

Data ingestion e anomaly detection:

Il ciclo di vita del dato

Giovanni Sorice

25 Settembre, 2019

Esame di laurea triennale in
Informatica

A.A 2018 / 2019



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



- Introduzione
- L'azienda
- Il progetto
- Tecnologie
- Conclusioni



Dato grezzo:

- Oggettivo
- Non elaborato
- Non organizzato

1000

Dato come informazione:

- Elaborato
- Comparabile
- Calato in un contesto

Utente	Accessi	Giorno
Rossi	150	19/09
Rossi	195	20/09
Rossi	1000	24/09
Rossi	210	25/09
Rossi	170	26/09

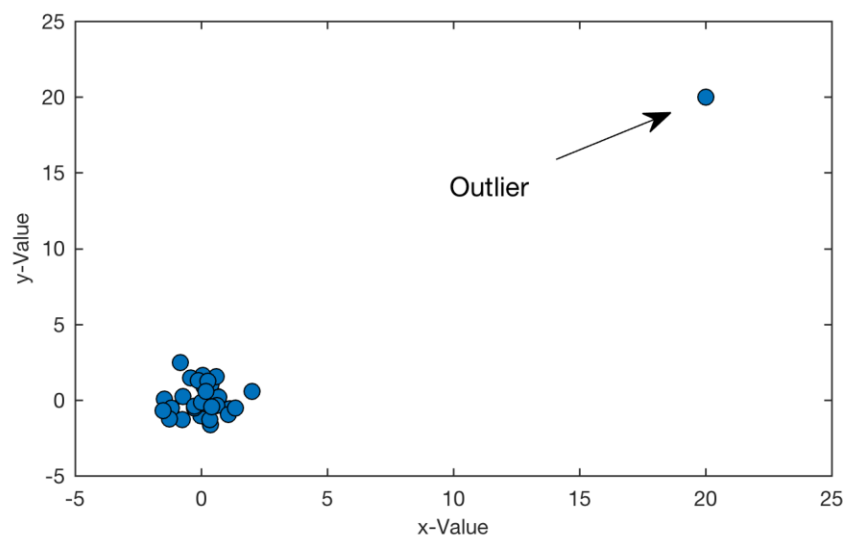
Fasi

- Data ingestion
- Data processing
- Data analysis
- Data integration



*"...difformità della regola
generale..."*

Vocabolario Treccani





- Fondata nel 2013
- Fatturato di 7,5 milioni nel 2017
- 150 dipendenti
- Innovazione e dinamismo

