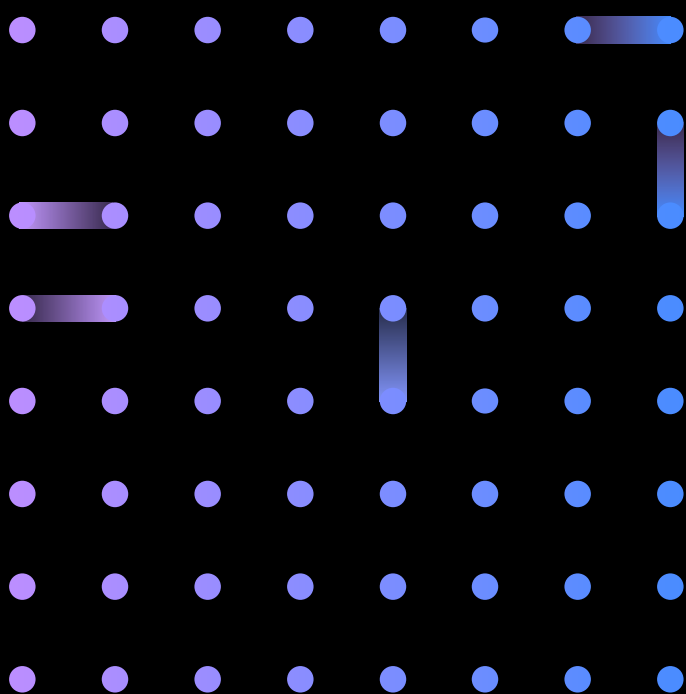


Ottieni e fornisci velocemente dati business-ready con DataOps

Introduzione a metodologia
e prassi di IBM DataOps



Sommario

Introduzione	3
Definizione di DataOps	3
Comparazione degli obiettivi	4
Programma IBM DataOps	7
La metodologia IBM Cloud Garage applicata a DataOps	7
L'impatto di una prassi di DataOps di successo	10
Conclusioni	11
Appendice: Modulo programma pilota di DataOps	12

Punti salienti

- DataOps significa orchestrazione di persone, processi e tecnologie per offrire rapidamente dati affidabili e di elevata qualità ai data citizen
- DataOps utilizza l'automazione e affronta le sfide relative alle inefficienze di valutazione, preparazione, integrazione e disponibilità dei dati
- La prassi di IBM DataOps fa uso di un livello molto elevato di automazione al fine di ottenere un impatto rilevante e misurabile sulle seguenti funzioni: servizi per l'accuratezza dei dati, gestione dei metadati, governance dei dati, master data management e interazione self-service
- IBM offre un percorso per la prassi di DataOps mediante una metodologia di tipo prescrittivo, un'automazione basata su intelligenza artificiale (AI) e l'IBM DataOps Center of Excellence
- Parte integrante della roadmap DataOps è il workshop DataOps, che aiuta l'organizzazione a valutare la propria maturità DataOps e a pianificare la realizzazione di un progetto pilota

Introduzione

I dati alimentano l'innovazione e sostengono un vantaggio competitivo. Sono l'ingrediente principale per favorire l'analytics e comprendere tendenze e opportunità di business. Disporre di nuove modalità per rivelare il valore dei dati può anche accelerare il percorso di un'azienda verso l'intelligenza artificiale.

Tuttavia, quando i progetti inerenti ai dati non riescono a fornire il ROI desiderato, gli stakeholder ne vogliono conoscere i motivi. Secondo il report del 2019 di Experian "Global Data Management Research", l'89% delle aziende ha dichiarato di avere difficoltà nella gestione dei dati. Tra queste figurano ritardi nella raccolta di informazioni utili e carenza di fiducia nei dati sottostanti.¹

La comprensione degli obiettivi di business di un'organizzazione è decisiva per lo sviluppo di una strategia dati efficace per l'analytics e l'AI. Affinché un qualsiasi modello di business funzioni, è necessario che soddisfi le esigenze dei clienti. Il successo dipende dalla razionalizzazione delle operazioni sui dati con una pipeline relativa ai dati business-ready e integrata, in grado di fornire una visione dell'azienda completa e coerente in qualsiasi momento.

Le aspettative di ottenere risultati più rapidamente sono in costante crescita. Ovunque, le aziende sono alla ricerca di modalità per migliorare la propria efficienza ed efficacia operativa, favorendo l'ottimizzazione dei processi decisionali, in particolare considerando i tanti silos interni. Questi due fattori costringono i leader aziendali a ricercare nuovi modi per affrontare le loro maggiori sfide all'interno di un unico framework.

La tecnologia di automazione può offrire un vantaggio competitivo alle organizzazioni che desiderano una trasformazione nella propria operatività sui dati. I dati diventano utili e preziosi se sono attendibili e business-ready, ovvero se consentono all'azienda insight differenziati ed eccellenza operativa.

Scopo del presente documento è evidenziare i vantaggi della metodologia, della pratica e della roadmap di DataOps.

Definizione di DataOps

DataOps (Data Operations - operazioni sui dati) è l'orchestrazione di personale, processi e tecnologie per fornire rapidamente dati affidabili e di alta qualità ai data citizen. La pratica è incentrata sul consentire la collaborazione all'interno di un'intera organizzazione, favorendo agilità, rapidità e iniziative su larga scala legate ai nuovi dati. Grazie alla potenza dell'automazione, DataOps è progettato per rispondere alle sfide relative alle inefficienze nell'accesso, nella preparazione, nell'integrazione e nella disponibilità dei dati.

