HP's Enterprise Directory: Integrating Open Standards and Open Source

Kartik Subbarao

Enterprise Directory Architect
Hewlett-Packard Company

HP WORLD 2003
Solutions and Technology Conference & Expo

What is the Enterprise Directory (ED)?

- Primary system of reference for information about HP Employees, Contingent Workers, Business Partners, Site Locations, Groups, and Business Organizations.
- Distributed, globally load-balanced, highly available repository of up-to-date data.
- Major authentication hub
- Provides multiple interfaces to broker data to customers throughout the company:
  - standard programmatic interfaces (LDAP)
  - outgoing data feeds

Enterprise Directory – Primary Roles

Information Retrieval

- White pages information
- Contact information (Name, phone #, address, email etc)

Authentication

- NT/AD passthrough authentication
  - DN and NT/AD Password
- Basic Auth. - username/password
- X.509 certificates and CRLs

Authorization

- Based on group membership
- Based on roles
- Policy store – Netegrity Siteminder

Group Management

- One group used for multiple purposes
  - mail enabled group
  - security group
  - news enabled group
- Dynamic groups
  - membership can be determined from a dynamic LDAP query

Messaging & Collaboration

- Email address translation
- Mail enabled groups, broadcast mailings
- Instant messaging

Enterprise Directory Customers

- Real Time Access
  - Anonymous
    - Applications / users that do not require access to privileged data
    - e.g. Messaging, PeopleFinder, end users, etc.
  - Authenticated
    - Applications that require access to privileged data (emp. number, etc)
    - 500+ applications
    - e.g. HP Portal, etc.
  - Operations Serviced:
    - ~20,000,000 / day or ~600,000,000 / month
- Data Files (data brokering)
  - 200+ data file customers

Guiding Principles

- Open Standards
- Open Source
- Suitability of Data
- Wide Readability
- Balanced Simplicity and Granularity
- Security

Open Standards

- Open standards maximize choice and interoperability, have extensive peer review, and usually produce the highest quality solutions to a given problem. We embrace open standards that contribute to our directory service vision.

Open Source

- Wherever possible, we use and contribute to open source implementations of open standards.
- Some open source software used by the directory service:
  - RPM – used to manage packages on both Linux and HP-UX
  - OpenSSH – used for secure remote access to all servers/devices
  - CVS – used to manage revision-controlled code and configuration files
  - rsync – used to synchronize data files between servers; uses OpenSSH as a transport mechanism
  - Perl – used for practically all programming tasks, from simple maintenance scripts to complex LDAP data feeds
  - Postfix – LDAP-aware mail server of choice
  - Samba – used for NT domain authentication
  - Sudo – used to delegate root privileges to administrators
  - Apache – our web server of choice

Suitability of Data

- Data should be well suited and formatted for inclusion in the directory; data will:
  - Follow open standards and industry-wide conventions wherever possible. IETF RFCs and other sources for schema definitions should be consulted when adding data to the directory.
  - Have a high read/write ratio.
  - Have descriptive names and values wherever possible. If abbreviations or codes are absolutely necessary, they can be stored in addition to the descriptive form of the name.
  - Be unique. Duplication of directory data must be avoided.

Wide Readability

- Data that goes into the directory should be useful to a broad audience.
- From both an architecture and performance standpoint, storing large amounts of application-specific data is not favorable. The directory should not be used as a private configuration repository for standalone applications.