Il progetto



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Gestione dei dati



Rilevamento delle anomalie



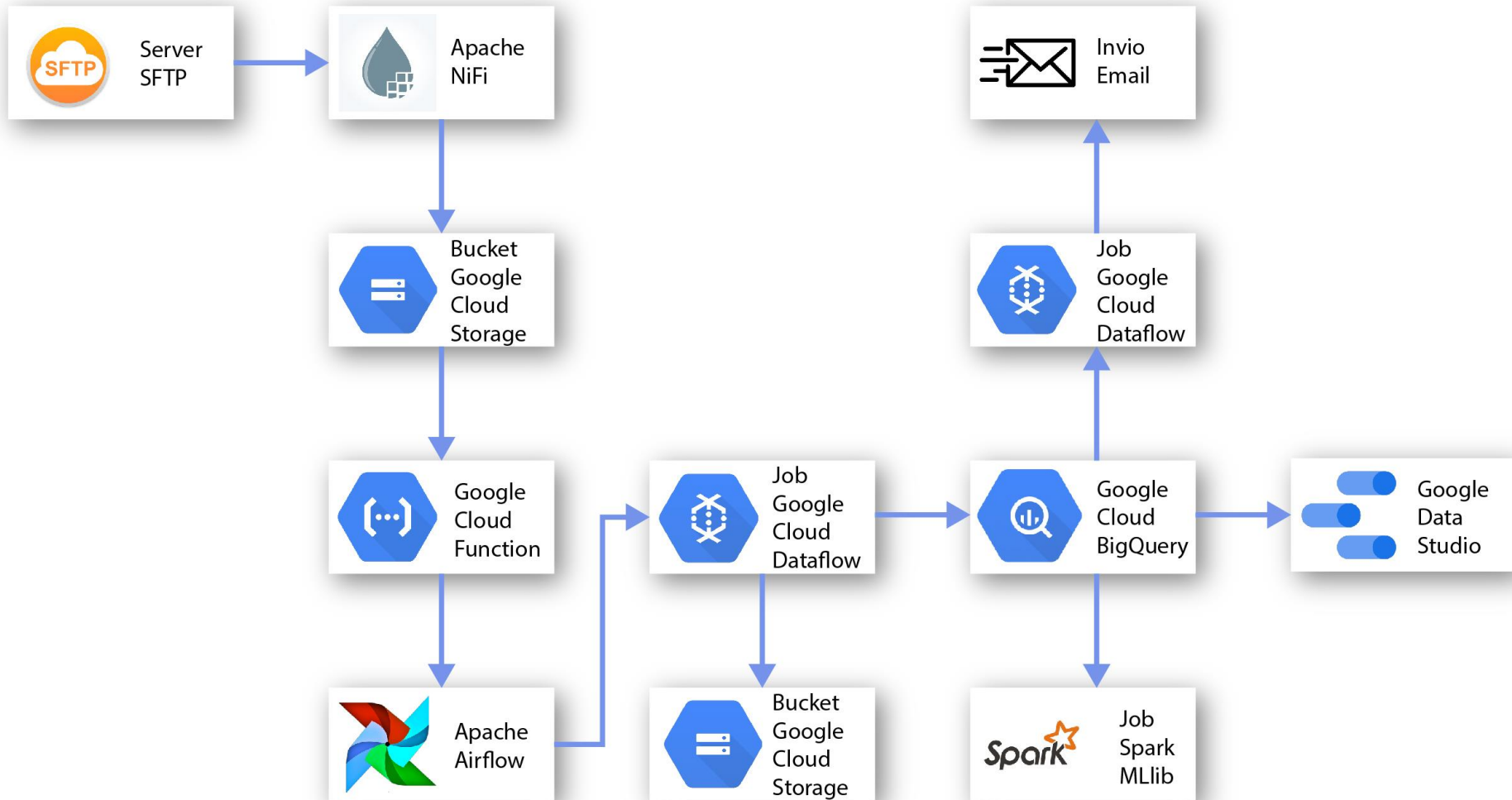
Analisi dei dati



Problemi



Integrazione

