

**APPENDICE 4 AL
CAPITOLATO TECNICO**

**DESCRIZIONE CONTESTO APPLICATIVO E TECNOLOGICO
LOTTO 1**

**GARA A PROCEDURA APERTA AI SENSI DEL D. LGS. 50/2016 E S.M.I., PER L'ACQUISIZIONE DI SERVIZI
PER LA MANUTENZIONE, EVOLUZIONE E GESTIONE DEI SISTEMI DI DATA WAREHOUSE E BUSINESS
INTELLIGENCE DEL MINISTERO DELL'ECONOMIA E DELLE FINANZE E DELLA CORTE DEI CONTI
ID 1973**

INDICE

1. MINISTERO DELL'ECONOMIA E DELLE FINANZE	4
1.1 INTRODUZIONE	4
1.2 DESCRIZIONE FUNZIONALITA'	4
1.2.1 Analisi e gestione attraverso i metadati	4
1.2.1.1 Caratteristiche della soluzione	4
1.2.1.2 Architettura	5
1.2.1.3 Evoluzione	5
1.2.2 Cruscotto di Monitoring Operativo	5
1.2.2.1 Caratteristiche della soluzione	5
1.2.2.2 Architettura	6
1.2.2.3 Evoluzioni	6
1.2.3 Dimensional Fact Model	7
1.2.3.1 Caratteristiche della soluzione	7
1.2.3.2 Architettura	8
1.2.3.3 Evoluzioni	10
1.2.4 Repository della qualita'	10
1.2.4.1 Caratteristiche della soluzione	10
1.2.4.2 Architettura	11
1.2.4.3 Processo	11
1.2.4.4 Evoluzioni	12
1.3 Master Data Management (MDM)	12
1.3.1.1 Caratteristiche della soluzione	12
1.3.1.2 Architettura	13
1.3.1.3 Evoluzioni	13
1.3.2 App Bilancio Aperto	14
1.3.2.1 Caratteristiche della soluzione	14
1.3.3 BDAP – BILAR (Bilanci Armonizzati)	14
1.3.3.1 Caratteristiche della soluzione	14
1.3.4 BDAP – GAD (Grafo Attributi Dimensionali)	14
1.3.4.1 Caratteristiche della soluzione	14
1.3.5 Open Data	15

Classificazione del documento: Consip Public

Gara a procedura aperta ai sensi del D. Lgs. 50/2016 e s.m.i., per l'acquisizione di servizi per la manutenzione, evoluzione e gestione dei sistemi di Data Warehouse e Business Intelligence del Ministero dell'Economia e delle Finanze e della Corte dei Conti - ID 1973

Appendice 4 al Capitolato tecnico - Descrizione contesto applicativo e tecnologico – Lotto 1



1.3.5.1	Caratteristiche della soluzione	15
1.3.6	Sistema di gestione delle forniture	15
1.3.6.1	GEFOD (Gestore delle Forniture Open Data)	15
1.3.6.2	GEPOD (Gestore Produzione e Trasporto Fornitura Open Data)	16
1.3.6.3	GEMOD (Gestore Metadati Forniture Open Data)	16
2.	CORTE DEI CONTI.....	18
2.1	INTRODUZIONE.....	18
2.2	DESCRIZIONE FUNZIONALITA'	18
2.2.1	Cruscotti Monitor Progetti	18
2.2.1.1	Caratteristiche della soluzione	18
2.2.2	Cruscotto Statistiche Utilizzo.....	18
2.2.2.1	Caratteristiche della soluzione	19
2.2.3	Cruscotto Monitoring Operativo	20
2.2.3.1	Caratteristiche della soluzione	20



1. MINISTERO DELL'ECONOMIA E DELLE FINANZE

1.1 INTRODUZIONE

Una importante categoria di sviluppi inerenti il Data Warehouse della RGS è quella dei cosiddetti software "trasversali".

Questi progetti hanno come obiettivo primario quello di realizzare sia dei sistemi indirizzati alle strutture organizzative e produttive del Data Warehouse, per agevolare la gestione di definiti aspetti operativi o metodologici di interesse comune all'interno dell'impianto del sistema, sia sviluppare sistemi a supporto dell'evoluzione normativa, dei quali il più eclatante è la BDAP (L.196/2009).

Tali sistemi software, che potranno essere realizzati all'occorrenza anche con tecnologie Object Oriented non essendo necessariamente di tipo conoscitivo, dovranno essere volti ad automatizzare, velocizzare ed ottimizzare la gestione sia del ciclo progettuale, sia il processo operativo del Data Warehouse, nonché favorire la migliore automatizzazione dei processi lavorativi dell'Amministrazione.

1.2 DESCRIZIONE FUNZIONALITA'

1.2.1 Analisi e gestione attraverso i metadati

1.2.1.1 *Caratteristiche della soluzione*

Tutti i metadati logici, di processo e tecnici sono integrati in un repository (Oracle) e resi disponibili attraverso query SQL, fogli xls o report BO.

In particolare, vengono collezionati tutti i metadati relativi al software realizzato, a quello che giornalmente gira e produce informazioni e a tutti i processi ed eventi interessati.

L'esecuzione dei Job è controllata da un piano di schedulazione gestito da un apposito prodotto (TWS). TWS schedula tutte le operazioni necessarie all'esecuzione del processo di alimentazione e pubblicazione dell'informazione verso l'utente. Tali operazioni sono classificabili secondo le seguenti tipologie:

- Procedure per l'estrazione dei dati dai sistemi sorgente;
- Procedure e batch per la predisposizione e l'alimentazione dell'area di staging;
- Batch di alimentazione di EDW e DMs;
- Procedure per i controlli di qualità;
- Procedure per il porting dati dall' ambiente di Back End a quello di Front End;
- Backup e svecchiamenti;
- Procedure per la manutenzione del DB (statistiche Oracle, drop e create di indici, ecc.)

La sequence (eseguibile di uno o più job) è l'elemento atomico sul quale possono essere posti vincoli in esecuzione (propedeuticità, stop su errore, dipendenza da un orario, ecc.) gestibili mediante schedatore.

Le attività necessarie al caricamento di una stessa tavola del DB o all'estrazione di dati dal DW per un utente "esterno", eseguite mediante il prodotto di ETL, sono raggruppate in una unica sequence. Una sequence è organizzata in unità elaborative (Job) che eseguono operazioni distinte sul target. Tutte le sessioni necessarie al popolamento del target sono nello stesso batch. Una sequence esegue uno ed un solo batch.



Per ogni batch sono definite le modalità di recovery mediante le quali è possibile ripristinare la configurazione dei dati precedente alla sua esecuzione. Se necessario viene predisposto un apposito script di recovery (undo).

Ogni batch è costruito in maniera tale che la sua mancata esecuzione o l'esecuzione della procedura di undo ad esso associata consenta la sua attivazione nel caricamento successivo senza ulteriori interventi manuali sui dati (nuovo requisito).

Le caratteristiche/requisiti del processo di pubblicazione, tenute presenti sono riassunti dalle considerazioni seguenti.

