

CAPITOLATO TECNICO

Servizio di upgrade del Sistema nazionale per l'elaborazione e la diffusione di dati climatici (SCIA)

1. Oggetto dell'affidamento

Oggetto dell'affidamento è l'adeguamento (upgrade) di alcune componenti software del Sistema Nazionale per l'Elaborazione e la Diffusione di Dati Climatici (SCIA), sviluppato e gestito dall'Ispra.

La necessità di adeguare il sistema SCIA deriva da diversi fattori: l'incremento e l'evoluzione delle fonti di dati; l'incremento della varietà e complessità delle variabili e dei prodotti climatici; l'evoluzione delle tecnologie hardware e software e di trasferimento o condivisione dei dati. Una particolare necessità deriva dall'avvenuto trasferimento del database da ambiente Oracle in ambiente PostgreSQL.

Le componenti del sistema soggette ad upgrade sono descritte nei capitoli che seguono. I requisiti generali del nuovo sistema, a cui dovranno rispondere tutte le componenti sviluppate attraverso l'affidamento, sono elencati e descritti nel Cap. 4.

L'Ispra è il soggetto unico titolare del sistema SCIA e della sua gestione; la configurazione attuale del sistema è totalmente e unicamente a carico del personale dell'Ispra e il personale stesso fornirà il supporto e la documentazione utili e necessari agli sviluppi oggetto del servizio.

Glossario:

Variabile climatica (o variabile): elemento che descrive il clima (es. temperatura, precipitazione)

Dato: valore numerico della variabile o dell'indicatore

Dati di ingresso: dati acquisiti dalle diverse fonti (orari o giornalieri prima del caricamento sul database)

Indicatore: nell'ambito del presente documento, si definisce convenzionalmente "indicatore" il risultato delle procedure di controllo, aggregazione, elaborazione del dato di ingresso; l'insieme degli indicatori costituisce il contenuto informativo del database del sistema (per es.: precipitazione cumulata giornaliera, decadale, mensile, annuale; temperatura massima assoluta giornaliera, decadale, mensile, annuale).

Stazione di monitoraggio: stazione fisica di osservazione e misurazione dei dati climatici

Fonte di dati: soggetto che rende disponibili, direttamente o attraverso il web, i dati di un insieme di stazioni di monitoraggio di una determinata rete

Rete di monitoraggio: insieme di stazioni di monitoraggio i cui dati distribuiti sono resi disponibili da una fonte.

Stazioni di monitoraggio 'duplicate': stazioni che fisicamente possono essere riconducibili, in un certo periodo di tempo, ad un'unica stazione, ma originano da fonti o reti diverse.

Serie duplicate: serie temporali di origine diversa o riferite a metadati diversi, che vengono riconosciute appartenenti alla stessa stazione di monitoraggio

GUI: Graphical User Interface - Interfaccia grafica che utilizza l'utente per eseguire procedure, effettuare ricerche, modifiche, inserimenti sulla base dati ecc.

Provider: fornitore o fonte dei dati di ingresso

SCIA_DB_GIORN: istanza Oracle contenente metadati ed indicatori giornalieri

SCIA_DB: istanza Oracle contenente metadati ed indicatori decadali, mensili e annuali

ADM_CLIMA: schema Oracle contenente i metadati

ANPA_CLIMA: schema Oracle contenente gli indicatori

SCIA_GG: schema PostgreSQL contenente gli indicatori giornalieri

SCIA_DMA: schema PostgreSQL contenente gli indicatori decadali, mensili e annuali

2. Breve descrizione del sistema

Il "vecchio" sistema è costituito dai seguenti elementi:

- un database relazionale in Oracle versione 10g r2 che viene alimentato da una serie di provider istituzionali;
- una serie di procedure e di programmi di caricamento, calcolo, validazione, controllo degli indicatori;
- un sito web dedicato (www.scia.isprambiente.it) che consente di visualizzare e scaricare diverse tipologie di indicatori e di prodotti climatici, mediante l'uso di un certo numero di funzioni interattive.

Come già detto, il database è stato recentemente trasferito da ambiente Oracle in ambiente PostgreSQL. Ciò comporta l'adeguamento delle procedure descritte nei capitoli seguenti e l'ottimizzazione della struttura logica e fisica del database, già in parte eseguita in house, ma che dovrà essere eventualmente adattata, in seguito allo sviluppo delle nuove procedure.