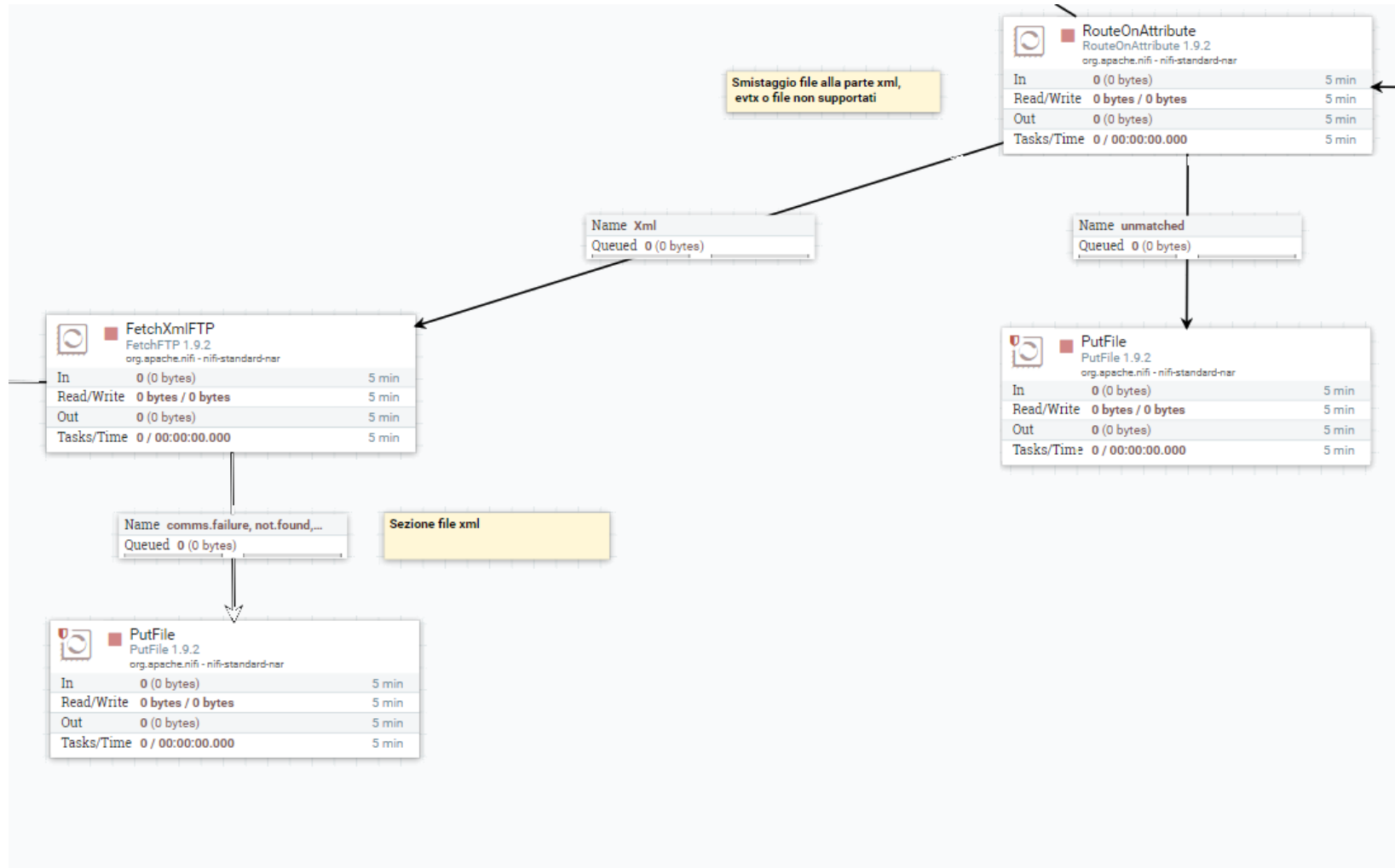


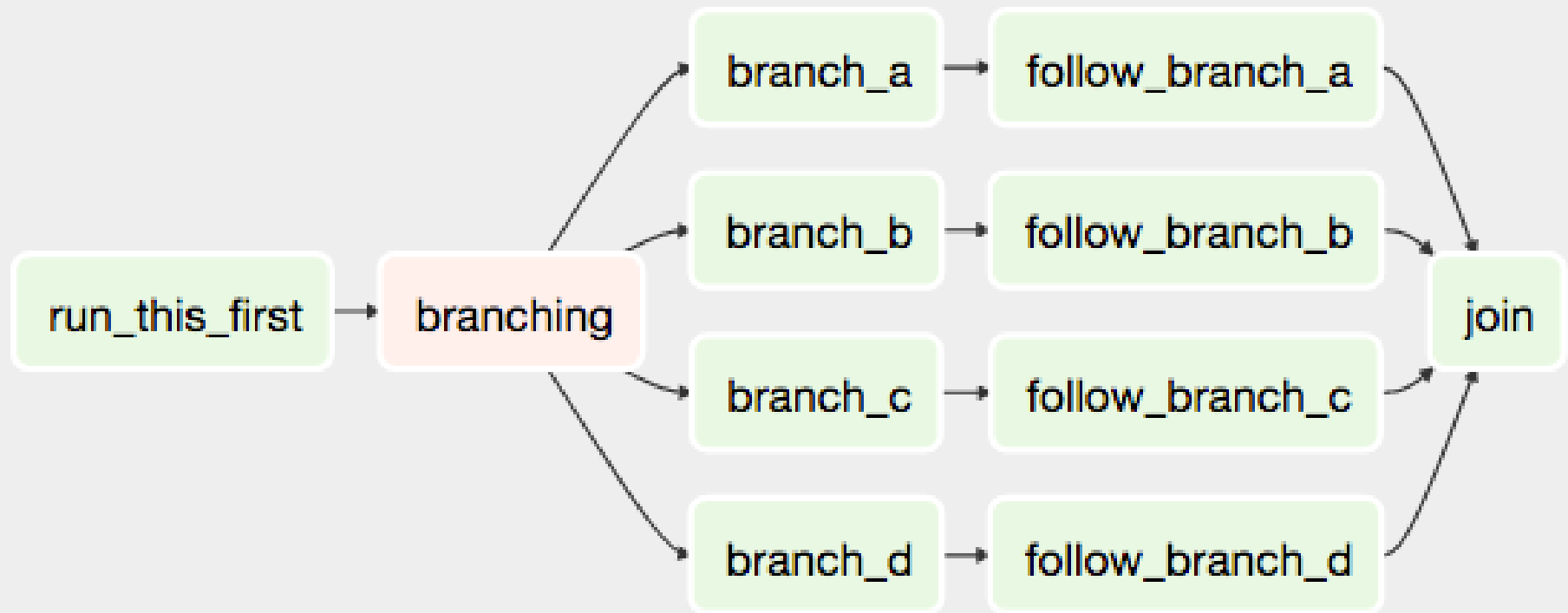
Integrazione

Apache NiFi

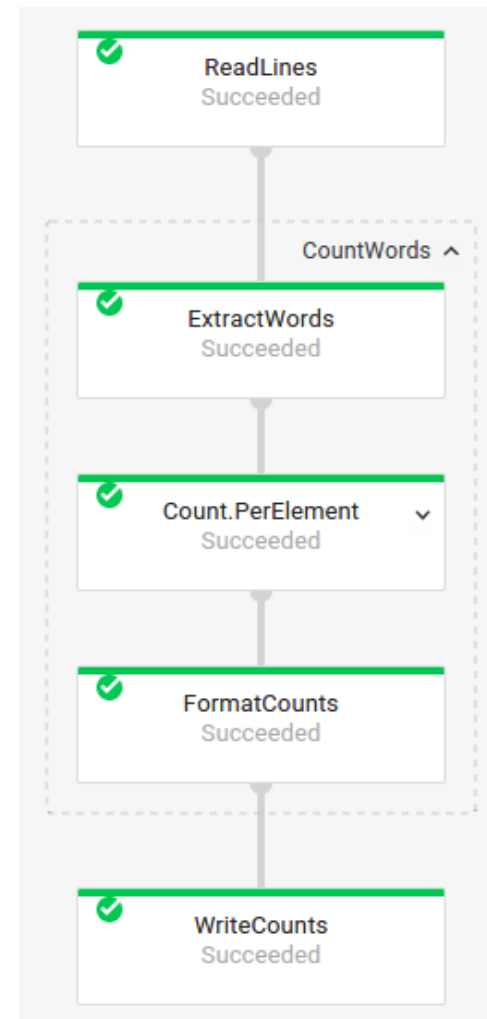


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

