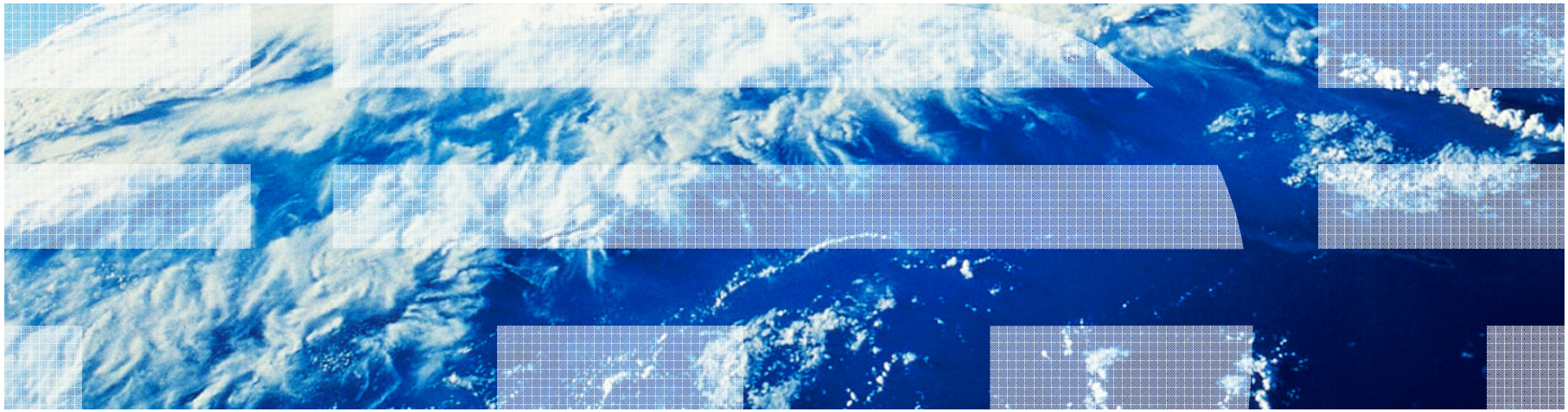
Exploratory Cloud Projects in IBM Research

Dr. Yaron Wolfsthal, Sr Manager, System Technologies

IBM Haifa Research Laboratory



EGI Technical Forum 2010

09/15/2010

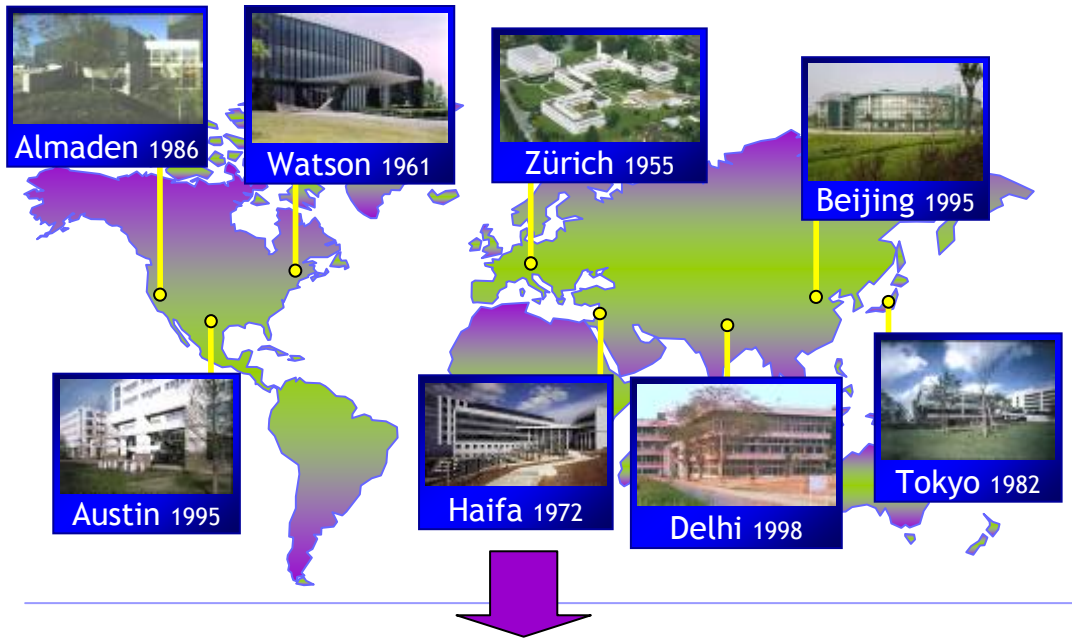
© 2010 IBM Corporation

In This Talk

- Introduction
- Federated Cloud Computing – the Reservoir Project
- Cloud-Based Data Storage – the VISION Project
- Validating Cloud Middleware – the Cloud Living Lab Project

© 2010 IBM Corporation

IBM Research



Research Disciplines

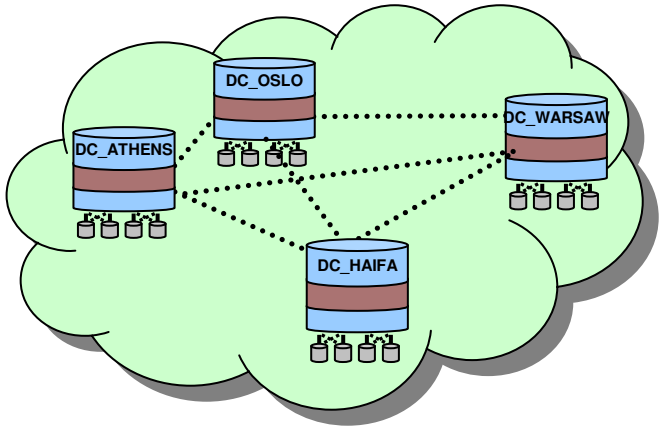
- Technology
- Systems
- Storage
- Software
- Industry Solutions
- Services
- Math Sciences
- Exploratory

Achievements

5 Nobel Prizes

6 Turing Awards

7 US National Medals of Technology

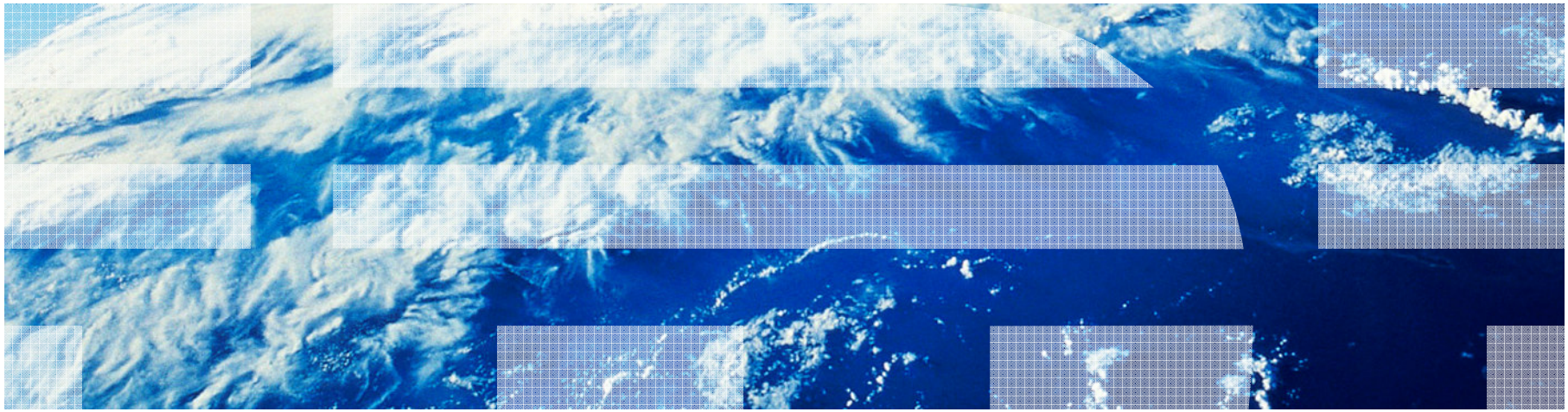


Cloud Computing

Mission:
Contributions to IBM products, and
… to science and technology
… to European research agenda

