

GOVERNANCE UNIFICATA DEI DATI CON AZURE



Come ottenere dati coerenti
e affidabili per decisioni
realmente data driven

SOMMARIO

LA DEFINIZIONE	PAG. 3
I VANTAGGI	PAG. 5
LE SFIDE	PAG. 6
FRAMEWORK PER LA GOVERNANCE DEI DATI	PAG. 7
LA GOVERNANCE DEI DATI IN AZURE	PAG. 9
MICROSOFT PURVIEW	PAG. 10
COME FUNZIONA MICROSOFT PURVIEW	PAG. 11
PROCEDURE CONSIGLIATE PER LA DISTRIBUZIONE: INDIVIDUARE OBIETTIVI E TRAGUARDI	PAG. 12
RACCOGLIERE DOMANDE	PAG. 13
MODELLI DI DISTRIBUZIONE	PAG. 13
CREARE UN PROCESSO PER PASSARE ALLA PRODUZIONE	PAG. 14
CICLO DI VITA DELLA PIATTAFORMA	PAG. 14
CONCLUSIONI	PAG. 15

LA DEFINIZIONE

In un contesto in continua evoluzione dove è sempre più importante avere informazioni puntuali e precise in tempi rapidi su una base dati potenzialmente infinita, la governance dei dati diventa fondamentale per rispondere a questa esigenza. La governance dei dati è prima di tutto un processo che gestisce la disponibilità, l'usabilità, l'integrità e la sicurezza dei dati nei sistemi aziendali, che si basa su standard e policy applicate sui dati. Una governance dei dati efficace garantisce che questi siano coerenti ed affidabili e non vengano utilizzati in modo improprio, seguendo le nuove normative sulla privacy dei dati.



La governance dei dati non può prescindere da un programma che gestisce i data silos attraverso azioni collaborative con gli stakeholder delle diverse unità aziendali disconnesse.

Questo programma dovrà eseguire operazioni complesse e strutturate in modo che i dati organizzati siano messi a disposizione nei sistemi in modo accurato ed usati correttamente.



I VANTAGGI

Un obiettivo chiave della governance dei dati è quello di **ridurre i data silos** presenti in un'organizzazione.

I data silos si accumulano quando le singole unità aziendali implementano sistemi di elaborazione delle transazioni separati senza un coordinamento centralizzato.

La **governance dei dati tende ad armonizzare i dati in tali sistemi** attraverso un processo collaborativo, con la partecipazione delle parti interessate delle varie unità aziendali.

La governance dei dati deve garantire che **i dati vengano utilizzati in modo corretto**, sia per evitare di introdurre errori nei sistemi sia per bloccare il potenziale uso inappropriato dei dati personali su informazioni sensibili. Ciò può essere ottenuto creando **policy**

sull'uso dei dati, insieme a **procedure per monitorarne l'utilizzo**, permettendo di trovare un equilibrio tra le pratiche di raccolta e di utilizzo dei dati e gli obblighi in materia di privacy.

Oltre ad analisi più accurate e una maggiore conformità alle normative, i vantaggi offerti dalla governance dei dati includono:

- **migliore qualità,**
- **minori costi di gestione,**
- **maggiore accesso ai dati** necessari per data scientist, data analyst ed altri utenti aziendali.

La governance dei dati, quindi, può **aiutare a migliorare il processo decisionale aziendale** fornendo ai data consumer informazioni migliori, portando quindi **vantaggi competitivi per l'organizzazione.**

LE SFIDE

Il percorso per ottenere i vantaggi citati in precedenza non è semplice e si devono tenere in considerazione alcuni aspetti fondamentali.

Tra le principali sfide troviamo quella di mancanza di leadership nei dati.

Le organizzazioni stanno valutando la necessità che ci siano **persone responsabili dei propri dati**, al di là della tecnologia necessaria per sfruttarli e proteggerli.