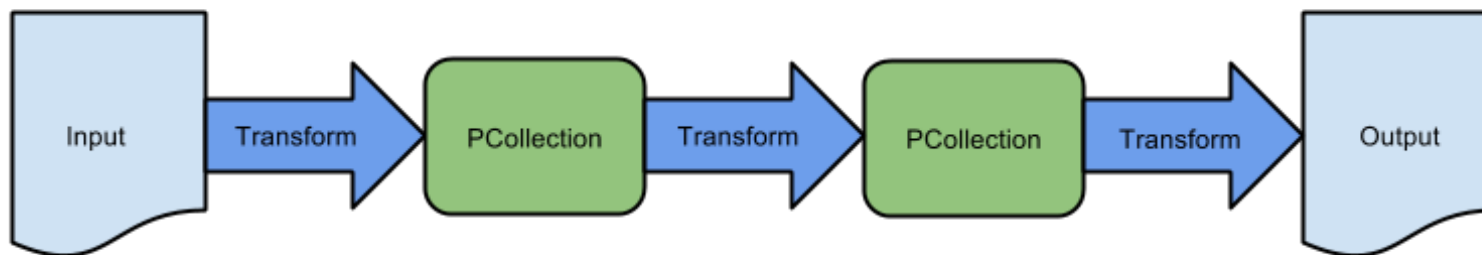




- Utilizzo dei modelli di pipeline
- Scalabilità automatica
- Utilizzo di molteplici fonti di dati
- Specificazione parametri pipeline
- Visualizzazione grafica