Il database viene aggiornato con cadenza annuale, generalmente nei mesi di febbraio e marzo dell'anno successivo all'ultimo anno di riferimento dei dati. Le serie temporali originano da varie fonti, elencate sul lato destro della home page del sito (www.scia.isprambiente.it). Sia le serie di dati di ingresso che quelle di indicatori vengono sottoposte a procedure standardizzate di controlli di validità, prima e dopo il caricamento sul database.

Viene quindi eseguita una procedura che in base ad una tabella di corrispondenza tra stazioni (compilata a cura di personale interno) assegna un unico codice stazione a serie caricate con codice stazione diverso, riconosciute appartenenti alla stessa stazione di monitoraggio.

I database Oracle sono:

- ✓ SCIA_DB_GIORN
- ✓ SCIA_DB

Il primo contiene i metadati e gli indicatori giornalieri, mentre il secondo contiene metadati ed indicatori decadali, mensili e annuali.

Ciascuno dei suddetti database è costituito da due schemi :

- ✓ ADM_CLIMA contenente metadati riguardanti le stazioni di monitoraggio, le variabili e gli indicatori
- ✓ ANPA_CLIMA contenente gli indicatori calcolati

Le figure 1.a e 1.b riportano gli schemi del database Oracle, la cui struttura è identica a quella del “nuovo” database PostgreSQL.

Il database si compone di tabelle di indicatori riferiti alle principali variabili climatiche (temperatura, precipitazione, vento, umidità, copertura nuvolosa, eliofania, ecc.). Alla base dati originale di indicatori (SCIA_DB) con passo temporale decadale, mensile o annuale si è aggiunto più di recente, un dataset di indicatori con passo giornaliero (SCIA_DB_GIORN), popolato prevalentemente con dati di temperatura minima e massima e precipitazione. Attraverso l'elaborazione delle serie di dati di ingresso (orari o giornalieri) provenienti dalle fonti esterne, sono popolate e aggiornate solo le tabelle di indicatori giornalieri; le tabelle di indicatori decadali, mensili e annuali sono aggiornate attraverso l'uso di procedure di calcolo interne al sistema sulla base delle tabelle di indicatori giornalieri.

L'integrità referenziale per le entità dello schema ANPA_CLIMA viene assicurata a livello applicativo.

Le procedure in esercizio, vengono utilizzate in modo manuale e non si avvalgono quindi di processi integrati ed automatizzati come dovrebbe essere per un sistema informativo. Le procedure, sviluppate in gran parte ‘in house’ in diverse tecnologie ormai considerate in parte obsolete ed in parte non omogenee dal punto di vista della tecnologia (Fortran, VB 2017, R ecc.), svolgono le funzioni di caricamento dei metadati delle stazioni; di pre-elaborazione, caricamento, controllo e post-elaborazione delle serie temporali di indicatori giornalieri; di calcolo degli indicatori decadali, mensili, annuali a partire dagli indicatori giornalieri.

