare un'applicazione

stand-up: Breve riunione giornaliera del team Agile

token: Unità base di testo processata dai modelli linguistici

7 Sitografia

- [1] CBT Nuggets, «What is COBOL and Who Still Uses It?». Consultato: giugno 2025. [Online]. Disponibile su: <https://www.cbtnuggets.com/blog/technology/programming/what-is-cobol-and-who-still-uses-it>
- [2] Version 1, «Legacy System Modernization: Challenges and Solutions». Consultato: maggio 2025. [Online]. Disponibile su: <https://www.version1.com/insights/legacy-system-modernization/>
- [3] DXC Luxoft, «How come COBOL-driven mainframes are still the banking system of choice?». Consultato: giugno 2025. [Online]. Disponibile su: <https://www.luxoft.com/blog/why-banks-still-rely-on-cobol-driven-mainframe-systems>
- [4] How-To Geek, «What Is COBOL, and Why Do So Many Institutions Rely on It?». Consultato: maggio 2025. [Online]. Disponibile su: <https://www.howtogeek.com/667596/what-is-cobol-and-why-do-so-many-institutions-rely-on-it/>
- [5] CAST Software, «Why COBOL Still Dominates Banking—and How to Modernize». Consultato: giugno 2025. [Online]. Disponibile su: <https://www.castsoftware.com/pulse/why-cobol-still-dominates-banking-and-how-to-modernize>