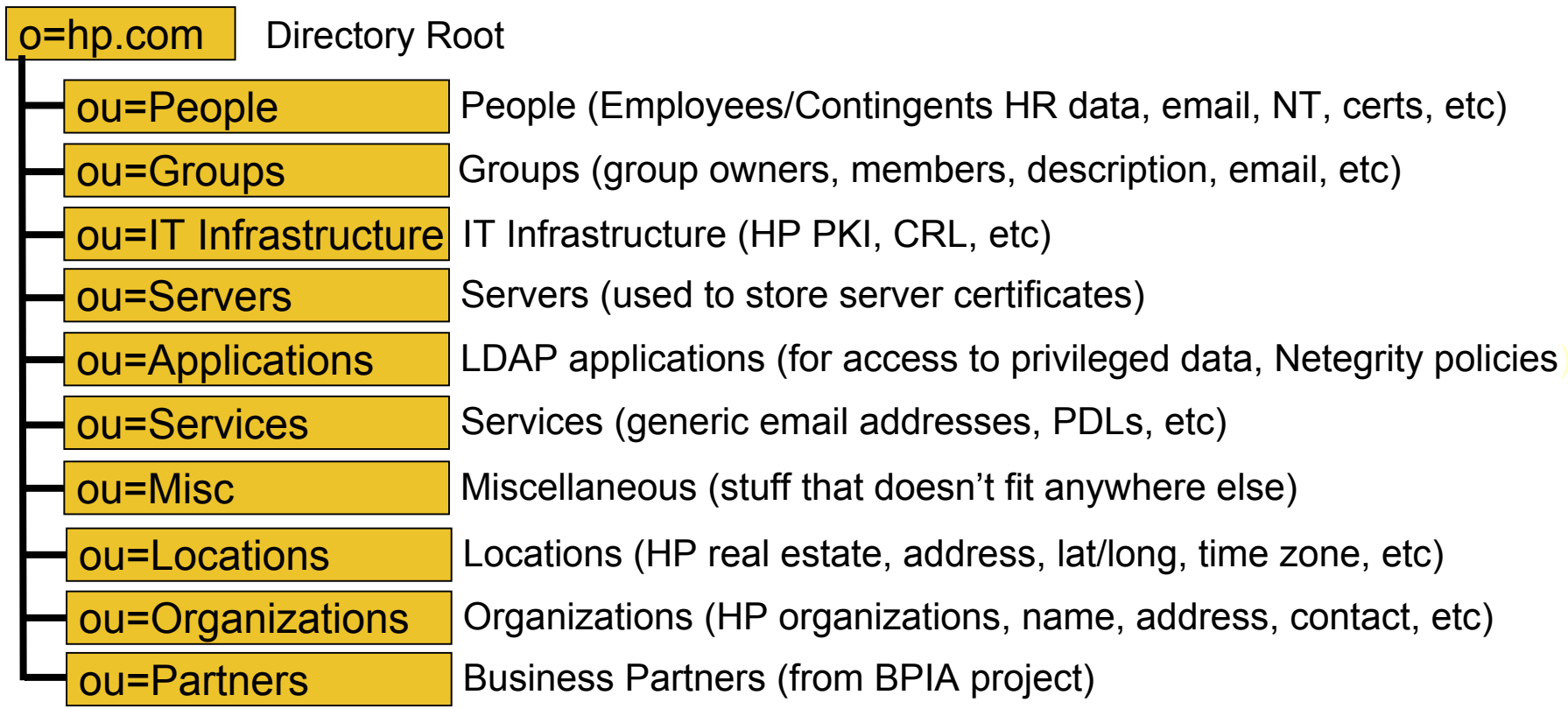
Balanced Simplicity and Granularity

- Access control rules to directory data is kept as simple as possible and as open as possible. Keeping access control rules simple is an important part of keeping the overall directory infrastructure scalable, flexible, and maintainable.
- Where necessary, high levels of granularity necessary for data security classifications are supported, but extensive granularity to support an application's proprietary attributes is discouraged. Applications with highly complex security requirements should make use of a private directory or database.

