

**Sistemi conoscitivi MEF
Architettura di riferimento**

TABELLA DELLE VERSIONI

Versione	Data	Descrizione delle modifiche
1.0	Novembre 2005	prima stesura

INDICE

1. INTRODUZIONE	4
2. PRINCIPALI PROGETTI	5
2.1 DATA WAREHOUSE R.G.S.	5
2.1.1 <i>Descrizione delle Applicazioni informatiche</i>	6
2.2 DATA WAREHOUSE IV DIPARTIMENTO	9
2.1 DATA WAREHOUSE DARPA	11
2.2 SISTEMA CONOSCITIVO IGRUE	14
3. SCHEMA DI RIFERIMENTO	16
3.1 PROCESSO DEL DATA WAREHOUSE	16
3.2 FRAMEWORK TECNOLOGICO	18
3.2.1 <i>ETL</i>	19
3.2.2 <i>Job scheduling</i>	20
3.2.3 <i>Database</i>	20
3.2.4 <i>Data quality management</i>	20
3.2.5 <i>Metadata management</i>	20
3.2.6 <i>Front-end</i>	21
3.2.6.1 <i>Reporting gestionale</i>	22
4. ARCHITETTURA ATTUALE	24
4.1 INFRASTRUTTURE DI CONSOLIDAMENTO	24
4.2 ETL	25
4.3 JOB SCHEDULING	28
4.4 DATABASE	30
4.5 DATA QUALITY MANAGEMENT	32
4.6 METADATA MANAGEMENT	33
4.7 FRONT-END	34
4.7.1 <i>BusinessObjects</i>	35
4.7.2 <i>Microstrategy</i>	39
4.7.3 <i>Suite Oracle</i>	41
4.7.3.1 <i>Oracle Discoverer</i>	41
4.7.3.2 <i>Oracle Reports</i>	42
4.7.3.3 <i>OFA</i>	42
4.7.4 <i>Actuate</i>	42
4.7.5 <i>Pacchetti statistico/econometrici</i>	44

1. Introduzione

Nel corso degli ultimi anni il MEF ha realizzato importanti progetti nell'ambito dei cosiddetti "sistemi conoscitivi", ovvero nell'ambito di quei sistemi informatici che, secondo una classica definizione, si pongono l'obiettivo di valorizzare il patrimonio di dati dell'azienda fornendo "l'informazione giusta, all'utente giusto, al momento giusto".

Tali progetti sono stati sviluppati in tempi e con finalità diversi, per le specifiche esigenze dei diversi dipartimenti dell'amministrazione; nel corso degli anni sono state quindi acquisite le opportune infrastrutture tecnologiche (sistemi hardware e prodotti software) e sono state realizzate le applicazioni necessarie.

Questa evoluzione, pur frammentata in base alle esigenze contingenti, è avvenuta all'interno di uno schema di riferimento comune, che costituisce quindi il "framework" dell'architettura attuale ¹.

Nel presente documento si intende fare il punto sullo stato dell'arte tecnologico dei progetti e dei sistemi realizzati nell'ambito di tale "framework", con l'obiettivo di dare una visione d'insieme dell'architettura complessiva.

¹ In questo contesto si fa riferimento a un framework di tipo tecnologico, tralasciando considerazioni di tipo metodologico sui modelli informativi che sottendono alla costruzione di un data warehouse.

2. Principali progetti

In questo paragrafo sono descritte a grandi linee la situazione attuale e le prospettive di sviluppo dei principali progetti conoscitivi del MEF.

2.1 *Data warehouse Ragioneria Generale dello Stato*

Il Data Warehouse R.G.S. è finalizzato alla realizzazione di un ambiente centralizzato destinato ad ospitare e rendere disponibili, mediante applicazioni di supporto alle decisioni, informazioni consolidate provenienti dai sistemi sorgente eterogenei.

E' infatti attraverso il Data Warehouse R.G.S. che ogni Amministrazione può disporre, in maniera rapida e sicura, di una fonte unica e certa del patrimonio informativo dei vari sistemi gestionali della Ragioneria a supporto del proprio processo decisionale.

Il Data Warehouse R.G.S. è inteso come Sistema costituito da dati, applicazioni, risorse tecnologiche ed umane, regole organizzative procedure deputate all'acquisizione, elaborazione, memorizzazione, scambio, reperimento e trasmissione delle informazioni.

Il Data Warehouse R.G.S., realizzato per soddisfare le esigenze conoscitive degli utenti della Ragioneria Generale dello Stato, consente di:

- rendere fruibili direttamente ed in modo semplice le informazioni riepilogative e di dettaglio relative al Bilancio dello Stato e più in generale inerenti alla Finanza Pubblica;
- garantire, per dette informazioni, la disponibilità ed un adeguato livello di qualità attraverso il meticoloso monitoraggio di correttezza e completezza dell'informazione gestita;
- realizzare un unico sistema per la fornitura di informazione ad uso conoscitivo che centralizzi i dati, ne certifichi la provenienza, ne tenga sotto controllo la proprietà e la diffusione.

L'utente di tale sistema è messo in grado di soddisfare le proprie specifiche esigenze di analisi, avvalendosi sia di prospetti predefiniti sia di interrogazioni estemporanee, senza dover necessariamente conoscere l'organizzazione logica e fisica della base dati né il suo linguaggio di accesso.

Il Data Warehouse R.G.S. è basato su un database centralizzato, l'Enterprise Data Warehouse (EDW) e su diversi Data Mart (DM). Le informazioni presenti sia sull'EDW che sui vari DM vengono sottoposte a processi di qualità in modo da certificare la correttezza del dato presente nel Data Warehouse rispetto al dato originario proveniente dal sistema sorgente.

Parallelamente all'alimentazione dell'EDW e dei Data Mart viene gestito l'archivio dei metadati, guida indispensabile dei processi di produzione e di sfruttamento delle informazioni contenute nel Data Warehouse.

Il Data Warehouse R.G.S., oltre a soddisfare le esigenze conoscitive dell'utente della Ragioneria Generale dello Stato, fornisce informazioni ad enti esterni; le modalità, la

tecnologia utilizzata e la frequenza con cui i flussi di dati in output vengono prodotti è differente a seconda delle esigenze del sistema che ospita tali informazioni.

Il portale, unica interfaccia verso l'esterno, mette a disposizione dell'utente finale sia la descrizione estesa di tutte le informazioni contenute nella reportistica predefinita, sia uno strumento di ricerca che, partendo dal concetto di business da analizzare e procedendo per affinamenti successivi, permette di individuare quali sono i report di interesse.

2.1.1 Descrizione delle Applicazioni informatiche

Enterprise Data Warehouse di Finanza Pubblica

L'Enterprise Data Warehouse (EDW) è realizzato in maniera incrementale ed iterativa e ne viene garantita la visione univoca e certificata dell'informazione. L'EDW è caratterizzato da una massima granularità e storicità dei dati, in accordo con le esigenze di analisi degli utenti.

L'informazione presente sull'EDW non è direttamente accessibile all'utente. La frequenza di aggiornamento dell'informazione è giornaliera con dati provenienti da diverse fonti (l'informazione viene prelevata dal sistema gestionale che ne ha la "proprietà"). I dati presenti sull'EDW, opportunamente filtrati ed aggregati in modo da soddisfare le esigenze specifiche di una classe d'utenti, vanno ad alimentare i vari Data Mart ai quali i vari utenti accederanno tramite il portale informativo.

Data Mart per IGPB (Ispettorato Generale per le Politiche di Bilancio)

Il Data Mart IGPB è il sistema conoscitivo attualmente utilizzato dal suddetto Ispettorato per ottenere le informazioni necessarie a supportare le decisioni che lo Stato è chiamato a prendere in materia di Bilancio Finanziario. Tale sistema, che utilizza una base dati storica contenente dati di Bilancio e di Spese opportunamente integrati, sfrutta tutte le potenzialità e la flessibilità delle moderne tecnologie per produrre, distribuire e condividere tali informazioni.

In particolare il sistema fornisce al personale dell'IGPB gli strumenti per effettuare analisi di tipo conoscitivo sui dati di Bilancio e di Spesa; inoltre vengono pubblicati sul portale una serie di report predefiniti organizzati in base alla fase di Bilancio di interesse (Formazione, Gestione, Consuntivo).

Data Mart per IGAE (Ispettorato Generale per gli Affari Economici)

Il Data Mart IGAE permette una rilevazione sistematica dei dati relativi all'avanzamento dell'attuazione del programma economico del Governo. L'obiettivo è rendere più efficiente ed efficace il sistema di controllo e monitoraggio della spesa al fine di controllare ed eventualmente correggere l'indirizzo del programma economico.

Tale sistema, che utilizza una base dati storica contenente dati di Bilancio e di Spese opportunamente integrati, sfrutta tutte le potenzialità e la flessibilità delle moderne tecnologie per produrre, distribuire e condividere tali informazioni.

Il sistema fornisce al personale dell'IGAE gli strumenti per effettuare analisi di tipo conoscitivo sui dati di Bilancio e di Spesa, in particolare dati relativi all'utilizzo dei fondi assegnati per mezzo di provvedimenti di tipo Delibera Cipe e dati relativi all'utilizzo degli stanziamenti assegnati ai capitoli di Bilancio per mezzo di Autorizzazioni di interesse dell'ispettorato IGAE, nonché gli aspetti legati all'evoluzione degli oneri delle autorizzazioni base dovuti alle eventuali autorizzazioni rifinanzianti/definanzianti collegate; inoltre vengono pubblicati sul portale una serie di report predefiniti organizzati in base alle aree di interesse dell'IGAE.

Data Mart per IGEP (Ispettorato Generale per la Finanza delle Pubbliche Amministrazioni)

Il Data Mart IGEP è il sistema conoscitivo dell'ispettorato che viene utilizzato per sovrintendere a quanto indicato nella Legge finanziaria 2003 art. 29 comma 13 e 17 in relazione al 'Monitoraggio del patto di stabilità interno' e a supporto della predisposizione della 'Relazione trimestrale di cassa' degli enti pubblici del settore pubblico allargato.

Il sistema, sfruttando tutte le potenzialità delle moderne tecnologie e di una base dati storica opportunamente integrata, fornisce al personale dell'IGEP gli strumenti per effettuare analisi di tipo conoscitivo sui prospetti acquisiti relativamente ai dati dei 'Flussi di Cassa' e del 'Patto di Stabilità Interno' aggregati e analizzati per territorio, ente, periodo, prospetto e singola voce.

Data Mart per UCB (Uffici Centrali di Bilancio)

Il Data Mart UCB realizza una base dati integrata e storica attraverso cui gli Uffici Centrali di Bilancio possono avere una visione completa ed esaustiva delle Spese dello Stato sia a livello di singolo titolo di spesa (Ordini di Pagare, Ordini di Accreditamento, Note di Imputazione) che aggregate a livello di Impegno, di Piano Contabile dei Residui e di Piano Gestionale. E' inoltre possibile monitorare la Contabilità Ordinaria e Speciale dei Funzionari Delegati che sono le figure istituzionali preposte alla distribuzione delle somme messe a disposizione dalle Amministrazioni.

