Catacomb WebDAV Server

Markus Litz

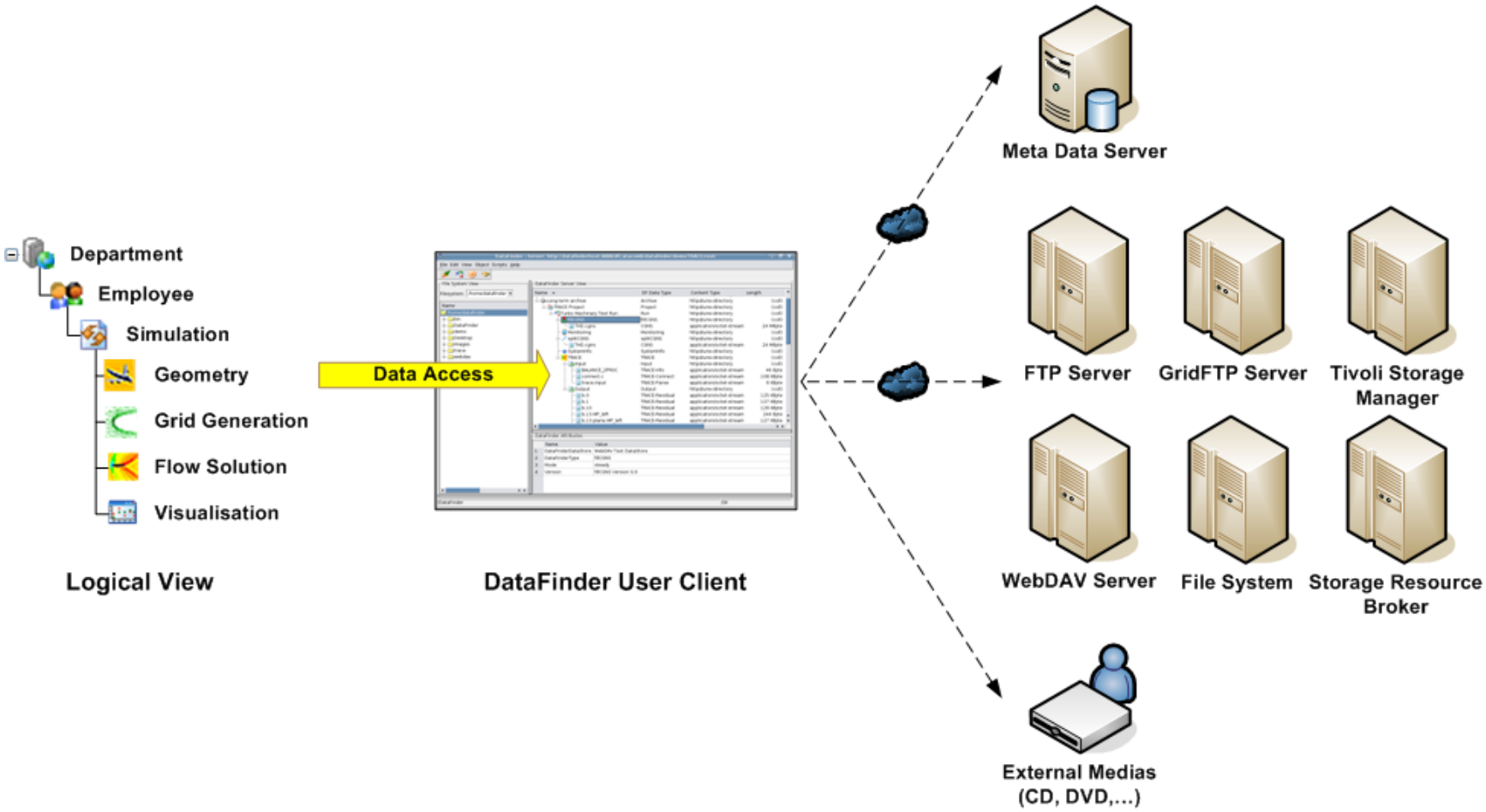
ApacheCon US 2008

Fast Feather Track

Motivation and Background at DLR

- DataFinder – a application for scientific data management
  - Storing and managing huge amounts of data
  - Search through the resource content and metadata
  - Various ways to store data, for example
    - ftp, network share, offline stores
  - Metadata management with the WebDAV protocol
    - Two supported WebDAV Server:
      - Tamino XML Server & Catacomb