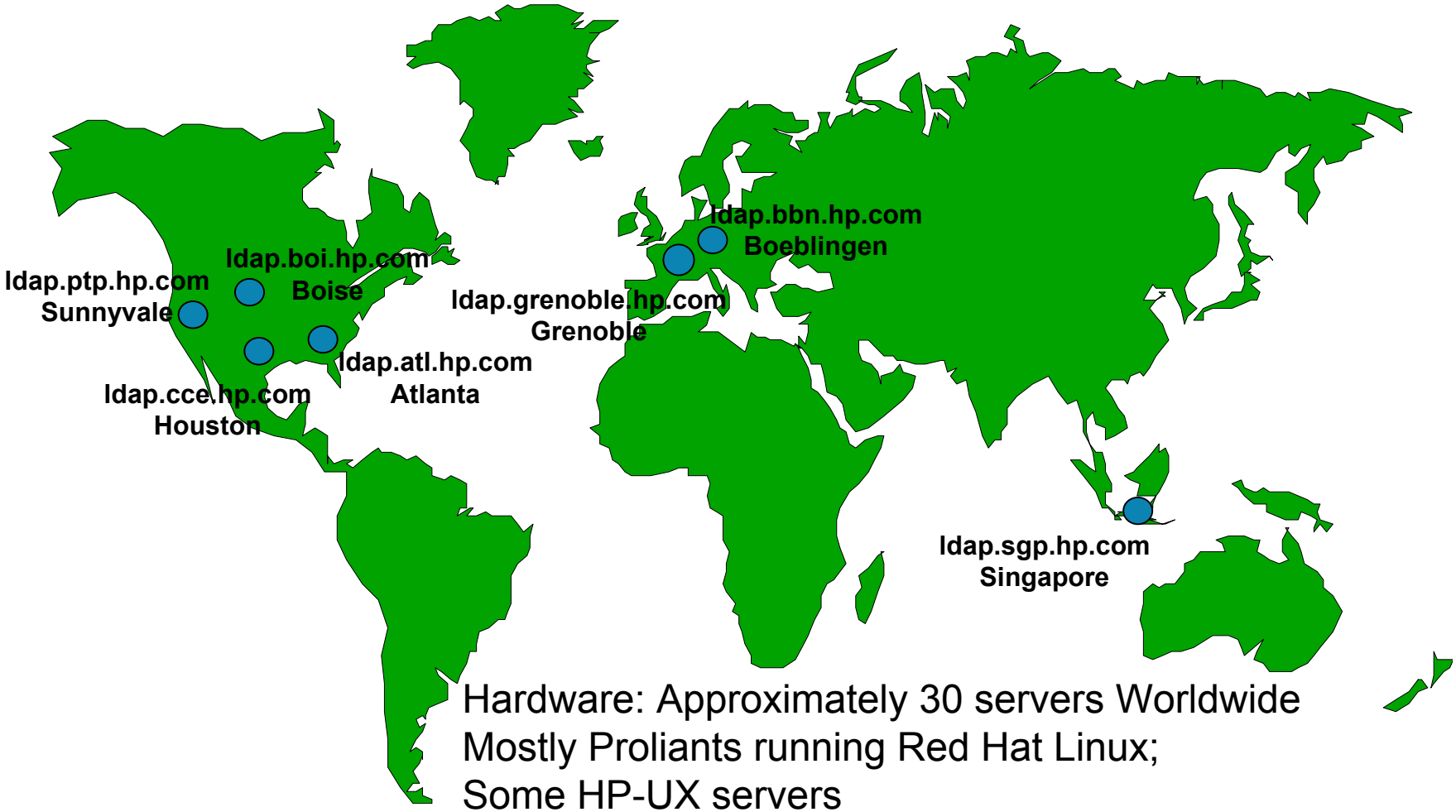
Security

- Security is a founding principle of the directory, as it is a core component of authentication and authorization decisions.
- Operational security is maintained to the highest specifications. The directory data itself is secured based on two factors: HP's legal and privacy requirements, and the guiding principle of ubiquitous access. Both factors are carefully considered when assigning access control rules. Some data, above a sufficiently sensitive level, is not appropriate to store in the directory.