© 2010 IBM Corporation

Cloud Computing
Definitions, Research Directions



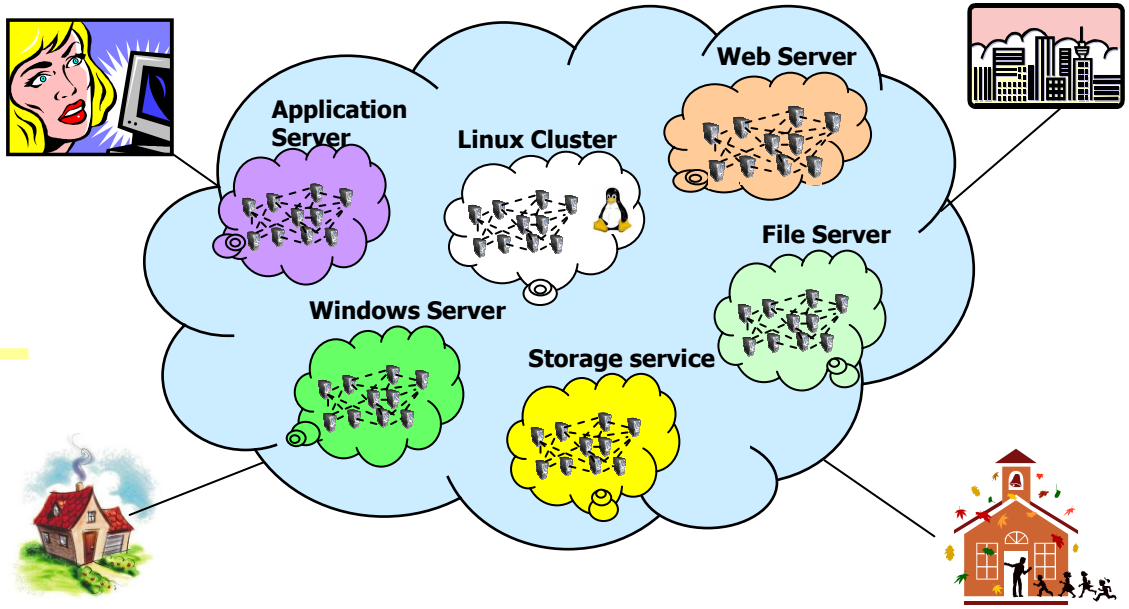
09/12/2010

© 2010 IBM Corporation

Cloud Computing Characteristics

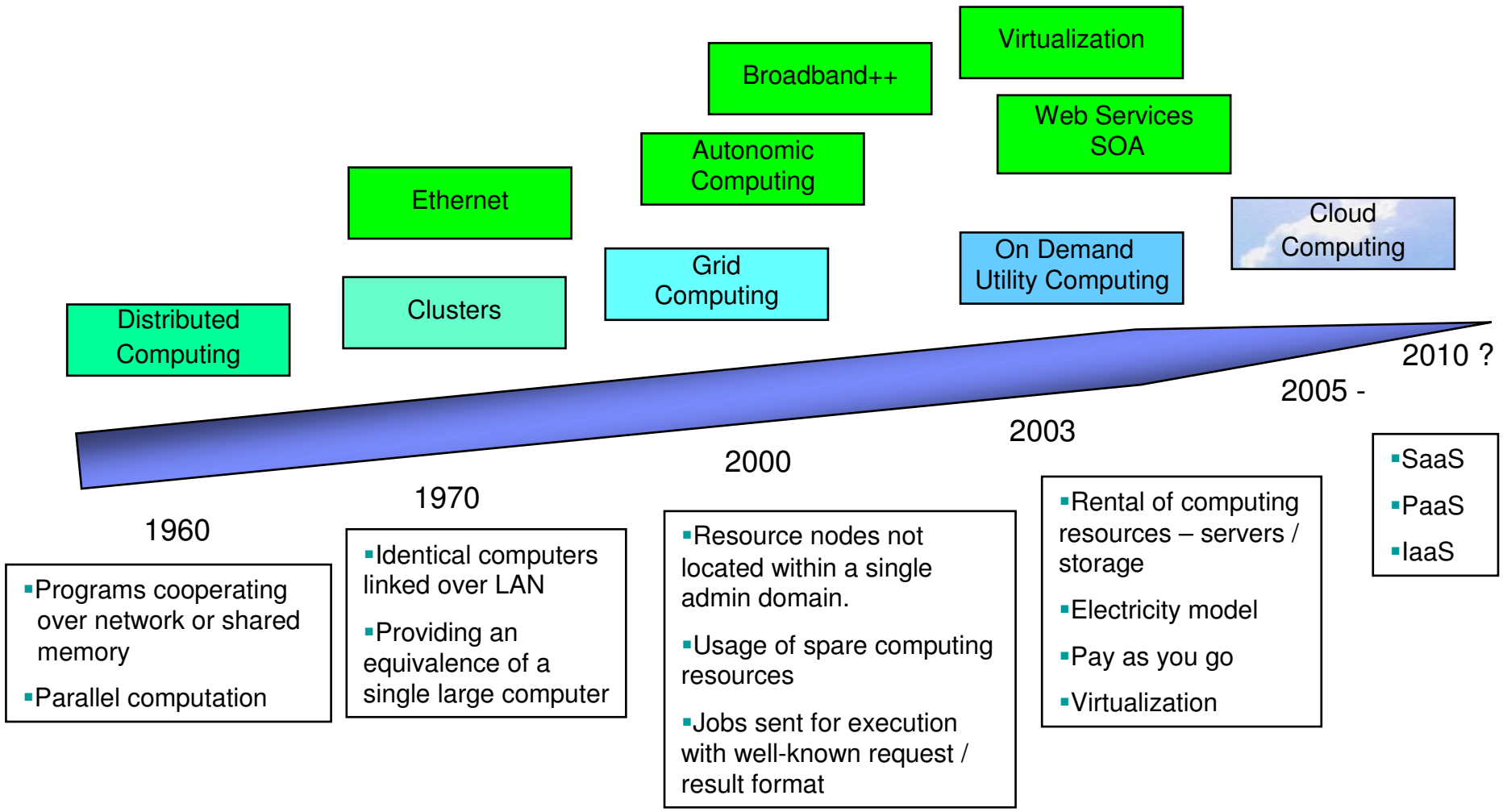
- Infrastructure – compute, storage and network
- Access to services/resources over web
- Utility model – provision on demand, charge back on use
- Ease of creation and delivery of services
- Simplified user experience
- User-composable services

Becoming a reality
-- Broadband
-- Virtualization
-- Strong Hardware



© 2010 IBM Corporation

The Evolution Towards Cloud Computing

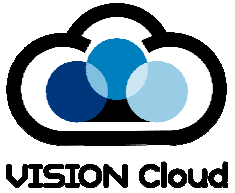


© 2010 IBM Corporation

Cloud Presents Significant Research Topics

- Massive Scale with highly distributed management, QoS challenges
  - Auto-scaling, elasticity management
  - Federation of compute resources (Reservoir)

  - Storage and data (VISION)



- Tools/platform to design and test services

- Cloud security
  - e.g., multi-tenancy

© 2010 IBM Corporation

The Reservoir Project:
Resources and Services Virtualization without Boundaries

- 3 Years EU FP7 project started in February 2008
- Budget: € 17 million
- 13 partners from across industry, academia and standards bodies
  - IBM is the technical leader as well as the project coordinator
- Public web site http://www.reservoir-fp7.eu/

Company/institute names, logos and trademarks are the property of their respective owners

© 2010 IBM Corporation

RESERVOIR Goals

- Develop and promote (thru standardization) an open architecture and technologies for federated cloud computing
  - **where resources and services can be transparently and dynamically managed, provisioned and relocated like utilities – virtually "without borders"**

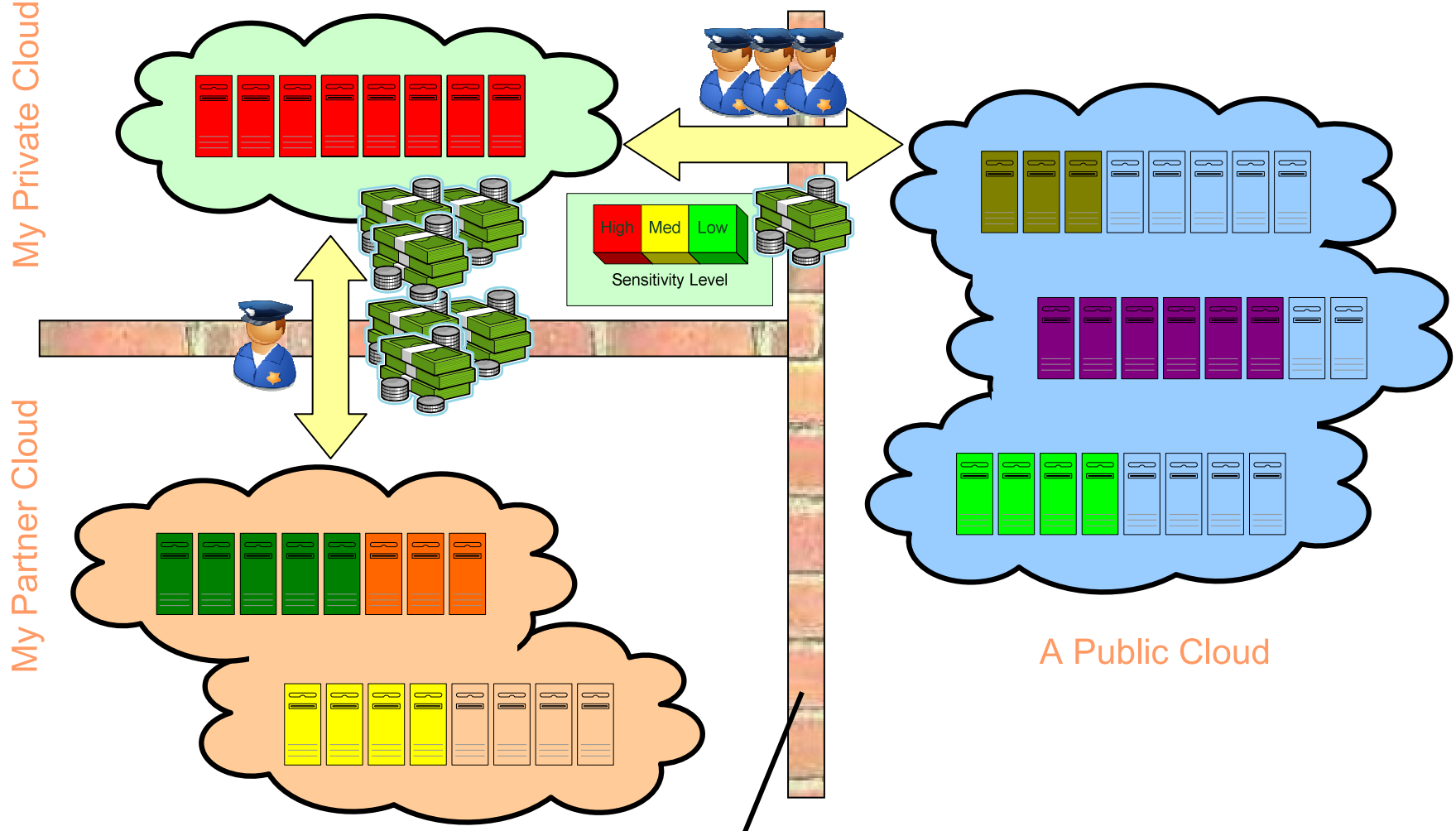
Workload Management

Infrastructure

- Identify the gaps in today's offerings, build (prototype) the bridges them
- Develop a proof of concept that validates the architecture and shows the viability and value of a federated cloud