Questo è il motivo per cui il **Chief Data Officer (CDO)** si sta facendo strada come ruolo fondamentale all'interno delle organizzazioni.

Un'altra sfida rilevante è la mancanza di fiducia e ownership dei dati.

La **qualità dei dati è la chiave per ottenere informazioni aziendali fruibili e risultati migliori**. Dati accurati, affidabili e coerenti sono importanti per **creare fiducia ed incoraggiare gli utenti ad essere innovativi** e scoprire nuovi modi per trasformare i dati in valore.

Diversamente, **quando le parti interessate diffidano dei dati, influenzano negativamente** il processo di reporting e analisi.





FRAMEWORK PER LA GOVERNANCE DEI DATI

Un **framework di governance dei dati** è costituito da **politiche, regole, processi, strutture organizzative e tecnologie** che vengono applicate come parte di un programma di governance, dove sono definiti gli obiettivi e come verrà misurato il suc-

cesso, nonché le responsabilità decisionali e la responsabilità per le varie funzioni che faranno parte del programma stesso.

Il programma di governance di un'organizzazione **deve essere documen-**

tato e condiviso internamente, in modo che sia chiaro a tutti i soggetti coinvolti, in anticipo, quale sarà il suo funzionamento.

Parlando di **tecnologia**, il software di governance dei dati può essere utilizzato per **automatizzare la gestione di un programma di governance**.

Gli strumenti di governance dei dati supportano la **gestione del programma e del flusso di lavoro**, la **collaborazione**, lo **sviluppo di policy** di governance, la **documentazione dei processi**, la **creazione di cataloghi di dati** e altre funzioni rilevanti.

Possono anche essere utilizzati in combinazione con strumenti di data quality, gestione dei metadati e gestio-

ne dei dati master (MDM).

Oltre ai processi e alla tecnologia, nel processo di governance dei dati devono essere **coinvolte diverse persone**, come dirigenti aziendali, professionisti della gestione dei dati e personale IT, nonché utenti finali che hanno familiarità con i domini di dati rilevanti nei sistemi di un'organizzazione.

Oltre ai data architect, i data modeler, i data analyst e i data engineer che fanno parte del processo di governance, hanno un ruolo importante **gli utenti aziendali e i team di analisi** che devono essere formati sulle politiche di governance dei dati e sugli standard dei dati, in modo che possano evitare di utilizzare i dati in modo improprio.



LA GOVERNANCE DEI DATI IN AZURE

I principali analisti di mercato citano le seguenti **capacità** come **disponibili in alcuni o tutti i fornitori presenti sul mercato**:



Repository di metadati (contiene **informazioni** su colonne, tabelle, viste, campi, file e **processi** utilizzati nei sistemi IT);



Business Glossary (contiene **termini commerciali** e relative **definizioni**, politiche, **procedure**);



Data Lineage (mostra il **movimento dei dati** dall'origine al loro utilizzo);



Impact Analysis (utilizza le informazioni sulla derivazione dei dati per mostrare **l'impatto** del **cambiamento** a monte ed a valle del consumo di dati);



Rules Management (**automatizza** l'applicazione delle regole e dei processi aziendali in relazione ai dati);



Semantic Frameworks (organizza i business glossary ed i data dictionary in strutture gerarchiche).

Uno dei vendor più rilevanti nel panorama della gestione ed organizzazione dei dati è **Microsoft**.

I servizi che mette a disposizione nel cloud con **Azure** per realizzare una **Modern Data Platform** sono diversi e coprono tutte le caratteristiche richieste che vanno dall'**ingestion del dato** (Azure Data Factory), allo **storage orientato alle analisi** (Azure Data Lake ed Azure Synapse), agli **advanced analytics** (Azure Machine Learning) ed alla **data governance** (Microsoft Purview).



MICROSOFT PURVIEW

Microsoft Purview è uno di questi servizi ed offre una **soluzione di governance dei dati unificata** per aiutare a gestire e governare i dati locali, multicloud e servizi SaaS.