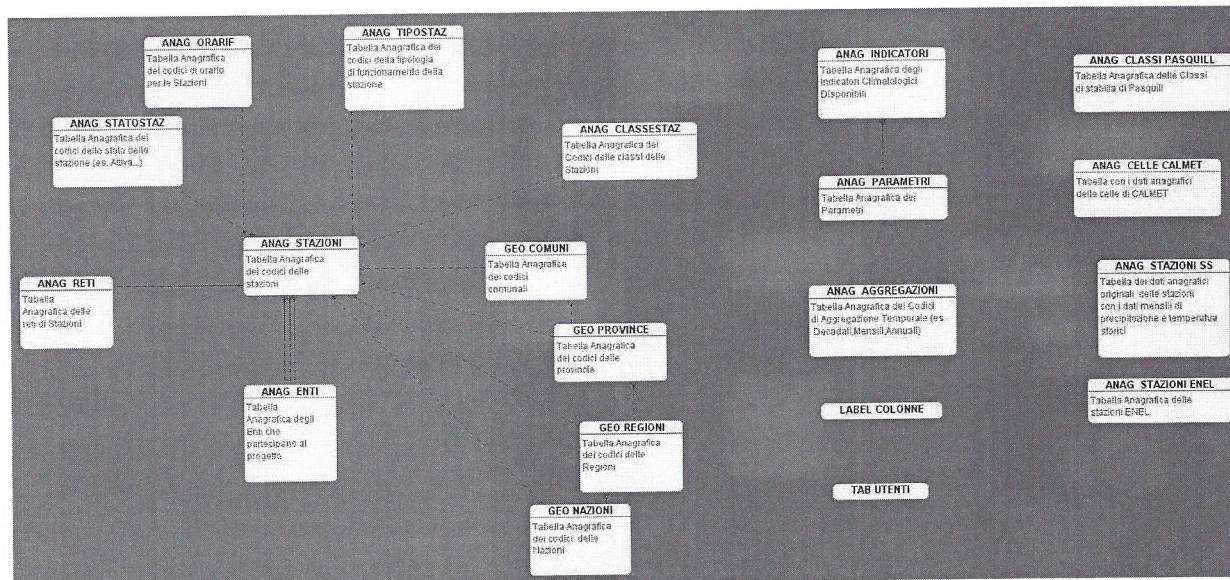


Fig. 1.a - Schema del database ADM_CLIMA

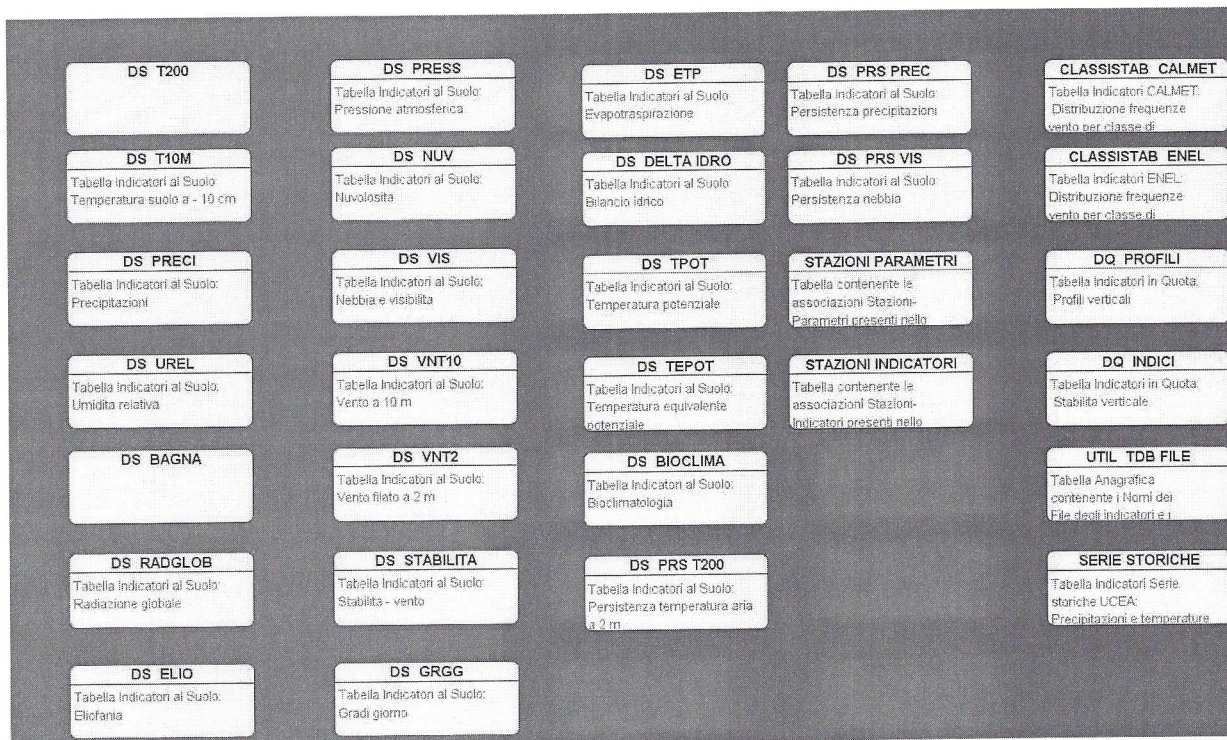


Fig 1.b - Schema del database ANPA_CLIMA

3 Descrizione dell'oggetto della fornitura

Nella figura 2 è rappresentato lo schema dei processi in esercizio. I riquadri in giallo rappresentano le procedure che sono oggetto del servizio e che dovranno essere sviluppate secondo l'analisi dei requisiti fornita dal personale ISPRA.