La pubblicazione dei dati di un Data Mart è organizzata per sottoinsiemi consistenti di informazioni (dati già per loro natura raggruppati in aree omogenee, ben identificabili, da noi definite "Aree di Pubblicabilità"). La caratterizzazione univoca di tali aree è sia

logica: un'Area di Pubblicabilità (AP) è costituita da tutte le informazioni necessarie a soddisfare il requisito di analisi dell'utente in merito ad una specifica tematica, che

applicativa: l'AP è costituita da tutte le informazioni accedibili mediante un "Universo".

L'analisi dei Metadati tecnici e di processo è orientata all'ottimizzazione del caricamento e della sua schedulazione, all'intercettazione, soluzione e monitoraggio degli errori dei caricamenti e provvede, fra l'altro, a gestire le situazioni di caricamenti parziali proprio nell'ottica di garantire la presenza e consistenza del dato per area di pubblicabilità. Esso fornisce, inoltre, adeguati strumenti per verificare gli esiti dell'esecuzione dei batch notturni per ciascuna Area di Pubblicabilità.

La fruizione del Metadato attraverso report BO si integra nel processo di fornitura e pubblicazione dell'informazione verso l'utente finale, lo arricchisce delle funzionalità necessarie a garantire una maggiore efficienza dei caricamenti, una migliore percentuale di successo degli stessi, un minor costo di gestione in caso di intervento manuale di ripristino per via del più alto numero di automatismi in fatto di recovery e ripartenze.

1.2.1.2 Architettura

Non è necessaria una architettura dedicata, in quanto tutte le informazioni provengono automaticamente attraverso appositi job schedulati, dalle stesse procedure di caricamento e dal repository del tool InfoSphere Information Server. La fruizione dei metadati è garantita dalla stessa architettura utilizzata dalla BI dei dati Front End.

1.2.1.3 Evoluzione

Le future implementazioni prevedono l'evoluzione dell'applicazione sia in termini di incremento delle funzionalità attualmente presenti sia come arricchimento grafico che permetta un più funzionale utilizzo dei metadati operativo/gestionali.

1.2.2 Cruscotto di Monitoring Operativo

Classe di rischio C, numero di utenti 30

1.2.2.1 Caratteristiche della soluzione

Il Data Warehouse RGS ha raggiunto nel tempo un'ampiezza rilevante per informazione gestita e moli di dati trattate ed una elevata complessità per numero di componenti che ne realizzano quotidianamente il processo di aggiornamento e messa in linea dell'informazione. La delicatezza e criticità assunte dalle attività di Data



Loading sono tali da rendere evidente la necessità di condurre in maniera organica e centralizzata le attività di monitoraggio per il processo di fornitura dei dati all'utente.

Il monitoraggio di elementi quali:

- prestazioni delle componenti applicative del sistema,
- stato di degrado dell'ambiente,
- risultati delle operazioni di caricamento e livello di qualità dei dati elaborati

richiede di una razionalizzazione e di una esecuzione sistematica e generalizzata su tutte le componenti interessate per essere efficace strumento di controllo e consentire l'intercettazione e la previsione di criticità potenzialmente bloccanti.

Il Cruscotto di Monitoring Operativo costituisce una soluzione applicativa che soddisfa l'esigenza di centralizzare e velocizzare gli interventi orientati alla fornitura del dato all'utente.

Lo strumento in oggetto si propone di rispondere alle esigenze descritte attraverso le seguenti funzionalità:

- Monitoring attraverso la valorizzazione di appositi indicatori e la produzione reportistica di dettaglio riguardante:
 - esiti di batch per l'alimentazione e procedure per il controllo della qualità dei dati;
 - tempi in esecuzione di tutte le componenti (Job ed Applicazioni TWS) del piano di schedulazione notturna e durata dell'intero piano;
 - spazi occupati per database interessati,
 - visualizzazione del trend storico delle metriche di qualità.
- Applicazioni di ausilio per il tracciamento delle attività:
 - gestione dell'agenda;
 - quaderno delle attività.

Le funzionalità presenti nel Cruscotto consentono la centralizzazione e la gestione attraverso un'unica interfaccia di molte delle attività al momento effettuate nell'ambito dell'operatività giornaliera (controllo esito caricamenti, verifica del livello di qualità dei dati processati, controllo esito refresh dell'ambiente utente, ecc...).

Il prodotto fornisce alle strutture preposte alla gestione gli elementi per intercettare più rapidamente le anomalie, prevedere e stimare le criticità, velocizzare gli interventi; il tutto avendo come obiettivo primario quello di favorire il data loading verso l'utente (massimizzando la disponibilità e la qualità dei dati) nel rispetto delle finestre temporali assegnate.

In tale ottica il cruscotto di monitoring operativo è la naturale prosecuzione di quanto già operato con la revisione delle politiche di caricamento e di controllo della qualità, per consentire pubblicazioni anche parziali delle informazioni.

Attraverso l'osservazione sistematica e oggettiva delle prestazioni della fase di aggiornamento dei dati del Sistema, si ha, inoltre, la possibilità di ottimizzare decidendo preventivamente gli interventi mirati ad ottenere il rispetto dei tempi di consegna ed aumentare la qualità del servizio.

1.2.2.2 Architettura

La fruizione dei metadati è garantita dalla stessa architettura utilizzata dalla BI dei dati Front End.

1.2.2.3 Evoluzioni

Le future implementazioni prevederanno l'evoluzione dell'applicazione sia in termini di incremento delle funzionalità attualmente presenti, sia come adeguamento architetturale per garantire migliori performance. Saranno inserite funzionalità di analisi sulla componente metadati nell'ambito del Front End.



1.2.3 Dimensional Fact Model

Classe di rischio C, numero di utenti 60

1.2.3.1 Caratteristiche della soluzione

Lo sviluppo del Data Warehouse RGS è organizzato in iterazioni guidate dalle esigenze informative degli utenti. L'impostazione incrementale è giustificata dall'impossibilità di definire preventivamente i requisiti conoscitivi nel loro complesso e dalla necessità di fornire nel minor tempo possibile i risultati all'utente.

Tali presupposti caratterizzano fortemente la gestione del ciclo di vita del SW che viene guidato da metadati di progetto come collante delle diverse fasi di un'iterazione e delle diverse iterazioni in cui viene suddiviso lo sviluppo del DW.

I metadati vengono prodotti, gestiti ed acceduti durante le attività di analisi, progettazione e sviluppo attraverso gli strumenti utilizzati dal gruppo di lavoro sulla base di definite linee guida metodologiche. In particolare la loro integrabilità è resa possibile dalla esistenza di un metamodello unico che collega i singoli metamodelli degli strumenti. Tale integrazione consente l'interazione fra attività progettuali appartenenti ad una medesima fase (attraverso i "collegamenti" stabiliti fra i prodotti di fase) e l'utilizzo di questi come input di nuove attività delle fasi successive.

La fase di Progettazione, finalizzata a definire le modalità di realizzazione di un progetto, è fortemente vincolata dai prodotti tecnologici scelti per lo sviluppo e l'esercizio del sistema (basi dati, ETL, front end, ...). Dal punto di vista metodologico ciò porta alla definizione di specifiche Best Practices tematiche, che devono tenere conto delle tecnologie adottate ed al tendenziale utilizzo dei prodotti anche per la formalizzazione e generazione della documentazione di progetto.