I potenziali vantaggi di DataOps includono significativi guadagni di produttività nella distribuzione di informazioni e dati alle persone e miglioramento dei processi per un aumento dell'efficienza e una maggiore ottimizzazione. Le operazioni automatizzate sui dati che comprendono iniziative AI possono favorire i seguenti risultati:

- Dati integrati business-ready in grado di favorire analytics e AI su larga scala
- Efficienza operativa
- Data privacy e conformità

89%

è la percentuale delle aziende che ha difficoltà nella gestione dei dati.¹ La comprensione degli obiettivi di business di un'organizzazione è decisiva per lo sviluppo di una strategia dati efficace per l'analytics e l'AI.



DataOps non è DevOps

La maggior parte delle aziende ha già implementato un qualche livello di DevOps nelle proprie discipline di sviluppo. La diffusa familiarità della prassi DevOps e le similitudini nelle convenzioni di denominazione hanno favorito i confronti con l'emergente prassi DataOps. Ma benché entrambe siano metodologie che favoriscono le migliori prassi, ognuna di esse svolge un diverso ruolo nell'ambito di un'organizzazione.

Nella tabella sottostante si può vedere il raffronto fra le due prassi in materia di obiettivi e vantaggi per l'organizzazione.

Comparazione degli obiettivi

	DataOps	DevOps
Focus primario	Dati attendibili, business-ready, di elevata qualità e rapidamente pronti all'uso.	Sviluppo di applicazioni e software
Obiettivi di trasformazione	<ul style="list-style-type: none">– Alimentare un'innovazione rapida e continua per le aziende, consentendo l'accesso self-service a dati affidabili e di alta qualità per tutti i data citizen– Consentire una distribuzione continua dei dati automatizzando la governance dei dati, l'integrazione e salvaguardando al tempo stesso gli aspetti normativi– Fornire un loop dei feedback dei data citizen per l'apprendimento continuo mediante il monitoraggio e l'ottimizzazione della pipeline dei dati	<ul style="list-style-type: none">– Velocizzare l'innovazione continua delle idee consentendo lo sviluppo e i test collaborativi lungo tutta la catena del valore– Consentire una diffusione costante di tali innovazioni automatizzando i processi distributivi del software ed eliminando gli sprechi, contribuendo nel contempo ad affrontare le problematiche normative– Fornire un loop dei feedback dei clienti per l'apprendimento continuo mediante il monitoraggio e l'ottimizzazione dell'innovazione basata su software
Obiettivi di efficienza	<ul style="list-style-type: none">– Correggere il disallineamento fra persone e obiettivi incoraggiando collegamenti più stretti fra supporto ai sistemi IT, operazioni e azienda– Accelerare la diffusione delle modifiche e migliorare la qualità della distribuzione introducendo l'automazione in ogni fase del ciclo di consegna dati– Migliorare la visibilità sul valore reale dei metadati e dei dati utilizzando i risultati per favorire l'ottimizzazione	<ul style="list-style-type: none">– Correggere il disallineamento fra persone e obiettivi incoraggiando collegamenti più stretti fra sviluppatori, operazioni e azienda– Accelerare ed eliminare gli errori dalla diffusione delle modifiche introducendo l'automazione in ogni fase del ciclo di sviluppo– Migliorare la visibilità sul valore reale delle applicazioni utilizzando il feedback dei clienti per favorire l'ottimizzazione

DataOps è l'orchestrazione di persone, processi e tecnologie e per sostenere l'impegno verso la metodologia DataOps è necessaria una profonda collaborazione fra tutte le funzioni. Inoltre richiede una particolare attenzione nel coltivare le prassi e i processi di gestione dati che migliorano la rapidità e l'accuratezza dell'analitica.

Persone e processi

DataOps supporta team altamente produttivi con la tecnologia dell'automazione per contribuire all'ottenimento di una maggiore efficienza in termini sia di risultati di progetto sia di tempo impiegato. Per ottenere tali risultati, tuttavia, è necessario che la cultura aziendale si evolva e diventi autenticamente basata sui dati. Sono sempre più numerose le unità di business che richiedono e desiderano gestire i dati per favorire insight contestuali, ecco perché è il momento giusto per fare quanto segue:

- Aumentare la qualità e la velocità del flusso dei dati verso l'azienda.
- Ottenere l'impegno della dirigenza per supportare e sostenere una visione basata sui dati in tutta l'azienda.

Questo tipo di trasformazione parte dalla comprensione degli obiettivi reali dell'azienda. *In che modo i dati influiscono sulle decisioni e sui servizi che hanno poi effetto sui clienti? Come possono i dati contribuire a mantenere un vantaggio competitivo sul mercato? Quali sono le priorità finanziarie che i dati possono contribuire a risolvere?*

I responsabili DataOps devono definire i ruoli di tutti i data citizen per favorire il progredire della cultura e della prassi di DataOps. Ogni azienda ha esigenze specifiche alle quali gli stakeholder in ambito IT, data science e linee di business devono poter dare valore aggiunto per favorire il successo dell'azienda. Inoltre, utilizzare i comitati esistenti per la governance dei dati e quanto rilevato dai programmi di governance dei dati per ruolo aiuta a stabilire tale cultura e impegno, in quanto la governance è una delle forze ispiratrici necessarie per supportare DataOps.