Directory Information Tree (DIT)



Geographic Clusters

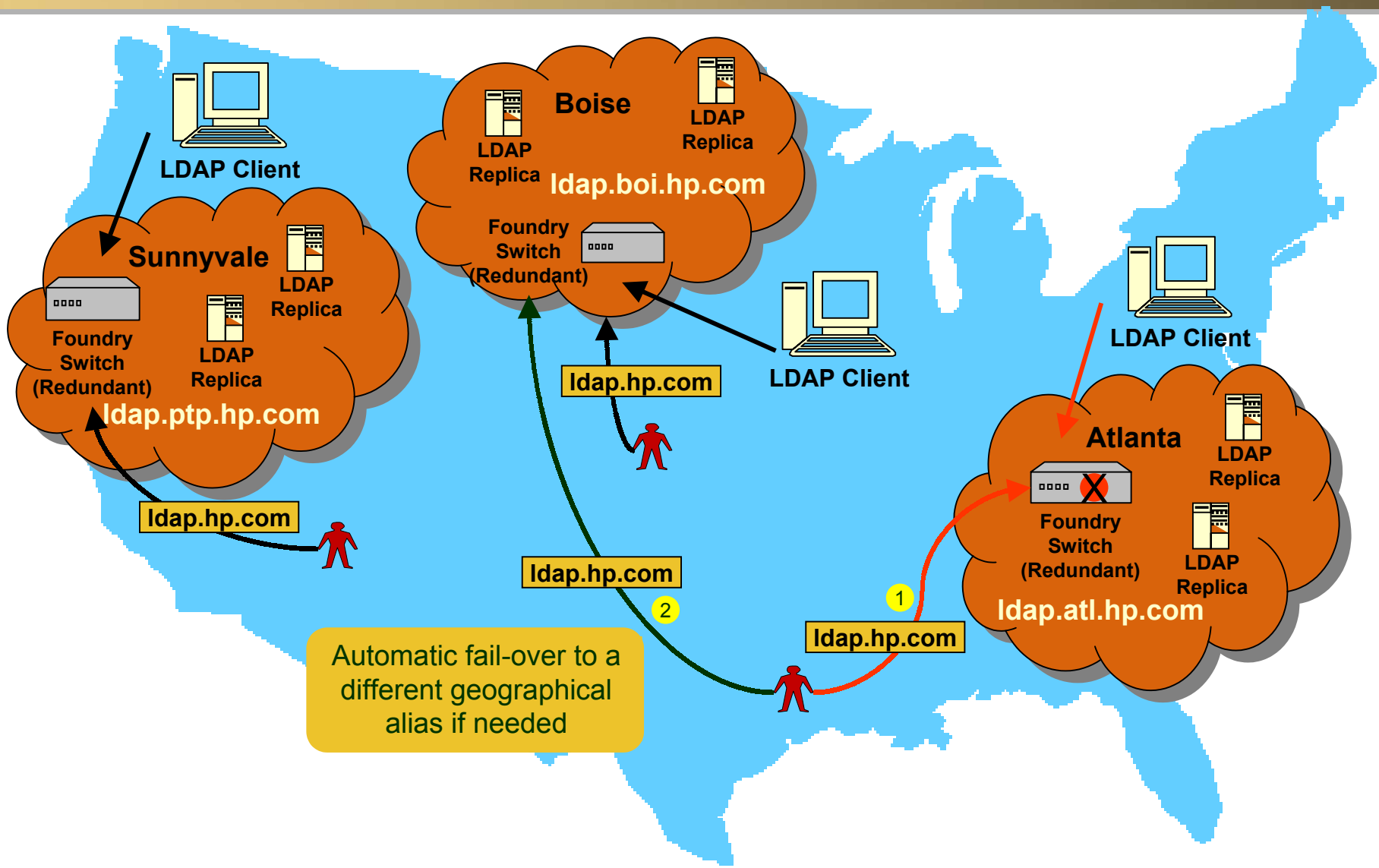


Hardware: Approximately 30 servers Worldwide
Mostly Proliants running Red Hat Linux;
Some HP-UX servers
Software: Sun ONE Directory Server 5.x

