



Grid File Systems

Requisiti, Funzionalità e Architettura



Grid File System: Requisiti

- Name Space Gerarchico "Logico"
- Storage Interface Uniforme
- Servizi di Replica Management
- Meccanismi di Data Access e Data Transfer
- Gestione della latenza di accesso
- Gestione dei metadati
- Security nel file system
- Ottimizzazione e Performance
- API per gestione di file e directory



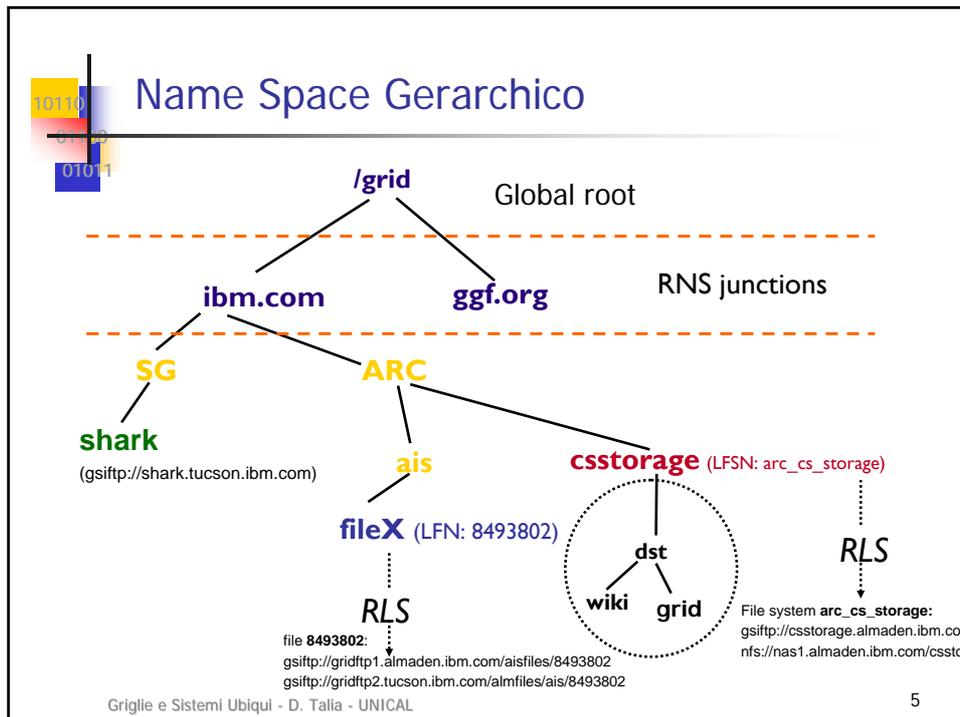
Name Space Gerarchico

- Name Space Logico
- Struttura Gerarchica
- Operazioni POSIX sui nomi logici
- Soft links
- Aggregazione di file in un singolo logical name space



Name Space Gerarchico

- Permette la realizzazione di un Grid File System con un namespace uniforme, globale e gerarchico
- Organizza gerarchicamente i file (in un network file systems accessibile tramite protocolli standard come NFS).
- Il Resource Name Space (RNS) può essere offerto come un data service.
- Permette di realizzare un namespace distribuito tra organizzazioni senza compromettere l'autonomia locale.

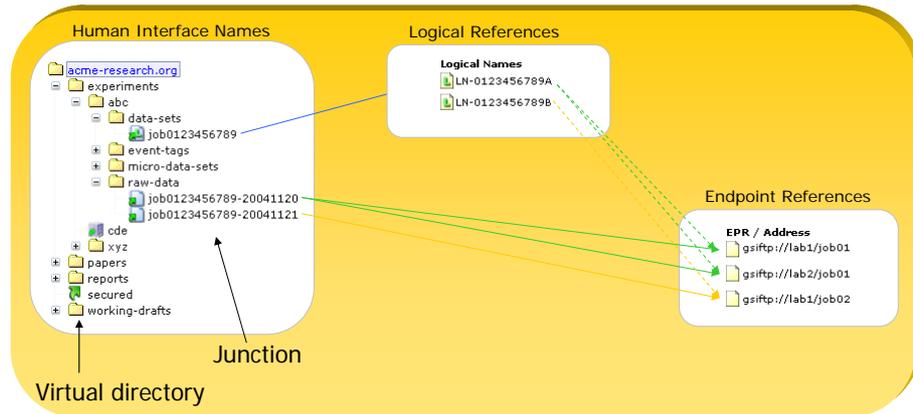


- ## Operazioni del Resource Name Space
- Naming Operations
 - create
 - delete
 - list
 - lookup
 - update
 - Iterator Context Operations
 - createIteratorContext
 - getIteratorContext
 - Extensibility Operations
 - deleteProperty
 - insertProperty
 - listProperties
 - updateProperty
- Griglie e Sistemi Ubiqui - D. Talia - UNICAL
- 6



RNS - Naming Architecture a Tre Livelli

- **Architettura a tre livelli:** Human Interface Names, Logical Reference Names, and Endpoint References.