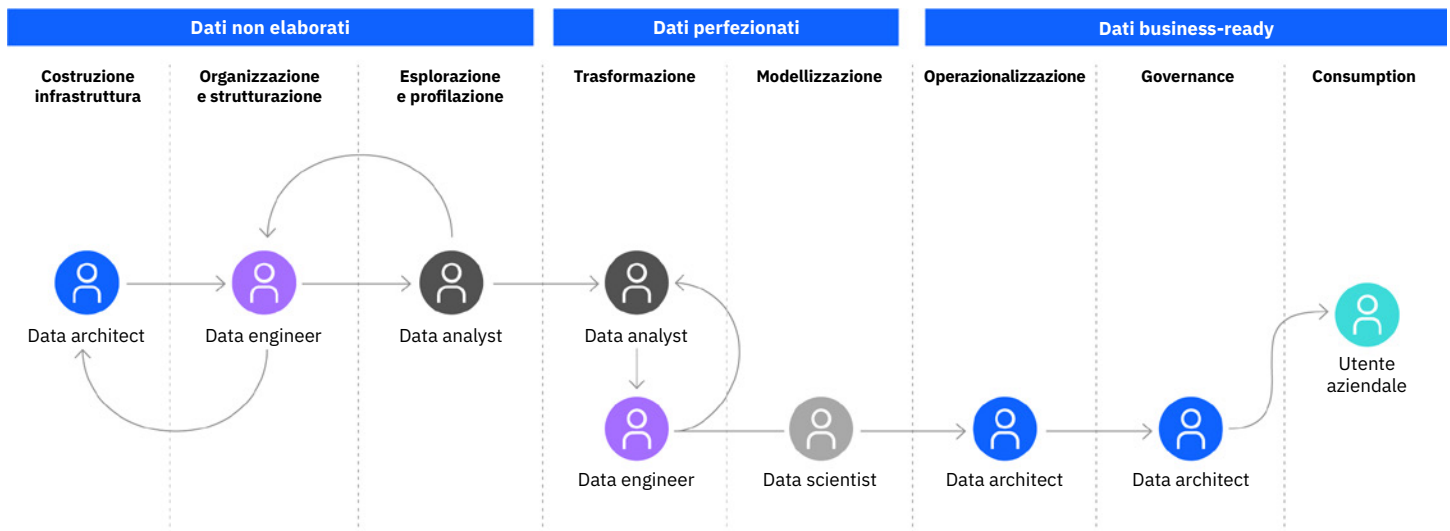


Figura 1: Esempio di workflow DataOps per ruolo

Tecnologia

Al centro di DataOps c'è l'**architettura delle informazioni** dell'organizzazione. *Conosci i tuoi dati? Ti fidi dei tuoi dati? Sei in grado di rilevare rapidamente gli errori? Sei in grado di effettuare modifiche progressive senza "interrompere" l'intera pipeline dei dati?* Per rispondere a queste domande, la prima cosa da fare è stilare un inventario dei propri strumenti e delle proprie prassi di **governance dei dati** e di **integrazione dei dati**. Il tooling è necessario per supportare qualsiasi prassi che si affidi all'automazione.

Considerando il tooling a supporto della prassi DataOps all'interno di un'organizzazione, pensa a come l'automazione può trasformare la pipeline dei dati in questi cinque ambiti critici:

1. Servizi di cura dei dati
2. Gestione metadati
3. **Governance dei dati**
4. **Master data management**
5. Interazione self-service