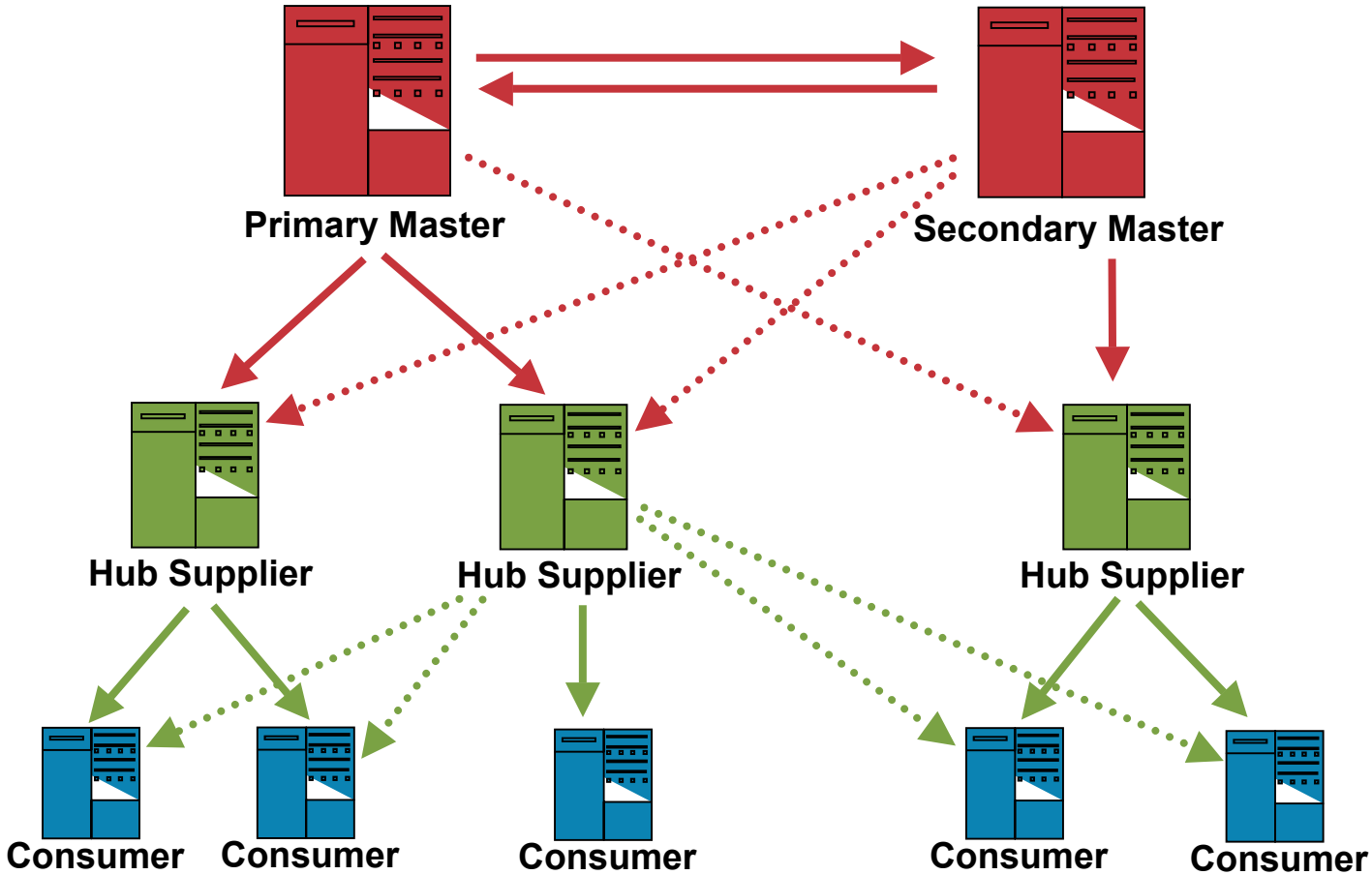
Load Balancing/Failover Implementation

- Replication is a natural strategy for making LDAP directories highly available
  - Capacity can be added simply by adding more servers
  - Easy to take down servers for maintenance
- Foundry ServerIron Switch
  - Layer 7 health checking for LDAP
  - Direct Server Return functionality
    - All real servers listen on the VIP address of the local cluster
    - Packets from LDAP clients go through the Foundry switch to a real server; responses go back directly to the client
    - Real servers see the actual source IP address, important for monitoring and log file analysis

