

Realizzazione in modalità Open Source di  
un software agile per la gestione di  
grandi quantità di informazioni non  
strutturate e semi strutturate  
organizzate in archivi digitali



Roma, dicembre 2006

*Gianni Sandrucci (Talete Informatica)*  
*Anna Maria Lomartire (Studio Hotbrain)*  
*Gianfranco Pesce (Studio Hotbrain)*

# L'esperienza dei DMS commerciali

- Il progetto nasce da cinque anni di esperienza con i sistemi di gestione documentale commerciali:
  - Mancanza di API standard e sviluppo di layer di indipendenza per evitare il lock-in
  - Utilizzo limitato a funzionalità core: imaging, foldering, metadati, security, versioning, checkout/checkin, full-text
  - Uso sempre mediato da applicativi
  - Alti costi delle licenze e problematica scalabilità verticale ed orizzontale



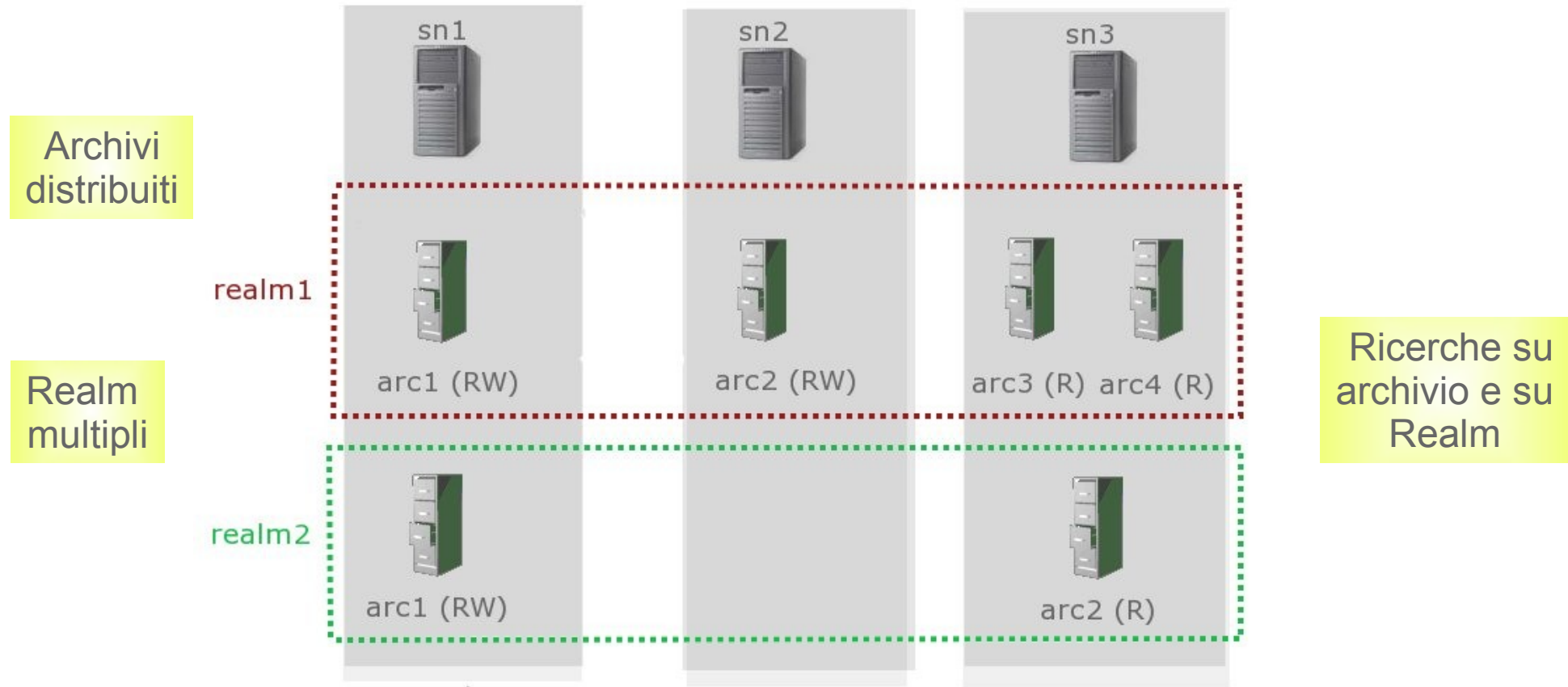
## Gli obiettivi della soluzione Open

- Indipendenza dalla piattaforma (100% java)
- Modello distribuito, overhead minimo, scalabilità lineare on demand e non on-project, configurazione centralizzata modificabile a caldo
- Corredo di test automatici di correttezza e test di carico
- Agilità e flessibilità dei metadati (metadati dinamici).  
Versioning completo: contenuto, metadati, security e foldering.