Viceversa per la fase di Definizione è possibile ed opportuno utilizzare linee guida metodologiche e strumenti che risultino indipendenti dalle scelte tecnologiche adottate. In tale fase si definiscono criteri e si operano scelte decisive sull'oggetto della realizzazione, per i quali l'aspetto centrale è la rispondenza alle esigenze individuate e la coerenza con quanto già in opera.

La crucialità della fase di Definizione viene amplificata nel caso del DW RGS dalla presenza di un unico attore, la Sogei, che deve gestire nel tempo l'evoluzione dell'intero sistema recependo le esigenze dell'utente, pianificando gli obiettivi progettuali, assegnandoli al Fornitore, verificando il loro collaudo e gestendo l'esercizio delle funzionalità realizzate.

Ad oggi la fase di Definizione viene condotta attraverso l'adozione di strumenti per l'analisi distinti, il cui uso è integrato in un processo regolamentato da linee guida. I metadati prodotti sono integrati "a posteriori" sulla base del metamodello definito.

Certamente la disponibilità di un ambiente unificato per l'analisi che, nel rispetto del processo e delle linee guida sperimentati, consente di realizzare l'integrazione "in fieri" dei metadati, aumenta la tempestività e garantisce l'organicità dei risultati della fase di definizione.

Il DFMCase è uno strumento di disegno che supporta l'attività di definizione di modelli dimensionali secondo il formalismo Dimensional Fact Model (DFM). Attraverso un'interfaccia utente "user friendly" consente di definire graficamente i modelli dimensionali utilizzando i costrutti fondamentali del formalismo DFM: gerarchie, misure, schemi di fatto e funzionalità.

La soluzione software proposta è costituita da un'applicazione CASE (Computer Aided Software Engineering) denominata DFMCase. Analogamente ad altri prodotti CASE come CA ERwin o SAP Sybase PowerDesigner, il DFMCase è un'applicazione "stand alone" che l'utente, tipicamente un analista o un progettista, utilizza in una o più fasi del ciclo di realizzazione/manutenzione del sistema.



Le principali attività supportate direttamente dal DFMCASE sono le seguenti:

disegno di modelli dimensionali DFM attraverso cui vengono formalizzati i requisiti di front-end e, a livello concettuale, la struttura dimensionale dell'informazione necessaria a rispondere a tali requisiti; la soluzione prevede un'organizzazione dei modelli dimensionali in progetti semanticamente coesi al loro interno (ad es. un progetto per ogni Data Mart);

produzione automatica di un documento (html) che raccolga in forma testuale e grafica i modelli dimensionali disegnati nella fase di specifica dei requisiti;

generazione di un file in formato XML che contenga i metadati prodotti attraverso il tool stesso affinché possano essere utilizzati, attraverso altri applicativi, nel ciclo di realizzazione del data warehouse; la grammatica secondo cui viene generato tale file è quella già in uso nel sotto sistema di gestione dei metadati realizzato nell'ambito del DW RGS.

Attraverso un successivo obiettivo progettuale, denominato Business Rules, il DFMCASE è stato aggiornato con nuove funzionalità al fine di descrivere in modo formale come il requisito utente, espresso mediante il modello dimensionale (DFM), debba essere soddisfatto attraverso i concetti che descrivono la realtà di interesse per l'iterazione, espresso mediante il modello Entità Relazione (E/R).

Il DFMCASE gestisce allo scopo le regole di business che descrivono a livello concettuale le regole che permettono di passare dalle informazioni modellate nello schema ER agli oggetti di analisi definiti nei diversi schemi di fatto; attraverso lo strumento l'utente, oltre ovviamente a disegnare gli schemi di fatto, può attualmente importare un modello concettuale ER disegnato con lo strumento ERWin v9.7 e definire, in accordo al linguaggio individuato, le regole di business.

1.2.3.2 Architettura

L'obiettivo è quello di procedere gradualmente verso un'evoluzione che preveda un processo di re-ingegnerizzazione del software esistente che specializzi il ruolo di ciascun componente separandolo quanto più possibile dal ruolo degli altri. Questo si otterrebbe tramite lo sviluppo e il perfezionamento del paradigma Model-View-Controller (MVC) che è oramai uno standard per le applicazioni object oriented e che la stessa Sun Microsystems ha cercato di evolvere nelle successive versioni della tecnologia Java 2 Standard Edition (J2SE) per applicazioni desktop.

L'architettura MVC si propone di distinguere i componenti software di un'applicazione in tre categorie o strati:

- il modello che rappresenta i dati trattati
- la vista che è la rappresentazione visuale dei dati
- il controllo che riceve gli input dell'utente dalla vista e li riporta nel modello.

Nella prima fase di sviluppo della soluzione si vuole porre un'enfasi particolare sulla evoluzione del modello. In particolare, come spiegato nei paragrafi precedenti, il DFMCASE v6.1 attuale utilizza archivi in formato XML per la persistenza dei dati, ma è evidente che si otterrebbe un significativo incremento di flessibilità e prestazioni laddove invece fosse utilizzato un DBMS relazionale, in particolare modo se si pensa ad una futura architettura multi-utente. L'evoluzione in questo senso che viene proposta si fonda sulle due attività descritte di seguito ed a ciascun rilascio successivo sarà possibile verificare il risultato raggiunto.

Definizione di un XML Schema che descriva la struttura e le regole di correttezza semantica degli archivi attualmente utilizzati. Ciò è necessario per definire in modo standard il "dizionario dei dati" e le regole di validazione dell'XML ed è vantaggioso per poter poi gestire il mapping tra XML ed oggetti Java utilizzando lo standard Java API for XML Binding (JAXB).



Adozione del pattern di disegno Data Access Object per i componenti del modello che consente poi di delegare ad un framework esterno la gestione dell'Object/Relational mapping tra gli oggetti Java trattati dall'applicazione ed un DBMS relazionale.

A titolo di esempio, un prodotto open source che implementa il pattern DAO ed è a larghissima diffusione nella comunità internazionale degli sviluppatori Java è Hibernate che fornisce tutte le funzionalità necessarie, un livello di prestazioni estremamente elevato e garantisce l'assoluta indipendenza dal DBMS utilizzato, il cui unico requisito è quello di consentire l'accesso attraverso le API standard Java Data Base Connectivity (JDBC). Dall'indipendenza dal particolare DBMS consegue una scalabilità notevole dell'applicazione che può così essere fruita sia in ambiente mono-utente, utilizzando come repository locale un DBMS open source e di minima complessità, sia in un ambiente web multi-utente, utilizzando come repository centralizzato un DBMS enterprise oppure un'alternativa open source.

In una fase successiva l'evoluzione dello strato di controllo potrebbe essere nella direzione di adottare uno dei frameworks open source più diffusi per l'MVC come ad esempio Struts oppure Spring. Peraltro una volta specializzato il ruolo di ciascun componente dello strato di controllo ed individuate le interfacce applicative verso gli altri strati, il passaggio al framework risulterà di certo semplificato ed a quel punto sarà effettivamente possibile avere due applicazioni, una Java desktop l'altra enterprise web, che condividono la logica applicativa ed il modello dei dati e si differenziano soltanto per l'interfaccia utente.