© 2010 IBM Corporation

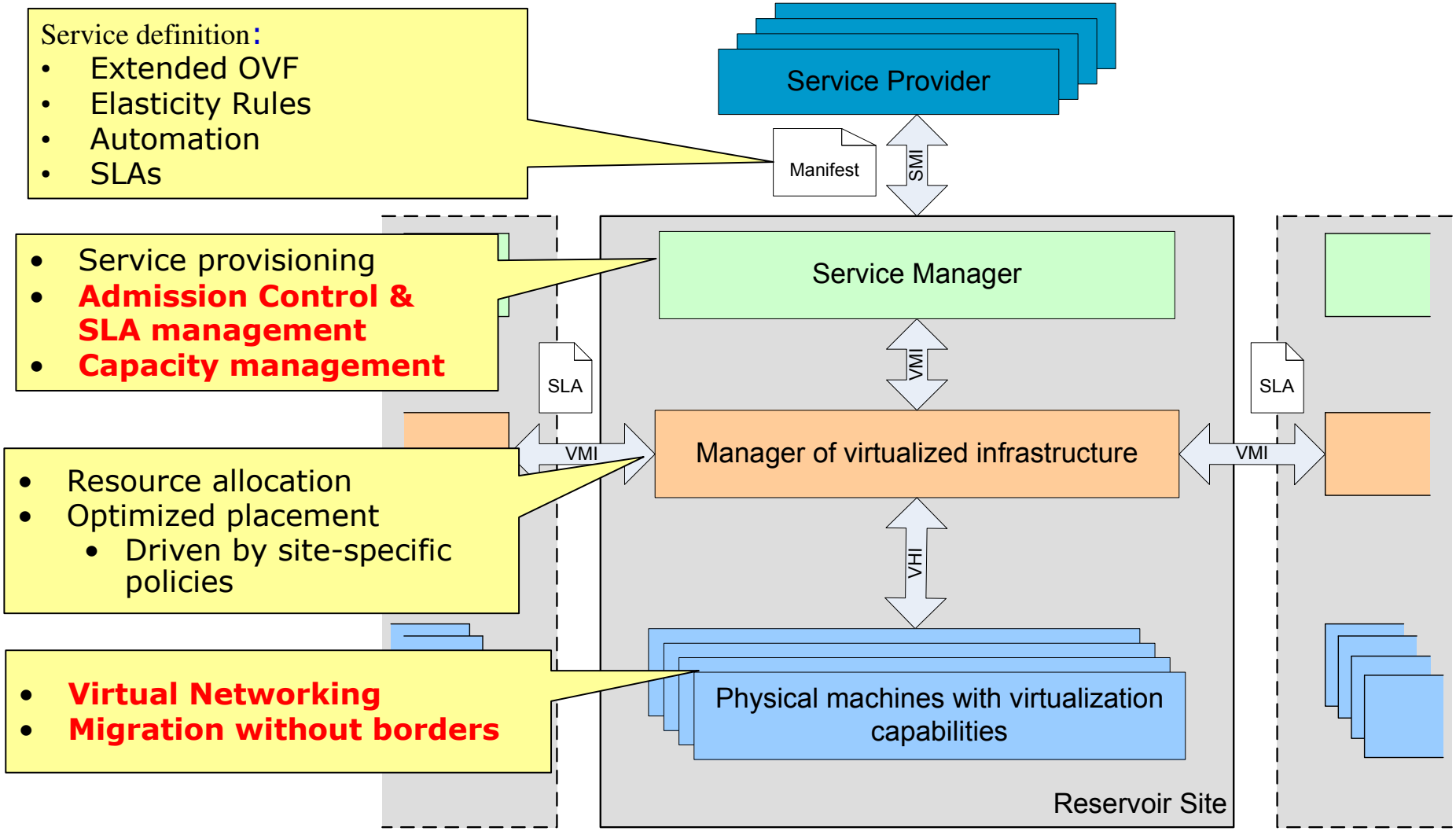
Federation of Cooperating Computing Clouds



Clouds differ in terms of cost and trust level, for while a public cloud maybe very cheap, companies will be reluctant to put in there sensitive data/compute

© 2010 IBM Corporation

The Reservoir Architecture



© 2010 IBM Corporation

Reservoir - Summary

- Federation is the only way to move cloud forward
- To make federation work, technologies are required – migration, autoscaling etc

CNNMoney.com
A Service of CNN, Fortune & Money

Symbol | Get Quote | Keyword

Home | Business News | Markets | Personal Finance | Real Estate | Technology | Small Business

IBM and European Union Launch Joint Research Initiative for Cloud Computing

February 05, 2008: 12:00 AM EST

marketwire

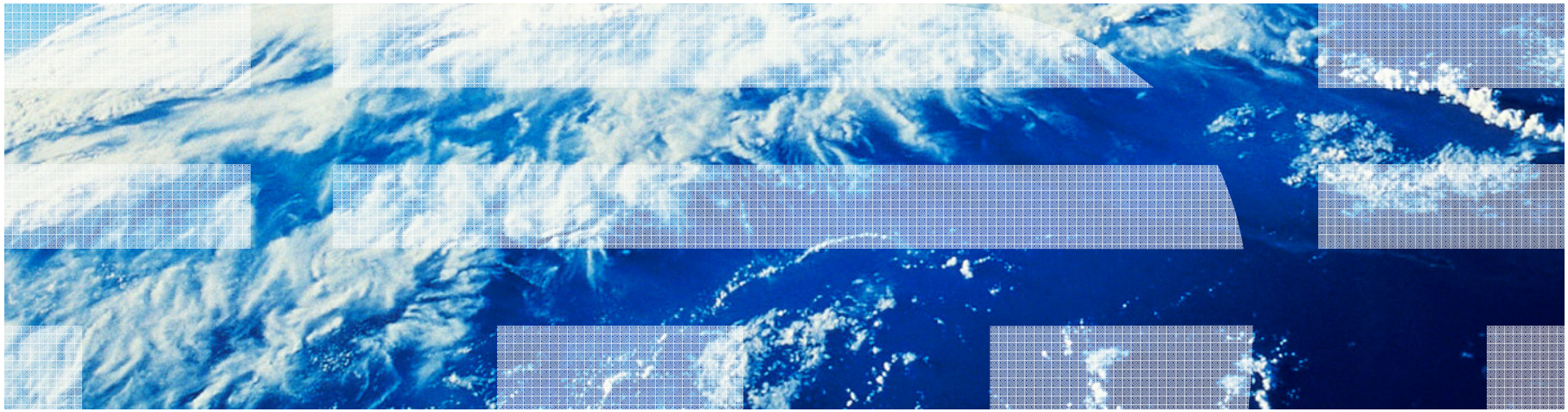
Expanding its cloud computing initiative, IBM (NYSE: IBM) today announced that it is leading a joint research initiative of 13 European partners to develop technologies that help automate the fluctuating demand for IT resources in a cloud computing environment.

The 17M Euro EU-funded initiative, called RESERVOIR -- Resources and Services Virtualization without Barriers -- will explore the deployment and management of IT services across different administrative domains, IT platforms and geographies. This cloud computing project aims to develop technologies to support a service-based online economy, where resources and services are transparently provisioned and managed.

Cloud computing is an emerging approach to shared infrastructure in which large pools of systems are linked together to provide IT services. The need for such environments is fueled by dramatic growth in connected devices, real-time data streams, and the adoption of service oriented architectures and Web 2.0 applications, such as mashups, open collaboration, social networking and mobile commerce. Continuing advances in the performance of digital components has resulted in a massive increase in the scale of IT environments, driving the need to manage them as a unified cloud.

IBM, which has been researching technologies related to cloud computing for more than a decade, kicked off a companywide cloud computing initiative in 2007 across its server, software, services, and R&D units. In November, IBM unveiled plans for "Blue Cloud," a series of cloud computing offerings that will allow corporate data centers to operate more like the Internet through improved organization and simplicity.

Cloud-Based Data Storage – the VISION Project



09/12/2010

© 2010 IBM Corporation

2010: Europe Has Made First Steps in Cloud Computing

- First research projects, e.g., RESERVOIR
- First cloud centers established, e.g. Amsterdam, Dublin
- First emergence of Cloud technology SMEs, e.g.,
- First steps in standards , e.g., OGF

- The challenges in Cloud Computing are becoming clear:

A large emerging gap in support for cloud storage and cloud data services

© 2010 IBM Corporation

The Data Deluge

Domain	Digital Data Characteristics[1]
Search	Google serves 300 million search requests and processes 20 petabytes of data per day
Email Services	Worldwide email traffic will total 247 billion messages per day in 2009 The working storage consumed by commercial email accounts exceeds 11 petabytes
Web 2.0	Facebook servers host 50 billion files and serves 50 million requests per second
Government	London's 200 traffic surveillance cameras capture and transmit 64 trillion bits per day
Commerce	Chevron, an US oil company, accumulates data at a rate of 2 terabytes a day
Science	The new Large Hadron Collider at CERN is generating about 10 petabytes of data per second, and stores 15 petabytes of data each year
"Born Digital" Visual Data	Information that is born digital, (e.g. images, video clips, digital TV, surveillance streams) is predicted to grow from 281 exabytes in 2007 to 1800 exabytes in 2011