Per avere una visione più sintetica di tali informazioni il sistema consente di aggregarle in base ad un set eterogeneo di dimensioni di analisi quali: Esercizio Finanziario, Amministrazione, Ragioneria, Capitolo/Articolo, Piano Gestionale, Classificazione Economica fino al 5° livello, Classificazione Funzionale, Centro di Responsabilità Amministrativa, Macroaggregato di Spesa, Unità Conoscitiva ed Unità di Spesa.

Gli Uffici Centrali di Bilancio hanno anche a disposizione un 'Sistema di Indicatori' che permette loro di effettuare un puntuale monitoraggio mensile delle loro attività al fine di poterle razionalizzare per rendere più efficace ed efficiente la loro attività lavorativa.

Oltre a permettere di effettuare interrogazioni di tipo generalizzato, per rendere più agevole l'utilizzo del sistema conoscitivo, sono a disposizione dell'utente un set di report predefiniti che rispecchiano le principali esigenze conoscitive dell'utente stesso.

Data Mart per IGICS (Ispettorato Generale Informatizzazione e Contabilizzazione dello Stato)

Il Data Mart IGICS è una base dati integrata e storica relativa alle Spese di Gestione delle Amministrazioni Centrali, attraverso cui gli utenti possono condurre analisi contabili a supporto delle decisioni.

Le classi di utenza interessate all'Applicazione sono:

- il personale amministrativo dell'IGICS;
- il personale di Amministrazioni Centrali (Ministeri, Dipartimento Amministrazione Penitenziaria, Agenzia delle Dogane, Agenzia del Demanio);
- il personale di Enti Locali.

L'analisi dei dati anagrafici e contabili viene effettuata avendo la possibilità di aggregare le informazioni relative alle Spese a diversi livelli di dettaglio: dal singolo titolo di spesa (Ordini di Pagare, Ordini di Accreditamento, Ruoli di Spesa Fissa, Note di Imputazione) sino alle aggregazioni contabili a livello di Piano di Ragioneria e Piano Contabile.

Per ciascun livello di dettaglio i dati sono inoltre analizzabili secondo le diverse classificazioni di cui sono oggetto (Classificazione Economica, Funzioni Obiettivo, Macroaggregato di Spesa, Unità Conoscitiva, Unità di Spesa, Centro di Responsabilità Amministrativa etc.).

Tali informazioni sono fruibili attraverso particolari strutture fisiche le quali facilitano le interrogazioni sia tramite operatori OLAP, sia attraverso reportistica predefinita.

La profondità storica dei dati giornalieri contenuti nel Data Mart è di 2 anni: quelli relativi all'esercizio finanziario in gestione e quelli relativi all'esercizio finanziario in consuntivazione. I dati relativi agli esercizi finanziari precedenti vengono inoltre storicizzati annualmente.

BDC (Banca Dati Conoscitiva)

L'Applicazione è costituita da due distinte aree funzionali:

Serie Storiche dell'Area Spese

Consente la prospettazione, a diversi livelli di aggregazione, dei dati riguardanti le serie storiche di Bilancio e di Spesa dello Stato provenienti da sottosistemi diversi, sincronizzati ed integrati in modo da renderli omogenei.

Sono a disposizione dati con valenza mensile che si dividono in dati che si modificano nel corso dell'anno registrati per la mensilità di interesse e in dati che si modificano nell'anno fino ad assumere un valore definitivo a chiusura del consuntivo per cui viene fornito il valore relativo al mese di osservazione, e dati con valenza annuale che si dividono in dati che non si modificano nel corso dell'anno e in dati che pur modificandosi nel corso dell'anno assumono un valore definitivo a chiusura del consuntivo.

Serie Storiche dell'Area Entrate

Consente la prospettazione, a diversi livelli di aggregazione, dei dati contabili delle Entrate, considerati nelle due diverse fasi di 'gestione' e di 'bilancio'. Sono a disposizione dati con valenza mensile che si dividono in dati che rappresentano fenomeni transitori tali da essere assorbiti a chiusura dell'esercizio finanziario da altre voci contabili e in dati che si modificano nell'anno fino ad assumere un valore definitivo a chiusura dell'esercizio finanziario, e dati con valenza annuale che si dividono in dati che non si modificano nel corso dell'anno, che vengono acquisiti in fase di consuntivazione, ed in dati che pur modificandosi nel corso dell'anno assumono un valore definitivo a chiusura dell'esercizio finanziario. Per quanto riguarda i dati osservati annualmente, sono state individuate 4 date corrispondenti a eventi salienti nella Contabilità dello Stato relativi al controllo e alla esazione delle Entrate: approvazione della Legge di Bilancio, presentazione del Provvedimento di Assestamento, approvazione del Provvedimento di Assestamento, elaborazione del Bilancio di Consuntivo.

2.2 Data warehouse Dipartimento Amministrazione Generale del Tesoro

Il Dipartimento dell'Amministrazione Generale, del Personale e dei Servizi del Tesoro (DAG) ha realizzato nel corso degli ultimi anni alcune applicazioni che per la loro valenza ed estensione costituiscono un punto di riferimento per la gestione nella Pubblica Amministrazione.

Allo stato attuale, i sistemi in esercizio sono:

- il sistema Service Personale Tesoro (SPT), che gestisce i processi di elaborazione, stampa e liquidazione degli stipendi (ca. 1.400.000 dipendenti gestiti) e delle pensioni di guerra;
- il Sistema Informativo per l'Amministrazione del Personale (SIAP), strumento di supporto per la gestione dei processi amministrativi riguardanti il personale del MEF, che gestisce gli aspetti giuridici ed amministrativi relativi al personale del Ministero (area Economia);
- la Banca Dati Direzionale del Personale (BDDP), il sistema conoscitivo alimentato dal SIAP che fornisce un quadro di riferimento del personale del MEF;
- il sistema di Controllo di gestione (CdG), che fornisce informazioni utili alla verifica del grado di efficienza della struttura ed alla impostazione delle nuove politiche gestionali.

I sistemi gestionali attualmente in esercizio hanno permesso, nel corso degli anni, di raccogliere una considerevole quantità di dati di ogni tipologia: lo sforzo di automazione dei processi ha quindi portato ad avere un patrimonio informativo che può costituire una vera e propria ricchezza da punto di vista del controllo e della programmazione.

Il passo successivo è quello di tradurre il dato in informazione concentrando l'attenzione non solo sul reperimento dei dati, ma sulla loro interpretazione e comprensione. Il DAG si

pone quindi l'obiettivo di realizzare un "Sistema Direzionale", che risponda alle esigenze conoscitivo-decisionali delle diverse organizzazioni che compongono il Dipartimento e che possa offrire eventualmente supporto ai diversi organismi istituzionali interni al Ministero stesso ma diversi dal DAG.

Per facilitare questo percorso e per rendere congruenti ed omogenee a livello di semantica i dati provenienti da differenti applicazioni realizzate per scopi diversi, la soluzione che la tecnologia suggerisce è la costituzione di un data warehouse.

Vista la complessità dell'ambito e dei sistemi all'interno dei quali vanno censite le informazioni, per la realizzazione del data warehouse del Dipartimento si seguirà un approccio di tipo "incrementale" o "federato" che prevede la creazione di un modello informativo comune, dal quale verranno sviluppati in maniera coerente modelli dati dell'Enterprise Data Warehouse (EDW) e/o dei Data Mart (dipendenti o indipendenti che siano). Vengono messi a fattor comune tra diversi progetti di Data Mart i processi di acquisizione dei dati dai sistemi transazionali.

Questo tipo di approccio comporta una maggiore rapidità di implementazione delle soluzioni verso gli utenti dal momento che non necessita di un modello dati enterprise disegnato completamente a priori, ma il modello stesso viene realizzato tramite un processo iterativo di aree tematiche di interesse prioritario.

Il data warehouse del DAG rappresenterà nel tempo un sistema:

- di supporto al controllo tattico-strategico al servizio del management del Dipartimento;
- di supporto alle attività istituzionali del Dipartimento;
- di supporto, quindi, alle attività del MEF nella sua interezza.

Attualmente è stato messo a disposizione del Dipartimento un primo Data Mart relativo all'Area Tematica "Controllo di Gestione" che raccoglie i dati sia dal sistema CDG che dal sistema Ermes.

Nell'ottica dell'approccio "incrementale" o "federato", sopra descritto, il Data Mart può essere a tutti gli effetti considerato il primo passo verso la creazione del Data Warehouse del Dipartimento e quanto realizzato può essere definito una prima release del modello informativo comune che, mediante le successive iterazioni, verrà poi ampliato nel corso del tempo a seconda delle implementazioni necessarie al nascere di nuove esigenze e nuove Aree Tematiche di interesse.

Il Data Mart, alimentato con i dati di tutti i semestri ad oggi elaborati all'interno del sistema CDG, è stato progettato al fine di rispondere ai seguenti requisiti:

- rendere disponibili le gerarchie anagrafiche sia dei centri di costo sia del piano dei conti. Tale scelta permette di sfruttare al meglio le funzionalità dello strumento di reportistica e di analisi;
- garantire l'esistenza e la stabilità del dato attraverso la corretta alimentazione (attualmente semestrale) dei dati all'interno del Data Mart.

E' stata pertanto realizzata una completa trasposizione del Modello Logico dei Dati, che mette a disposizione degli utenti tutte le dimensioni di analisi (Gerarchie) disponibili, insieme ad una serie di funzioni di misurazione (Metriche) sui relativi fatti di interesse.

L'architettura applicativa è composta da:

- un Data Mart che raccoglie i dati provenienti dal controllo di gestione;
- procedure custom per l'alimentazione;
- report Microstrategy al fine di analizzare e diffondere i dati.

Le procedure di caricamento provvedono ad aggregare il dato, in modo da rendere le informazioni estratte in una struttura tale da agevolare il processo di analisi con il tool di Microstrategy.

Ad oggi i sistemi alimentanti del Data Mart sono essenzialmente "Controllo di Gestione" (Budget, Consuntivo, Prodotti, Centri di Costo, Nature), "Ermes" (Volumi Attività), SIAP (anagrafica dei dipendenti per centro di costo) ed SPT (costo relativo allo stipendio del dipendente e tempo lavorato).

2.1 Data warehouse DARPA

In aderenza al mandato assegnatole e secondo le più avanzate esperienze europee e mondiali di centrali acquisti nel settore pubblico, Consip ha istituito un'apposita direzione, la DARPA, con la responsabilità di individuare e promuovere soluzioni per la razionalizzazione della spesa per beni e servizi della Pubblica Amministrazione, di definire e diffondere le metodologie di e-procurement, di avviare e gestire, in accordo con il Ministero dell'Economia e delle Finanze, i progetti rivolti alla semplificazione dei processi di acquisto, formazione, pagamenti e monitoraggio della spesa.

A supporto delle attività della DARPA è stato realizzato un sistema di Data Warehouse che attualmente comprende:

- un Enterprise Data Warehouse (EDW), che comprende ed integra i dati di tutte le iniziative rilasciate, alimentato con informazioni provenienti dai fornitori e da Consip stessa;
- alcuni "Data Mart" alimentati con le informazioni provenienti dal EDW;
- reporting specializzati per le diverse funzioni della direzione.

Oltre ai suddetti data base, ove si è reso necessario per attività di bonifica o di controllo, sono state utilizzate delle staging area.

I modelli dati relativi a tutte le iniziative realizzate sono stati documentati con lo strumento ErWin. In particolare per semplificare l'integrazione delle strutture dei dati, sono stati realizzati due soli modelli : uno per gli schemi EDW, uno per gli schemi Data Mart.