È possibile **creare una mappa aggiornata dell'intero sistema dati** con tutte le **caratteristiche richieste ad una piattaforma di questo tipo**, come il **rilevamento automatico**, la **classificazione dei dati sensibili** ed il **lineage dei dati end-to-end**, per generare **informazioni dettagliate** su come vengono archiviati ed utilizzati i dati.

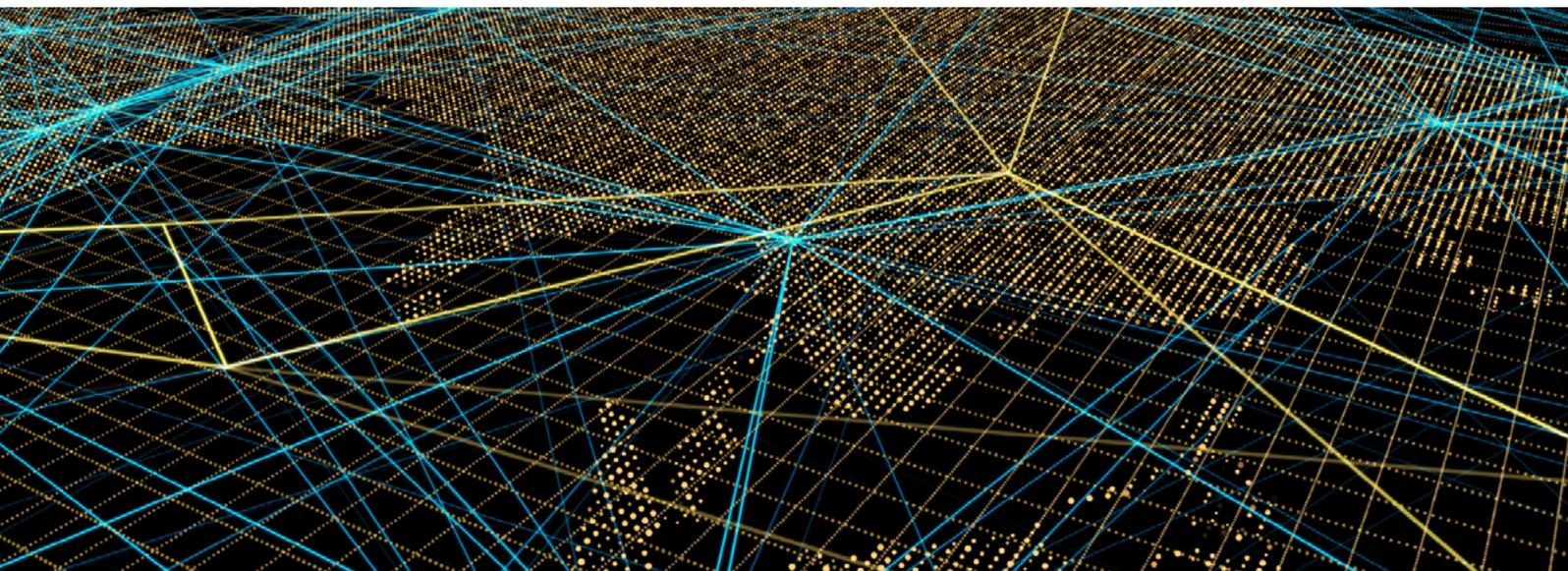
Il prodotto ha diversi componenti:

- **Microsoft Purview Data Map** è possibile creare una mappa unificata dei dati dell'intero patrimonio dati per stabilire le basi di una governance ed un utilizzo efficace dei dati stessi.
- **Microsoft Purview Data Catalog** i dati sono facilmente individuabili, massimizzando quindi il valore aziendale della gestione dei dati per i data consumer.
- **Microsoft Purview Data Estate**

Insights si ottiene una visione d'insieme dei dati sensibili per una comprensione completa delle attività di gestione dei dati.