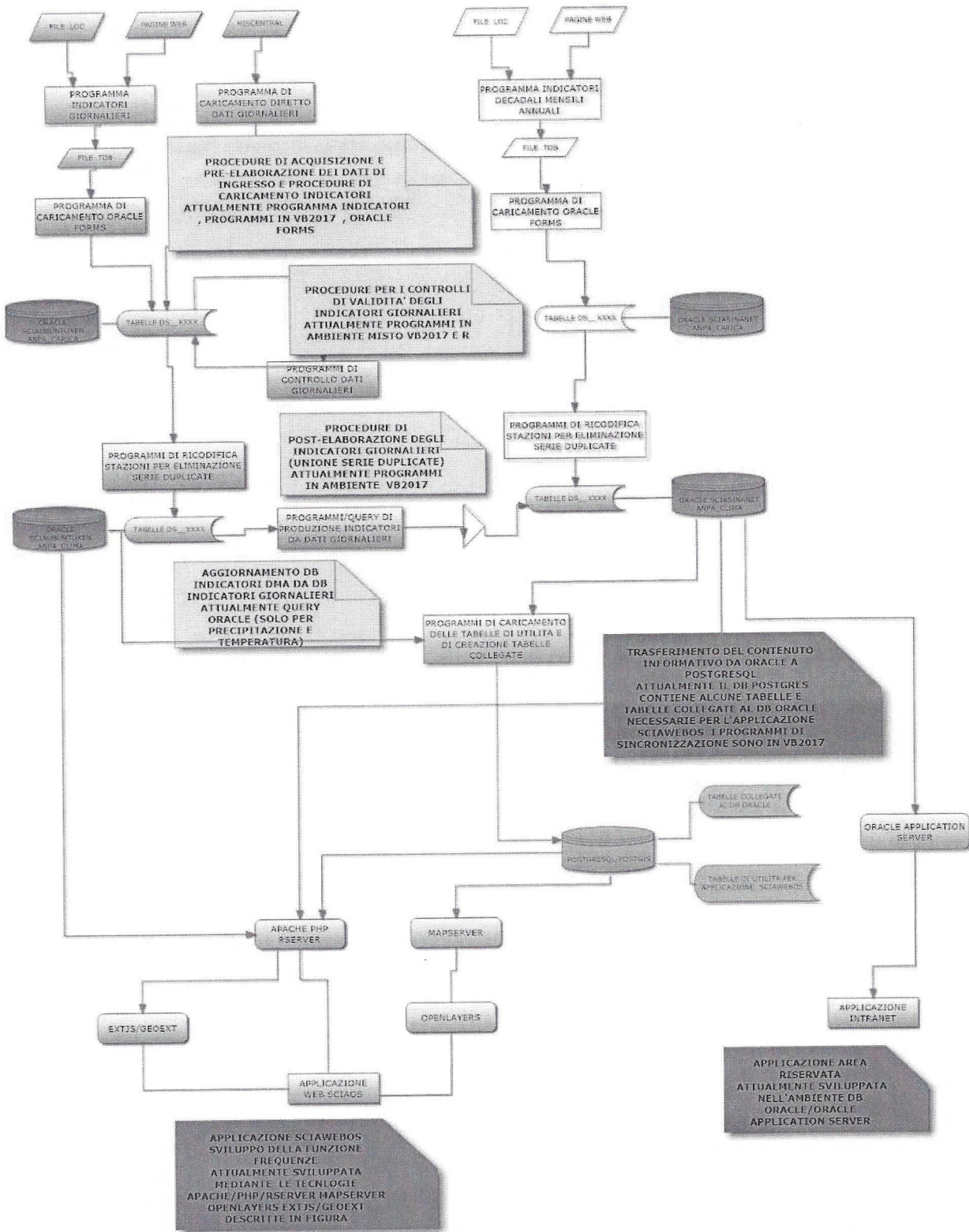


Fig 2 - Schema dei processi in esercizio. Gli sviluppi oggetto del servizio sono rappresentati nei riquadri con sfondo giallo. Il ramo in alto sulla destra non sarà operativo nel nuovo sistema.

In particolare si possono individuare i seguenti macro processi:

- A. Pre-elaborazione dei dati di ingresso orari
- B. Caricamento degli indicatori giornalieri sul database
- C. Controlli di validità degli indicatori giornalieri
- D. Post-elaborazione degli indicatori giornalieri
- E. Aggiornamento DB indicatori decadali, mensili e annuali da DB giornalieri

Di seguito vengono descritti in dettaglio i requisiti delle suddette procedure.

A - Pre-elaborazione dei dati di ingresso orari

Le serie temporali di dati in ingresso al sistema sono fornite o acquisite da varie fonti con cadenza annuale. Le serie di dati hanno formati e passo temporale diverso a seconda della fonte.

Nel caso di dati di ingresso orari (file di testo .loc) occorre sviluppare una procedura che effettui dei controlli sul dato orario di ciascuna variabile (superamento di alcune soglie, inversione e duplicazione di data/ora) e che calcoli il corrispondente indicatore giornaliero, che andrà caricato su delle tabelle di appoggio con struttura predefinita. Si richiede la generazione di un report contenente gli eventuali errori riscontrati.