Funzionalità DataOps

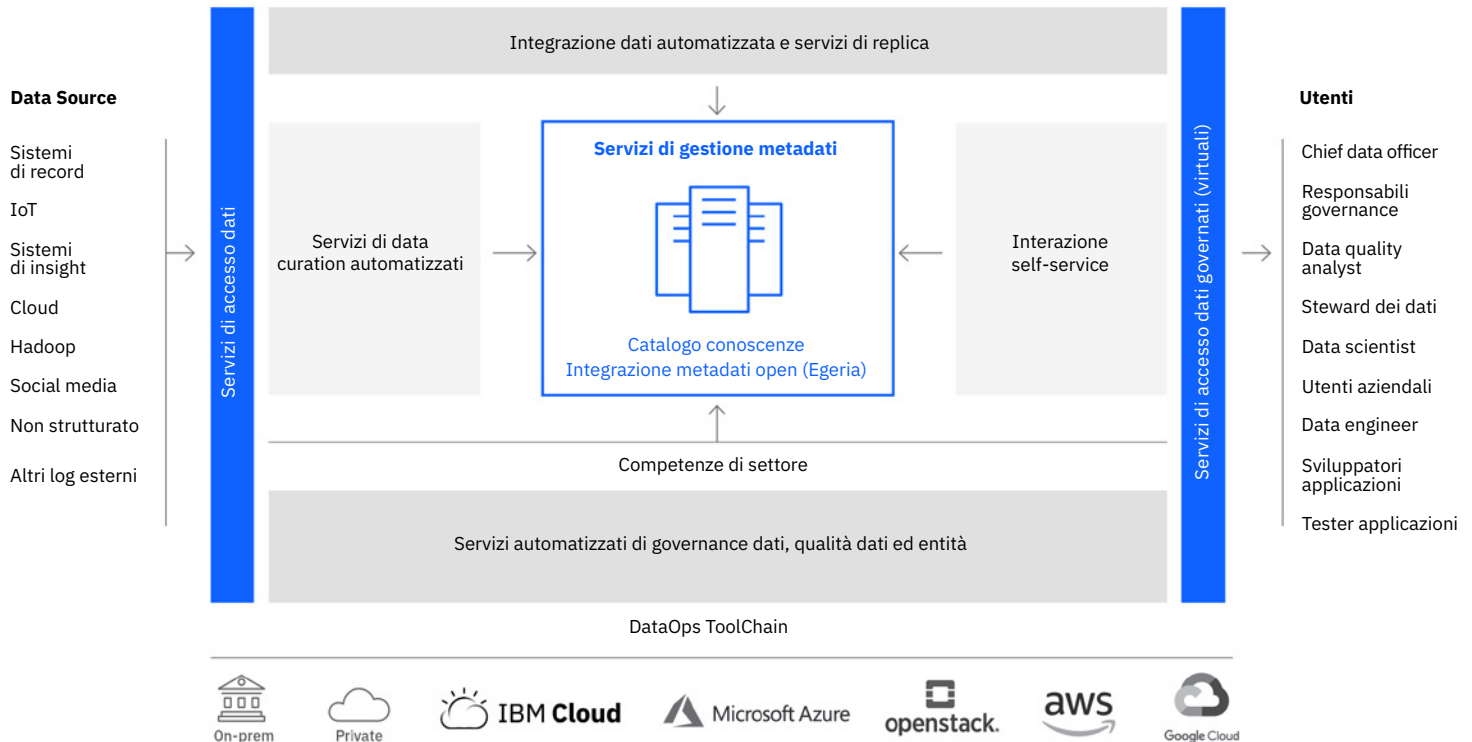


Figura 2: Disamina di un'architettura delle informazioni a supporto di DataOps

La fornitura di dati business-ready include tutti questi aspetti, e qualsiasi prassi di DataOps deve comprendere un approccio olistico che incorpori tutti e 5 gli aspetti. Le aziende che si concentrano su un elemento della pipeline dati a scapito di altri probabilmente non riusciranno a ottenere i vantaggi derivanti dall'implementazione delle metodologie di DataOps. La discussione e l'implementazione delle tecnologie non va svincolata dalla pianificazione corrente in merito a persone e processi. Il tooling aiuta a supportare e sostenere la cultura.

Conclusione

DataOps può sembrare scoraggiante se l'azienda è ancora alle prese con problemi di base quali la definizione dei ruoli degli steward dei dati o la creazione di regole di convalida dei dati. La metodologia DataOps, in realtà, offre la soluzione a molti

problemi che le aziende hanno riscontrato nelle loro iniziative di trasformazione digitale.

L'esempio prevalente di errore che le aziende riconoscono è all'interno dei data lake. Molti si trovano al secondo, terzo o anche quarto tentativo di ottenere il successo tecnologico e hanno "rivitalizzato" le proprie dirigenze proprio per adottare i cambiamenti di mentalità necessari. *Ma perché queste implementazioni di data lake hanno fallito finora?*

Molti di questi progetti si concentravano unicamente sull'immissione nel data lake di dati non ripuliti e non regolati. La cosa più probabile è che il fallimento si sia verificato a causa dei limiti nell'affrontare efficacemente le problematiche relative a persone, processi e tecnologie.

Programma IBM DataOps

Il passaggio all'adozione di DataOps è un dato di fatto. Secondo un recente sondaggio, [il 73% delle aziende ha in programma di investire in DataOps](#).² IBM offre un percorso verso la [prassi di DataOps](#) che prevede una metodologia prescrittiva, una tecnologia d'avanguardia e l'[IBM DataOps Center of Excellence \(CoE\)](#). Nel CoE, gli esperti IBM collaborano con le aziende per personalizzare un approccio che sia basato sugli obiettivi aziendali e identificare i progetti pilota ottimali per generare valore per gli stakeholder.

Le funzionalità di IBM DataOps aiutano a ottenere dati business-ready grazie a tecnologie leader di settore che operano con l'automazione AI, una governance integrata e un importante catalogo di conoscenze per rendere operativi dati continui e di elevata qualità in tutta l'azienda. Consente di aumentare l'efficienza, la qualità dei dati, la reperibilità e infonde regole di governance per fornire una pipeline dati self-service alle persone giuste al momento giusto, da qualsiasi fonte.

Dalle soluzioni che facilitano la governance dei data lake allo sviluppo di applicazioni e alla conformità normativa, IBM DataOps aiuta le aziende a mostrare il valore dei dati nell'ottimizzazione delle decisioni e del tempo. L'ottenimento di dati aziendali di elevata qualità che consentano l'AI è un obiettivo a portata di mano, quando un'azienda conosce, si affida e usa i dati per generare valore in cloud e in qualsiasi ambiente critico.

La metodologia IBM Cloud Garage applicata a DataOps

La metodologia IBM Cloud Garage è un approccio che consente all'azienda, allo sviluppo e alle operazioni di progettare, fornire e convalidare in modo continuo nuove funzionalità. Le prassi, l'architettura e le toolchain coprono l'intero ciclo di vita del prodotto, dall'adozione alla raccolta e alla risposta ai feedback dei clienti e all'evoluzione del mercato. L'architettura Open Toolchain è stata progettata per facilitare la combinazione dei servizi IBM Cloud™ Platform, quali Continuous Delivery (CD), con strumenti open-source e di terze parti leader, in una toolchain integrata allineata alle pratiche di DataOps. Tali modelli sono condivisibili fra team sotto forma di template per promuovere con successo l'adozione di DataOps nell'intera azienda.

IBM ha identificato sei fasi nel ciclo di vita di DataOps, oltre le necessarie considerazioni "culturali", per un'implementazione di successo della metodologia DataOps. Questo si basa sull'adozione interna di DataOps come parte del percorso di trasformazione per IBM.

La metodologia IBM Cloud Garage descrive queste sei fasi come:

- **Think.** Concettualizzazione, affinamento e gerarchizzazione delle funzionalità
- **Code.** Generazione, potenziamento, ottimizzazione e test delle funzioni
- **Deliver.** Produzione automatizzata e fornitura dell'offerta
- **Run.** Servizi, opzioni e funzionalità necessarie al funzionamento
- **Manage.** Monitoraggio e supporto costante, ripristino dell'offerta
- **Learn.** Apprendimento continuo e feedback basato sui risultati degli esperimenti



Figura 3. Le sei fasi della metodologia IBM Cloud Garage

Think: Valutazione continua della maturità DataOps e allineamento agli obiettivi aziendali

DataOps può essere trasformativa per un'organizzazione esistente e per processi consolidati. La funzione di DataOps è quella di automatizzare molte delle attività manuali esistenti e razionalizzare il processo di creazione delle pipeline di dati. Sia che si tratti di avviare o di sostenere una prassi di DataOps, è importante valutare la capacità del team di fornire rapidamente dati business-ready e di elaborare un piano di miglioramento in linea con la creazione del valore aziendale.