Ogni modello è organizzato in viste logiche tematiche; queste viste logiche, pur utilizzando oggetti comuni, rappresentano o una particolare tipologia di iniziativa di risparmio (es.

telefonia fissa, global services) o una specifica grandezza per il monitoraggio del programma di razionalizzazione della spesa (es. spesa, risparmio, movimenti, etc.).

Nell'ambito del modello relazionale sono state integrate tutte le informazioni comuni alle diverse iniziative, finora individuate, e ne vengono storicizzati i cambiamenti (anagrafe delle amministrazioni, informazioni territoriali, catalogo prodotti, etc). Questo in modo da costituire la base comune ed il punto di partenza per la definizione ed il popolamento di tutte le principali dimensioni di analisi presenti negli schemi a stella (Data Mart).

Relativamente ai modelli a stella (star schema), particolare attenzione è stata posta nella definizione di dimensioni e di fatti standard al fine di consentire eventuali analisi "drill accross" tra diversi Data Mart o la semplificazione di eventuali realizzazioni di modelli riassuntivi (trasversali) di tutte le iniziative.

Sono attualmente in esercizio le seguenti applicazioni conoscitive:

- Anagrafiche comuni - Sistema per la gestione centralizzata delle anagrafi dei clienti Consip (PA), del territorio, etc.
- Reporting PM - Sistema per il monitoraggio dei principali indicatori del programma degli Acquisti in Rete di Consip.
- Fonia Mobile - Sistema di monitoraggio della specifica iniziativa di risparmio "Fonia Mobile".
- Fonia Fissa - Sistema di monitoraggio della specifica iniziativa di risparmio "Fonia Fissa".
- Global Service - Sistema di monitoraggio della specifica iniziativa di risparmio "Global Services".
- Marketplace - Sistema per il monitoraggio del marketplace del sistema di Acquisti in Rete di Consip.

Le linee evolutive del Data warehouse DARPA prevedono il potenziamento e l'arricchimento del livello della conoscenza in termini di:

- Integrazione nel sistema di Data Warehouse di ulteriori informazioni, sia interne che esterne, d'interesse per lo sviluppo del programma (es. monitoraggio delle gare telematiche, monitoraggio del market Place, etc);
- Evoluzioni dei DM esistenti e sviluppo di nuovi DM specializzati, sia per tipologia di convenzione che per funzione aziendale;
- Definizione di ulteriori misuratori di direzione (KPI, OPI) e adozioni di ulteriori modalità di analisi dei dati.

Sono in fase di studio iniziative di analisi specializzate, più precisamente:

- profilazione dei clienti Consip sull'analisi del comportamento;
- soluzioni di CRM analitico;

Inoltre sono previsti sviluppi anche nei seguenti contesti già esistenti:

- Anagrafiche comuni - Adeguamento del sistema al Marketplace ed alle Gare telematiche quindi inserimento ed integrazione dei fornitori dei mercati Consip, etc.,
- Reporting PM - Adeguamento del sistema ai nuovi obiettivi del programma Mef/Consip.
- Marketplace - Evoluzione del Sistema con i nuovi obiettivi direzionali e realizzazione di nuove analisi di contesti specifici (comportamenti dei fornitori, PPAA, etc.).
- Revisione architetturale del sistema di front-end e realizzazione di cruscottistica personalizzata che faciliti il reperimento e la fruizione delle informazioni.
- Realizzazione di un sistema di monitoraggio del servizio di invio dei dati dai fornitori esterni e di controllo della qualità dei dati ricevuti.

2.2 Sistema conoscitivo IGRUE

Il Sistema conoscitivo dell'IGRUE è un insieme di funzionalità finalizzate a rendere disponibile, in modo flessibile e dinamico, l'intero patrimonio informativo gestito dall'Ispettorato Generale per i Rapporti con l'Unione Europea.

Le prime implementazioni (Data mart) sono nate dall'esigenza di disporre di strumenti per estrarre ed organizzare le informazioni, in modo diversificato, ai fini del processo di divulgazione dei dati competenza dell'Ispettorato. Successivamente le funzionalità conoscitive hanno assunto particolare rilevanza nella fase di controllo e di analisi qualitativa e statistica del dato; e grazie alle nuove soluzioni architetturali, le funzionalità inizialmente dedicate solo all'interno del MEF, sono state rese disponibili anche alle Amministrazioni che alimentano/fruiscono del Sistema informativo gestionale, nel rispetto di profili di utenza differenziati.

Quota parte delle informazioni di competenza dell'Ispettorato, quelle afferenti la programmazione ed attuazione degli interventi di carattere strutturale finanziati dalla Ue, alimentano anche il Data Warehouse R.G.S., ambiente centralizzato destinato ad ospitare e rendere disponibili, mediante applicazioni di supporto alle decisioni, informazioni consolidate provenienti dai sistemi sorgente eterogenei.

E' infatti attraverso il Data Warehouse R.G.S. che ogni Amministrazione può disporre, in maniera rapida e sicura, di una fonte unica e certa del patrimonio informativo dei vari sistemi gestionali della Ragioneria a supporto del proprio processo decisionale. Il Data Warehouse R.G.S. è inteso, infatti, come Sistema costituito da dati, applicazioni, risorse tecnologiche ed umane, regole organizzative, procedure deputate all'acquisizione, elaborazione, memorizzazione, scambio, reperimento e trasmissione delle informazioni. L'utente di tale sistema è messo in grado di soddisfare le proprie specifiche esigenze di analisi, avvalendosi sia di prospetti predefiniti sia di interrogazioni estemporanee, senza dover necessariamente conoscere l'organizzazione logica e fisica della base dati né il suo linguaggio di accesso.

Obiettivi di sviluppo e manutenzione evolutiva

Le attività di sviluppo e manutenzione del sistema conoscitivo, comprendono i seguenti obiettivi:

Enterprise DWH - dati della nuova gestione finanziaria del Fondo di rotazione

Si tratta di implementare l'EDW della Ragioneria Generale dello stato con i dati dei movimenti contabili gestiti nell'ambito dell'IGRUE, sia per quanto riguarda i contributi che affluiscono dalla Unione europea, sia per le quote di partecipazione nazionale.

Enterprise DWH - dati relativi alle quote di partecipazione al Bilancio comunitario

Si tratta di implementare l'EDW della Ragioneria Generale dello Stato con i dati che concorrono alla determinazione delle quote di versamento mensile per l'alimentazione del Bilancio della Unione europea e sulle evoluzioni del bilancio comunitario.

Enterprise DWH - dati relativi ai controlli comunitari

Si tratta di implementare l'EDW della Ragioneria Generale dello Stato con i dati relativi alle verifiche effettuate sul territorio dagli uffici di controllo.

Data mart analitico sui dati di partecipazione al bilancio comunitario

Realizzazione di un Data Mart per IGRUE sui dati che concorrono alla determinazione delle quote di versamento mensile per l'alimentazione del Bilancio della Unione europea e sulle evoluzioni del bilancio comunitario.

Data mart integrato sui dati della gestione finanziaria dell'IGRUE

Realizzazione di un Data Mart di sintesi per l'Ispettore Generale Capo in cui si rendano disponibili, in forma aggregata, le informazioni che caratterizzano la gestione finanziaria dell'Ispettorato, con particolare riguardo alle previsioni di movimentazione nel medio periodo.

Data mart di supporto ai modelli di previsione dei flussi Ue

Data Mart in cui storicizzare le versioni dei dati sulla base dei quali vengono elaborate le previsioni di flussi comunitari, annuali e trimestrali.

Data mart Monitoraggio dei controlli comunitari

Data Mart in cui rendere disponibili i dati relativi al monitoraggio dei controlli comunitari.

Sistema di reportistica orientato alle pubblicazioni

E' esigenza dell'Ispettorato implementare una soluzione tecnica che, a partire dai dati contenuti nei diversi Data mart, consenta di predisporre, in modo flessibile, componenti da utilizzare nelle pubblicazioni, sia in forma tabellare, sia in forma di grafico.

Reportistica sui dati di monitoraggio per gli uffici della Commissione UE

Realizzazione di una soluzione che consenta di mettere a disposizione degli uffici comunitari i dati "ufficiali" di monitoraggio che si consolidano alle scadenze prestabilite, a valle del processo di verifica operato dal MEF sui dati validati dalle Autorità di Gestione.

Introduzione nel Data Warehouse e nei Data Mart esistenti dei concetti di Qualità dei Dati e di Metadati.

Questa attività è finalizzata all'introduzione, nelle applicazioni in esercizio, delle procedure di controllo della Qualità del dato per ottenere una certificazione dell'informazione messa a disposizione dell'utente finale.

E' anche prevista la realizzazione di un sistema di Metadati sia Tecnici che di Business, volto ad essere d'ausilio sia al personale specializzato che gestisce il Data Warehouse, sia all'utente finale per l'utilizzo delle applicazioni e per la sua corretta comprensione dei dati stessi.

3. Schema di riferimento

3.1 Processo del data warehouse

Le caratteristiche peculiari di un sistema conoscitivo richiedono la creazione di un “magazzino” di dati specifico (Data Warehouse), distinto dalle basi dati dei sistemi operazionali.

Secondo una definizione di Ralph Kimball (The Data Warehouse toolkit), un Data Warehouse è “una copia di dati transazionali strutturati specificamente per attività di query e analisi”; più in generale, si può affermare che il datawarehouse è un insieme di dati integrati, ottenuti da fonti interne o esterne, che, a diversi livelli di aggregazione, rappresenta eventi o fatti avvenuti nel tempo e che viene utilizzato per soddisfare le esigenze di conoscenza dell’organizzazione.

Nei sistemi conoscitivi del MEF, la costruzione di un datawarehouse avviene secondo lo schema generale descritto nella figura seguente.

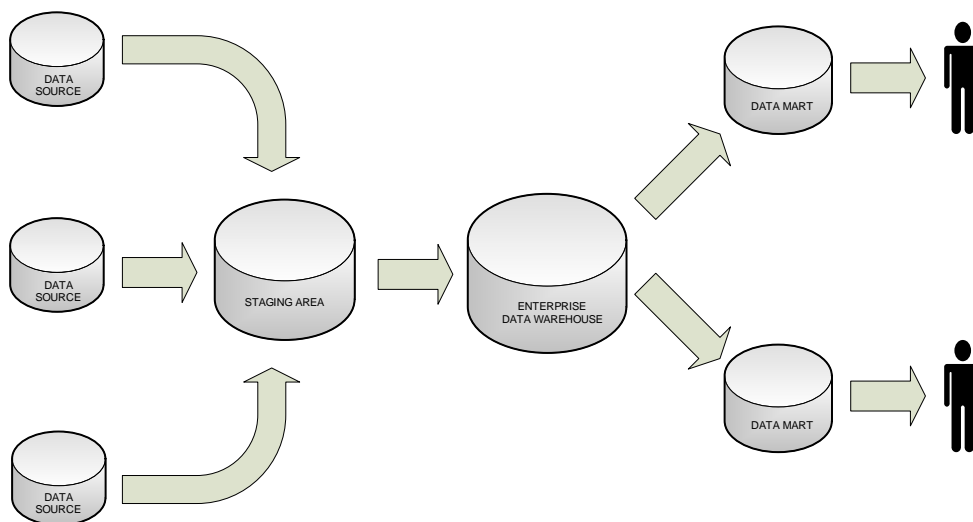


Fig. 1 - Flusso logico informazioni

Le informazioni sono raccolte da varie applicazioni legacy, integrate e storicizzate all’interno dell’Enterprise Data Warehouse (EDW).