Tutto ciò: la separazione delle interfacce, l'adozione di framework esterni, l'utilizzo di standard consolidati per lo scambio di informazioni tra diversi componenti, sono i presupposti fondamentali per un'evoluzione progressiva verso un'architettura orientata ai servizi (Service Oriented Architecture o SOA) che è quella proposta in ultima analisi per l'integrazione del SW Trasversale. Gli obiettivi prioritari dell'architettura di servizi e di processi sono la flessibilità e la scalabilità. La massima flessibilità può essere ottenuta solo garantendo l'autonomia funzionale ed implementativa dei componenti interni ed esterni del sistema (accoppiamento debole) e, nel contempo, la massima interoperabilità di tali componenti. L'autonomia deve potersi applicare a livello tecnologico (i sistemi operativi, i linguaggi di programmazione, i sistemi di gestione di basi di dati, i contenitori di applicazioni, i motori di eventi e processi): tale esigenza è vitale per i componenti esterni, i sistemi partner con cui il DFM Case deve poter dialogare, ma è utile anche per l'architettura interna. Nei limiti delle caratteristiche dei diversi SW Trasversali e delle esigenze di integrazione, il sistema proposto è comunque conforme alle specifiche Sun Java Enterprise Edition, versione 1.4 o superiore; e ne viene garantita la portabilità sui più diffusi application server commerciali (IBM, Oracle, ecc...) oppure open source (Sun, JBoss, ecc...).

La figura che segue rappresenta un'ipotesi di architettura futura del sistema DFM Case, in particolare nel diagramma di deployment vengono integrate un'applicazione Java desktop ed un'applicazione enterprise web utilizzando esclusivamente prodotti open source; l'esempio raffigurato ha pertanto il vantaggio di non richiedere l'acquisto di alcuna licenza commerciale.

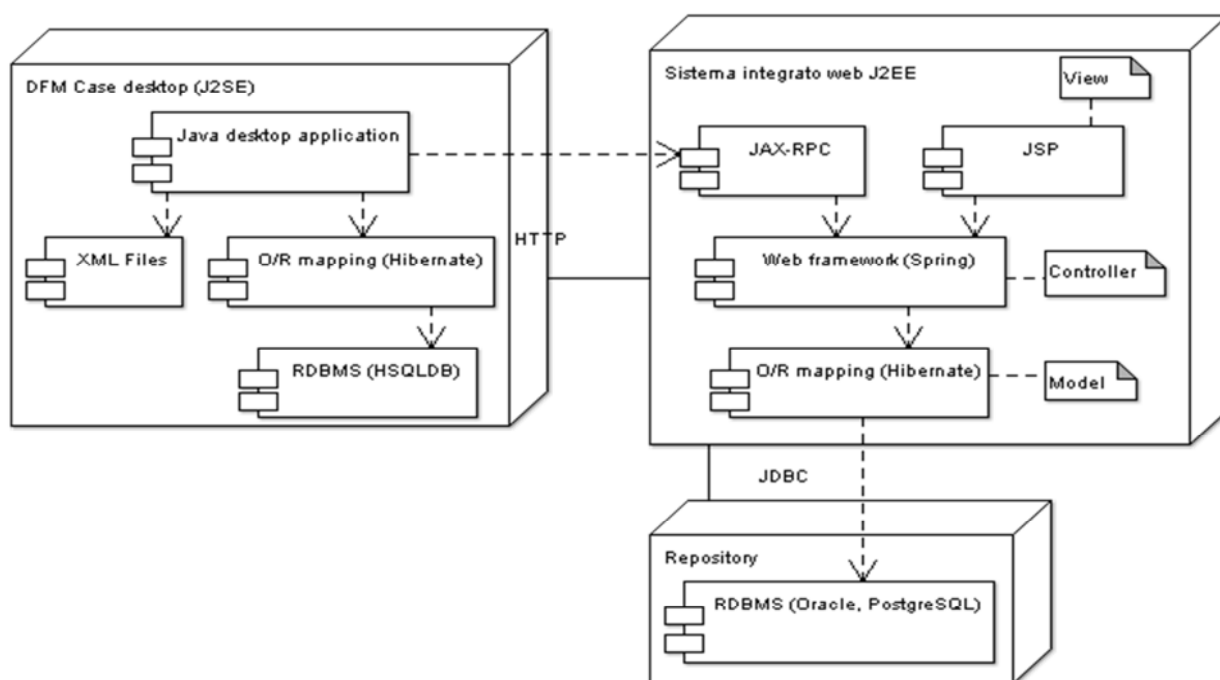


Figura 1 - Ipotesi di architettura futura

1.2.3.3 Evoluzioni

La complessità del sistema DW RGS in esercizio e degli obiettivi progettuali di prevista attuazione, congiuntamente alla complessità di un ciclo di vita incrementale ed evolutivo, focalizzano l'attenzione di Sogei sull'insieme degli strumenti al momento utilizzati in sede di analisi da diversi fornitori che si avvicinano nello sviluppo del sistema.

L'evoluzione del DFMCase si inserisce in un percorso che prevede l'ottimizzazione, l'integrazione e l'automazione degli attuali strumenti (SW Trasversali) e delle linee guida (Best Practice) al momento utilizzati in sede di analisi da diversi fornitori che si avvicinano nello sviluppo del DW RGS.

Si propone infatti di far evolvere gli strumenti di supporto all'analisi (quali il DFMCase) verso una soluzione integrata al fine di mettere a disposizione dei gruppi di lavoro una soluzione completa per la gestione della fase di Definizione.

1.2.4 Repository della qualità'

Classe di rischio C, numero di utenti 30

1.2.4.1 Caratteristiche della soluzione

Il Repository della Qualità (Oracle) si propone come obiettivo la centralizzazione dei metadati delle metriche di qualità dati sviluppate per i Data Mart del Data Warehouse RGS e della BDAP, al fine di fornire un supporto alle fasi di definizione e progettazione delle iterazioni future.

Il Repository della Qualità permette di fornire una visione unificata dell'intero processo di qualità, integrando in un unico repository centralizzato il patrimonio informativo della qualità dei dati in termini di:

- metriche, loro definizione e informazioni a corredo, come, ad esempio, i valori attesi nei diversi periodi dell'anno;
- associazione delle metriche al modello concettuale ER;
- Regole utilizzate e quindi riutilizzabili;



- Quali job utilizzano le regole e quali regole.

Per l'accesso al Repository della Qualità (Oracle), previsto sia per i gruppi di lavoro (per verificare ad esempio se una metrica è stata o meno sviluppata e come) sia per Sogei (per avere una visione globale delle metriche sviluppate), è stata realizzata una modalità che permetterà di fruire dei dati, attraverso report o utilizzando interrogazioni SQL. E' inoltre possibile avere a disposizione le informazioni sulle regole utilizzate, attraverso il Glossary e la Console di InfoSphere Information Server e in particolare sulla componente Analyzer. Sarà possibile :

- Fruire dei dati nel repository consentendo l'analisi e una più agevole divulgazione delle informazioni relative alle metriche;
- produrre la documentazione di progetto per la fase di definizione e utilizzare il Glossary e Fasttrack per la progettazione della fase realizzativa;
- esportare facilmente e velocemente il contenuto del repository (Oracle);
- utilizzare un universo BO per condurre analisi statistiche e report sui metadati contenuti nel repository di qualità.

L'implementazione della Qualità non differirà dal modello/processo disegnato per il Software ETL, bensì ne ricalcherà tutti gli schemi e, utilizzando la stessa piattaforma, ne condividerà i metadati.