- **Microsoft Purview Data Sharing** utile per condividere i dati all'interno o tra organizzazioni senza dover effettuare duplicazioni.
- **Microsoft Purview Data Policy** consente l'accesso ai dati segmentato ed in modo intuitivo per i diversi utenti che devono utilizzare i dati.





COME FUNZIONA MICROSOFT PURVIEW

Creando una directory per tutti i dati, chiamata **mappa dei dati** (Microsoft Purview Data Map), raccoglie e compila il patrimonio di dati in modo che sia più gestibile e possa essere **facilmente scansionato, scoperto e classificato**.

La mappa dei dati è la base per tutte le operazioni di Microsoft Purview poiché, al di sopra di essa, operano **le app** che **lavorano per creare un ambiente ottimizzato per l'archiviazione, l'accesso e l'acquisizione di informazioni sui dati**.

Il **Data Catalog** (Microsoft Purview Data Catalog) opera sulle basi fornite dalla mappa dei dati e funge da **catalogo per tutte le banche dati**. La

classificazione e l'ordinamento sono gestiti in base a parametri e variabili come termini del glossario, etichette di riservatezza e altro. L'**ordinamento e l'archiviazione dei dati con i diversi filtri** semplifica la ricerca.

A differenza del Data Catalog, l'**applicazione degli insights (Microsoft Purview Data Estate Insights)** funziona al di fuori della mappa dei dati. Fornisce una **situazione completa della proprietà dei dati** insieme ad indicatori ed informazioni, che può essere utilizzata per produrre risultati tangibili. Una caratteristica utile di questi indicatori è che **l'applicazione funziona automaticamente da sola per generarli** e fornirli mentre gli

utenti lavorano e consente quindi, principalmente a chief data officer o data manager, di prestare attenzione alle informazioni invece di perdere tempo per la creazione manuale di report.

È possibile condividere i dati (Microsoft Purview Data Sharing) in modo sicuro e protetto con utenti interni o esterni all'organizzazione. Una caratteristica fondamentale relativa alla sicurezza è che **il fornitore di dati può coordinare, monitorare**

e controllare la condivisione dei dati. Ciò consente di limitare l'accesso in qualsiasi momento se si rilevano attività sospette.

Il **servizio di gestione delle policy (Microsoft Purview Data Policy)** consente di organizzare l'accesso a diverse data source nell'intero patrimonio di dati. Le policy di accesso ai dati possono essere applicate tramite Purview sui sistemi di dati che sono stati registrati per la policy.



PROCEDURE CONSIGLIATE PER LA DISTRIBUZIONE: INDIVIDUARE OBIETTIVI E TRAGUARDI

Il primo passaggio è quello di **individuare obiettivi e traguardi.**

Molte organizzazioni hanno iniziato il loro programma di governance dei dati **sviluppando soluzioni individuali** che soddisfano requisiti specifici di gruppi isolati e domini di dati all'interno dell'organizzazione e possono trovare **difficile mantenere controlli e policy coerenti per questi tipi di soluzioni.**

Alcuni degli **obiettivi comuni che vengono identificati nelle fasi iniziali** per creare un'esperienza comple-

ta di governance dei dati **sono:**

- la **massimizzazione del valore aziendale dei dati;**
- **l'abilitazione di una cultura dei dati** in cui i data consumer possono facilmente trovare, interpretare e fidarsi dei dati;
- **aumentare la collaborazione tra le varie unità aziendali** per fornire un'esperienza di dati coerente;
- **diminuire il tempo per scoprire i dati** tramite opzioni self-service.



RACCOGLIERE DOMANDE

Nella **seconda fase si devono raccogliere domande dai gruppi di lavoro**. È fondamentale raccoglierle per elaborare un piano utile ad affrontare tutte le criticità ed i dubbi.