- Supporto per i requisiti internazionali sul Record Management (MoReq, Nara) ed alla normativa italiana



# Modello distribuito: Archivi e Realm

ServerNodes, Archives & Realms



# Requisiti funzionali

- Document management (documenti, allegati e folder),  
Record management (classi, file, volumi, record),  
Archive management (supporto per archivi correnti, di deposito e storici)
- Metadati dinamici (XML). Nessun vincolo di tipizzazione.
- Versioning totale (security, metadati, foldering)
- Ricerche su archivio singolo e su realm: range date e numeri, and, or, not, radice. Full text PDF e Office (pluggable text extractor)
- Checkin/chekout, Import/export. Auditing.
- Security ACL, RBAC, gerarchica (Top Secret...)
- Autenticazione con supporto JAAS.



# Agilità: metadati dinamici

```

<invoice>
  <date>2005-11-05</date>
  <number>1</number>
  <customer>
    <id>AZ-02354</id>
    <name>John The Buyer</name>
    <address>One stop beyond the world's end</address>
  </customer>
  <item>
    <sku>SAB.342</sku>
    <descr>Ironmade yellow ant killer</descr>
    <quantity>100</quantity>
    <price>19.50</price>
  </item>
  <item>
    <sku>XX.342</sku>
    <descr>Special black mostard</descr>
    <quantity>7</quantity>
    <price>2.90</price>
  </item>
</invoice>