Il successo di [DataOps](#) inizia dalla [catalogazione](#) delle risorse dati mediante l'acquisizione dei metadati e l'assegnazione di policy alle classi di dati, la valutazione e l'attribuzione di punteggi di [qualità dei dati](#), e l'utilizzo di strumenti di [integrazione dati](#) anziché fogli di calcolo, la conoscenza "tribale" o la codifica manuale. Una volta definita la maturità del team, gli obiettivi diventano quelli di accrescere funzionalità e potenzialità su più aspetti di DataOps possibile.

I team di DataOps devono concentrarsi sull'allineamento della distribuzione dei dati con il valore che essi possono apportare all'azienda. Occorre domandarsi: *Quanto si potrebbe risparmiare o guadagnare se questo dato fosse disponibile più rapidamente?*

Code: Utilizzo di un sistema di controllo versione: gestione controllo sorgente

Una pipeline dati è il codice sorgente responsabile della conversione di contenuto non elaborato in informazioni utili. Tale pipeline è fondamentale nella data analytics e può essere automatizzata end-to-end per produrre un codice sorgente che può essere utilizzato in forma riproducibile. File, configurazioni e parametri associati all'analytics vengono spesso distribuiti in diverse parti e ambienti di un'azienda senza un controllo regolamentare, il che determina un'implementazione incoerente. Uno strumento di controllo versione come GitHub aiuta ad archiviare e gestire tutte le modifiche al codice e alla configurazione. Una repository centralizzata, inoltre, aiuta le aziende a disporre di informazioni coerenti e affidabili in ogni circostanza, su diversi ambienti, inclusa l'eventualità di emergenza con ripristino affidabile. Il controllo versione consente ai team di parallelizzare i propri sforzi di sviluppo e di essere agili nella delivery pipeline grazie al metodo branch e merge.

Per accertarsi che la pipeline di data analytics funzioni adeguatamente, essa va testata. Grazie alla CI/CD (integrazione e sviluppo continuo), unita alla parametrizzazione, può essere implementata e testata in modo totalmente automatico. Ad ogni fase della data analytics va applicato il testing di input, output e logica di business, nonché verificata in termini di accuratezza o potenziale deviazione, oltre agli errori o agli avvisi prima del rilascio per assicurare una qualità costante. Il testing manuale non va applicato alle aziende ad alte prestazioni, è soggetto a errori, è dispendioso in termini di tempo ed è laborioso. Una suite di test solida e automatizzata è un elemento importante per ottenere la CI/CD ed è essenziale in un'economia on-demand.

Deliver: Processo di DataOps e automazione del workflow: tecnologie dati

Per una metodologia di DataOps di successo, l'automazione è essenziale e richiede una pipeline di data analytics progettata con flessibilità di runtime. Un requisito chiave per l'ottenimento di dati attendibili e affidabili è una pipeline regolamentata e coerente basata sull'accurato trattamento dei dati, immissione, cataloghi e classificazione mediante metadati e tecniche di campionamento dati.

Una pipeline dati ripetibile e solida per la distribuzione di dati attendibili e governati richiede un meccanismo che consenta le seguenti attività:

- Definizione ed esecuzione congruenti di policy di governance e di riservatezza dei dati.
- Supporto a un efficiente spostamento di dati.
- Avvio correzione o best practice e template specifici per settore con glossari predefiniti.

Questo processo può implementare una pipeline di dati governati su piattaforme differenti e in modo coerente senza codici sorgente o modifiche alla configurazione, e fornire dati interamente governati e attendibili. Il processo di DataOps deve inoltre essere supportato mediante appropriati strumenti di correzione per gestire le eccezioni. La tracciabilità e la

verificabilità di qualsiasi modifica è un requisito de facto per queste pipeline dati governate.

IBM offre funzionalità innovative che includono machine learning (ML) integrato, automazione AI, una governance integrata e un importante catalogo per dati operativi continui e di alta qualità per tutta l'azienda. L'efficienza di DataOps dipende fortemente dall'automazione dei componenti di tecnologia dati utilizzati per la pipeline dati.

- IBM InfoSphere® and IBM Watson® Knowledge Catalog (WKC) possono rispondere a queste esigenze in modo efficiente, solido, automatizzato e ripetibile.
- IBM InfoSphere Information Server affronta le esigenze relative a trasferimento, pubblicazione e utilizzo dei dati all'interno di una pipeline, e contribuisce ad assicurare qualità dei dati e applicazione delle policy. Con una gestione efficiente del controllo codice sorgente, può essere automatizzato ed eseguito in modo efficiente nell'ambito della pipeline CI/CD.
- Il ML integrato all'interno di IBM WKC e IBM Information Server, completa il processo di automazione e lo ottimizza a ogni iterazione per una solida pipeline.
- IBM Cloud™ DevOps Insights aiuta a fornire insight e visualizzazione operativi per la pipeline dati. Contribuisce all'applicazione di misure di sicurezza e qualità che vengono costantemente monitorate, al rilevamento di qualsiasi variazione inattesa e alla generazione di statistiche operative basate sull'automazione e sull'integrazione personalizzata con IBM WKC e IBM Information Server.
- Apache Airflow e NiFi possono aiutare con la progettazione del flusso di lavoro e la sua orchestrazione.
- L'utilizzo dell'automazione mediante endpoint REST e parametrizzazione può aiutare a selezionare data set specifici o un ambiente in modo dinamico e alterare il comportamento senza impatto sul codice sorgente della pipeline, soddisfacendo le esigenze quotidiane dei professionisti di data analytics.