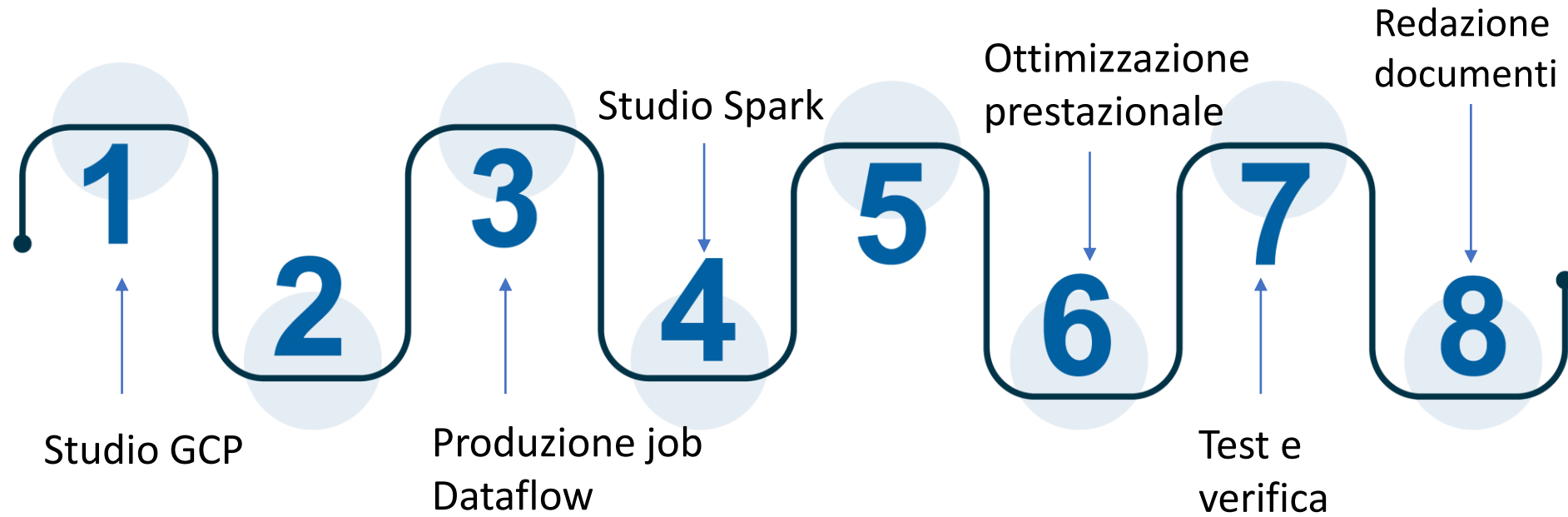
- Pipeline ETL batch e streaming
- Trattamento dei dati unificato



Impegno orario



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



8 settimane
8 ore al giorno
320 ore complessive



- 4 job dataflow
- 1 architettura producer/consumer kafka
- 15 tecnologie usate
- 10 classi per l'ingestion dei dati
- 1 container docker



- Manuale sviluppatore
- Documentazione delle tecnologie usate
- Presentazione interna

Conclusioni



Sviluppi futuri:

- Miglioramento fase data analysis
- Miglioramento gestione flussi streaming
- Aggiunta gestione file con diversi schemi XML
- Aggiungere acquisizione file non evtx o XML



Considerazioni personali:

- Acquisizione di molteplici competenze
- Sviluppo di un sistema funzionante
- Ampliata rete sociale
- Gestione di un intero progetto