```

Metadati  
in XML  
standard

Libera definizione degli  
elementi ricercabili

```

<FieldDefConf name="date" type="date" />
<FieldDefConf name="number" type="number" />
<FieldDefConf name="id" type="key" />
<FieldDefConf name="name" type="text" />
<FieldDefConf name="address" type="text" />
<FieldDefConf name="sku" type="key" />
<FieldDefConf name="descr" type="text" />
<FieldDefConf name="quantity" type="number" />
<FieldDefConf name="price" type="number" />

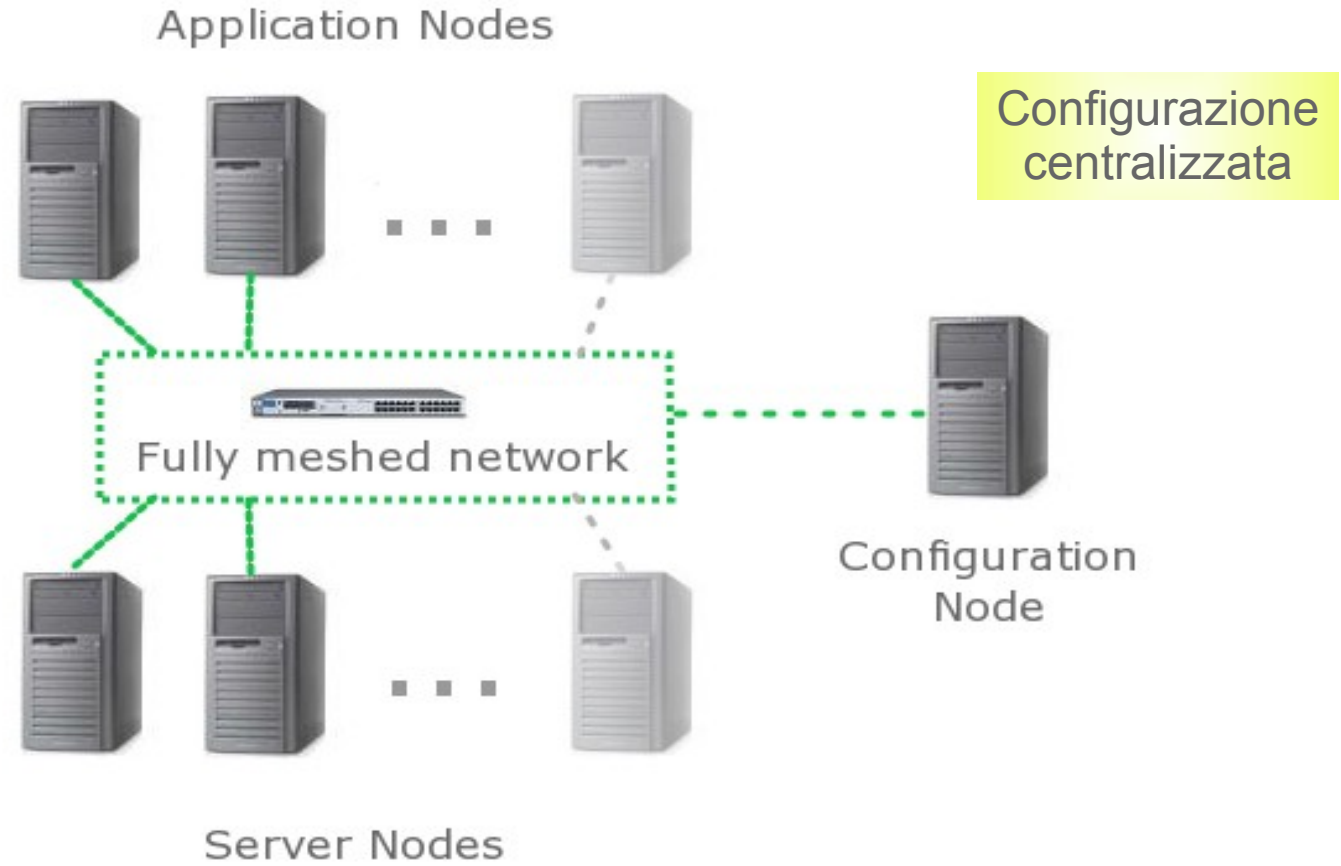
```

- Search full text su "address" o su "invoice.customer.address"
- Search di un range su "price" o su "invoice.item.price"