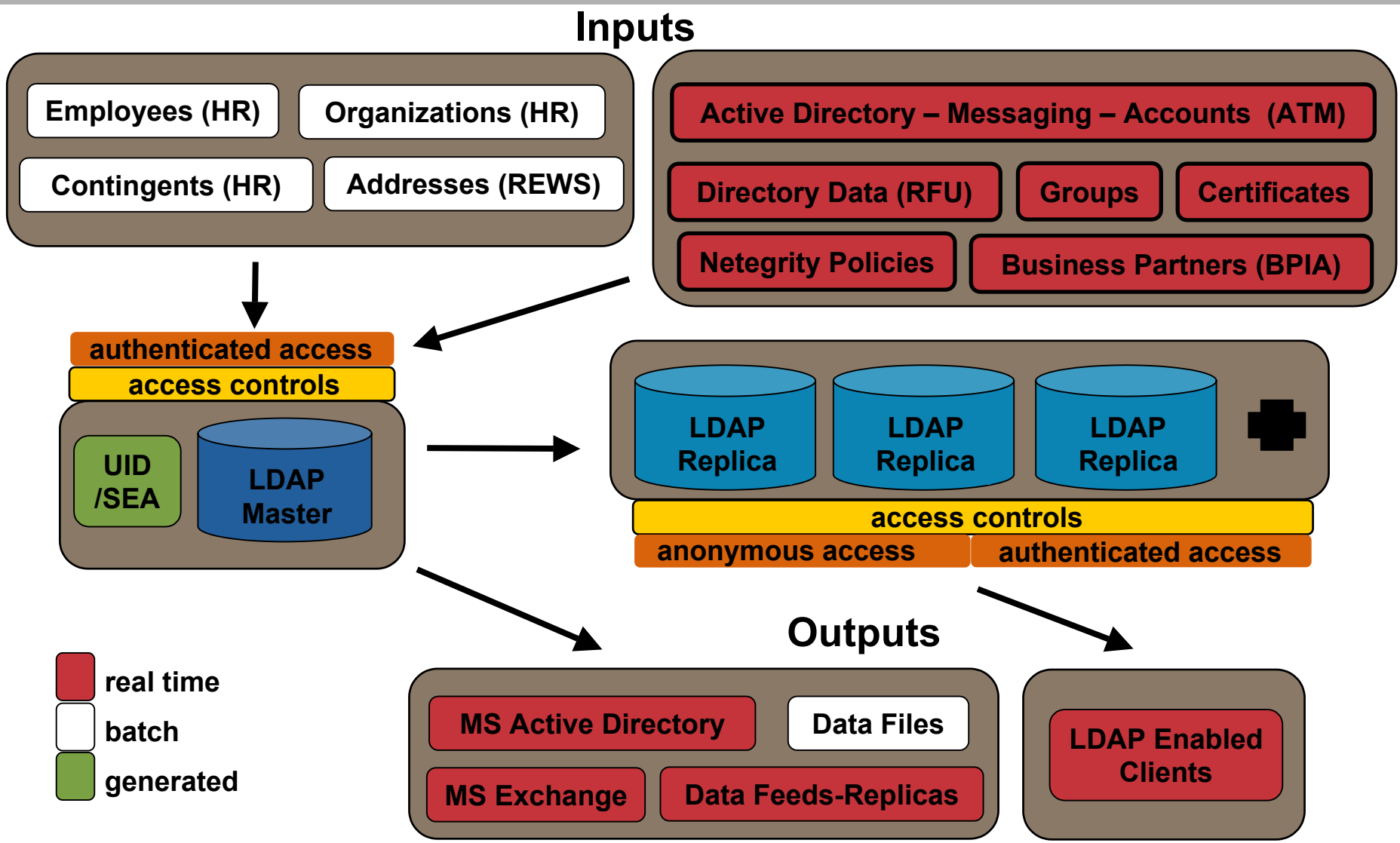
Global Load Balancing and Failover Scenarios -- ldap.hp.com