B - Caricamento degli indicatori giornalieri sul database

Le procedure di caricamento sulla base dati riguardano sia i metadati delle stazioni che gli indicatori giornalieri. Tali indicatori sono ottenuti attraverso l'elaborazione sia delle serie orarie dei dati di ingresso (cfr. sezione A) che di quelle fornite dalle fonti direttamente con passo giornaliero (tabella 1).

Le serie di dati giornaliere devono essere sottoposte a controlli (inversione e/o duplicazione di data/ora) della stessa tipologia di quelli effettuati nella fase di calcolo degli indicatori provenienti dai dati orari.

Per il processo di caricamento dei dati vi è l'esigenza, vista la molteplicità delle fonti dei formati, di sviluppare uno strumento di acquisizione customizzabile, da parte degli operatori abilitati (profilo utente: amministratore). Tale customizzazione dovrà consentire di poter gestire eventuali modifiche del tracciato dati e/o del formato dei campi senza dover modificare il codice del sistema. Il sistema dovrà consentire di impostare la sorgente dei dati/metadati (file o web services), il formato dei dati, la loro mappatura con i corrispondenti campi della base dati, eventuali trasformazioni da eseguire e le validazioni da effettuare prima del caricamento. L'insieme dei parametri che definisce la customizzazione dei processi di acquisizione e di controllo qualità definisce un template. Il sistema dovrà prevedere un meccanismo di schedulazione automatica dei processi di acquisizione configurabile da parte dell'operatore e consentire, allo stesso tempo, l'esecuzione manuale degli stessi processi. Il processo di ingestione dei dati nella base dati prevede, in seguito al caricamento,

l'esecuzione di una pre-validazione che dovrà evidenziare, attraverso la restituzione di opportuni messaggi di errore, eventuali problemi di validazione dei dati.

Tutte le informazioni relative all'esecuzione dei processi suddetti verranno salvate in dei log (su file o su database) che dovranno contenere tutte le informazioni atte a diagnosticare eventuali problemi: l'utente che ha avviato l'operazione, il provider utilizzato, il template utilizzato, il file caricato e l'esito del caricamento con l'indicazione di eventuali errori. Si richiede, inoltre, che il sistema sia dotato di una funzionalità di dashboard per monitorare l'esito delle operazioni eseguite senza dover consultare necessariamente i log. In alcuni casi potrebbe essere richiesto che il file validato venga approvato da un operatore prima di essere caricato nel database.

Provider	Serie	Formato del file	Modalità di trasmissione
HISCENTRAL	giornaliere	.txt; .csv	Web services
NOOA-AM	giornaliere	.txt	Tramite ftp
Servizi regionali (ARPA, Centri funzionali ecc.)	orarie	.loc	Fornite dalle fonti
Servizi regionali (ARPA, Centri funzionali ecc.)	giornaliere	.xls, csv	download

Tabella 1- Tipologie di dati di ingresso, formato e modalità di trasmissione per le diverse fonti di dati

Caricamento dei metadati delle stazioni

L'appalto ha ad oggetto:

- La realizzazione di una GUI con cui poter consultare ed aggiornare il contenuto delle tabelle riguardanti i metadati (anag_reti e anag_stazioni). Tale procedura dovrà anche permettere il caricamento sul database di un insieme di stazioni a partire da un file di testo (.csv) contenente le informazioni su ogni stazione (codice rete, codice utente, coordinate, regione, provincia, ecc.) ed eseguire alcune verifiche:
 - ✓ Verifica che la stazione che si sta inserendo non sia già presente nella tabella anagrafica, controllando che il codice utente-codice rete non sia nella base dati.
 - ✓ Verifica che le coordinate della stazione ricadano all'interno del territorio nazionale. Nel caso in cui i metadati della stazione comprendano la regione di appartenenza verificare la correttezza delle coordinate rispetto ai limiti regionali.

La procedura di caricamento dei metadati dovrà inoltre produrre un file di log contenente informazioni relative al caricamento effettuato.

Caricamento degli indicatori

Sviluppo di un'applicazione dotata di GUI per il caricamento sul database degli indicatori giornalieri. Tale procedura di caricamento, oltre ad effettuare i necessari controlli di integrità referenziale tra le tabelle contenenti i metadati e le tabelle contenenti gli indicatori (presenza della stazione in anagrafica, validità del codice di aggregazione), dovrà aggiornare le relative tabelle di popolamento degli indicatori e delle variabili, preparare un file di log con gli esiti del caricamento e gestire eventuali situazioni di aggiornamento di indicatori già caricati, in base alle richieste dell'operatore.