L'innovazione che si intende perseguire pone le sue fondamenta sull'assunzione di poter definire la Qualità dei Dati a partire dai singoli oggetti dei modelli concettuali E/R e DFM.

Tale novità guida una nuova strutturazione degli obiettivi di qualità in obiettivi di:

- Profilazione dei dati propri delle fonti sorgente dell'informazione;
- Audit dei dati dell'ambiente EDW;
- Certificazione dell'area di pubblicabilità di un DM e validazione del suo processo di caricamento giornaliero.

Mentre le prime due tipologie di obiettivo, tra l'altro schedabili in base alla necessità, hanno per oggetto il modello concettuale E/R (o meglio delle sue viste), la terza tipologia di obiettivo, l'unica che a nostro avviso dovrebbe essere oggetto di una schedazione giornaliera, è incentrata sul modello concettuale DFM (schemi di fatto con relative gerarchie e dimensioni) di una determinata area di pubblicabilità.

Sul fronte della qualità dei dati, l'evoluzione focalizza l'attenzione esclusivamente sugli obiettivi di certificazione proponendo una loro formalizzazione in termini di corrispondenza tra area di pubblicabilità ed Obiettivo, tra schema di fatto e una lista di Question possibili, tra ciascun oggetto di business e un predefinito set di controlli di qualità.

Risultato fondamentale è la possibilità di riuso delle metriche/regole implementate; in questo modo verranno apportate indiscutibili ottimizzazioni sul fronte dei tempi/costi di sviluppo e della manutenibilità del sistema.

1.2.4.2 Architettura

L'architettura a sostegno di tutto il processo di qualità, è la stessa che governa l'implementazione del Software ETL.

1.2.4.3 Processo

La figura sotto riportata rappresenta il processo per l'implementazione della Qualità.

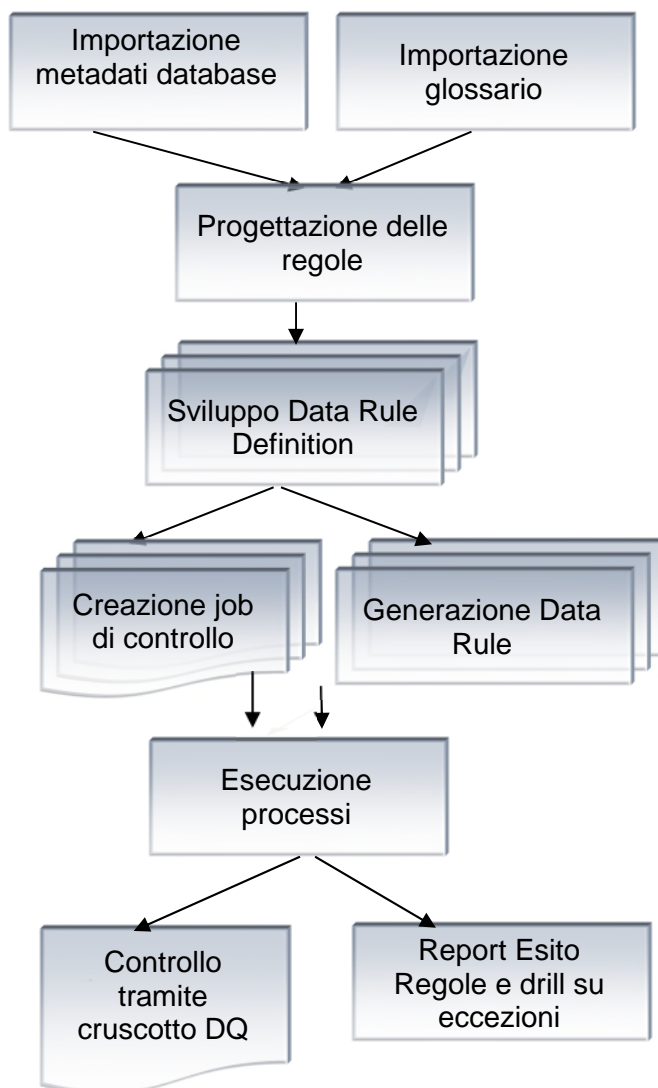


Figura 2 - Processo per l'implementazione della qualità

1.2.4.4 Evoluzioni

Il Software di Qualità è stato realizzato utilizzando il tool Analyzer; quest'ultimo è una delle componenti della piattaforma InfoSphere Information Server.

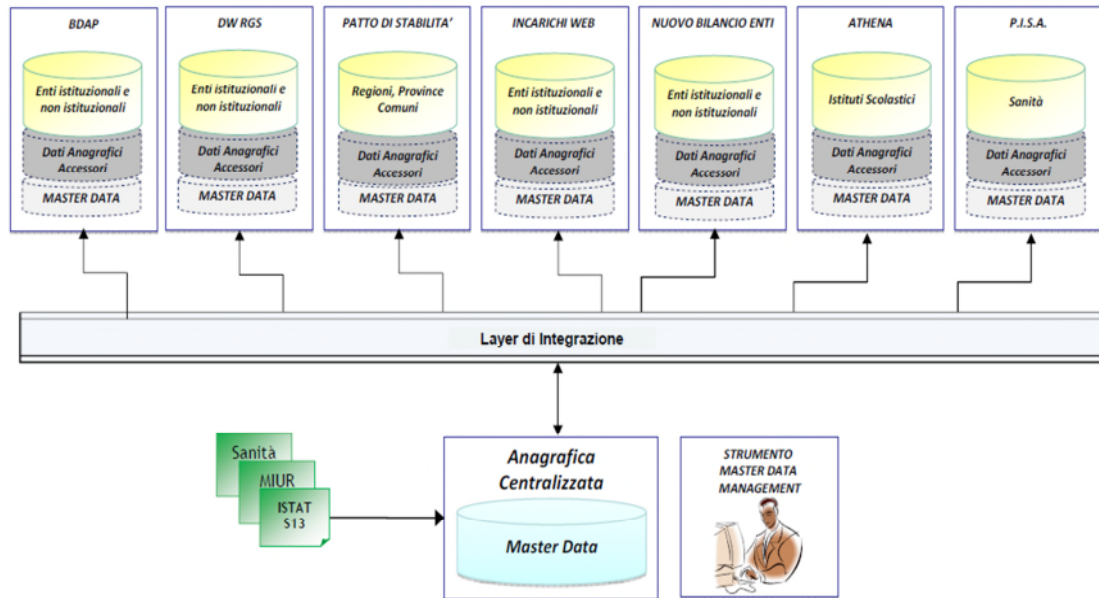
1.3 Master Data Management (MDM)

Classe di rischio C, numero di utenti 20

1.3.1.1 Caratteristiche della soluzione

Il Master Data Management, come insieme di processi, politiche, servizi e tecnologie, in ambito RGS è stato implementato per superare le criticità di disallineamento e ridondanza delle informazioni anagrafiche definite come dati Master, gestite all'interno delle diverse applicazioni.

Nella sua prima implementazione ci si è focalizzati nel consolidamento del patrimonio informativo relativo all'anagrafica degli enti del settore pubblico allargato gestiti in ambito Ispettorato IGF, con l'obiettivo di creare la miglior fonte possibile di informazioni anagrafiche attendibili e di qualità.