L’EDW è caratterizzato da massima granularità, storicità e frequenza di aggiornamento dell’informazione, in accordo con le esigenze di analisi degli utenti.

Ogni EDW alimenta diversi Data Mart (DM); i DM, caratterizzati da dati di sintesi, sono tematici ed orientati a classi di utenza specifiche. Essi filtrano ed aggregano i dati provenienti dall'EDW al fine di soddisfare le esigenze della classe di utenza cui i DM sono destinati. Le informazioni immagazzinate nei DM sono infine rese disponibili all'utente finale tramite gli strumenti di front-end.

3.2 Framework tecnologico

Lo schema di riferimento, o “framework”, tecnologico è sintetizzato nella seguente figura, in cui i componenti del sistema conoscitivo sono rappresentati mediante macro-blocchi funzionali. Tale framework, implementato in Consip nell’ambito dei sistemi conoscitivi, è coerente con quanto rappresentato nel documento di Gartner Group (2001), “Data Warehouse Architecture” .

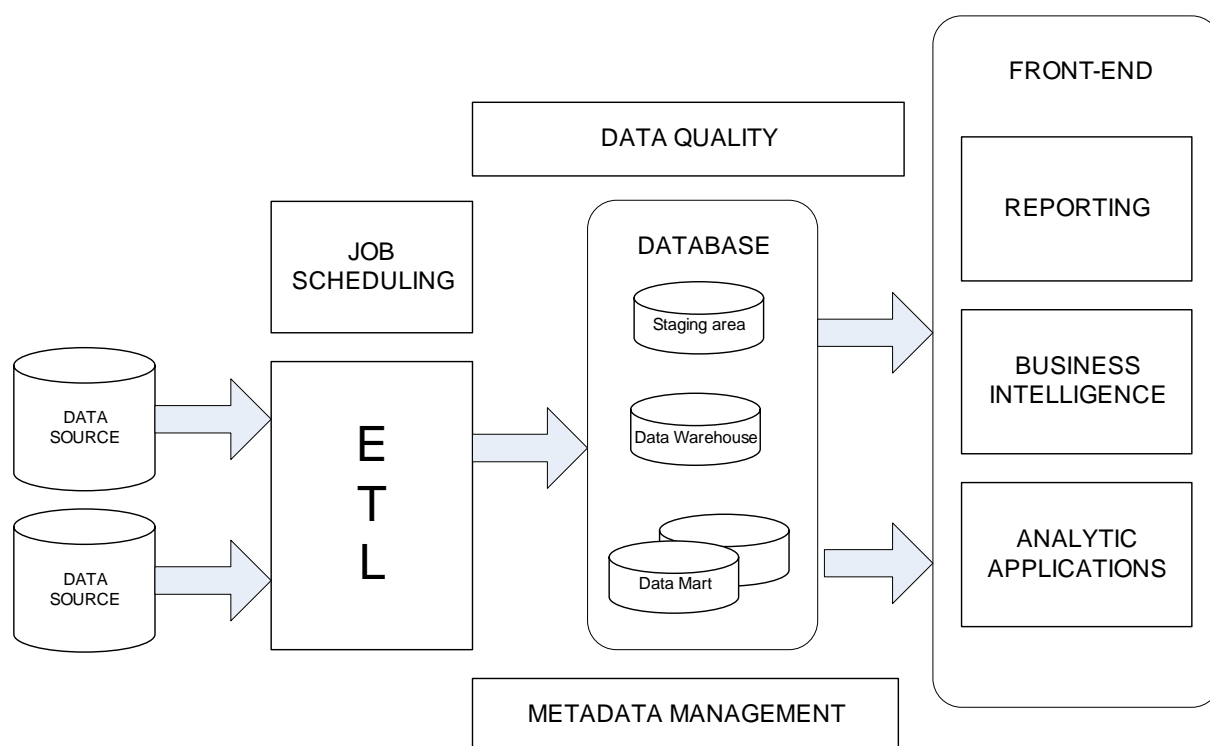


Fig. 2 – Framework tecnologico

3.2.1 ETL

Il processo di ETL (Extract, Transform and Load) è quello mediante il quale i dati vengono estratti dai sistemi gestionali (e in generale da fonti esterne) ed elaborati per renderli disponibili agli utenti finali in una forma adeguata al tipo di interrogazione richiesta.

Nell'ambito del processo sono svolte le seguenti attività:

- **Estrazione dei dati dai sistemi source**

Permette di acquisire dai sistemi operazionali e da fonti esterne i dati utili al sistema conoscitivo. Dopo il primo caricamento, l'estrazione deve essere in grado di rilevare ed estrarre le variazioni dei dati nei database sorgenti.

A seconda dei requisiti delle singole iniziative, le estrazioni successive possono essere attivate dagli eventi di modifica oppure essere eseguite a determinati intervalli di tempo.

- **Cleaning**

I dati estratti vengono raccolti in aree di staging e qui sottoposti a controllo di qualità tecnica e funzionale.

La qualità tecnica riguarda soprattutto la presenza di codifiche disomogenee per dati che hanno lo stesso contenuto informativo ma provengono da sistemi diversi, o formati da convertire tra sistemi operativi diversi.

La qualità funzionale è notevolmente più complessa e coinvolge più direttamente l'utente finale del sistema di Data Warehouse in quanto è l'unico a poter certificarne la correttezza semantica.

Il processo di cleaning che mira alla omogeneizzazione dei dati, può evidenziare carenze e/o incompletezze che inducono un processo di correzione a livello dei dati sorgente.

- **Trasformazione**

L'attività di trasformazione applica alle informazioni estratte e bonificate l'insieme di regole che li rendono rispondenti ai requisiti funzionali e tecnici richiesti dall'EDW e dai DM.

Si eseguono pertanto attività di riorganizzazione e aggregazione di dati.

I dati vengono inoltre arricchiti attraverso elaborazioni fatte rispetto alle diverse dimensioni di aggregazione previste dal disegno dei singoli Data Mart, in particolare per la caratterizzazione della dimensione temporale.

- **Caricamento**

I dati estratti e trasformati vengono caricati sugli ambienti di Data Warehouse o Data Mart per l'accesso da parte degli utenti. La fase di caricamento può avvenire secondo differenti modalità di aggiornamento: inserimento o modifica.

3.2.2 Job scheduling

Sotto l'etichetta di job scheduling vengono raggruppate le tecnologie utilizzate per l'esecuzione programmata di processi (batch); in ambito conoscitivo l'esecuzione batch viene utilizzata prevalentemente per i processi ETL, ma è utilizzata anche, p.es., per l'esecuzione di report predefiniti e per automatizzare l'esecuzione di processi relativi alla qualità.

3.2.3 Database

Nel database sono memorizzate le strutture dati relative ad aree di staging, EDW e data mart, nonché i metadati tecnici (repository Powercenter, BusinessObjects, ecc.).

3.2.4 Data quality management

Con il termine di "data quality management" si indicano i processi attraverso cui si controllano la completezza e l'omogeneità dei dati provenienti dalle diverse fonti; dati inaccurati o incompleti, infatti, si traducono nella presentazione agli utenti finali di informazioni non "veritiere", con ovvie conseguenze sulle decisioni e sui processi di business.

Questa attività si presenta logicamente come un ciclo ininterrotto di profilazione dei dati, integrazione/correzione/arricchimento degli stessi e monitoraggio dell'intero processo; da un punto di vista operativo, essa è integrata con i processi di ETL e può utilizzare tools specifici.

3.2.5 Metadata management

Nella costruzione di un sistema conoscitivo vengono sempre prodotte "informazioni sui dati", comunemente chiamate metadati: queste riguardano, per esempio, la descrizione delle strutture del database, le modalità attraverso cui i dati vengono estratti dai data source, trasformati e caricati nel data warehouse, la catalogazione dei report in relazione ai dati interrogati o ai processi di business cui sono dedicati.

Queste informazioni, se utilizzate e gestite in modo appropriato, sono importanti sotto diversi aspetti, in quanto:

- migliorano la comprensione della realtà da parte degli utenti finali;
- sono di ausilio nei processi di gestione/evoluzione del sistema stesso (analisi di impatto).

Sotto il termine "metadata management" sono compresi gli standard e gli strumenti dedicati alla gestione di questo particolare aspetto.

3.2.6 Front-end

Gli strumenti di front-end sono quelli attraverso cui gli utenti accedono alle informazioni contenute nel sistema.

Nel framework di riferimento riportato in precedenza, gli strumenti di front-end sono suddivisi in tre categorie:

1. reporting;
2. business intelligence;
3. analytic applications.

Questa suddivisione deriva dal fatto che gli strumenti di front-end devono essere utilizzati da una grande varietà di utenti finali, con esigenze conoscitive di tipo diverso; secondo una definizione dell'analista Mike Ferguson, gli utenti e le esigenze conoscitive possono essere raggruppati in tre classi, come schematizzato nella seguente figura:

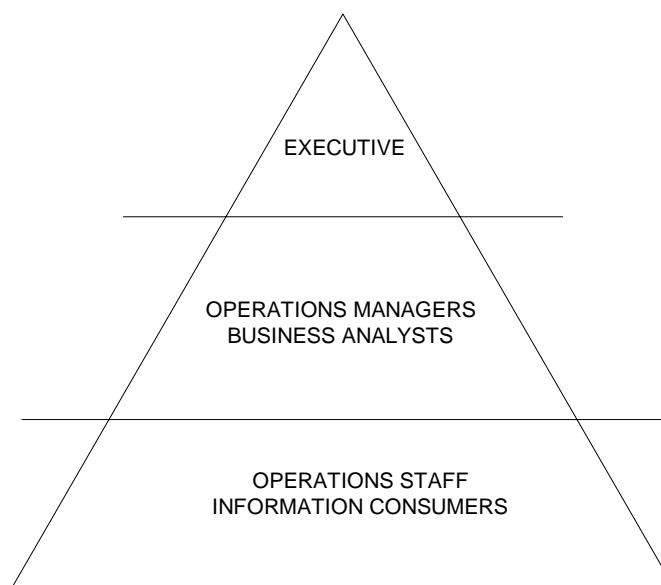


Fig. 6 – Schema esigenze conoscitive

Nella fascia inferiore sono compresi gli utenti che hanno bisogno di informazioni definite, di tipo statico e tipicamente con elevato livello di dettaglio; un'altra caratteristica tipica è quella di richiedere elevati livelli di formattazione del layout di presentazione, orientati alla produzione di stampe (reporting).

Nella fascia intermedia sono compresi gli utenti con requisiti di analisi più elevati; essi tipicamente hanno dimestichezza con gli strumenti di front-end e richiedono la possibilità di

“navigare” tra i dati, selezionando diversi livelli di dettaglio e diversi livelli di aggregazione, per analizzare un fenomeno di business da più punti di vista (business intelligence).

Nella fascia superiore rientrano gli utenti che hanno necessità di una visione sintetica dei fenomeni aziendali, tipicamente attraverso strumenti evoluti di analisi e indicatori raggruppati in un cruscotto (analytic applications).