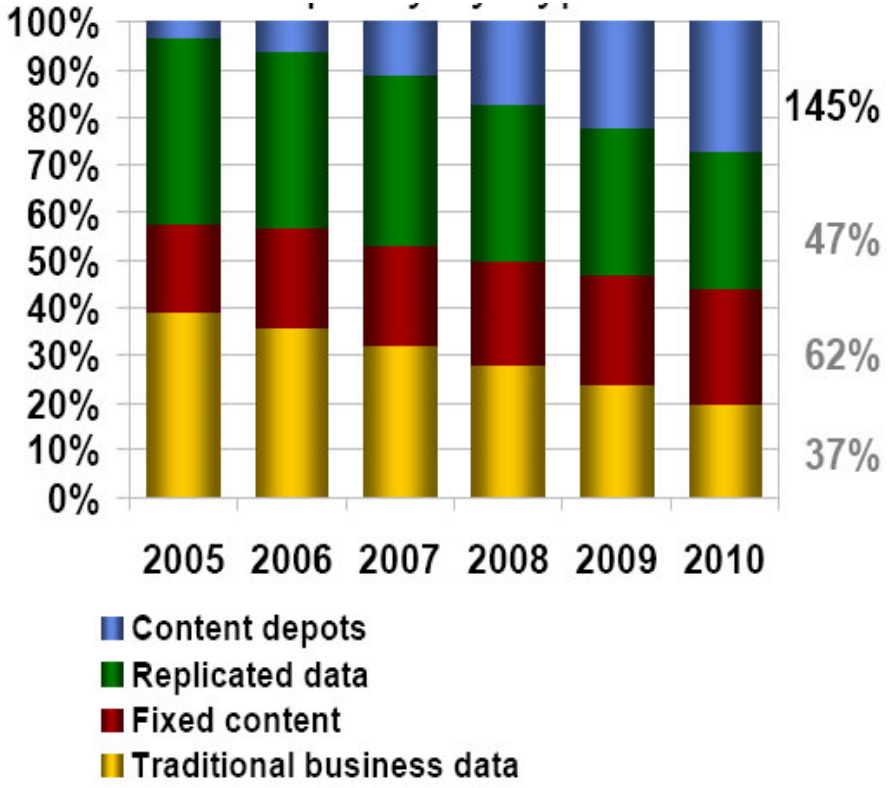
Prefixes for bit and byte multiples

Decimal			Binary		
Value	SI		Value	IEC	JEDEC
1000	k	kilo-	1024	Ki	kibi-	K	kilo-
1000^2	M	mega-	1024^2	Mi	mebi-	M	mega-
1000^3	G	giga-	1024^3	Gi	gibi-	G	giga-
1000^4	T	tera-	1024^4	Ti	tebi-		
1000^5	P	peta-	1024^5	Pi	pebi-		
1000^6	E	exa-	1024^6	Ei	exbi-		
1000^7	Z	zetta-	1024^7	Zi	zebi-		
1000^8	Y	yotta-	1024^8	Yi	yobi-		

The Emerging Petabyte Age

- The explosion of personal and organizational digital data is one of the most significant characteristics of this decade.
- Volume of data in enterprises is doubling approximately every three years (Forrester Research).
- According to IDC report :
  - Data created surpassed available storage in 2007, and this trend is expected to double in five years
  - Unstructured data accounts for more than 90% of the digital universe
  - Internet-scale photo sharing sites store billions of images, petabytes of data

- The digital universe will grow 10-fold in five years, from ~177 exabytes in 2006 to over 1,700 exabytes in 2011.



SOURCE: IDC, The Exploding Digital Universe, 2008

© 2010 IBM Corporation

Data Availability and Accessibility

- There are basic requirements that must be met to satisfy both private and enterprise clients
  - **Access from any location**
    The data must be accessible from anywhere.
    The data retrieval time should be minimized.
  - **Access from any device**
    Access the same software, media and data items from any web-enabled device, regardless of its type.
    E.g. "smart" mobile and car phones, lightweight laptops and notebooks.
  - **Access at any time**
    The data must always be available.
  - **Access from any application**
    Compute close to storage enabling a variety of applications each with its own selection and presentation model.

© 2010 IBM Corporation

Enter Cloud Storage

- Definition
  - Model where data is stored on multiple virtual servers, generally hosted by 3rd parties, rather than being hosted on dedicated servers

- Storage coming from the cloud facilitates:
  - Content depot for application and services
    - e.g. public data, Web 2.0 user-generated content
  - Information retention services for businesses
    - e.g. backup and archive
  - Platform as a service
    - The storage aspect of the platform
  - eScience

- All of the above can share a common infrastructure

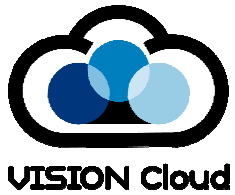
© 2010 IBM Corporation

Attributes of Cloud Storage

- Provide access to seemingly infinite storage capacity
- Provide access to heterogeneous types of storage (disks vs. tapes, high vs. low end) regardless of the physical location of the user of the data, client or application
- Can locate the services at low-cost sites, potentially turning off some copies of the data
- Built-in disaster recovery services
- Better utilization of capacity, power, bandwidth

© 2010 IBM Corporation

VISION – Virtualized Storage Services for the Future Internet



DATA / Storage Services will be at the heart of the Future Internet and must be properly supported by infrastructure

© 2010 IBM Corporation

VISION – An IBM-led European Government Initiative on Cloud Storage

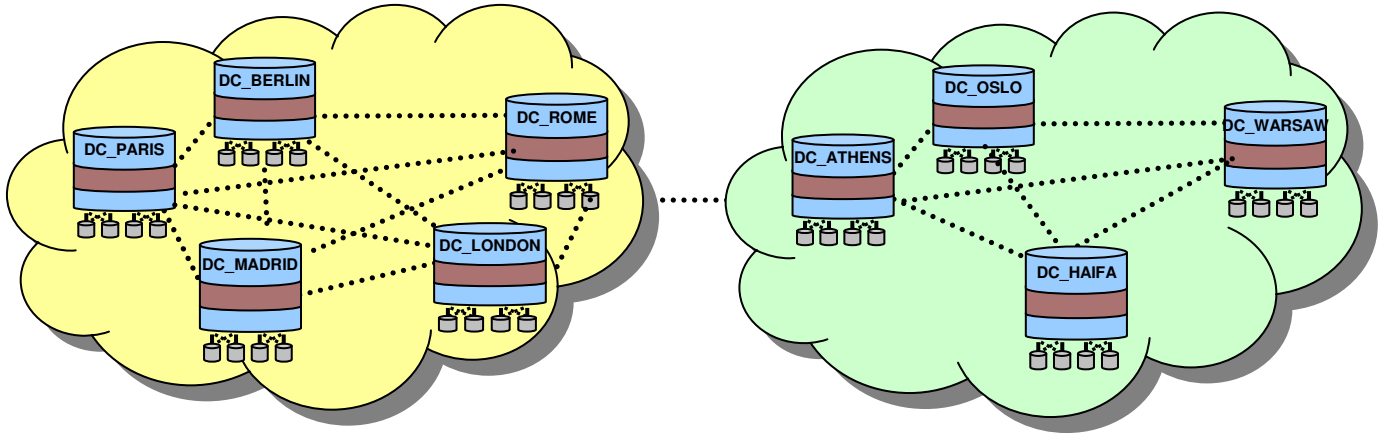
- Motivation for initiative
  - Enables research on future cloud storage features
  - Gather and understand requirements through collaboration with potential customers
  - Lay the foundation for Future Internet from a storage perspective
- Consortium
  - IBM leads
  - Participants include SAP, Siemens, Telefonica, Engineering, Deutche Welle, RAI, France Telecom, Telenor, NTUA, UMEA, SICS, UniMe, SNIA, iTricity
- Status
  - Negotiation completed
  - Kickoff meeting October 2010

Company/institute names, logos and trademarks are the property of their respective owners

© 2010 IBM Corporation

VISION's Goal

Infrastructure for delivery of data-intensive storage services, facilitating the convergence of ICT, media and telecommunications



Technology Components

1. **Raise Abstraction Level of Storage**: objects with user-defined and system-defined metadata
2. **Data Mobility and Federation**: enable comprehensive data migration and interoperability across remote locations
3. **Computational Storage**: technology for specifying and executing computations close to storage
4. **Content-Centric Storage**: facilitate access to data by content and its relationships
5. **Advanced Capabilities for Cloud-based Storage**: support delivery of data-intensive services securely, at the desired QoS, at competitive costs

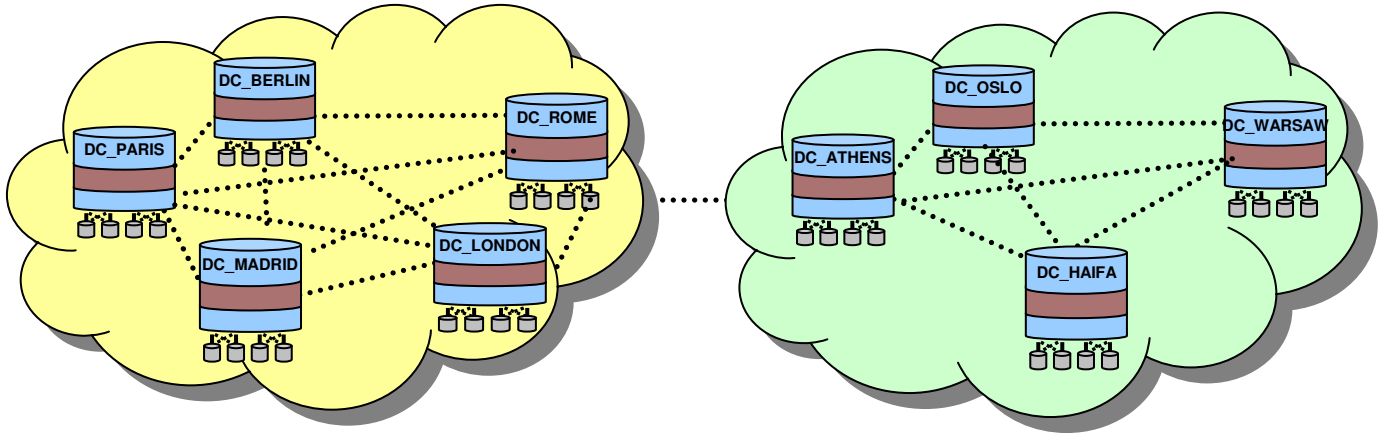
Validation Scenarios

- **Media**
- **Telco**
- **Healthcare**
- **Enterprise**

© 2010 IBM Corporation

VISION's Goal

Infrastructure for delivery of data-intensive storage services, facilitating the convergence of ICT, media and telecommunications