Le domande che possono essere fatte sono di diversa natura, come “Quali sono le principali fonti di dati e sistemi di dati nell'organizzazione?” oppure “Chi utilizzerà Microsoft Purview e quali ruoli avrà?”. Per garantire il successo dell'implementazione di Microsoft Purview per l'intera organizzazione, è importante **coinvolgere tutte le parti interessate**.

Solo poche persone sono coinvolte nella fase iniziale, tuttavia, man mano che l'ambito si espande, si aggiungeranno più persone per contribuire al progetto e fornire feedback.



MODELLI DI DISTRIBUZIONE

Vanno poi **disegnati i modelli di distribuzione** che dovranno tenere conto dei gruppi dai lavoro.

Se ci fosse solo un piccolo gruppo che utilizza Microsoft Purview con casi d'uso di consumo ridotti, l'approccio potrebbe essere quello di disporre di una sola istanza per servire l'intero gruppo.

Nella maggior parte dei casi comunque **dovrebbe esserci un solo account Microsoft Purview per l'intera organizzazione**, per sfruttare al massimo gli effetti della **collaborazione**, in cui il valore della piattaforma aumenta in modo esponenziale in funzione dei dati che risiedono all'interno della piattaforma stessa.



CREARE UN PROCESSO PER PASSARE ALLA PRODUZIONE

Dopo la definizione dei modelli si deve creare un **processo per passare alla produzione**.

Alcune organizzazioni potrebbero decidere di semplificare il processo lavorando con un'unica versione di produzione di Microsoft Purview. Tuttavia, la

maggior parte delle organizzazioni che desiderano distribuire Microsoft Purview in varie unità aziendali e che vorranno disporre di una qualche forma di processo e controllo, potranno **suddividere in quattro fasi differenti: progetto pilota, mvp, pre-produzione, produzione**.



IL CICLO DI VITA DELLA PIATTAFORMA

Ovviamente vanno fatte poi delle **considerazioni sul ciclo di vita della piattaforma**. In particolare, qui l'impatto è su **Microsoft Purview Data Map** dove ci sono diverse aree in cui gli amministratori del catalogo devono garantire la coerenza e le migliori pratiche di manutenzione durante il ciclo di vita:

- **Data Asset** (le origini dati dovranno essere nuovamente analizzate in tutti gli ambienti).
- **Regole di scansione** (raccolta di regole assegnate a scansioni specifiche come il tipo di file e le classifica-

zioni da rilevare).

- **Custom classification** (le classificazioni potrebbero non cambiare regolarmente).
- **Glossario** (è possibile esportare e importare i termini del glossario tramite l'UX).
- **Gestione delle policy** (manutenzione delle regole dei set delle sorgenti dati).
- **Assegnazione dei ruoli** (manutenzione degli accessi e dei ruoli).

CONCLUSIONI

Le funzionalità fornite da Microsoft Purview lo rendono un **componente vitale dell'ambiente cloud Azure** di ogni organizzazione.

Le funzionalità di **rilevamento e tracciabilità dei dati** consentono ai team di **operare in modo più efficiente ed eliminare i tempi e i costi** associati

alla creazione di dashboard e pipeline di dati ridondanti.

Microsoft Purview è progettato per aiutare le funzioni aziendali a **superare le criticità descritte nella governance dei dati e migliorare l'utilizzo degli asset informativi** con conseguente valorizzazione ed incremento dell'efficienza e della redditività dell'azienda.



L'AUTORE



Simone Bocchi

È “Data Platform Practice Director” in Horsa S.p.A.: con più di 15 anni di esperienza nel settore data & analytics, con rilevanti competenze in sviluppo business e project management. Il suo lavoro è nell’ambito della gestione di complesse architetture per avviare e migliorare il processo di digitalizzazione della gestione dei dati, dei processi e delle informazioni in aziende strutturate operanti in diversi settori.

Horsa[®] insight

Horsa Insight è la Business Unit di Horsa che si occupa di Advanced & Predictive Analytics. Un consulente di fiducia che accompagna e guida le aziende nel proprio percorso di data strategy.