Uniform Storage Interface

- **Accesso a**
 - File systems
 - Database objects
- **Interfaccia verso il middleware di gestione dello storage**
- **Meccanismi di interfacciamento comuni per files e database objects**



Replica Management

- Replica Catalog Distribuito/Gerarchico
- Creazione e gestione delle Repliche
- Meccanismi di consistenza
- Load balancing tra le repliche
- Replica di frammenti di un file/object



Data Access/Transfer

- Supporto di GridFTP per l'accesso e lo spostamento di file tra nodi di una Griglia.
- Altri meccanismi di trasferimento (selezionabile dall'utente)
- RFT: Reliable transfer
- Parallel I/O



Gestione della Latenza

- Streaming di dati/file/blocchi
- Disk Caching
- Pre-fetching
- Remote I/O proxies



Metadata Management

- Metadata Catalog
 - Gerarchico
 - Distribuito
 - Federato
- Metadati da gestire
 - File level (size, creation/modification/access time, creator ...)
 - Storage metadata
 - Access control metadata
 - Provenance metadata
- Consistenza dei metadati



Security

- GSI Authentication
- Altri meccanismi di autenticazione
- Meccanismi di autorizzazione
- Access Control Lists
- Gestione proprietari dei file



Ottimizzazioni

- Optimal replica selection
- Istanze di servizi pre-avviate
- Altre ottimizzazioni?



File System API

- File API (semantica POSIX)
 - Apertura/chiusura
 - Accesso
 - Modifica
 - Controllo

- Object level API

- Web service API



Autonomous Administrative Domain

- Una Grid Entity :
 - Gestisce grid resources
 - Usa le proprie politiche
 - Può adottare una politica "superiore"
 - Può agire come fornitore e/o cliente di una risorsa

 - Esempi:
 - Un centro di ricerca
 - Un finance department di una società (sub-organization)
 - Un centro di servizi di storage

- GFS contiene uno o più domini amministrativi con risorse distribuite eterogenee.



Digital Entities

- GFS fornisce una vista logica indipendente dalla locazione di entità eterogenee distribuite (dati, metadati, servizi, utenti).

- Dal punto di vista queste entità sono raggruppate in tre categorie:
 - Contesto
 - Contenuto
 - Commodity



GFS/Grid Resources

- Context (Informazione)
 - Informazione su digital entities (locazione, dimensione, proprietari, ..)
 - Relazioni tra digital entities (repliche, collezioni, ..)
 - Comportamento di digital entities (servizi)
- Content (Dati)
 - Strutturati e non strutturati
 - Virtuali o derivati
- Commodity (Produttori e consumatori)
 - Risorse di storage
 - Providers, brokers e requestors



GFS - Organizzazione di Risorse

- Risorse e WSRF
 - URIs per denotare risorse (dati, servizi, ...)

- Organizzazione di Grid Resources
 - Human readable naming system
 - Sistema unico per l'organizzazione dello stato distribuito di dati persistenti

- Data Model per aggregare e organizzare
 - Mapping di URIs / WS-Addresses su collezioni di dati digitali
 - Meta-dati associati con ogni digital entity



GFS - Logical Naming

- Infrastruttura Data Grid Distribuita
 - Applicazioni Data-intensive possono essere in esecuzione dovunque su una grid

- Logical naming indipendente dalla locazione
 - Il sistema offre accesso ai dati trasparente rispetto alla locazione fisica

- Logical Data Identifiers
 - Un namespace logico in cui gli identificatori dei dati sono mappati su locazioni fisiche.



GFS – Componenti dell'Architettura

- GFS Resource Provider
 - Fornisce content / context / commodity storage

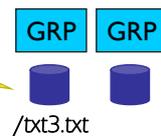
- GFS Administrative Domain
 - Una sub-organization che possiede una o più GFS resources (file, directory, ...)

- GFS Service Provider
 - Fornisce la GFS standard service interface per uno o più GFS Administrative domains



GFS Resource Providers

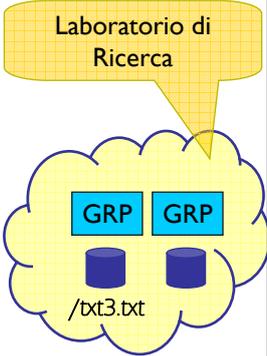
GFS Resource Providers (GRP) forniscono contenuti e/o storage





GFS Administrative Domain

Dominio Amministrativo
GFS con uno o più GFS
Resource Providers



GFS Administrative domains

Fornitore di Servizi di
Storage
data + storage (50)

Finance Department
data + storage (30)

Laboratorio di
Ricerca
data + storage (20)