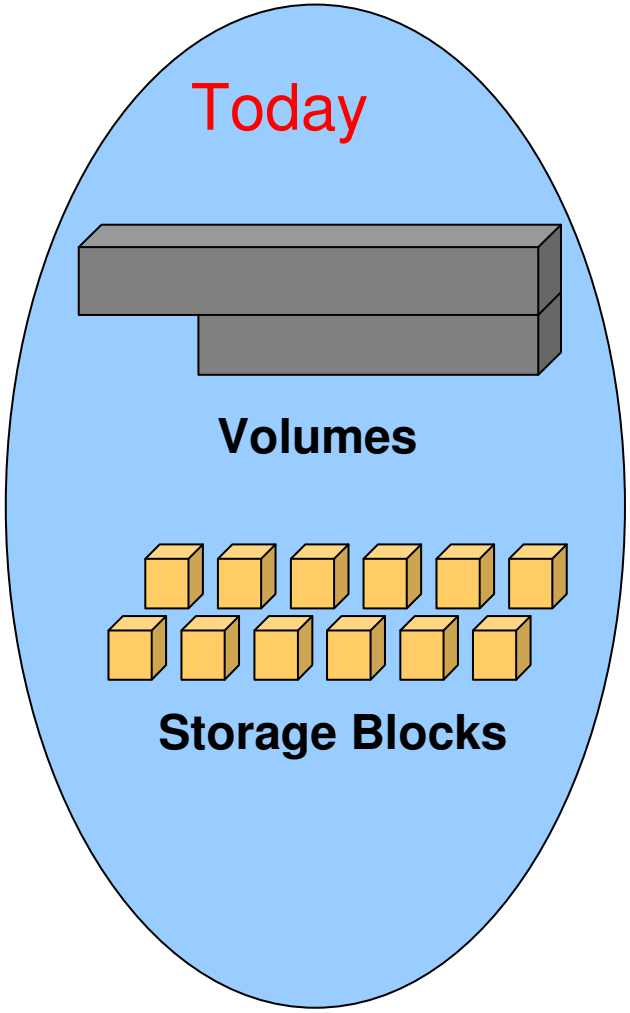
Technology Components

1. **Raise Abstraction Level of Storage**: objects with user-defined and system-defined metadata
2. **Data Mobility and Federation**: enable comprehensive data migration and interoperability across remote locations
3. **Computational Storage**: technology for specifying and executing computations close to storage
4. **Content-Centric Storage**: facilitate access to data by content and its relationships
5. **Advanced Capabilities for Cloud-based Storage**: support delivery of data-intensive services securely, at the desired QoS, at competitive costs

© 2010 IBM Corporation

Why Should We Raise the Abstraction Level of Storage?

- **Improved device and data sharing**
  - Platform-dependent metadata moved to device
  - Systems need only agree on naming
- **Improved scalability & security**
  - Devices directly handle client requests
  - Security w/ application-level granularity
- **Improved performance**
  - Applications can provide hints, QoS, policy
  - Data types can be differentiated at the device
- **Improved storage management**
  - Self-managed, policy-driven storage
  - Storage devices become more autonomous

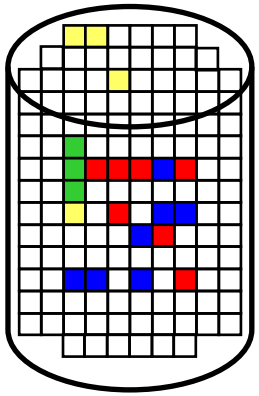


© 2010 IBM Corporation

Enter Object Storage
Encapsulates Data and Metadata in a Persistent Manner

- Raises level of abstraction to present object and not individual blocks
  - Off-loads physical allocation of blocks
  - Fine grain, object-level security
  - Groups objects into collections
- T10 Standard

Today's Block Device

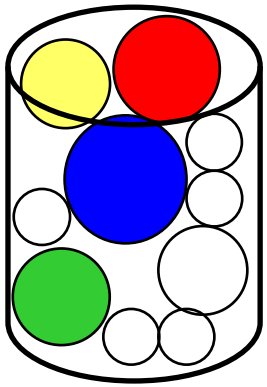


Operations
- read block
- write block

Security
- Weak
- Full disk

Allocation
- External

Object Store



Operations
- read object offset
- write object offset
- create object
- delete object

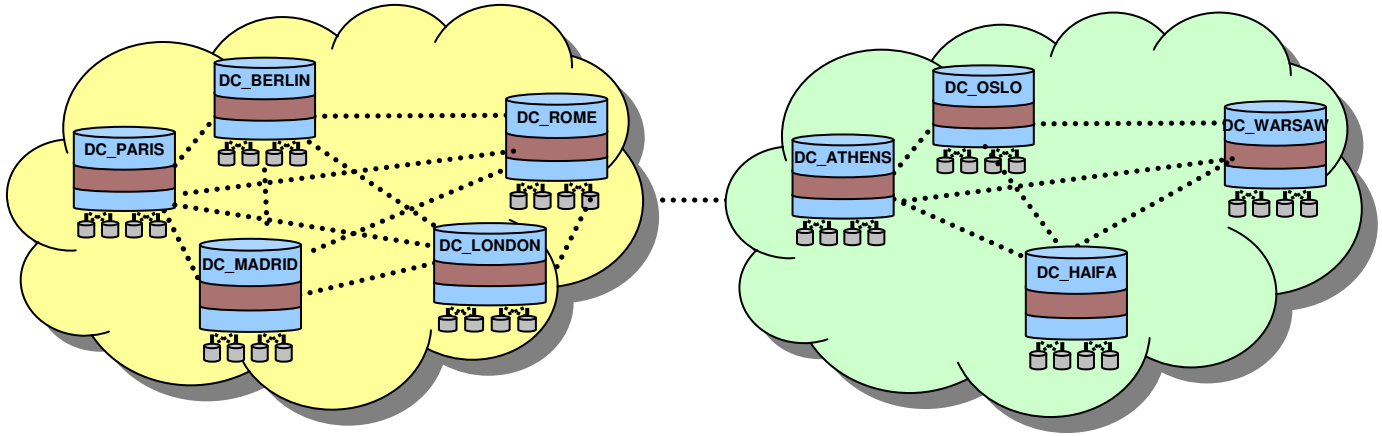
Security
- Strong
- Per Object

Allocation
- Local

© 2010 IBM Corporation

VISION's Goal

Infrastructure for delivery of data-intensive storage services, facilitating the convergence of ICT, media and telecommunications



Technology Components

1. **Raise Abstraction Level of Storage**: objects with user-defined and system-defined metadata
2. **Data Mobility and Federation**: enable comprehensive data migration and interoperability across remote locations
3. **Computational Storage**: technology for specifying and executing computations close to storage
4. **Content-Centric Storage**: facilitate access to data by content and its relationships
5. **Advanced Capabilities for Cloud-based Storage**: support delivery of data-intensive services securely, at the desired QoS, at competitive costs

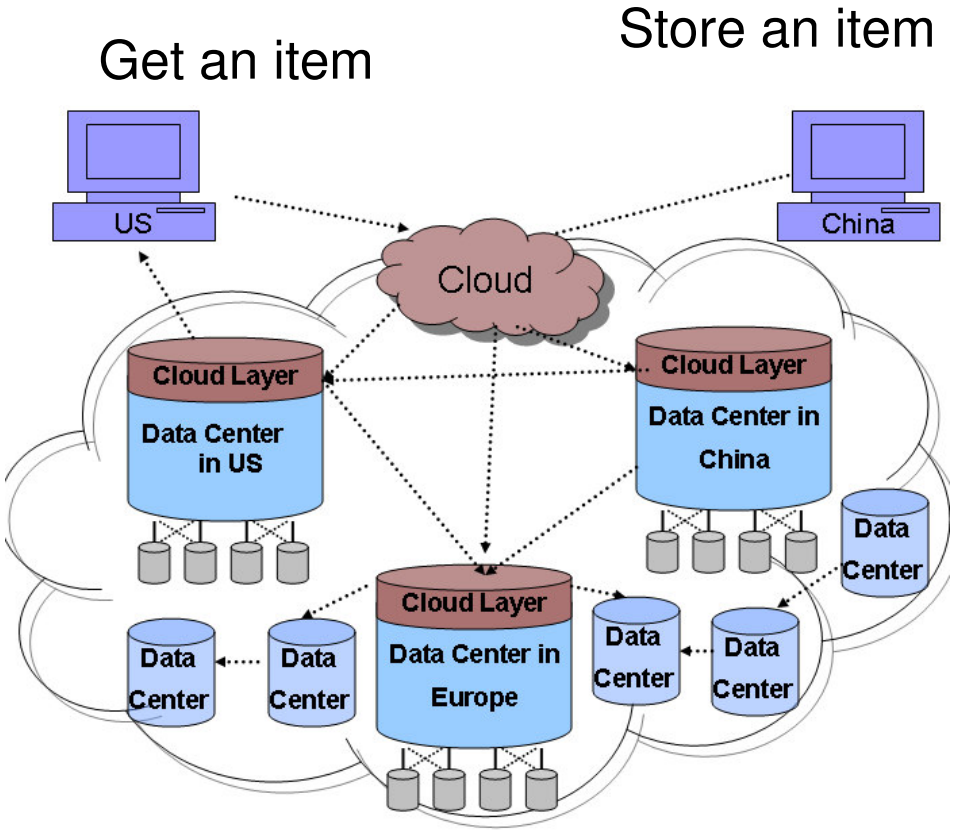
CCS – Cloud Content Store

© 2010 IBM Corporation

CCS is secure cloud storage for unstructured data and content depots that is continuously available everywhere