Run: Integrazione e implementazione continua Integrazione continua

Ingegneri e proprietari di pipeline dati possono effettuare aggiornamenti o modifiche alla pipeline in qualsiasi momento e mantenerla, nell'ambito del sistema di controllo versione, in forma di copia privata all'interno del ramo di sviluppo o del ramo privato. Diversi ingegneri possono operare in parallelo e fornire modifiche a sviluppo o private branch contemporaneamente e moltiplicare l'ottimizzazione della produttività. Quando le modifiche alla pipeline sono ultimate e testate all'interno del ramo, il codice sorgente può essere integrato nel codebase o trunk principale e distribuito alla linea di produzione. Nel caso in cui il codice integrato non funzioni, la pipeline dati può sempre tornare alla precedente versione funzionante del codice sorgente della pipeline. La ramificazione e la fusione consentono ai team di data analytics di svolgere i propri test in autonomia, effettuare modifiche, assumersi rischi e annullare se un set di modifiche risultasse inefficace.

Implementazione continua

I professionisti di data analytics richiedono dati rilevanti utilizzati da pipeline separate dalla copia privata del codice sorgente e un ambiente adeguato per eseguire tali pipeline. Operare direttamente sul database o sull'ambiente di produzione non è efficiente e porta spesso a conflitti. Per ridurre conflitti e dipendenze, le pipeline dati richiedono:

- Gestione efficiente del controllo codice sorgente
- Disponibilità di opzioni flessibili di deployment dell'ambiente
- Testing del comportamento dei dati

Jenkins Pipeline è uno strumento che funge da complemento alla pipeline di distribuzione IBM. Red Hat® OpenShift® offre una piattaforma di deployment ripetibile e congruente da utilizzare per la convalida di istanze simultanee di una pipeline dati con valori differenti forniti in fase di runtime.

Manage: Operare con deployment continui e costanti

I professionisti di data analytics vogliono evitare di implementare modifiche che interrompano l'attuale pipeline dati in produzione. Vi sono due workflow chiave per affrontare la situazione:

- **Pipeline del valore.** Creare valore per le aziende ora con flussi dati in produzione.
- **Pipeline dell'innovazione.** Considerare il futuro della nuova analytics che è in fase di sviluppo e viene aggiunta alla pipeline di produzione.

Entrambe queste pipeline si intersecano in produzione, dove le aziende di DataOps gestiscono l'orchestrazione dei dati verso la produzione e l'implementazione di nuove funzioni, mantenendo sempre una qualità impeccabile. Il controllo della pipeline dati, come il controllo di processo statistico che monitora i dati e le nuove pipeline di sviluppo, consente al team di sviluppo di implementare senza interrompere i sistemi di produzione. Con uno sviluppo agile e DevOps, la velocità della nuova analytics è massimizzata. Aiuta a ridurre al minimo il tempo e l'impegno necessari per trasformare un'esigenza aziendale in un'idea di analytics e rilasciarla come processo di produzione ripetibile e riutilizzabile.

Learn: Gestione comunicazione e processi

Una notifica efficiente e automatica è fondamentale per la comunicazione e per i processi di correzione nell'ambito di una metodologia di DataOps. Quando vengono effettuate modifiche a qualsiasi codice sorgente o quando la pipeline è triggerata, errata, completata o implementata, può generarsi una notifica. In caso di malfunzionamento, l'informazione può essere veicolata in modalità push insieme alla notifica per consentire di affrontare il problema. Il processo post correzione può essere attivato automaticamente per convalidare la pipeline, implementarla nella prossima fase e aggiornare la dashboard con le ultime informazioni e qualità dei dati. Strumenti come Slack, Apache Kafka, PagerDuty e Trello sono comunemente usati per facilitare la comunicazione, la collaborazione, l'acquisizione dei feedback e la condivisione fra stakeholder nell'ambito della toolchain di DataOps.