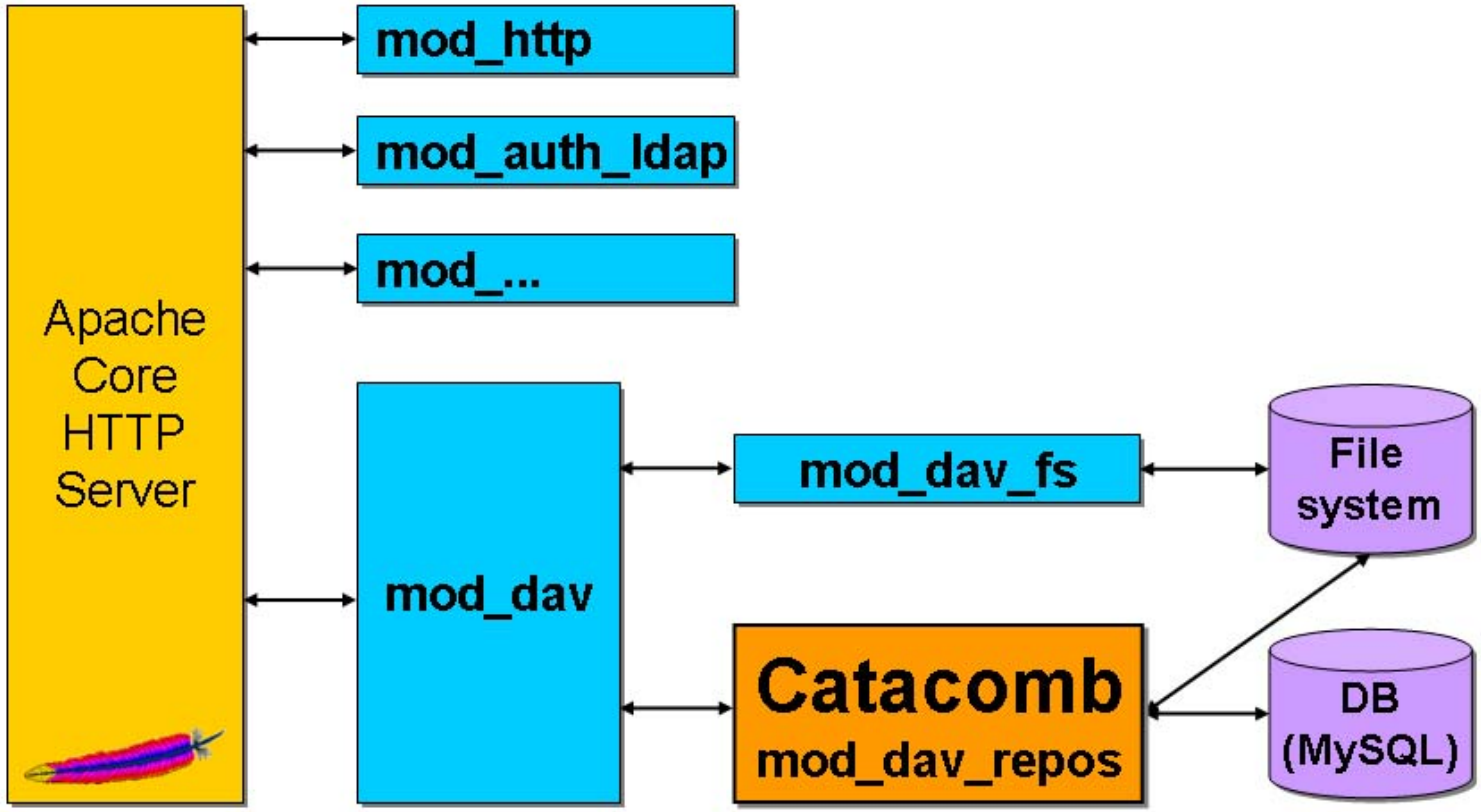
Motivation and Background at DLR



The WebDAV Protocol

- WebDAV – suite of protocol extensions for HTTP and supports collaborative authoring
- A Web-based network file system
- Several protocol extensions, for example:
  - DeltaV ➔ Versioning of resources
  - DASL ➔ Server side searching through content and its metadata
  - ACP ➔ Access Control Protocol