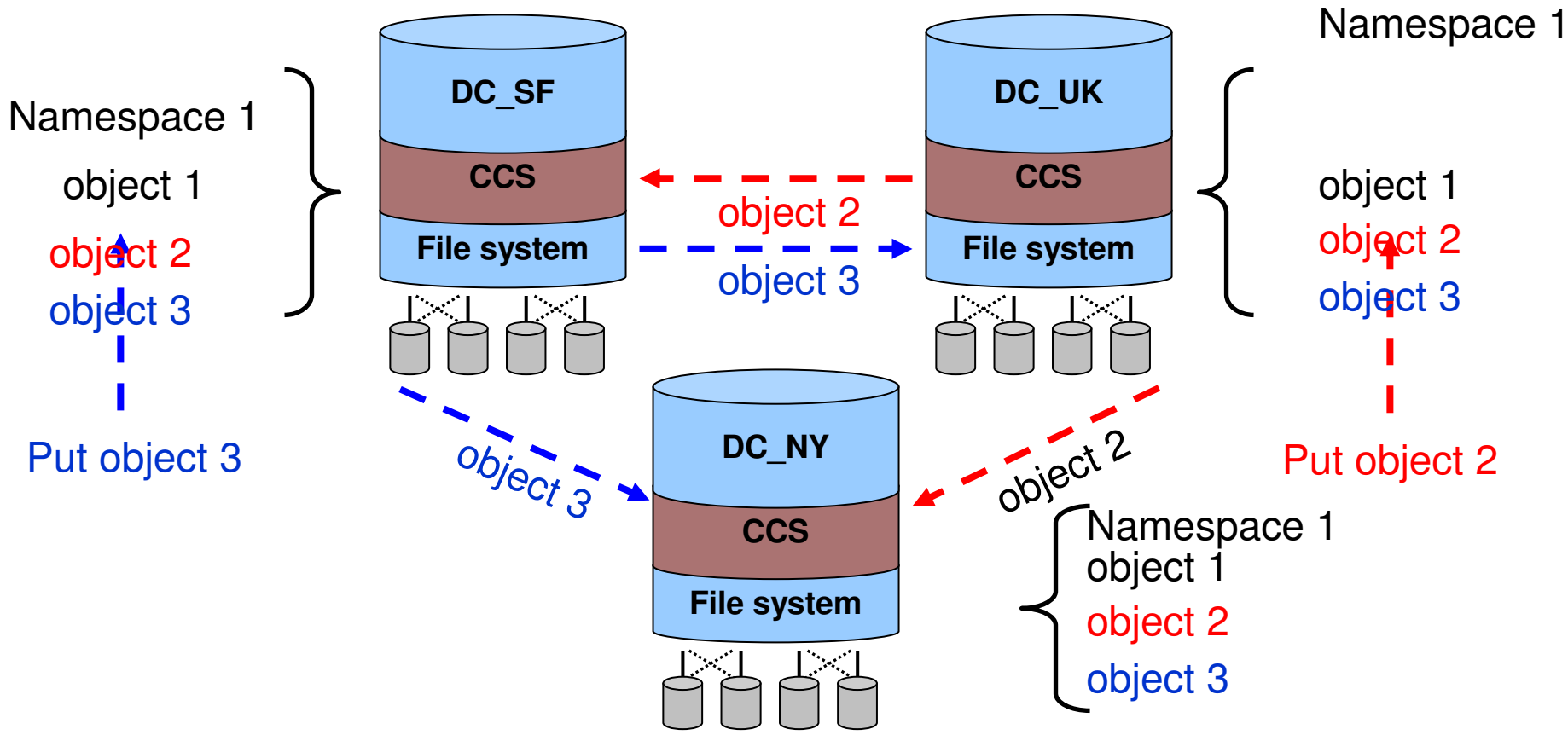
Scalable, available and cost efficient differentiated storage service that spans multiple data-centers, storing large, fixed content data

- **Replication**: Store several replicas of each item
  - Increased availability
  - Backup and fault tolerance
  - Disaster recovery
  - Consistency controlled by the application
- **Location awareness**: Access the closest data center
  - Increased availability
  - Minimal access time
  - Load balancing
- **Resource utilization**: Efficiently manage a large set of heterogeneous resources
  - Scalability
  - Load balancing



© 2010 IBM Corporation

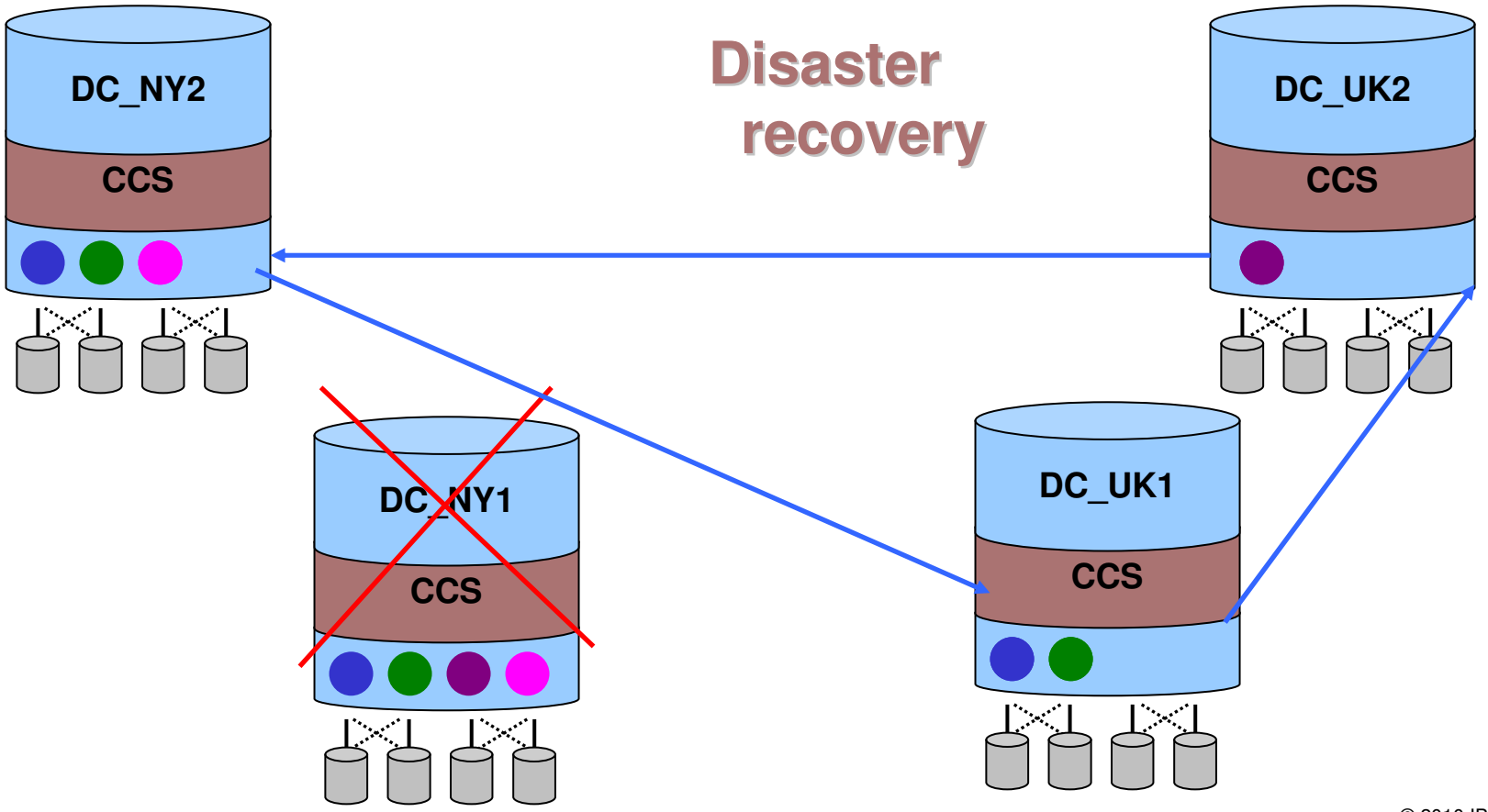
CCS: Multi-Master Optimistic Replication



© 2010 IBM Corporation

CCS provides continuous availability and parallel disaster recovery so data is available in spite of data center failures

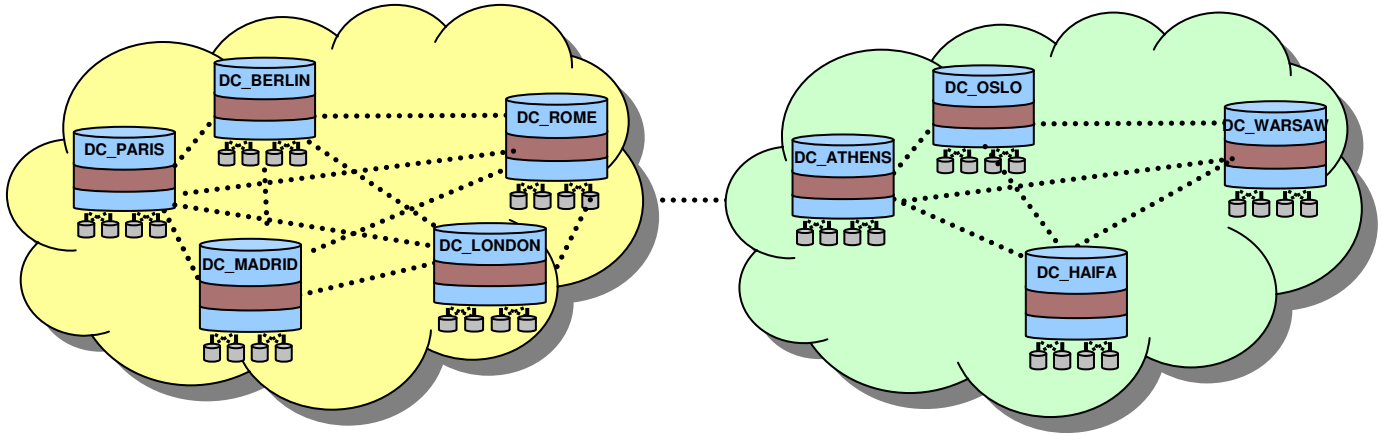
– The replication degree for namespaces on a failed DC is restored in parallel (each replica can be sourced from a different DC and sent to a different destination DC).
  - Balances load across DC's and connections to minimize disruption