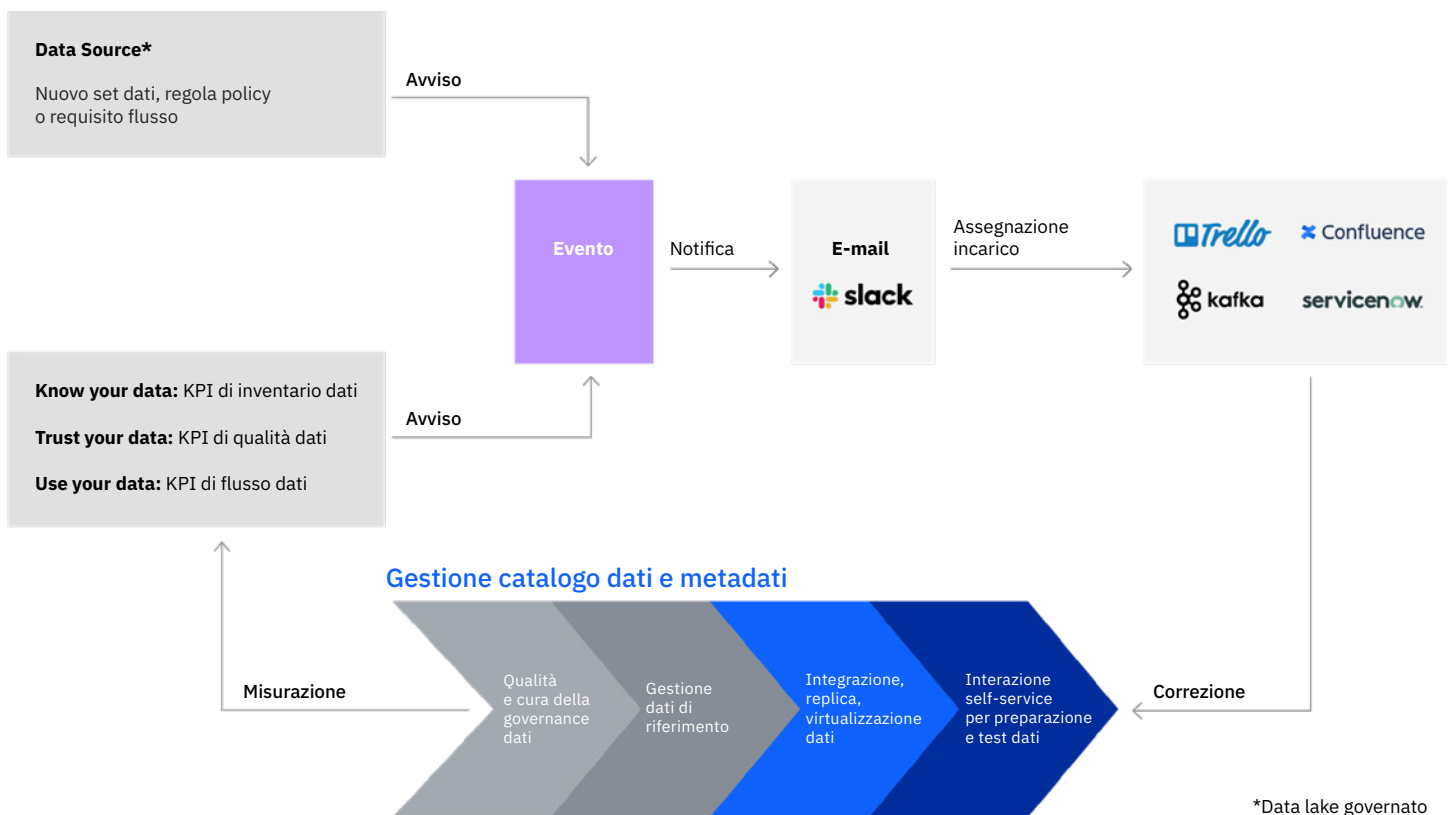


Figura 4: Visualizzazione della gestione di comunicazione e processi all'interno di un ambiente data lake governato

L'impatto di una prassi di DataOps di successo

Adottando la metodologia di DataOps, un operatore nel retail ha ottenuto miglioramenti in tutte le pipeline dati, in modo tale da rendere le modifiche ai dati applicabili in meno di due minuti nell'intera azienda, mentre prima richiedevano tre settimane. Di conseguenza l'azienda ha potuto utilizzare dati business-ready per effettuare analisi di affinità clienti in meno di un giorno: un processo che prima avrebbe precedentemente richiesto 20 giorni. In più, ha ridotto la quantità di tempo necessaria per segnalare la situazione delle giacenze a un sesto del tempo.

Tra i criteri per una prassi di DataOps di successo ricordiamo:

1. **Creare un data office.** Questo processo implica una netta definizione dei confini del ruolo nel fornire i dati come risorsa, identificare gli stakeholder principali, comprendere gli impegni che ogni stakeholder nella pipeline dati si assume nei confronti di un'operatività e cultura collaborativa.
2. **Allineamento con gli obiettivi dell'azienda.** Per rimanere competitivi, occorre rispondere rapidamente alle nuove opportunità di mercato utilizzando un approccio informato e basato su dati. In altre parole, a meno che non vi sia una comunicazione eccellente fra business e data delivery, i leader aziendali sanno che la propria organizzazione non prospererà.
3. **Far crescere il valore dei dati.** In qualsiasi iniziativa basata sui dati, i responsabili devono far sì che i dati prodotti possano essere utilizzati e riutilizzati continuamente, e che il loro valore cresca ogni volta. Tale risultato può essere ottenuto unicamente se è condiviso a livello centrale, se è ricercabile e se è allineato con il linguaggio aziendale.



Conclusioni

Le organizzazioni che hanno implementato DataOps con successo sanno a quali risorse dati hanno accesso, confidano nel significato dei dati e nella loro qualità, e usano i dati sfruttandone il massimo potenziale. I dati hanno valore se sono business-ready, attendibili, e contribuiscono a favorire insight differenziati, eccellenza operativa, collaborazione e vantaggi competitivi.

Per adottare una [prassi di DataOps](#) è necessario:

- *Investire nella comprensione delle competenze esclusive e delle sfide specifiche dell'organizzazione attivando un progetto pilota*
- *Utilizzare il successo del progetto pilota per espandere e far crescere le capacità di DataOps e l'azienda stessa*
- *Promuoverne il successo per affidare a più team la partecipazione alla metodologia di DataOps*
- *Condividere gli apprendimenti e iniziare a costruire un centro di eccellenza DataOps*

Prevedere i passi successivi, programmare il proprio IBM DataOps Garage Workshop, accelerando il percorso verso i dati business-ready contattando Siro.Corriga@ibm.com.

Le organizzazioni scoprono di essere ben posizionate per l'adozione di DataOps se hanno già lavorato a un catalogo dati, a un data lake o a iniziative su master data. Scopri di più sul supporto DataOps con tecnologie leader su ibm.com/it-it/analytics/dataops.



