



Cinque miti sul data lake

Quando le aziende iniziano a rendersi conto dell'importanza dei propri dati e della loro capacità di aprire nuove opportunità e modelli di fatturato, si trovano davanti a un oceano di tecnologie che offrono tutte l'approccio "migliore" alla gestione di tutti questi dati al fine di ricavarne informazioni importanti.

Se al momento il data lake è un'architettura di riferimento sotto i riflettori, un motivo c'è. Tuttavia, esistono diversi miti sul data lake che portano le aziende e i responsabili IT a perdere tempo prezioso durante la scelta della soluzione che meglio si adatta alle proprie esigenze.

1.

Il data lake può essere distribuito in un solo cloud → Falso

Il data lake non è limitato a una sola posizione o un solo cloud. Non è limitato a una distribuzione on-premise. È possibile creare un data lake in più cloud con opzioni ibride. Un data lake è un'architettura di riferimento che è indipendente dalla tecnologia. Si tratta di un approccio ideale per quelle organizzazioni che vogliono mettere i dati al centro delle proprie operazioni al fine di facilitare l'accesso a una serie di tipi di dati su vasta scala e offrire agli utenti la possibilità di analitica self-service.

2.

Hadoop è l'unico data lake → Falso

Sebbene il termine data lake sia spesso associato a Hadoop, o a un archivio di oggetti orientato ad Hadoop, è possibile sviluppare e utilizzare un data lake in modo efficace senza incorporare Hadoop. Ad esempio, un data lake efficace potrebbe essere basato su diversi sistemi relazionali di gestione del database. Un data lake combina una serie di tecnologie per costruire sistemi di informazioni utili e fornire così un'esplorazione dei dati agile che consenta agli scienziati dei dati di soddisfare le esigenze aziendali.

3.

È possibile aggiungere data lake per il dump di tutti i dati, senza la necessità di governance → Falso

Sebbene software e hardware siano componenti fondamentali di una soluzione di data lake, altrettanto importante è la catalogazione, la qualità e la governance dei dati, nonché i processi di gestione.

Così come i data warehouse sono diventati degli enormi buchi neri che risucchiano gigantesche quantità di dati, senza opportune politiche di governance un data lake può diventare una palude di dati.

4.

La misura dell'efficacia dei data lake è l'accesso → Falso

Il dump dei dati in una posizione centrale non è una vera soluzione di analitica. L'obiettivo è eseguire analisi dei dati che restituiscano informazioni commerciali significative, nonché scoprire nuovi flussi di fatturato, modelli di ritenzione dei clienti o estensioni dei prodotti.

Ma quei dati devono essere affidabili, pertinenti e disponibili per tutti i consumatori di dati. Un data lake richiede una catalogo di metadati intelligente che sia legato alla terminologia aziendale: una contestualizzazione che consente di rendere più comprensibili tutti quei dati codificati che in genere risultano criptici. Inoltre un data lake assegna ai dati asset informativi – siano essi strutturati o meno – e infrastruttura di governance in base all'origine e alla qualità per garantire protezione, standardizzazione, gestione efficiente e affidabilità delle informazioni.

5.

Il data lake è un sostituto dei data warehouse → Falso

I data lake possono incorporare diversi enterprise data warehouse (EDW), insieme ad altre origini dei dati, come quelli provenienti da social media o IoT. Tutte queste informazioni confluiscono nei data lake dove è possibile integrare la governance, agevolando un rilevamento dei dati affidabile per gli utenti all'interno dell'organizzazione.

Pertanto, un data lake potenzia gli ambienti EDW per consentire agli scienziati dei dati e agli analisti di esplorare in modo semplice i propri dati e scoprire nuove prospettive, informazioni e opportunità per accelerare l'innovazione e la crescita aziendale.



I vantaggi dei data lake multicloud

Evitate di impantanarvi nella palude delle informazioni fuorvianti; un data lake governato può fornire, all'interno di un'azienda, l'accesso a una vasta gamma di dati strutturati e non strutturati, contribuendo a garantirne così affidabilità e sicurezza, senza eccezioni.

Se calibrato sulle esigenze aziendali, un data lake governato può accelerare l'analisi e migliorare l'accuratezza delle informazioni in quanto:



I dati poggiano sulle fondamenta sicure e affidabili dell'infrastruttura.



I feed di dati controllati compilano il data lake con informazioni affidabili per poi documentare gli asset informativi, i relativi metadati e il contesto aziendale fornendo un flusso di dati in tempo reale nel data lake.



La qualità, l'origine e la derivazione dei dati sono chiare.



I dati vengono espressi nel linguaggio aziendale per consentire agli scienziati dei dati di mettersi subito al lavoro anziché perdere tempo prezioso a cercare di capire il significato di termini criptici.



I dati vengono classificati, protetti e governati in modo adeguato.

La verità è sotto gli occhi di tutti

Il [portafoglio di gestione dei dati IBM](#) è riuscito ad aiutare i clienti a evitare le trappole e a sfatare i miti sul data lake, aiutandoli a superare le fasi critiche nel percorso verso un'implementazione efficace di data lake governati.

Scopri le soluzioni di data lake IBM

Prendi decisioni più smart sfruttando diversi tipi di dati da un maggior numero di origini.



Leggi l'ebook

Scopri in che modo i data lake governati creano opportunità per ottenere informazioni commerciali strategiche.