© 2010 IBM Corporation

VISION's Goal

Infrastructure for delivery of data-intensive storage services, facilitating the convergence of ICT, media and telecommunications

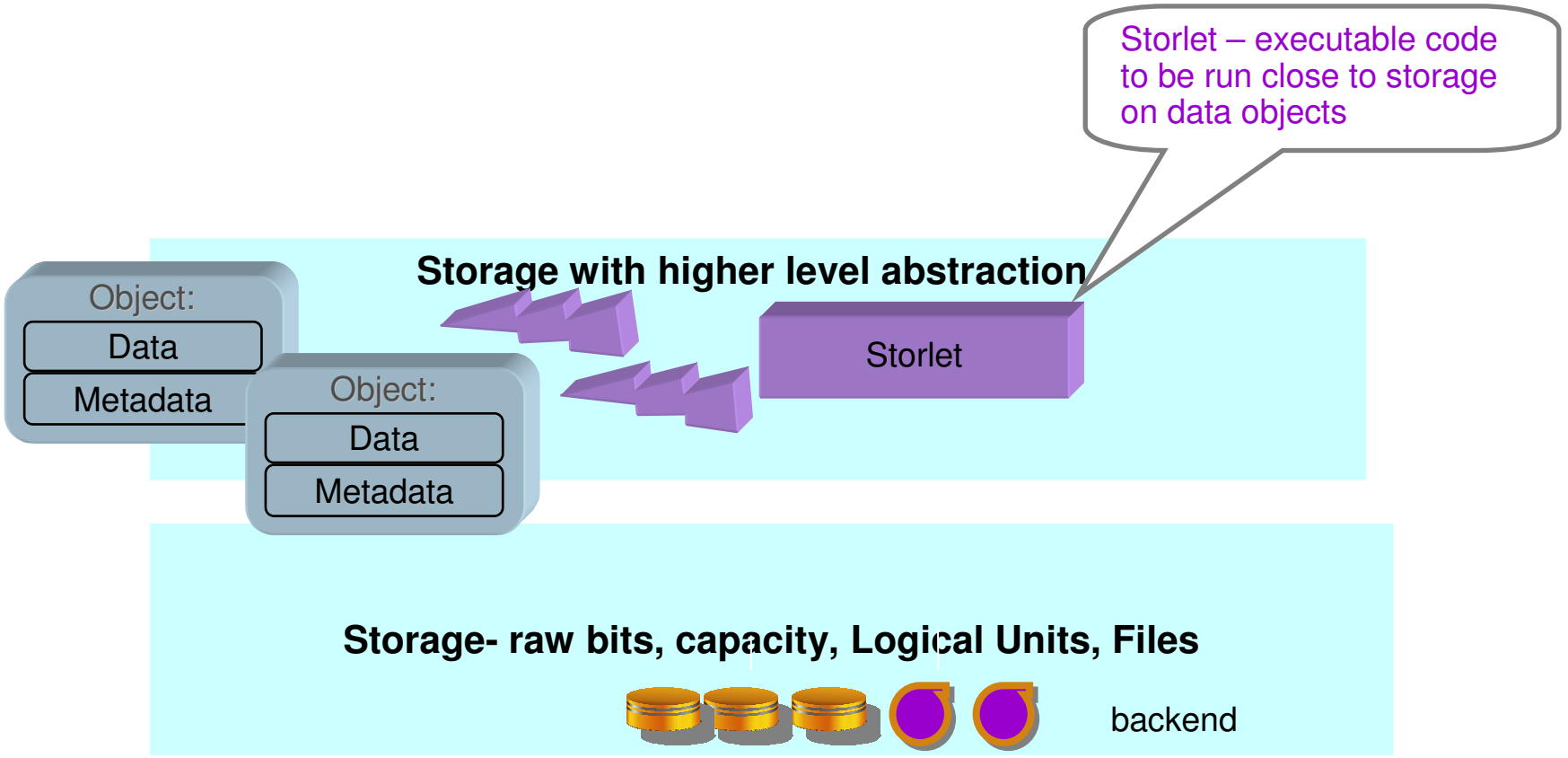


Technology Components

1. **Raise Abstraction Level of Storage**: objects with user-defined and system-defined metadata
2. **Data Mobility and Federation**: enable comprehensive data migration and interoperability across remote locations
3. **Computational Storage**: technology for specifying and executing computations close to storage
4. **Content-Centric Storage**: facilitate access to data by content and its relationships
5. **Advanced Capabilities for Cloud-based Storage**: support delivery of data-intensive services securely, at the desired QoS, at competitive costs

© 2010 IBM Corporation

Storage Cloud Data Model – Add Storelets



© 2010 IBM Corporation

VISION Should Support Computing Near Data

Example

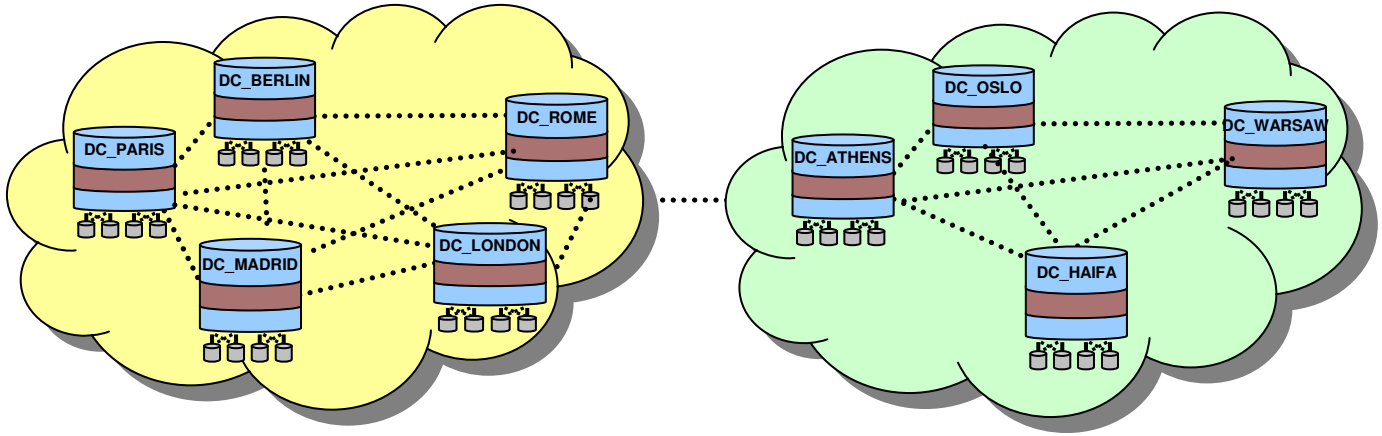
– A service provider wants to enable to adjust video to the device on which it is being viewed

– **Video retargeting** is the process of transforming an existing video to fit the dimensions of an arbitrary display. A compelling retargeting aims at preserving the viewers' experience by maintaining the information content of important regions in the frame, while keeping their aspect ratio.

© 2010 IBM Corporation

VISION's Goal

Infrastructure for delivery of data-intensive storage services, facilitating the convergence of ICT, media and telecommunications



Technology Components

1. **Raise Abstraction Level of Storage**: objects with user-defined and system-defined metadata
2. **Data Mobility and Federation**: enable comprehensive data migration and interoperability across remote locations
3. **Computational Storage**: technology for specifying and executing computations close to storage
4. **Content-Centric Storage**: facilitate access to data by content and its relationships
5. **Advanced Capabilities for Cloud-based Storage**: support delivery of data-intensive services securely, at the desired QoS, at competitive costs

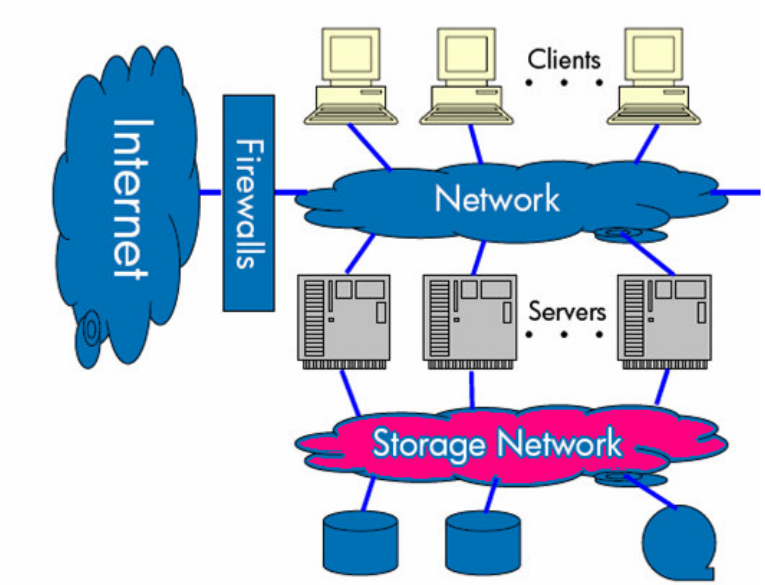
© 2010 IBM Corporation

The Problem – Providing Secure Access in the Cloud

- Not a new problem

- Critical in a multi-tenant environment
  - Different users
  - Different organizations