# Scalabilità e semplicità di gestione

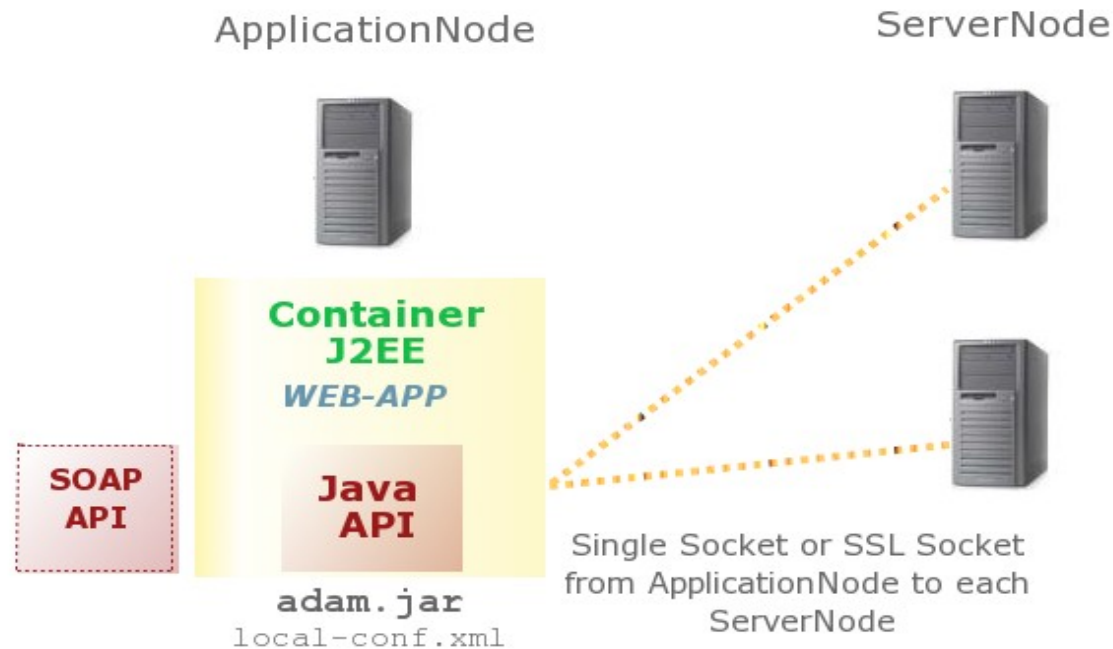
Aggiungere e modificare nodi *a caldo*





# API

- Accesso in contesto J2EE: diretto o tramite SOAP (accesso da DMZ)
- Accesso diretto ad alta efficienza da ApplicationNode (API Java)



# Tecniche di ottimizzazione

- Massimo utilizzo della banda di comunicazione disponibile sfruttando il naturale *non-determinismo* nelle comunicazioni (Java 5).
- Minimizzazione dell'overhead dovuto al context switching (uso ridotto dei thread)
- Fairness basata sulla nuova gestione della sincronizzazione di Java 5



# 100% Open Source

- Metodologie e modello di sviluppo Open Source
- Tutto il software è interamente funzionale usando esclusivamente componenti Open Source, quindi indipendente da qualsiasi prodotto commerciale.



- Supporto di Oracle RDBMS e degli application server J2EE commerciali.



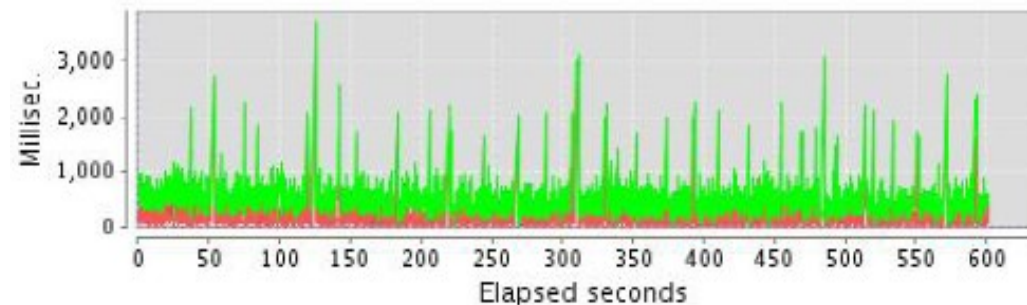
# Esempio di prestazioni raggiunte

- Portatile Toshiba Satellite M60 – Pentium 1.6 Ghz, HD interno 7200 rpm, HD firewire 7200 rpm.
- Documenti di 100 Kbyte, metadati di una fattura (randomizzati)
- Base dati con un archivio di circa un milione di documenti, 20 utenti concorrenti per 10 minuti, think time = 0, RAM 512 MB

Operations	Count	Op/sec.	Min.	Max.	Aver.	Time 90	Op. over 90
create doc.	9,957	16.6	0	3.7	0.5	3.3	9
search	9,956	16.6	0	2	0.1	1.8	4
read content	9,960	16.6	0	3.5	0.4	3.2	10

**50 operazioni documentali al secondo**

**Response time: create doc.**





Grazie per l'attenzione

[www.adam-project.com](http://www.adam-project.com)