Replication Topology



Data Flows


Inputs
Employees (HR)
Organizations (HR)
Contingents (HR)
Addresses (REWS)
Active Directory – Messaging – Accounts (ATM)
Directory Data (RFU)
Groups
Certificates
Netegrity Policies
Business Partners (BPIA)
authenticated access
access controls
UID /SEA
LDAP Master
LDAP Replica
LDAP Replica
LDAP Replica
access controls
anonymous access
authenticated access
Outputs
MS Active Directory
Data Files
MS Exchange
Data Feeds-Replicas
LDAP Enabled Clients
real time
batch
generated

Data Sync Processes

- Lightweight, modular approach as opposed to a monolithic meta-directory solution
- Some data sources update ED directly via LDAP
- For others, we pull flat XML or ASCII delimited files and generate incremental LDAP updates
- Extensive use of LDIF, Perl scripts in data manipulation
- Example of a lightweight tool – ldifdiff.pl

ldifdiff.pl

- Given source and target LDIF files sorted by the same key attribute, generates LDIF change output to update the target

source.ldif

```
dn: cn=Notebook
color: Purple

dn: cn=Printer
color: Blue

dn: cn=Scanner
color: Silver
```

target.ldif

```
dn: cn=Notebook
color: Red

dn: cn=Printer
color: Blue
```

ldifdiff.pl output

```
% ldifdiff.pl -k dn source.ldif target.ldif
dn: cn=Notebook
changetype: modify
replace: color
color: Purple

dn: cn=Scanner
changetype: add
color: Silver
```

Infrastructure Management Processes

- The directory has a GUI interface, but can also be manipulated via LDAP commands. LDAP can be used to add/delete replication agreements, modify server parameters, monitor information, etc.
- Example command line tool – replicamonitor.pl

```
% replicamonitor.pl
ldap-master.hp.com 3ea9ad3b000000290000: 3f11a94d0000000a0000
(2003071314:47:41#0#10#0)
 edhub.atl.hp.com: 3f11a94d0000000a0000 (2003071314:47:41#0#10#0)
 edhub.boi.hp.com: 3f11a94d0000000a0000 (2003071314:47:41#0#10#0)
 edhub.cce.hp.com: 3f11a94d0000000a0000 (2003071314:47:41#0#10#0)
 edhub.ptp.hp.com: 3f11a94d0000000a0000 (2003071314:47:41#0#10#0)
 edhub.sgp.hp.com: 3f11a94d0000000a0000 (2003071314:47:41#0#10#0)
```

Data Download Tool

- For apps that are not LDAP-aware, or for those that need to download large volumes of data on a periodic basis, we offer a web-based data download tool.
- Download options include LDIF and tab-delimited files (easy to import into spreadsheets, databases, etc).
- Users can perform selective queries, and/or select a subset of attributes, to slice and dice various sections of data in the directory (e.g. return a tab-delimited file containing the telephone numbers and email addresses of all US employees, one employee per line).
- The tool can be used interactively through any web browser, or by any standards-based HTTP client (e.g. curl, wget, Perl LWP). Example:

```
wget http://directory.hp.com/download/download.cgi?param...
```

Some Contributions to Open Source Software

- NT authentication plug-in – provides pass-through authentication to NT/AD
  - http://sourceforge.net/projects/dsntauth
- Net::LDAP – utility tools, other enhancements
- Jabber – directory integration
- Postfix – enhanced LDAP groups integration
- OpenSSH – alternate key formats, certificate/directory integration
- Stunnel – X.509 integration

Interex, Encompass and HP bring you a powerful new HP World.