Catacomb – A WebDAV Server Module for Apache



Catacomb – The Difference to mod_dav_fs

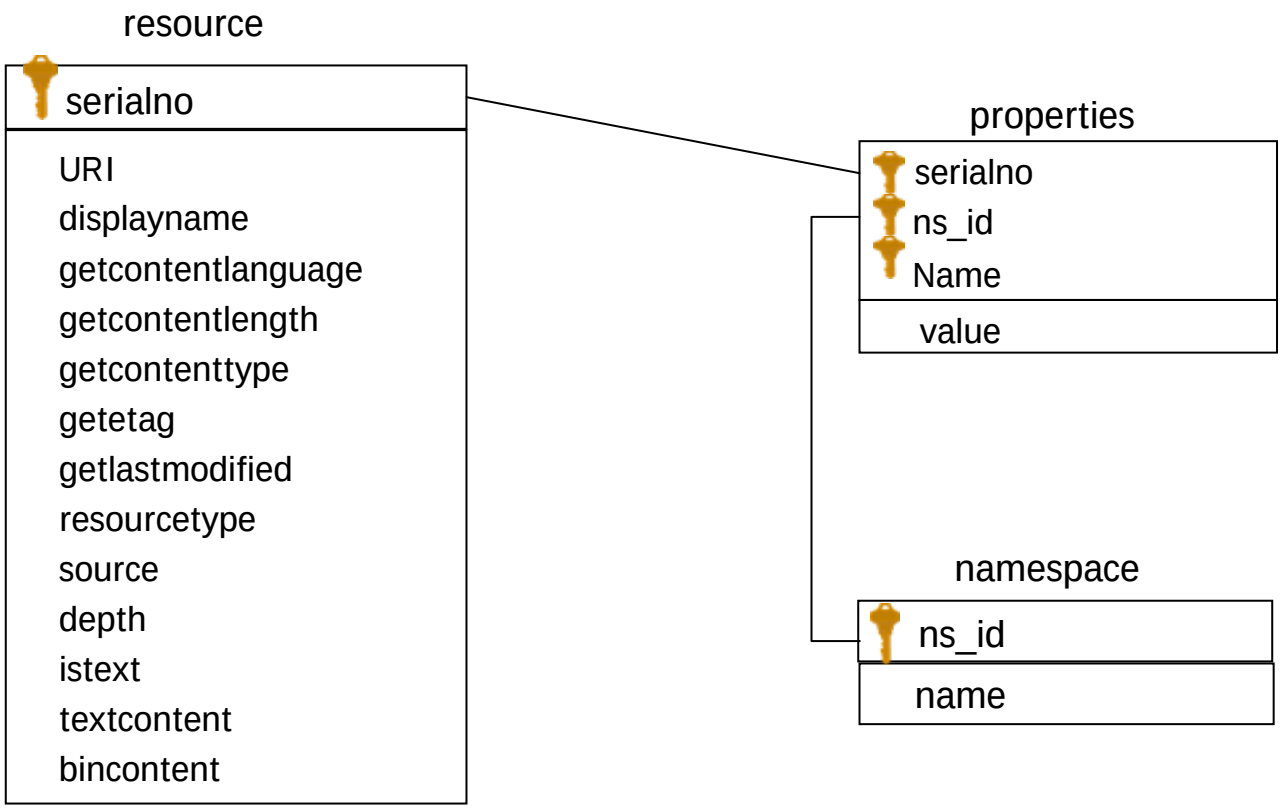
- Saving the resources
  - mod_dav_fs save content and properties in files on the filesystem
  - mod_dav_fs creates for every resource, and also for every collection, their own property file

- Consequence:
  - A single DASL query needs to open many files
  - Implementation of complex queries is difficult
  - Full text search is expensive

Catacomb – A WebDAV Server Module for Apache

- WebDAV repository module for mod_dav
- Catacomb uses relational databases to store the metadata
  - Strong search performance through SQL statements
- Catacomb is:
  - Good for Content management
  - Good for Collaborated web authoring
    - Support locks, avoid the “lost update” problem
  - Capable of searching (DASL) and versioning (DeltaV) resources

Catacomb – Database Schema



Catacomb – A WebDAV Server Module for Apache

- Advantages of using a DBMS?
  - Facilitates management of data/metadata and containment relations
  - Supports SQL-based searching
  - Full text searching
    - Text content and metadata could be searched at the same time
  - Not a hierarchical structure
    - Only URIs represent the hierarchy
    - Supports referential containment
  - Fast “depth infinity” operations
    - *Select * from resource where URI like ‘/repository/%’*

Catacomb – The Search Query Parser

```
<d:searchrequest xmlns:d="DAV:">
  <d:basicsearch>
    <d:select>
      <d:prop>
        <d:displayname/>
        <d:year/>
        <d:author/>
      </d:prop>
    </d:select>
    <d:from>
      <d:scope>
        <d:href>/dbms</d:href>
        <d:depth>infinity</d:depth>
      </d:scope>
    </d:from>
    <d:where>
      <d:gt>
        <d:prop><d:author/></d:prop>
        <d:literal>Markus Litz</d:literal>
      </d:gt>
    </d:where>
  </d:basicsearch>
</d:searchrequest>
```

```
SELECT
        dasl_resource.displayname,
        t.name, t.value
FROM
      dasl_resource
      LEFT JOIN
        dasl_property t USING (serialno)
      LEFT JOIN
        dasl_property bar_t USING (serialno)
WHERE
        ( bar_t.name = 'author' AND
        bar_t.value > 'Markus Litz' )
AND
        ( t.name = 'year' OR t.name = 'author' )
```

Catacomb – History and Current State

- Initial development at the University of California under the chair of Jim Whitehead
- Open Source project since 2002
- DeltaV and DASL implementation
- Since 2006 contribution of the DLR
  - ACP support
  - Database abstraction using mod_dbd
  - License changed to ASL2.0

Catacomb – Future Works and Milestones

- Getting though the incubator in order to become a official apache project
- Improvements of the Search support – at this time only basic search is supported
- Stable version with Access Control Protocol
- Full implementation of the DeltaV-protocol (advanced versioning) in order to support transactions

Questions?

http://www.walle-derfilm.de/