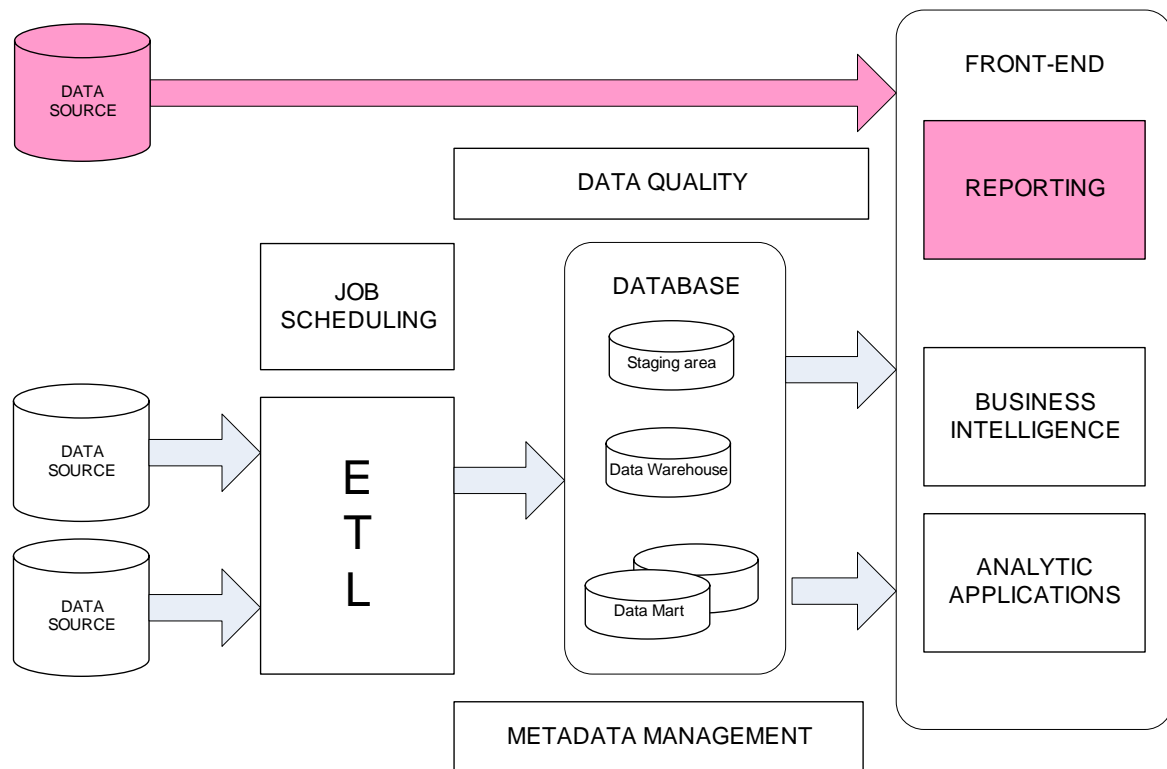
Secondo questa visione, la maggior quantità di utenti e di informazioni prodotte rientra nella fascia inferiore, mentre la percentuale si riduce salendo verso l’alto.

3.2.6.1 Reporting gestionale

Come già detto nel par. 2.1, l’impostazione standard dei sistemi informativi è quella di separare gli ambienti destinati alle attività conoscitive da quelli utilizzati invece per le attività gestionali.

Esiste tuttavia un’“area grigia”, in cui questa divisione non è così netta: nel campo delle attività gestionali, infatti, è comune la necessità di produrre stampe o report; da una parte, queste operazioni sono effettuate con strumenti e prodotti tipici di un ambiente conoscitivo, dall’altra, tuttavia, esse accedono direttamente alle fonti dati, contraddicendo il principio secondo il quale le attività di tipo conoscitivo devono utilizzare una base dati derivata e distinta da quella gestionale.

Questa situazione è evidenziata nella figura seguente, in cui i blocchi e la freccia più scuri rappresentano gli elementi suddetti; lo schema evidenzia l’ambiguità descritta.



Nel seguito del documento si accennerà anche a questa particolare applicazione degli strumenti di tipo conoscitivo; l'avvertenza è comunque quella di considerare il reporting gestionale come un'attività a parte rispetto al sistema conoscitivo vero e proprio.

4. Architettura attuale

Nei successivi paragrafi è descritta la situazione attuale degli strumenti utilizzati nei sistemi conoscitivi MEF e il modo in cui sono stati implementati nello sviluppo dei progetti applicativi.

Inizialmente viene fornito un quadro sintetico dei processi di consolidamento generale delle infrastrutture elaborative, quindi vengono trattati in dettaglio i vari componenti del sistema conoscitivo; la descrizione segue lo schema rappresentato nel framework tecnologico precedente.

4.1 Infrastrutture di consolidamento

Sottosistemi Storage Area Network

Nei CED che ospitano i sistemi elaborativi del MEF sono stati recentemente introdotte nuove infrastrutture di Storage Area Network; tali sottosistemi garantiscono una grande efficienza nella gestione dello storage richiesto dai sistemi di elaborazione e consentono di rispondere con la massima flessibilità alla crescente richiesta di risorse da parte delle applicazioni: per i nuovi server da acquisire, infatti, sarà possibile dimensionare in modo minimale i dischi interni, mentre si potrà acquisire lo storage necessario nell'ambito del sottosistema SAN.

Consolidamento applicazioni gestionali Websphere/Oracle II dipartimento

Il consolidamento delle applicazioni gestionali del II dipartimento, iniziato con l'installazione del sistema MIZAR, proseguirà con l'acquisizione di un server di caratteristiche prestazionali analoghe e con lo stesso sistema operativo (AIX). I due sistemi saranno configurati in bilanciamento di carico, sfruttando le tecnologie di Oracle 10g RAC e di Websphere AS.

Su questa infrastruttura vengono consolidate le applicazioni e i database gestionali in origine ospitati sui server ERA, ZEUS e sul sistema Superdome (istanza GEST su SDPART2).

Consolidamento server Microsoft CED Rustica

E' previsto un processo di consolidamento dei server Microsoft installati nel CED di Rustica (81 server censiti). Questo processo prevede di trasferire dati e applicazioni su pochi server di taglia medio-grande, per ottimizzare l'utilizzo delle risorse elaborative e ridurre i costi di manutenzione e gestione.

Saranno acquisiti alcuni server con funzioni di DB server in tecnologia Microsoft e altri destinati ad ospitare il workload applicativo; per questa seconda categoria si prevede di

utilizzare il prodotto VMware ESX per realizzare, sui server fisici, delle “macchine virtuali” nelle quali isolare le diverse applicazioni.

Questo consolidamento coinvolge per il momento i server del progetto SICO (Microstrategy) ed esclude i server del progetto “cruscotto FP” (di acquisizione recente) e i server BusinessObjects (oggetto di consolidamento specifico, come descritto in seguito).

Razionalizzazione server Microsoft CED XX settembre

E' previsto un processo di razionalizzazione dei server installati nel CED di via XX settembre, con l'obiettivo di sostituire gli apparati non adeguati ed eliminare le situazioni di “single point of failure”.

Si prevede l'utilizzo di server di tipo Blade, che consentono l'abbattimento degli spazi occupati dai server, cosa molto importante per i problemi di spazio presso il CED del Centro Comunicativo.

Il progetto coinvolge tre server BusinessObjects (G2, G10, WEBIGPB), che saranno migrati su server di tipo blade.

4.2 ETL

Il prodotto ETL utilizzato nell'infrastruttura DW del MEF è Power Center della società Informatica. La piattaforma è installata sui sistemi ERA (esercizio) e ZEUS (sviluppo/collauda/manutenzione), collocati entrambi nel CED di Rustica. I server ERA e ZEUS sono sistemi IBM S80 con sistema operativo AIX.

La figura seguente riporta uno schema di massima dell'architettura del prodotto.

Architettura Powercenter

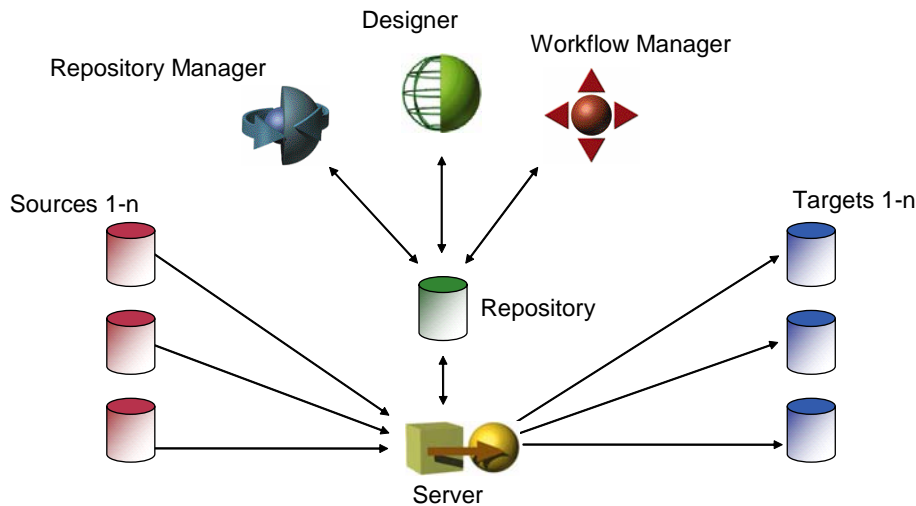


Fig. 4 - Architettura Powercenter

I componenti fondamentali sono:

- i tool client, utilizzati per lo sviluppo e la gestione del sistema; si utilizzano in particolare:
 - designer, con il quale si creano le definizioni dei source e dei target e i “mappings” che descrivono le regole di trasformazione; tutte queste informazioni sono memorizzate nel repository;
 - repository manager, con il quale si gestiscono i metadati contenuti nel repository (mappings, utenti, gruppi, privilegi, ecc.);
 - workflow manager, con il quale si creano, si schedulano, si eseguono e si controllano le sessioni batch di trasformazione;
- il repository, un database nelle cui tabelle sono memorizzati i metadati prodotti dai tool client e utilizzati dal server;
- il server, ovvero il “motore” che estrarre i dati dai source, applica le regole di trasformazione e carica i dati trasformati nei target, basando la sua attività sui metadati del repository.

La connettività tra i diversi componenti si realizza via ODBC o con driver nativi; in particolare, per la connessione tra il server e i data source, i driver nativi disponibili nel MEF sono quelli per Oracle e per DB2 (attraverso un componente specifico di connessione su sistema mainframe).

Le attività di realizzazione dei processi di ETL si articolano sui quattro ambienti canonici: sviluppo, correzione, collaudo ed esercizio.

Le politiche architetturali si diversificano da ambiente ad ambiente, in relazione soprattutto alle politiche di licenza del prodotto; in particolare:

- per gli ambienti di sviluppo/correzione si prevede un rapporto 1:1 tra “motore” e repository ETL, cioè in questi ambienti ogni progetto ha un proprio motore e un proprio repository. Lo sviluppatore ha piena libertà di modificare ed eseguire i propri mapping.
- Per l’ambiente di collaudo si utilizza un unico motore ed un unico repository; in quest’ambiente non è consentito l’editing degli oggetti e il lancio dei lavori viene effettuato dal personale di gestione tecnica tramite richiesta.
- Per l’ambiente di esercizio si utilizzano due motori e due repository; questa configurazione consente un bilanciamento di carico ed un grado di parallelizzazione dei lavori fino a 200 sessioni in parallelo. La schedulazione dei lavori ETL avviene mediante due meccanismi: TWS nella maggior parte dei casi e “Crontab”, specialmente per le schedulazioni estemporanee.