© Copyright IBM Corporation 2020
IBM Italia S.p.A.
Circonvallazione Idroscalo
20054 Segrate (Milano)
Italia

Prodotto negli Stati Uniti d'America
gennaio 2020

IBM, il logo IBM, ibm.com, IBM Cloud, IBM Cloud Garage, IBM Watson e InfoSphere sono marchi di International Business Machines Corp., registrati in diversi Paesi del mondo. Altri nomi di prodotti e servizi potrebbero essere marchi di proprietà di IBM o di altre società. Un elenco aggiornato dei marchi IBM è disponibile sul web come "Copyright and trademark information" alla pagina ibm.com/legal/copytrade.shtml

Microsoft e Windows sono marchi di Microsoft Corporation negli Stati Uniti, in altri Paesi, o entrambi.

Red Hat e OpenShift sono marchi registrati di Red Hat, Inc. o delle sue affiliate negli Stati Uniti e in altri Paesi.

Le informazioni contenute nel documento sono aggiornate alla data della prima pubblicazione e potrebbero essere modificate da IBM senza alcun preavviso. Non tutte le offerte sono disponibili in tutti i Paesi in cui IBM opera.

Gli esempi citati relativi a clienti sono presentati unicamente a scopo illustrativo. Gli attuali risultati in termini di performance possono variare a seconda delle specifiche configurazioni e delle condizioni operative. LE INFORMAZIONI FORNITE NEL PRESENTE DOCUMENTO SONO DA CONSIDERARSI "NELLO STATO IN CUI SI TROVANO", SENZA GARANZIE, ESPLICITE O IMPLICITE, IVI INCLUSE GARANZIE DI COMMERCIALIZZABILITÀ, DI IDONEITÀ PER UN PARTICOLARE SCOPO E GARANZIE O CONDIZIONI DI NON VIOLAZIONE. I prodotti IBM sono coperti da garanzia in accordo con termini e condizioni degli accordi che li regolamentano.

Il cliente è responsabile della conformità con leggi e regolamenti applicabili. IBM non fornisce consulenza legale né dichiara o garantisce che i propri servizi e prodotti assicureranno che il cliente sia conforme a qualsivoglia legge o regolamento.

Dichiarazione di conformità alle procedure di sicurezza: La sicurezza dei sistemi IT richiede la protezione di sistemi e informazioni tramite prevenzione, identificazione e risposta agli accessi impropri di origine interna o esterna alle aziende. L'accesso improprio può causare l'alterazione, la distruzione, l'appropriazione indebita o l'uso improprio delle informazioni; può inoltre provocare danni e uso improprio dei sistemi, che possono essere utilizzati per attaccare altri sistemi. Nessun prodotto o sistema IT può essere considerato completamente sicuro e nessun prodotto, servizio o misura di sicurezza è del tutto efficace nel prevenire l'uso o l'accesso improprio. Sistemi, prodotti e servizi IBM sono progettati come elementi di un approccio di sicurezza completo, nel rispetto delle normative, che richiederà necessariamente procedure operative aggiuntive e il probabile impiego di altri sistemi, prodotti o servizi per raggiungere la massima efficienza. IBM NON GARANTISCE IN ALCUN MODO CHE SISTEMI, PRODOTTI O SERVIZI SIANO ESENTI DA O RENDERANNO LE AZIENDE ESENTI DA CONDOTTA MALEVOLA O ILLEGALE DI UNA QUALSIASI PARTE.

1 "2019 Global data management research: Taking control in the digital age." Experian, 2019.

2 Jarah Euston, "The DataOps Trend is Real: 73% of Companies Plan to Invest in DataOps to Manage Data Teams in 2018," Nexla.

10028810USEN-01

Appendice: Modulo programma pilota di DataOps

Nome progetto:

Data:

Reparto o business unit:

Responsabili del programma pilota:

Nome e cognome	Ruolo	E-mail	Telefono

Stakeholder estesi in tutta l'azienda:

Nome e cognome	Ruolo	E-mail	Telefono

Definizione del problema:

Checklist causa principale:

Sfida	Applicabile? S/N	Note aggiuntive

Metriche per la riuscita:

Data inizio:

Data finale sprint:

Valutazione:

Implementazione	Stato attuale	Risultato desiderato dello sprint	Azioni da intraprendere per ottenere il risultato desiderato
Immissione risorse dati, rilevamento automatico e classificazione			
Valutazione della qualità dei dati e correzione			
Assegnazione termini aziendali			
Riservatezza dati, conformità normativa, definizione policy aziendale e applicazione			
Elaborazione richiesta definizione requisiti dei consumatori di dati			
Comunicazione e notifica richiesta dati, incluse gestione e correzione eccezioni ed errori			
Pubblicazione dati curati nel catalogo			
Derivazione dei dati e reporting			
Collaborazione, feedback e audit			

Domande campione da porre durante un audit di implementazione:

- Immissione risorse dati, rilevamento automatico e classificazione
- Eseguiamo repliche ad alto volume e bassa latenza ai fini della business continuity?
 - Utilizziamo analytics di flusso avanzata per analisi in tempo reale a bassa latenza?
 - Ci colleghiamo facilmente a qualsiasi fonte di dati ed eseguiamo trasformazioni e integrazioni di dati complesse?
 - Siamo in grado di fornire dati da social media, dati meteo o altre fonti di dati in cloud pubblico?
 - I nostri data consumer dispongono di accesso in tempo reale ai nostri repository di metadati da qualsiasi applicazione desktop?
 - I nostri data consumer dispongono di un accesso in tempo reale al nostro catalogo dati che consenta self-service e assistenza nel rilevamento di set dati pertinenti al loro lavoro?
 - Utilizziamo strumenti di profilazione dati per comprendere i dati, convalidarne i valori, i rapporti fra tabelle e colonne e trovare e analizzare eventuali anomalie?
 - La nostra gestione delle regole di business è integrata con la nostra infrastruttura di gestione dei metadati?