1.3.1.2 Architettura

La figura sotto riportata rappresenta l'attuale architettura del Master Data Management.

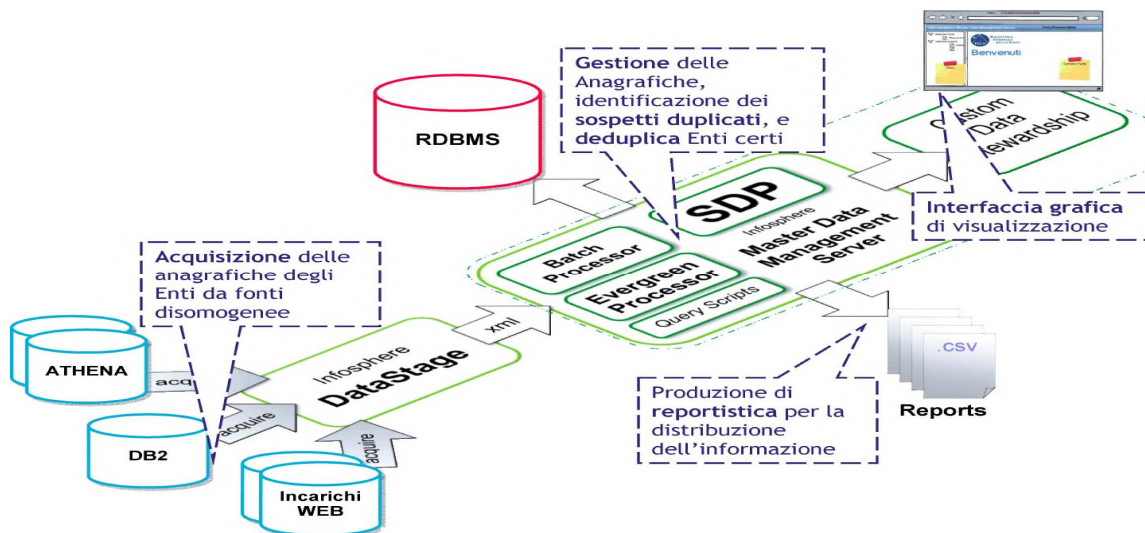


Figura 4 - Architettura del Master Data Management

1.3.1.3 Evoluzioni

Ora deve essere definito uno scenario complessivo per la gestione del ciclo di vita e la diffusione dell'Anagrafe degli Enti del settore pubblico allargato, ovvero:

- un'architettura informativa di riferimento;
- dei pattern standard di integrazione;
- delle politiche di gestione e dei referenti per l'attuazione.



1.3.2 App Bilancio Aperto

Classe di rischio A, numero di utenti (Aperto ad Internet)

1.3.2.1 Caratteristiche della soluzione

Bilancio Aperto è l'applicazione che consente a tutti i cittadini di consultare il Bilancio dello Stato, secondo i principi di trasparenza e accessibilità dei dati e delle informazioni.

L'app è a cura della Ragioneria Generale dello Stato. È rivolta sia agli "addetti ai lavori", Professionisti e Parlamentari, che ai Cittadini e ha l'obiettivo di accompagnare l'utente nell'evoluzione annuale dell'intero ciclo di bilancio.

In particolare, è possibile consultare, scaricare e condividere la seguente documentazione, a mano a mano che questa diventa disponibile:

- **Il Bilancio di Previsione** che fissa la fonte e i limiti delle Entrate e delle Spese per l'anno e stabilisce la destinazione delle risorse disponibili. Il Bilancio di Previsione viene seguito in tutto il suo processo di formazione: parte con il Disegno di legge di Bilancio
- presentato dal Governo al Parlamento, prosegue con la discussione parlamentare che si riflette nelle Note di variazione e termina con l'adozione della Legge di Bilancio.
- **L' Assestamento** di metà esercizio che aggiorna le previsioni delle Entrate e delle Spese contenute nella Legge di Bilancio. Anche il provvedimento di Assestamento è seguito nella sua evoluzione dalla fase di presentazione del Disegno di legge al Parlamento, all'adozione della Legge di Assestamento.
- **Il Rendiconto** che presenta i risultati della gestione dell'anno precedente con riferimento alle Entrate, alle Spese e al Patrimonio dello Stato.

1.3.3 BDAP – BILAR (Bilanci Armonizzati)

Classe di rischio A, numero di utenti 8.000 c.a

1.3.3.1 Caratteristiche della soluzione

Il nuovo sistema "BDAP - bilanci armonizzati" nasce dall'esigenza del Ministero dell'economia e delle finanze di acquisire i bilanci armonizzati degli Enti avvalendosi di un formato dati aperto per la rappresentazione dei dati XBRL (extensible business reporting language) secondo tassonomie standard che raccolgono le informazioni previste dalla normativa (d.lgs 118/2011 e smi).

Il nuovo sistema è predisposto al fine di acquisire in maniera graduale le tre diverse tipologie di bilancio (preventivo, rendiconto di gestione e consolidato); per gli enti in contabilità finanziaria, il primo step è la trasmissione dei dati del bilancio di previsione 2016-2018.

1.3.4 BDAP – GAD (Grafo Attributi Dimensionali)

Classe di rischio C, numero di utenti 40

1.3.4.1 Caratteristiche della soluzione

Il Grafo degli Attributi Dimensionali (GAD) è una web application attraverso cui viene rappresentato concettualmente il patrimonio informativo dei dati contenuti nei sistemi DW RGS e BDAP.

Il formalismo concettuale adottato per la modellazione grafica è il Dimensional Fact Model (DFM).

L'applicazione GAD permette agli utenti dei Portali dei sistemi BDAP e DW RGS di ricercare e visualizzare, per Area Tematica, le dimensioni di analisi e le misure dei processi istituzionali di Finanza Pubblica.

Classificazione del documento: Consip Public

Gara a procedura aperta ai sensi del D. Lgs. 50/2016 e s.m.i., per l'acquisizione di servizi per la manutenzione, evoluzione e gestione dei sistemi di Data Warehouse e Business Intelligence del Ministero dell'Economia e delle Finanze e della Corte dei Conti - ID 1973

Appendice 4 al Capitolato tecnico - Descrizione contesto applicativo e tecnologico – Lotto 1



Il GAD fornisce, inoltre, un servizio per la metadattazione dei report/cruscotti prospettati nell'area riservata dei Portali dei sistemi BDAP e DW RGS.

1.3.5 Open Data

Classe di rischio A, numero di utenti (Aperto ad Internet)

1.3.5.1 Caratteristiche della soluzione

Il Ministero dell'Economia e delle Finanze è impegnato a realizzare obiettivi di trasparenza e condivisione del suo patrimonio informativo attraverso la pubblicazione di Open Data. Gli Open Data sono informazioni e dati pubblici resi disponibili in formato digitale non proprietario, senza copyright, brevetti o altri limiti all'utilizzo ed alla riproduzione.