C - Controlli di validità degli indicatori giornalieri

Dopo il caricamento dei “nuovi” indicatori giornalieri, le serie complete di indicatori aggiornati sul database dovranno essere sottoposte ad una sequenza di controlli di validità. Tali controlli sono attualmente limitati agli indicatori delle variabili temperatura e precipitazione e sono descritti in dettaglio in:

www.scia.isprambiente.it/wwwrootscia/Documentazione/Rapporto_controlli_qualit%C3%A0_0_clima.pdf

Il predetto documento costituisce parte integrante del presente capitolato; ulteriori dettagli verranno forniti in fase di esecuzione del contratto.

Le procedure dedicate ai controlli degli indicatori dovranno essere eseguite sulle serie complete di indicatori presenti sul database e non solamente su quelli appena inseriti. Si presume, salvo diverse soluzioni proposte dalla ditta appaltante, che le procedure per i controlli di validità degli indicatori giornalieri verranno eseguite sugli indicatori già caricati sul database.

Inoltre, i controlli devono essere eseguiti secondo un ordine predefinito; a seguito di ciascuna tipologia di controllo, l'indicatore riconosciuto come non-valido non viene cancellato dal database; bensì, gli viene attribuito un codice identificativo corrispondente al controllo non superato, tramite un flag di validazione presente come attributo nella tabella della base dati.

Ciascun controllo deve essere effettuato sulla tabella contenente gli indicatori, escludendo gli indicatori già invalidati dai controlli precedenti.

La procedura attualmente in esercizio si limita ai controlli sugli indicatori delle variabili temperatura e precipitazione, ma si richiede che venga estesa agli indicatori di tutte le altre variabili, sulla base di specifiche e algoritmi definiti e documentati a cura dell'ISPRA. Il numero di controlli delle nuove variabili sarà inferiore a quello di temperatura e precipitazione (non più di 5 step).

La nuova procedura dovrà essere il più possibile automatizzata e dovrà consentire all'operatore di scegliere la/le variabile/i su cui eseguire i controlli e/o l'insieme di stazioni (ad esempio, una

singola stazione, un elenco predisposto dall'operatore, le stazioni di una rete, tutte le stazioni). La procedura dovrà infine produrre un file di log con i risultati dei controlli effettuati.

D - Post-elaborazione degli indicatori giornalieri

A causa della varietà delle fonti utilizzate per popolare il database, accade che serie di indicatori attribuiti a stazioni diverse e caricati sul DB con identificativo di stazione diverso, si riferiscano in realtà alla stessa stazione. Pertanto è stata compilata 'in house' una tabella di corrispondenza tra stazioni "duplicate". Si richiede lo sviluppo di una procedura che, sulla base della tabella di corrispondenza, elimini le eventuali serie duplicate e unisca le serie da riferire ad una stessa stazione, attribuendo la serie completa alla stazione con i dati più recenti.

La procedura dovrà essere estesa agli indicatori di altre variabili (oltre a temperatura e precipitazione), utilizzando le relative tabelle di corrispondenza prodotte dall'ISPRA.

E - Aggiornamento DB indicatori decadali, mensili e annuali da DB giornalieri

Analogamente alla precedente struttura Oracle, il nuovo database PostgreSQL prevede due differenti schemi: uno per gli indicatori giornalieri (SCIA_GG) e uno per gli indicatori decadali, mensili e annuali (SCIA_DMA).

Lo schema SCIA_GG verrà alimentato con nuovi dati di ingresso dalle procedure descritte precedentemente nella sezione B. Per alimentare lo schema SCIA_DMA dovranno essere sviluppate in ambiente PostgreSQL delle query di calcolo analoghe a quelle già sviluppate in ambiente Oracle (o una procedura), che consentano di popolare le tabelle di indicatori decadali, mensili e annuali a partire dai giornalieri. Inoltre, tale procedura dovrà essere estesa agli indicatori di altre variabili, oltre a temperatura e precipitazione, su specifiche prodotte dall'ISPRA.

4 Requisiti generali del nuovo sistema

Nella figura 3 è stata schematizzata la configurazione hardware attualmente in esercizio e che, presumibilmente, sarà anche quella relativa al nuovo sistema sostituendo, ovviamente ORACLE con PostgreSQL.