4.3 Job scheduling

Le tecnologie di job scheduling utilizzate nel MEF sono:

- IBM Tivoli Workload Scheduler (TWS);
- tool di schedulazione integrati nei sistemi operativi Unix e Windows.

Il Tivoli Workload Scheduler (TWS) è un completo sistema di schedulazione multi-piattaforma e multi-livello; la sua architettura è descritta nella figura seguente.

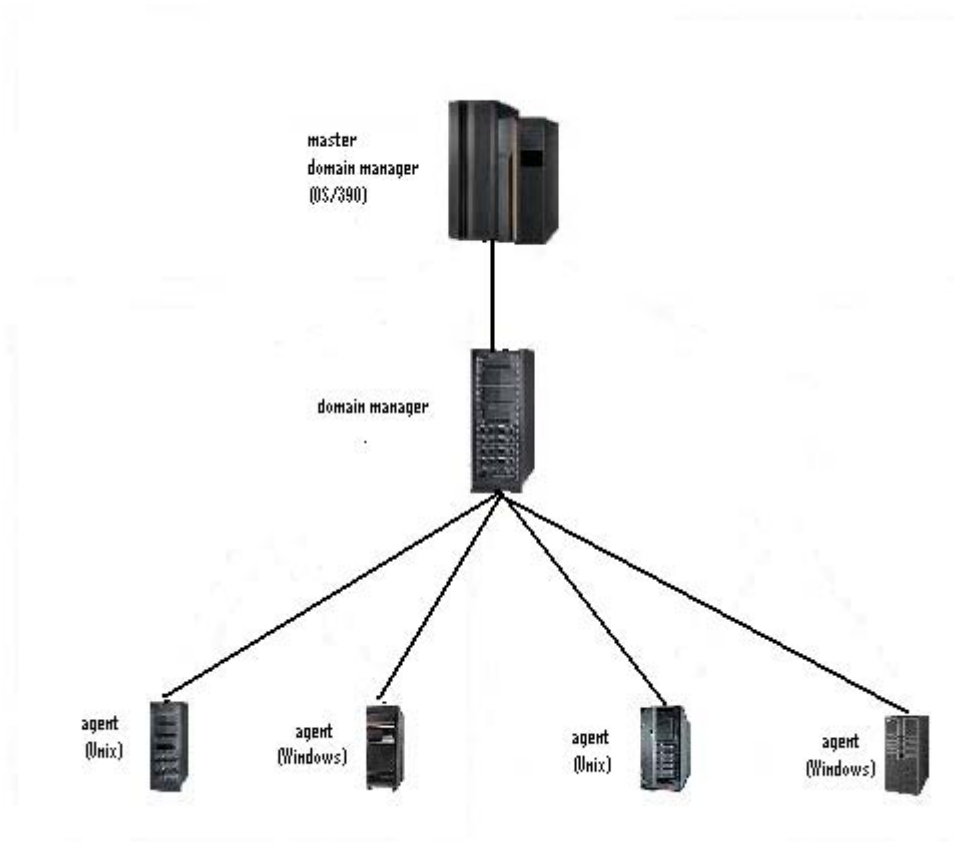


Fig. 5 - Architettura TWS

La definizione e il controllo della schedulazione vengono realizzate mediante un “master domain manager” centralizzato, ospitato su sistema mainframe; attraverso un “domain manager” intermedio ospitato su un sistema dipartimentale (in realtà ne sono presenti due, uno principale e uno di backup), la schedulazione viene quindi propagata in modo coordinato a tutti i sistemi Unix e Windows sui quali sono installati gli “agent” che eseguono i job batch.

Il TWS consente sofisticate funzioni di schedulazione, che comprendono, tra le altre:

- la definizione di criteri temporali per la schedulazione dei job;
- la definizione di vincoli di esecuzione dei job, del tipo “esegui B quando termina A”;
- il coordinamento tra sistemi differenti (come il caso precedente, con A e B eseguiti su sistemi distinti);
- la gestione dell’esito dei job (corretto o in errore);
- funzioni di recovery/restart dei job;
- una console di gestione centralizzata per il controllo dei job eseguiti su tutti i sistemi.

Il TWS viene correntemente utilizzato per la schedulazione di batch ETL, di batch relativi a processi di qualità e per la produzione di report Actuate.

I tool di schedulazione integrati nei sistemi operativi Unix (CRONTAB) e Windows forniscono capacità di schedulazione rudimentali (tipicamente limitate ad un esecuzione periodica basata su giorno e ora), con scarse possibilità di controllo e monitoraggio. Ogni sistema gestisce la propria schedulazione in modo indipendente. Questi strumenti sono utilizzati in alcuni casi per la schedulazione di batch ETL.

4.4 Database

Il database utilizzato per le esigenze specifiche dei sistemi conoscitivi è Oracle.

I dati sono distribuiti su due siti: il CED di Rustica e il CED del centro comunicativo di via XX settembre.

Nel CED di Rustica sono disponibili due sistemi:

- ZEUS per le attività di sviluppo/collauda/manutenzione;
- ERA per le attività di esercizio.

Le istanze Oracle definite sui due sistemi sono:

server ERA:

- | | |
|----------|----------------|
| • DWHE | data warehouse |
| • DWERGS | data warehouse |
| • DTME | data mart |
| • DTME2 | data mart |
| • REPE | repository ETL |

Server ZEUS:

- | | |
|---------|----------------|
| • DWHS | data warehouse |
| • DWHS2 | data warehouse |
| • DTMS | data mart |
| • DTMS2 | data mart |
| • REPS | repository ETL |
| • REPS2 | IGED |
| • REPO | repository ETL |

Nel CED del centro comunicativo sono presenti due sistemi di esercizio, Leonardo e Archimede (server IBM H70 con sistema operativo AIX), che ospitano data mart di progetti specifici e repository BusinessObjects.

In particolare, per quanto concerne il server Archimede, sono presenti due istanze Oracle, una dedicata al sistema gestionale PIT ed una -DMARC- sulla quale sono residenti le basi dati dedicate ai seguenti progetti:

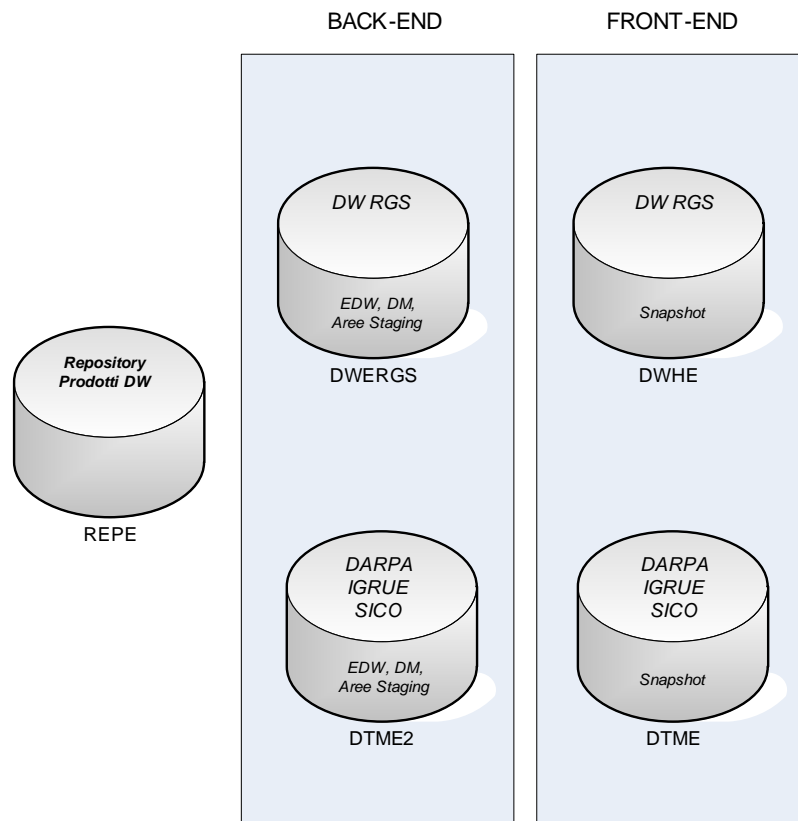
- Spese di Lite;
- Banca dati del personale;
- Area CEE;
- Investimenti Pubblici;
- Scambio Dati

Il server Leonardo presenta invece una sola istanza Oracle (DMLEO), dedicata ai progetti:

- CA;
- IGESPES;
- Winter;

- DbContatti.

Attualmente la disposizione dei processi conoscitivi sulle istanze Oracle del sistema ERA è in fase di razionalizzazione e di riordino. La configurazione a regime delle istanze RDBMS sarà suddivisa logicamente in due tipologie: Back-End (base dati non interrogabile) costituita dalle Aree di staging, EDW e dai DM; Front-End costituita dai cosiddetti “snapshot” consultabili:



Il naming delle istanze è puramente indicativo; infatti è stato mantenuto quello delle istanze attuali per limitare il più possibile le modifiche dei datasource powercenter, delle shell unix e delle configurazioni dei software di front-end (Business Objects, Microstrategy).

4.5 Data quality management

Nei processi ETL sviluppati con Powercenter vengono utilizzate funzioni base di data cleansing, che consentono di standardizzare, correggere e validare i dati.

Nell'ambito del Data Warehouse RGS, l'area tematica Finanza Pubblica ha sviluppato procedure di controllo della qualità dei dati utilizzando il prodotto IBM/Ascential AuditStage.

Il prodotto AuditStage opera in ambiente Windows ed è stato installato sui seguenti server per l'ambiente di esercizio:

- CONSVIL
- NSC_LR
- LR_BO1
- LR_BO2

E' inoltre presente un'installazione sul server GIOVE per l'ambiente di collaudo.

I processi sono eseguiti in modalità batch per mezzo dello schedatore TWS.

4.6 *Metadata management*

Nella realizzazione del Data Warehouse della RGS, l'area tematica Finanza Pubblica ha sviluppato una metodologia per la gestione dei metadati e un sistema di gestione dei metadati a supporto della stessa. Il sistema di gestione dei metadati è stato costruito attraverso l'integrazione di un prodotto commerciale per la gestione dei metadati (IBM/ Ascential MetaStage) con componenti realizzate ad-hoc.

L'architettura del sistema consente di integrare i metadati prodotti dai tool Erwin, Powercenter e BusinessObjects. Questi metadati sono quindi pubblicati su portale per la navigazione da parte degli utenti del sistema conoscitivo.

4.7 Front-end

Considerando la classificazione degli strumenti di front-end descritta in precedenza, per il MEF sono state sviluppate soluzioni che rientrano in tutte e tre le categorie suddette; gli strumenti utilizzati sono:

- reporting
 - Actuate.
 - Oracle Reports;
 - Oracle Financial Analyzer (OFA);
- Business intelligence
 - BusinessObjects;
 - Microstrategy;
 - Oracle Discoverer;
- Analytic applications
 - BusinessObjects(Application Foundation)
 - Microstrategy
 - pacchetti statistico/econometrici (SAS, Troll, ecc.)