La Ragioneria generale dello Stato ha avviato un percorso di "apertura" dei dati relativi alla finanza pubblica, a partire dai seguenti temi:

- **Il Bilancio dello Stato**
Dati sulle spese e sulle entrate contenuti negli Atti Dovuti del Bilancio dello Stato:
Disegno di legge di bilancio, Legge di bilancio, Legge di assestamento e del Rendiconto generale dello Stato
- **Enti della Pubblica Amministrazione**
L'Anagrafe degli Enti che costituisce, in base all'articolo 2 della legge 31 dicembre 2009, n. 196, il perimetro di riferimento del patrimonio informativo della Banca Dati delle Amministrazioni Pubbliche (BDAP)
- **Gestione della Spesa**
Dati sulle spese effettuate dalle amministrazioni centrali, costituite da tutte le uscite registrate contabilmente nel bilancio dello Stato

1.3.6 Sistema di gestione delle forniture

Il **Sistema di Gestione delle Forniture** è nato per centralizzare la gestione della definizione e del ciclo di vita di qualsiasi dataset (risorsa) che, originato dal DW RGS (modellabile tramite una funzionalità DFM e quindi esprimibile con una query SQL sul DW RGS), voglia essere reso disponibile ad un sistema destinatario esterno (sia esso un applicativo, un sistema di BI, un portale di pubblicazione, un dataset open data, un destinatario di posta, ecc. ecc.) veicolandolo su canali differenti (E-mail, FTP, ETL Datastage, REST API CKAN, File System) in formati differenti (csv, xml, json, ecc. ecc.).

Le funzioni realizzate sono ripartite sulle due diverse applicazioni GEFOD e GEPOD di seguito descritte

1.3.6.1 GEFOD (Gestore delle Forniture Open Data)

L'applicativo nasce per razionalizzare ed estendere il modello di gestione dei dati di tutte le forniture che la RGS deve veicolare verso l'esterno, generalizzando il concetto stesso di fornitura, disaccoppiando la logica di validazione e produzione dei dati da quella del loro trasporto verso sistemi terzi e della loro eventuale trasformazione in altri formati.

Il modello logico delle forniture gestito da GEFOD può essere sintetizzato nella figura seguente.

Un **evento** (es: di natura temporale o di business) afferente ad una certa area informativa (es: il Bilancio dello Stato) viene ripartito in **sezioni** (es: sezione spese, sezione entrate, sezione saldi) ad ognuna delle quali possono essere associate diverse forniture.

Le **forniture** sono individuate da una query parametrica sul DW RGS e rappresentano l'insieme dei dati che il "Committente MEF" ha specificato nei requisiti della fornitura stessa e che il "Referente del MEF" ha il compito di autorizzare ad una precisa *data di osservazione* da lui scelta in modo che possano essere pubblicati.



I referenti SOGEI modellano i **prodotti** che derivano dai dati delle forniture effettuando dei tagli orizzontali e verticali sui dati autorizzati e distribuendo questi prodotti in un formato predefinito (risorse) su più destinatari.



1.3.6.2 GEPOD (Gestore Produzione e Trasporto Fornitura Open Data)

L'applicativo nasce per gestire il trasporto e il monitoraggio delle risorse (ovvero dei dataset originati dalle forniture) a valle del workflow approvativo gestito da GEFOD.

Le funzionalità principali sono suddivise nel modo seguente:

- **Generazione:** le sezioni di una fornitura pubblicate relativamente ad un evento per un certo esercizio finanziario ed autorizzate ad una certa data di osservazione generano prodotti da distribuire in diversi formati a diverse destinazioni. GEPOD si occupa dell'esecuzione delle query (anche schedulandole), della trasformazione dei dati e del loro salvataggio nelle strutture dati del DB applicativo.
- **Trasporto:** il trasporto (schedulabile) dei dati attraverso diversi canali (ftp, mail, file system, DB Datastage, SAS-OD) verso diverse destinazioni definite per i prodotti viene effettuato da GEPOD a valle della generazione.
- **Monitoring:** funzione essenziale di GEPOD è il monitoraggio di tutti i processi di generazione e trasporto dei dati ma anche dei passaggi di stato gestiti da GEFOD al fine di fornire all'amministratore una visione complessiva dello stato di lavorazione delle forniture e delle attività svolte dai diversi utenti durante le lavorazioni.
- **Utilità:** è possibile effettuare delle modifiche al WF delle forniture utilizzando funzioni di utilità: cancellazione di una pubblicazione di un evento-sezione, rigenerazione di risorse, reinvio di risorse.

1.3.6.3 GEMOD (Gestore Metadati Forniture Open Data)

L'applicativo nasce per gestire i metadati propri di ciascuna fornitura con l'obiettivo di descrivere ciascun dataset pubblicato sul portale OpenBDAP.

All'interno del ciclo di vita delle forniture, in fase di definizione anagrafica (GEFOD), è prevista l'immissione del legame tra la fornitura e la funzionalità DFM (Dimensional Fact Model) che descrive il modello concettuale dimensionale dei dati del dataset.

Per ciascun dataset pubblicato nell'apposita pagina Open Data del portale OpenBDAP, unitamente ai metadati richiesti dalle linee guida dell'AGID, il GEMOD rende disponibile un documento contenente i metadati estratti dal repository del tool di modellazione DFM.



Attraverso tale documento per ciascun dataset è presentato in forma grafica il modello DFM che descrive i livelli di aggregazione gerarchica degli eventi oggetto di analisi negli schemi di fatto di riferimento ed evidenziati gli attributi dimensionali di interesse.

Nel documento sono inoltre descritti in forma tabellare i fatti e gli attributi dimensionali presenti nel grafo, seguiti da informazioni di dettaglio sui campi (attributi descrittivi e misure d'analisi) che costituiscono il dataset esposto.



2. CORTE DEI CONTI

2.1 INTRODUZIONE

Una importante categoria di sviluppi inerenti il Data Warehouse della Cdc (ConosCo) è quella dei cosiddetti software “trasversali”. Questi progetti hanno come obiettivo primario quello di realizzare dei sistemi indirizzati alle strutture organizzative e produttive del Data Warehouse, per agevolare la gestione di definiti aspetti operativi o metodologici di interesse comune all’interno dell’impianto del sistema.

Tali sistemi software dovranno essere volti ad automatizzare, velocizzare ed ottimizzare la gestione sia del ciclo progettuale che del processo operativo del Data Warehouse.

2.2 DESCRIZIONE FUNZIONALITA’

2.2.1 Cruscotti Monitor Progetti

Classe di rischio C, numero di utenti 10

2.2.1.1 Caratteristiche della soluzione

ConosCo ha raggiunto nel corso del tempo un’evoluzione crescente sia in termini di Datamart disponibili sia in termini di numero degli utenti.

Nasce quindi l’esigenza di rendere disponibili dei cruscotti attraverso i quali sia possibile monitorare l’andamento del numero di oggetti presenti nel sistema e le statistiche di utilizzo dei sistemi.

I dati di interesse verrebbero estratti attraverso il Progetto Enterprise Manager di Microstrategy che provvede in modo automatico alla raccolta dei dati a seguito di tutte le operazioni effettuate sui Progetti da parte degli amministratori e dagli utenti di ConosCo.

In una fase successiva, come riportato in maniera più dettagliata di seguito, tali informazioni potranno essere integrati con altre informazioni relative all’utilizzo dei Sistemi gestionali Sorgenti.

I Datamart di ConosCo nel corso del tempo hanno visto crescere il numero degli oggetti che li costituiscono (Attributi, Metriche, Filtri, Report condivisi, Report personali etc.). Si rende pertanto necessaria la definizione e l’implementazione di un cruscotto che consenta il monitoraggio dell’andamento del numero di oggetti e il loro utilizzo da parte degli utenti.