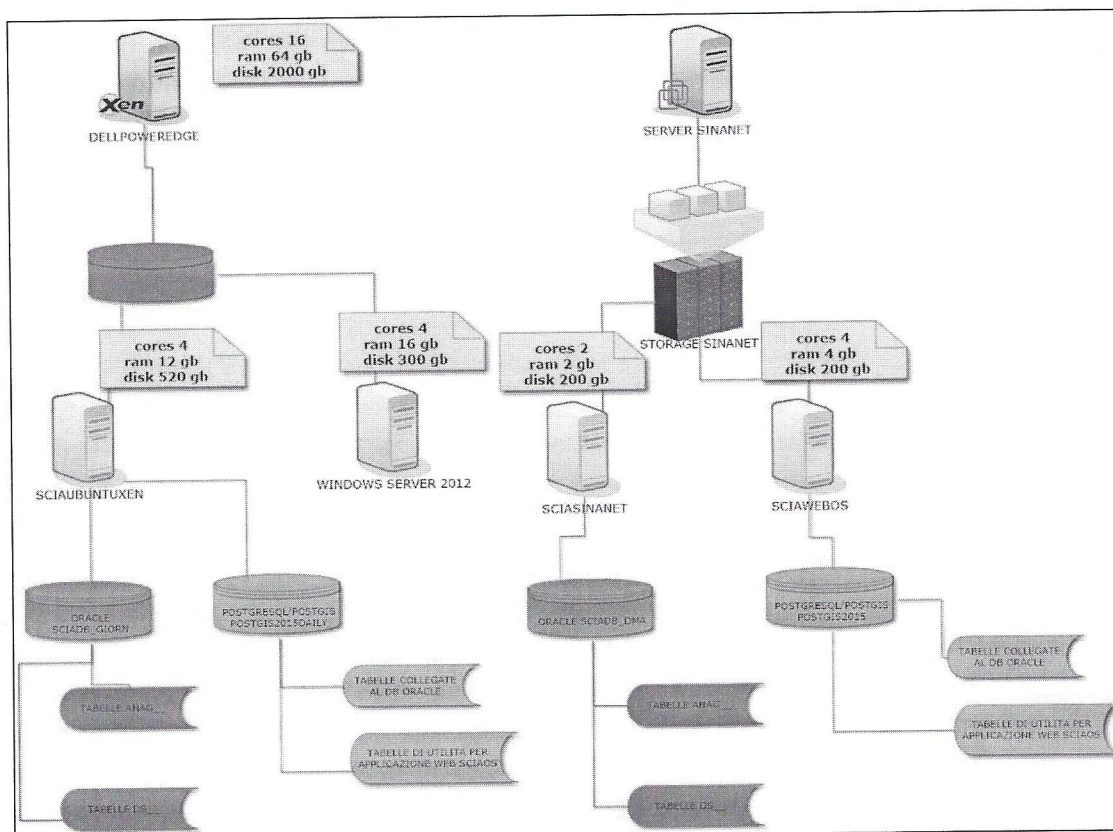


Fig 3 – Configurazione hardware

- 1) Per il processo di acquisizione e caricamento di dati e metadati si dovranno prevedere tre profili di utenza:
 - ✓ *Amministratore* - accede a tutte le funzionalità di amministrazione del sistema (abilitazione e profilazione degli utenti, definizione dei template ecc.)
 - ✓ *Utente operatore esperto*: accede alle funzionalità che permettono l'acquisizione, l'aggiornamento e la modifica dei dati
 - ✓ *Utente ISPRA privilegiato*: accede alla consultazione dei dati

- 2) Si dovranno prevedere due ambienti identici, uno di sviluppo e uno di produzione. Al termine del contratto i due ambienti saranno perfettamente allineati, salvo che per le procedure di backup del database non necessarie nell'ambiente di sviluppo.