Nei successivi paragrafi viene fornita una breve descrizione delle tecnologie suddette e dei progetti che le hanno utilizzate.

4.7.1 BusinessObjects

L'architettura generale della piattaforma BusinessObjects è rappresentata nella seguente figura.

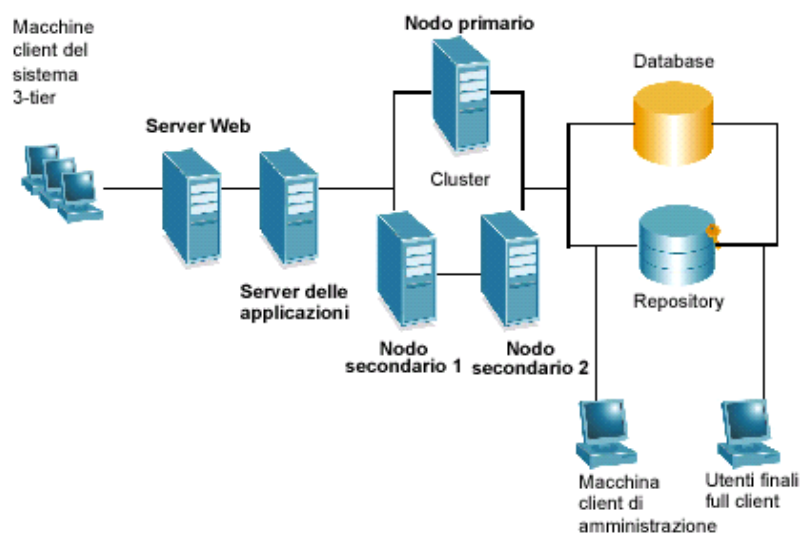


Fig. 7 – Architettura BusinessObjects

La soluzione BusinessObjects è stata utilizzata in diversi progetti, sia in modalità client/server a due livelli (client – database), sia in modalità web multi-tier (browser – web server/application server/nodi BO – database).

L'accesso ai dati è filtrato attraverso uno strato semantico (universo) che fornisce una visione dei dati orientata alle misure e alle dimensioni; su questi universi si costruiscono i report. Universi e report sono memorizzati nel repository.

Nell'architettura a più livelli il carico elaborativo si può distribuire tra più nodi (cluster).

La logica prevalente in queste installazioni è stata in passato quella di realizzare infrastrutture dedicate per i diversi progetti. Ciò ha portato alla proliferazione di installazioni server su numerose macchine di taglio medio-piccolo (mono o bi-processor) in modalità "single-node".

La maggior parte di progetti realizzati utilizzano come interfaccia web il portale standard della piattaforma (Infoview), che consente la creazione di report ad hoc, la scelta da lista e l'esecuzione di report predefiniti, l'invio di report ad altri utenti del sistema, ecc.

Per alcuni progetti è stata realizzata un'interfaccia web personalizzata, utilizzando le potenzialità di sviluppo del tool Dashboard Manager fornito con il componente Application Foundation. Il componente Application Foundation è stato inoltre utilizzato per la realizzazione di cruscotti decisionali.

Per alcuni progetti è stato infine utilizzato il componente Broadcast Agent, che consente la schedulazione periodica dei report, al fine di presentare all'utente i dati aggiornati senza che questi debba richiederlo esplicitamente.

Per i più recenti progetti conoscitivi del II dipartimento è stata realizzata un'infrastruttura BusinessObjects consolidata su sistemi Windows e condivisa da tutti i progetti, che prevede l'accesso via web sia per gli utenti intranet, sia per gli utenti internet; la configurazione logica del sistema è rappresentata nella figura seguente: i nodi primari del sistema BO sono installati sui server LR-BO1 e LR-BO2 (CED di Rustica), ai quali gli utenti intranet accedono direttamente dalla LAN interna; due nodi secondari, installati sui server TEVERE e TAMIGI (CED XX settembre), consentono l'accesso web da internet e fungono da tramite verso i nodi primari suddetti. I local director garantiscono il bilanciamento del carico elaborativo sui nodi.

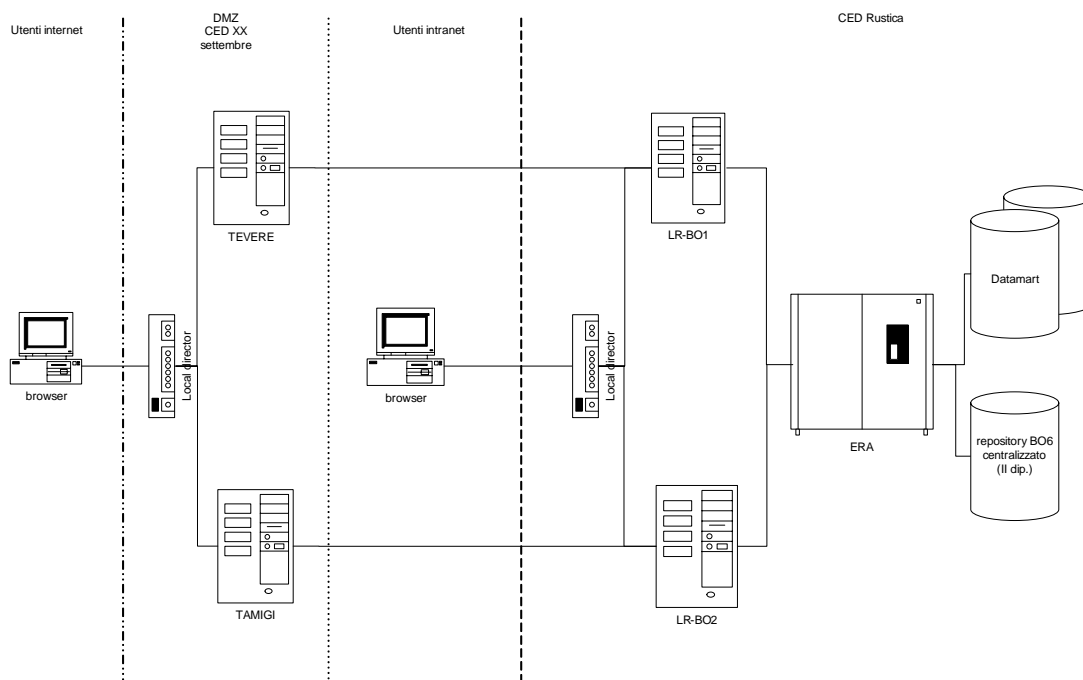


Fig. 11 - infrastruttura BusinessObjects II dipartimento

Gestione ambienti

Il fatto che un'infrastruttura tecnologica sia condivisa da più applicazioni, rende necessario definire ambienti elaborativi in cui i singoli progetti possano lavorare in completa autonomia e senza interferenze; questo si realizza principalmente con un'opportuna strutturazione del repository.

Nel sistema di esercizio, il repository multi-progetto è organizzato secondo un'architettura distribuita, caratterizzata da un singolo dominio di sicurezza e da più domini degli universi e dei documenti, come nella figura seguente.

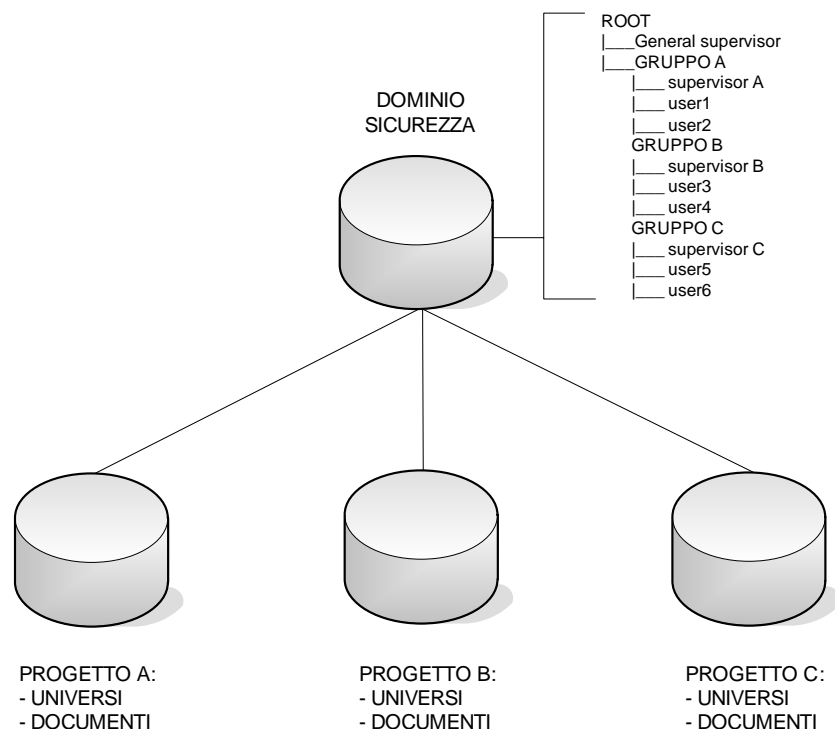


Fig. 12 - schema repository BusinessObjects

Più precisamente, vengono creati un dominio degli universi e un dominio dei documenti per ogni progetto di Data Mart. Nel dominio di sicurezza vengono definiti gruppi specifici per progetto, con i relativi utenti; ogni gruppo contiene inoltre un utente con privilegi di gestione delle risorse del progetto stesso.

Per gli ambienti di sviluppo, collaudo e manutenzione si definisce uno schema analogo.

Per trasferire gli oggetti verso l'ambiente di esercizio devono invece essere utilizzate tecniche diverse:

- gli oggetti sviluppati con tool client (universi o report BO) utilizzano il client stesso come tramite per la pubblicazione verso il repository di esercizio;
- gli oggetti sviluppati con interfaccia web (report webintelligence) devono invece essere trasferiti in tre fasi (esempio da collaudo ad esercizio):
 1. copia del report dal repository di collaudo alla cartella personale dell'utente supervisor di gruppo;
 2. copia fisica del report dalla cartella del sistema di collaudo alla cartella corrispondente del sistema di esercizio;
 3. pubblicazione del report nel repository di esercizio.

Controllo accessi

Il meccanismo base di controllo accessi della piattaforma BO prevede l'uso del repository interno di sicurezza, in cui sono memorizzati gli utenti, le password e i privilegi di accesso. L'accesso al sistema richiede quindi una fase di identificazione/autorizzazione, in cui l'utente fornisce al sistema le sue credenziali.

Caratteristica fondamentale dei sistemi informatici MEF è invece quella di garantire ad un utente l'accesso a tutte le applicazioni autorizzate, attraverso un'unica fase di identificazione iniziale a cui fa seguito la propagazione automatica e trasparente delle credenziali a tutti i sistemi acceduti (single signon). Questo meccanismo si basa sul SSO server Oracle, che gestisce in modo centralizzato le utenze.

Per la piattaforma BO il meccanismo di single signon è stato realizzato sfruttando la feature "external application" del SSO server: l'applicazione BO viene definita come external application al SSO server e viene poi creata la tabella di corrispondenza tra le utenze del SSO server e le utenze/password del repository BO; vengono quindi create delle pagine JSP di login e di logout personalizzate rispetto a quelle standard BO.

Per consentire il single signon, infine, la URL standard di accesso al portale BO viene sostituita con una opportuna chiamata al SSO server: questo provvede ad effettuare la fase di autenticazione in modo trasparente per l'utente, sfruttando la login personalizzata.