Il cruscotto dovrebbe quindi fornire indicazioni relativamente ai seguenti aspetti:

- andamento di “crescita” di ogni singolo progetto ConosCo a partire dal numero di oggetti creati per mese e distinti per tipologia. Tali indicatori permetteranno di tenere sotto controllo il numero di oggetti presenti sul sistema
- indicazione dell’utilizzo degli oggetti da parte degli utenti. In tal modo infatti sarà possibile l’individuazione degli oggetti non utilizzati e quindi mettere in atto delle azioni al fine di sensibilizzare l’utente ad utilizzarli (ad esempio, durante i corsi di formazione si potrebbe dedicare più spazio all’importanza di tali oggetti in termini di semplificazione delle query) oppure procedere alla loro cancellazione dal sistema evitando così la proliferazione di oggetti non utilizzati od obsoleti.

2.2.2 Cruscotto Statistiche Utilizzo

Classe di rischio C, numero di utenti 20



2.2.2.1 Caratteristiche della soluzione

Nel corso degli anni il numero di utenti dei Datamart di ConosCo è cresciuto e con esso anche il numero di oggetti creati dagli utenti e dall'amministratore del sistema per gli utenti (report personali, report di consultazione e cruscotti).

Nasce quindi l'esigenza di monitorare le attività di svolte dagli utenti in termini di utilizzo del sistema.

Il cruscotto per le statistiche di utilizzo dovrebbe quindi prospettare, in formato sintetico attraverso dei micrografici, alcuni indicatori (di seguito un possibile elenco), espressi per il progetto selezionato.

Indicatori:

- **n° Utenti Attivi:** identifica il numero di utenti che hanno effettuato almeno un accesso al progetto selezionato.
- **Max Utenti Attivi:** identifica il massimo numero di utenti che hanno effettuato almeno un accesso al progetto
- **n° Connessioni:** l'indicatore rappresenta il numero di connessioni (sessioni) effettuate complessivo.
- **Media durata Sessione:** nell'ambito del numero di connessioni sopra descritte, viene calcolata la media complessiva della durata di ogni connessione (\sum durata complessiva sessioni/n° sessioni).
- **Media tempo d'esecuzione per i Report:** tale indicatore di performance, indica la media tempo trascorso nel Sistema per esporre i risultati interrogati. Chiaramente tanto più è basso tale indicatore, tanto più è veloce il Sistema a restituire i dati.
- **n° Report Eseguiti:** l'indicatore mostra il numero di report (tabellari e/o grafici) eseguiti
- **n° Documenti Eseguiti:** l'indicatore mostra il numero di documenti eseguiti.
- **n° Report Eseguiti per Documenti:** i dati mostrati in ogni Document MicroStrategy, sono provenienti da interrogazioni reimpostate di report definiti secondo i criteri necessari (normalmente denominati dataset). Essi non sono altro che report tabellari e grafici, utilizzati appunto come fornitori di dati di prospetti, adatti alla stampa e preformattati. Questo indicatore mostra i report eseguiti in maniera implicita all'esecuzione di Document.
- **n° Report Utente realizzati:** l'indicatore mostra complessivamente i report realizzati (personali e/o condivisi) dagli utenti.

L'andamento dei dati dovrà mostrare i valori puntuali per gli ultimi quindici mesi trascorsi, compreso quello corrente e per ogni indicatore sono da presentare, oltre alla linea d'andamento tra i valori mensili, il valor medio ed esplicitamente il valore del mese corrente.

Dovrà essere possibile interrogare il cruscotto selezionando i gruppi di utenti definiti sul sistema.

I dati prospettati, anche in formato tabellare, dovranno mostrare il confronto mensile tra anno solare ed anno solare precedente.

Si dovrà prevedere inoltre la possibilità di integrare i dati dell'utilizzo del sistema conoscitivo con quelli di utilizzo di altri sistemi della Corte dei conti in particolare quelli dei gestionali da cui i datamart di ConosCo si alimentano.

Questo cruscotto, ha tra gli obiettivi, quelli di:

- individuare i periodi di picco di attività degli utenti sul sistema per meglio organizzare le attività a supporto degli utenti
- fornire una lista di report non utilizzati (sia personali che condivisi) al fine di fare pulizia di oggetti non più utili
- individuare i report che hanno un tempo di estrazione eccessivamente lungo per migliorarne le performance



- promuovere l'utilizzo del sistema presso gli utenti che hanno seguito i corsi e per i quali non risulta un utilizzo del sistema.

2.2.3 Cruscotto Monitoring Operativo

Classe di rischio C, numero di utenti 10

2.2.3.1 Caratteristiche della soluzione

ConosCo ha raggiunto nel tempo un'ampiezza rilevante per informazione gestita e moli di dati trattate ed una elevata complessità per numero di componenti che ne realizzano quotidianamente il processo di aggiornamento e messa in linea dell'informazione. La delicatezza e criticità assunte dalle attività di Data Loading sono tali da rendere evidente la necessità di condurre in maniera organica e centralizzata le attività di monitoraggio per il processo di fornitura dei dati all'utente.

Il monitoraggio di elementi quali:

- prestazioni delle componenti applicative del sistema,
- stato di degrado dell'ambiente,
- risultati delle operazioni di caricamento e livello di qualità dei dati elaborati

richiede una razionalizzazione e di una esecuzione sistematica e generalizzata su tutte le componenti interessate per essere efficace strumento di controllo e consentire l'intercettazione e la previsione di criticità potenzialmente bloccanti.

Il Cruscotto di Monitoring Operativo costituisce una soluzione applicativa che soddisfa l'esigenza di centralizzare e velocizzare gli interventi orientati alla fornitura del dato all'utente.

Lo strumento in oggetto si propone di rispondere alle esigenze descritte attraverso le seguenti funzionalità:

- Monitoring attraverso la valorizzazione di appositi indicatori e la produzione reportistica di dettaglio riguardante:
 - esiti di batch per l'alimentazione e procedure per il controllo della qualità dei dati;
 - tempi in esecuzione di tutte le componenti (Job ed Applicazioni TWS) del piano di schedulazione notturna e durata dell'intero piano;
 - spazi occupati per database interessati,
 - visualizzazione del trend storico delle metriche di qualità.

Le funzionalità presenti nel Cruscotto consentono la centralizzazione e la gestione attraverso un'unica interfaccia di molte delle attività al momento effettuate nell'ambito dell'operatività giornaliera (controllo esito caricamenti, verifica del livello di qualità dei dati processati, controllo esito refresh dell'ambiente utente, ecc...).

Il prodotto fornisce alle strutture preposte alla gestione gli elementi per intercettare più rapidamente le anomalie, prevedere e stimare le criticità, velocizzare gli interventi; il tutto avendo come obiettivo primario quello di favorire il data loading verso l'utente (massimizzando la disponibilità e la qualità dei dati) nel rispetto delle finestre temporali assegnate.

Attraverso l'osservazione sistematica e oggettiva delle prestazioni della fase di aggiornamento dei dati del Sistema, si ha, inoltre, la possibilità di ottimizzare decidendo preventivamente gli interventi mirati ad ottenere il rispetto dei tempi di consegna ed aumentare la qualità del servizio.