- 3) Il sistema dovrà garantire un'adeguata sicurezza applicativa e funzionale.
- 4) Con riferimento alla configurazione hardware in esercizio (fig. 3), le nuove procedure dovranno avere dei tempi di esecuzione migliori o al massimo uguali a quelli delle procedure attuali. In particolare si richiede un miglioramento sui tempi di esecuzione del pre- e post- processamento dei dati rispetto all'attuale. La tabella più "grande" da caricare o aggiornare ha dimensioni di ca. 100 milioni di record e i tempi di esecuzione delle procedure in esercizio sono: circa 30 ore per la procedura che effettua i controlli di validità, circa 4 ore per la procedura che effettua l'eliminazione delle serie duplicate, circa 5 ore per le query che effettuano il calcolo degli indicatori decadali, mensili e annuali a partire dagli indicatori giornalieri.
- 5) Dovrà essere garantita la continuità dei servizi o il recupero dei dati (database) in tutti i casi in cui è necessario, garantendo sempre la ricostruzione del sistema informativo a partire dal momento dell'ultimo salvataggio.
- 6) Dovrà essere previsto un sistema schedato (settimanale) di back-up della base dati.
- 7) Allo scopo di massimizzare l'efficienza nel mantenimento della soluzione realizzata da parte del personale interno ISPRA, il sistema dovrà essere basato su scelte tecnologiche quanto più omogenee in termini di linguaggi di programmazione e frame.
- 8) Le procedure dovranno essere sviluppate in ambiente open-source.
- 9) Le procedure dovranno essere sviluppate in un linguaggio adatto ad algoritmi di calcolo matematico (per es. Python).
- 10) Tutti i codici sorgente delle applicazioni sviluppate, se non disponibili in forma libera online, devono essere caricati sul server GIT dell'Istituto in modo che il personale tecnico ISPRA possa verificare lo stato di avanzamento dei lavori e testare il software sviluppato durante lo svolgimento del contratto.
- 11) Si richiede RDBMS PostgreSQL versione 11 o successive
- 12) Si richiede sistema operativo Linux Ubuntu 18 o successive
- 13) L'applicazione, composta di tutte le parti e comprensiva della base dati, deve risiedere all'interno della rete intranet ISPRA e nessuno dei servizi o parte di essi devono essere in hosting presso altre strutture.
- 14) Il software deve essere sviluppato seguendo i più recenti standard di programmazione e documentazione e deve essere compatibile con le ultime versioni del software disponibile.

5 Modalità di esecuzione dell'appalto

- 1) La società appaltatrice dovrà prevedere la formazione del personale ISPRA sulle funzionalità dell'applicazione in ogni sua parte, nonché almeno 3 (tre) giorni di formazione sul linguaggio di programmazione utilizzato per lo sviluppo del software. Dovrà, inoltre, produrre la documentazione necessaria sulle procedure di installazione e gestione sistemistica per la corretta manutenzione dell'applicazione.

- 2) Tutte le procedure e il codice sorgente dovranno essere commentate in modo comprensibile.
- 3) Deve essere garantita la manutenzione correttiva (risoluzione di eventuali bug del software) di sei mesi dopo la verifica di regolare esecuzione.
- 4) Il software realizzato e tutta la documentazione di progetto richiesta saranno rilasciati all'ISPRA che ne sarà il proprietario e ne detiene i diritti di utilizzo, divulgazione ecc.
- 5) La società appaltatrice fornirà il manuale tecnico e operativo del nuovo sistema.
La manualistica e la documentazione applicativa dovranno essere ben descritte e consegnate in formato digitale al termine di ciascuna fase del progetto (cfr. punto 9).
- 6) Il tempo previsto per l'espletamento del servizio è di 10 (dieci) mesi a partire dalla data del verbale di inizio attività.
- 7) Il contratto avrà efficacia dalla data della stipula e le relative attività decorreranno dalla data del verbale di inizio attività a cura del direttore dell'esecuzione del contratto.
- 8) Con riferimento all'elenco dei processi (cfr. cap.3) l'attività oggetto del contratto viene suddivisa in due fasi. La fase I avrà la durata di mesi 4 (quattro) a partire dalla data di inizio attività e comprenderà lo sviluppo delle procedure relative ai processi A e B. La Fase II avrà la durata di mesi 6 (sei) e inizierà del termine della fase I (mese cinque) fino alla fine del contratto. Durante la Fase II verranno espletate le attività dei processi C, D ed E. Al termine di ogni fase dovranno essere eseguiti i test nell'ambiente di sviluppo insieme al personale ISPRA e dovrà essere verificata la congruenza dei risultati delle nuove funzionalità sviluppate con i risultati forniti (sulla base di un set di dati storici) dal sistema attualmente in esercizio. Il personale ISPRA metterà a disposizione gli strumenti ed i metodi di analisi per lo svolgimento dei test di congruenza. L'esito positivo dei test comporterà la validazione del rilascio di quanto realizzato in ciascuna fase.
- 9) Con riferimento alle fasi definite sopra, l'espletamento del servizio avverrà secondo il seguente cronoprogramma.

Fase/Attività	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
FASE I										
FASE II										

- 10) I pagamenti sono previsti come segue:
 - 40% al completamento della FASE I (4 mesi dal verbale di inizio attività)
 - 60% al completamento della FASE II (10 mesi dal verbale inizio attività)
- 11) Per ogni pagamento è prevista una fattura che verrà pagata entro 30 (trenta) giorni decorrenti dalla data di ricevimento della stessa, previo rilascio del certificato di regolare esecuzione da parte del responsabile unico del procedimento.

Per tutto quanto non espressamente indicato nel presente Capitolato, la società appaltatrice dovrà attenersi alla normativa vigente in materia.