4.7.2 Microstrategy

Nella figura seguente è presentata l'architettura generale della piattaforma Microstrategy.

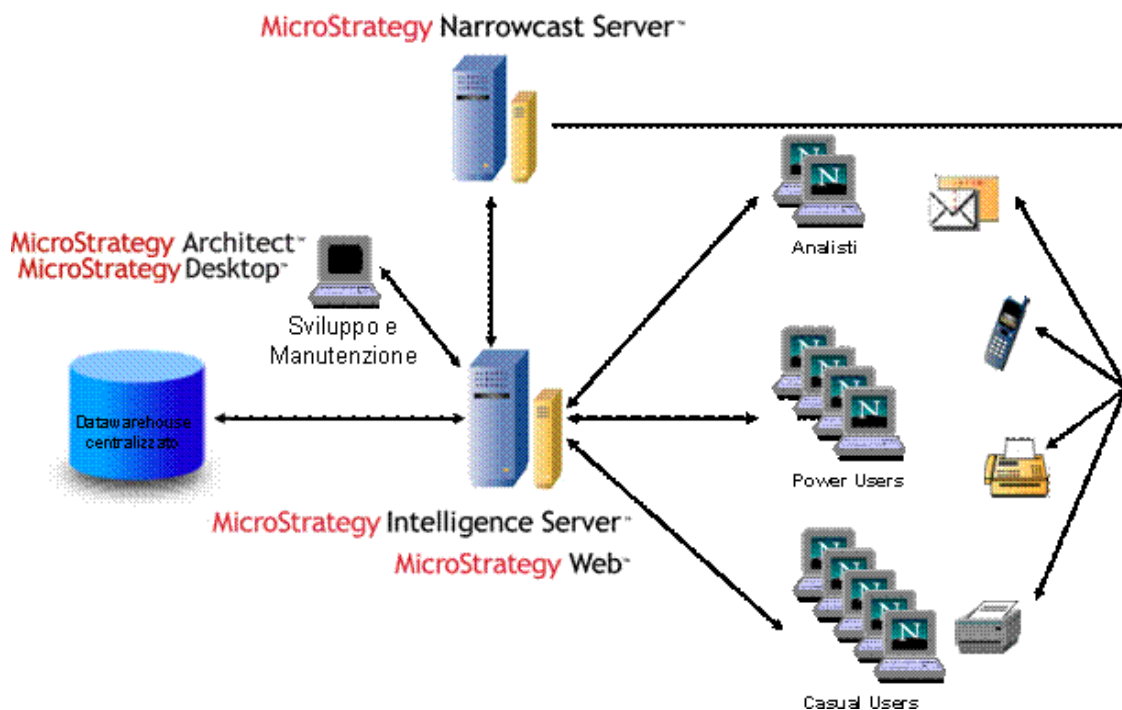


Fig. 8 – Architettura Microstrategy

L'architettura del sistema Microstrategy è analoga a quella del sistema BusinessObjects; la differenza fondamentale consiste nel fatto che Microstrategy non prevede la configurazione client-server a due livelli, ma esclusivamente configurazioni multi-tier con interfaccia web.

Il sistema è stato utilizzato per diversi progetti, che prevedono la realizzazione di applicazioni di business intelligence e cruscotti direzionali.

I progetti interessati sono attualmente:

- Sistemi decisionali direzionali (cruscotto Finanza Pubblica);
- Pubblico Impiego (SICO);
- Cruscotto IV dipartimento (area tematica controllo di gestione).

Ciascuno di questi ha utilizzato una propria infrastruttura, basata su server Windows:

- per il progetto cruscotto FP sono utilizzati i server CASTORE, POLLUCE, ALTAIR, ANTARES, RIGEL;
- per il progetto SICO sono utilizzati i server SICOSVIL e SICOPROD;
- per il cruscotto IV dipartimento si utilizza il server SVILCRU.

4.7.3 Suite Oracle

Della suite di prodotti Oracle dedicati ai sistemi conoscitivi, nei progetti del MEF sono stati utilizzati i componenti Discoverer , Reports ed OFA (Oracle Financial Analyzer).

4.7.3.1 Oracle Discoverer

Nella figura seguente è riportato uno schema della piattaforma Discoverer.

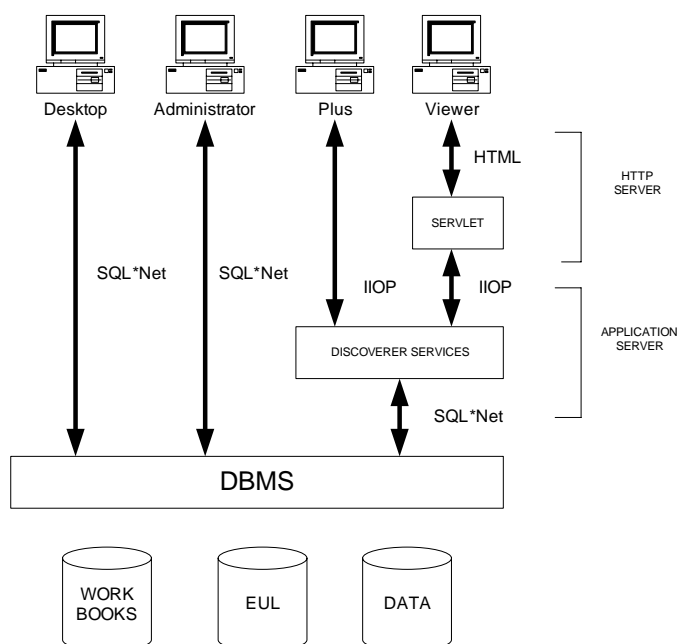


Fig. 9 – Architettura Oracle Discoverer

Con le funzioni di Administrator si generano i metadati (End User Layer) che nascondono all'utente la struttura fisica del database attraverso una visione logica orientata alle misure e alle dimensioni; con le funzioni desktop (client-server) o Plus/Viewer (web) si eseguono le analisi; queste ultime possono essere salvate nel work book per successive elaborazioni o per distribuirle ad altri utenti.

Il prodotto è utilizzato in alcuni progetti, sia in modalità client-server che in modalità multi-tier; la modalità client-server è utilizzata nel progetto BDDP; i componenti di application server e di http server sono installati sui server ERA (esercizio) e ZEUS (sviluppo/collauda/mantenimento) per il progetto "controllo di gestione" e su server Windows dedicati per il progetto SICO.

4.7.3.2 Oracle Reports

Il tool Oracle Reports è utilizzato per la reportistica di applicazioni gestionali e non rientra nell'ambito dei sistemi conoscitivi veri e propri; viene qui descritto solo per completezza di esposizione.

Attraverso gli Oracle Report Services è possibile sviluppare applicazioni di carattere gestionale sia in modalità client server, sia in Java per applicazioni multi-tier.

Questa soluzione è indicata nella produzione di report predefiniti, in forma tabellare non finalizzata all'analisi multidimensionale ma piuttosto alla produzione di stampe dettagliate in funzione dei parametri specifici.

Il tool di sviluppo di Oracle Reports mette a disposizione un Wizard per la creazione di report, in cui l'estrazione dei dati è programmata accedendo direttamente ai data source e costruendo i relativi comandi SQL.

Al runtime, in formato client-server o su piattaforma web (mediante Oracle Reports Server), l'utente ha a disposizione una interfaccia grafica per l'impostazione dei parametri e la generazione del report (predefinito) di interesse.

Il server Oracle Reports è installato sul sistema Superdome e sui sistemi ERA e ZEUS.

Su Superdome è utilizzato da applicazioni realizzate con "Oracle Applications" (SIAP, Controllo Gestione, Contabilità Economica) e dall'applicazione gestionale NSF (area IGRUE).

Su ERA e ZEUS è utilizzato dalle applicazioni gestionali Monitweb (area IGRUE) e GEBEMO per la costruzione di report che accedono direttamente alla base dati operativa.

4.7.3.3 OFA

Questo prodotto, appartenente alla suite delle "Oracle Application", viene utilizzato attualmente solo per produrre le stampe di Contabilità Economica. Tali stampe vengono realizzate utilizzando OFA a partire da una base dati multidimensionale (cubi) ospitata dal RDBMS Express di Oracle.

4.7.4 Actuate

Il tool Actuate è utilizzato per la reportistica di applicazioni gestionali e non rientra nell'ambito dei sistemi conoscitivi veri e propri; viene qui descritto solo per completezza di esposizione.

La caratteristica fondamentale del prodotto Actuate, attualmente utilizzato nei progetti Bilancio, Spese, CdC e Conto Patrimonio, è quella di consentire la produzione di report con elevate esigenze di formattazione.

Il sistema opera in architettura web a più livelli e prevede l'accesso diretto ai data source via SQL, senza l'utilizzo di strati semantici intermedi.

Per i progetti che utilizzano Actuate sono state predisposte due infrastrutture tecnologiche:

- per il progetto Bilancio il prodotto è installato sui server NBNODE1 e NBNODE2 (esercizio), NODE1 (collaudo) e NODE2 (manutenzione); i report vengono prodotti a partire da dati Oracle e DB2 (via DB2Connect);
- per gli altri progetti, il prodotto è installato sui server Afrodite (esercizio) e ZEUS (collaudo/manutenzione); anche questi accedono a dati Oracle e DB2.

4.7.5 Pacchetti statistico/econometrici

L'unità organizzativa "Modelli di Previsione" svolge, all'interno della Consip, attività di consulenza specialistica in materia di sistemi di supporto alle decisioni per conto del Dipartimento della Ragioneria dello Stato e del Dipartimento del Tesoro.

Tale attività è finalizzata principalmente alla costruzione di modelli di previsione, sia di breve che di lungo periodo, funzionali ai compiti istituzionali del MEF.

Per lo sviluppo dei modelli di previsione vengono utilizzati i seguenti pacchetti statistico/econometrici:

- SAS
- Troll
- QMS Eviews
- OUP Microfit
- StataCorp Stata 8/SE
- Estima Winrats-32.

Le attività di sviluppo e gestione dei modelli e degli altri progetti di supporto alle decisioni sono effettuate utilizzando prevalentemente personal computer collegati in rete fra di loro in ambiente Microsoft Windows.

L'hardware e il software utilizzati, nonché le applicazioni sviluppate dall'unità "Modelli di previsione", risiedono su macchine installate all'interno degli uffici della Consip. Inoltre, il software prodotto non viene dato in gestione all'Amministrazione, né si utilizzano applicazioni, sistemi o spazi di memoria presenti nei centri di elaborazione dati del MEF, ad eccezione di alcuni flussi di dati, di proprietà di altre applicazioni, utili alla alimentazione delle basi dati dei modelli. In particolare, i flussi in questione provengono principalmente dalle interrogazioni del sistema conoscitivo del Bilancio dello Stato (Business Object - Webintelligence) e dalle banche dati dei seguenti ispettorati:

- IGESPES
- IGRUE
- IGOP.

Si utilizzano inoltre informazioni provenienti da fonti dati esterne, tra cui INPS, ISTAT, IRPET, Banca d'Italia, MIUR; queste informazioni sono acquisite in formato cartaceo, come fogli excel, e-mail, via web, ecc.

I dati elaborati vengono forniti al MEF in formato excel o cartaceo; gli utenti finali non dispongono quindi di strumenti di front-end specifici.