- Special needs in virtualized environments
  - Application/data servers access their storage via networks
  - Virtualized servers
    - Sharing physical I/O adapters
    - Moving from one adapter to another
  - Server repurposing



© 2010 IBM Corporation

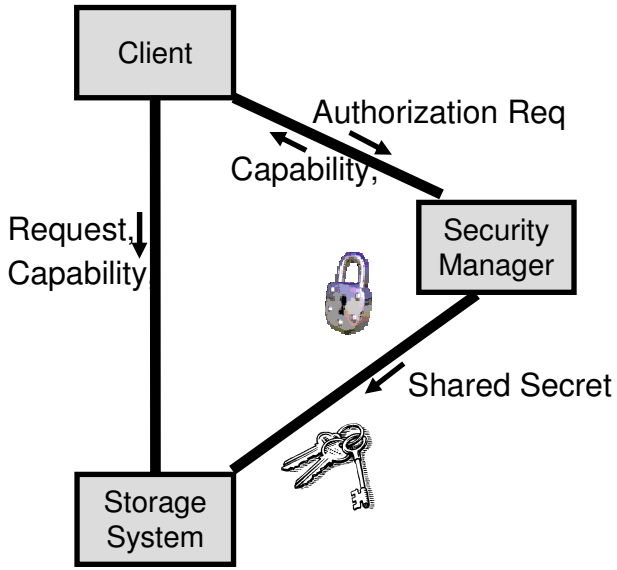
VISION's Approach to Cloud Storage: Dynamic Secure Access

- Every access to a LUN must provide a credential, obtained from a security/policy manager
- The storage system grants/denies access based on the credential
- Credentials are cryptographically secured against forging/replay

➢ Purely logical, not physical

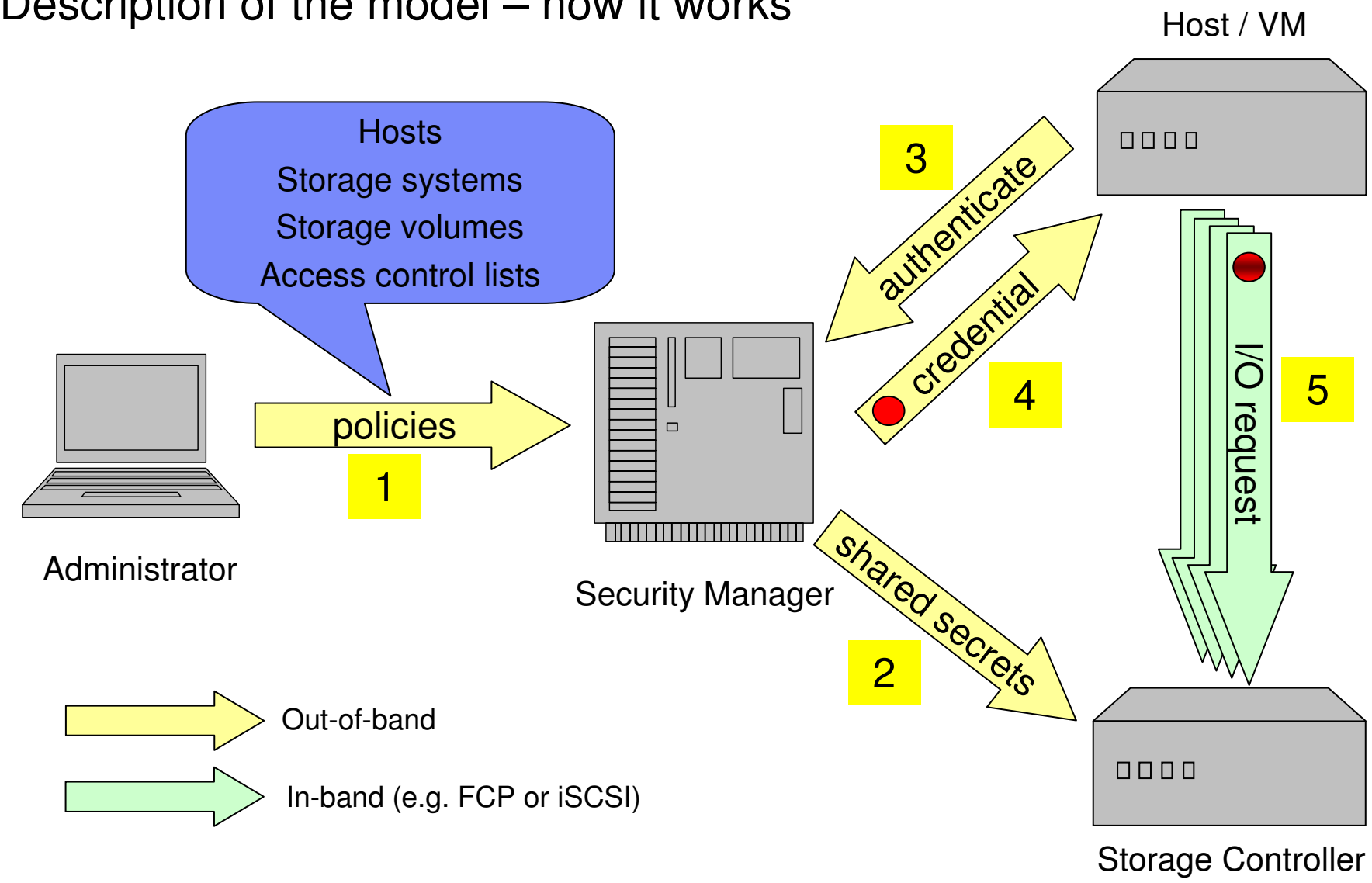
➢ Provides secure segregation between independent VMs

➢ End-to-end, dynamic, integrated security involving servers and storage



© 2010 IBM Corporation

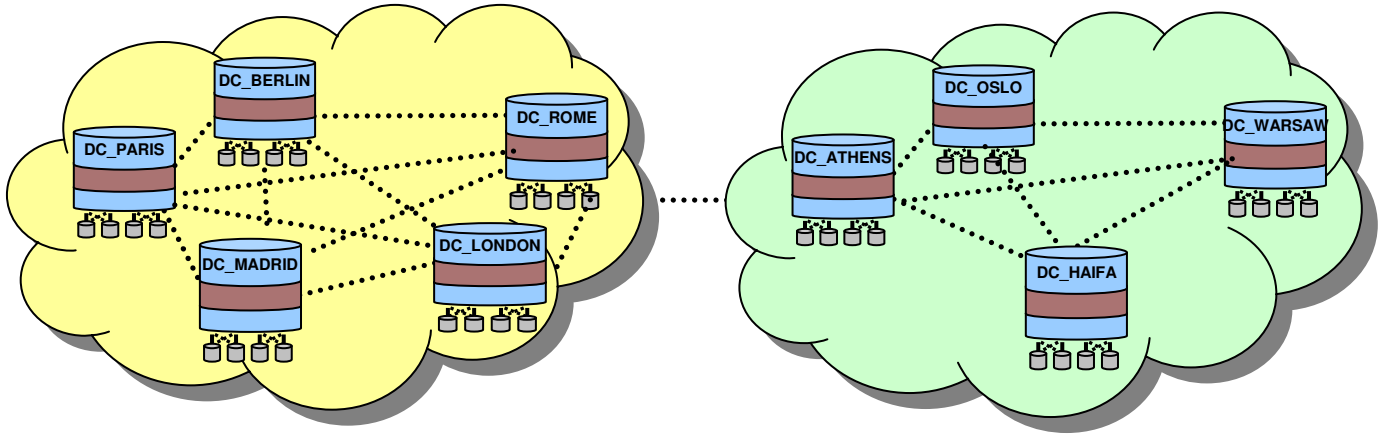
Description of the model – how it works



© 2010 IBM Corporation

VISION's Goal

Infrastructure for delivery of data-intensive storage services, facilitating the convergence of ICT, media and telecommunications



Technology Components

1. **Raise Abstraction Level of Storage**: objects with user-defined and system-defined metadata
2. **Data Mobility and Federation**: enable comprehensive data migration and interoperability across remote locations
3. **Computational Storage**: technology for specifying and executing computations close to storage
4. **Content-Centric Storage**: facilitate access to data by content and its relationships
5. **Advanced Capabilities for Cloud-based Storage**: support delivery of data-intensive services securely, at the desired QoS, at competitive costs

For more info… ask

© 2010 IBM Corporation

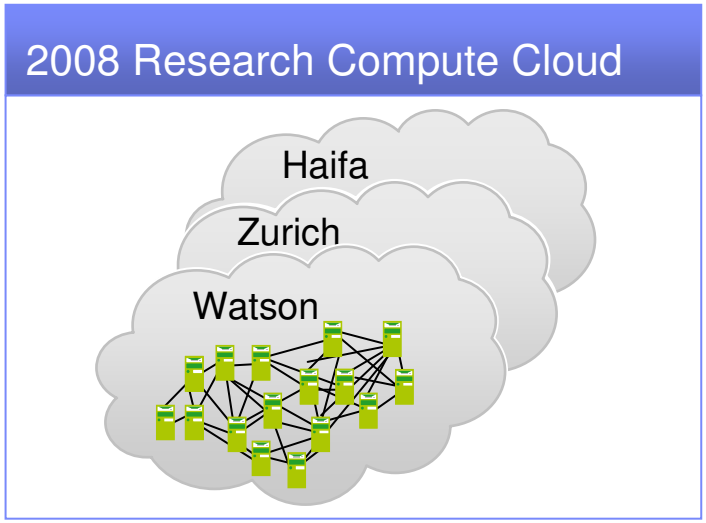
Summary

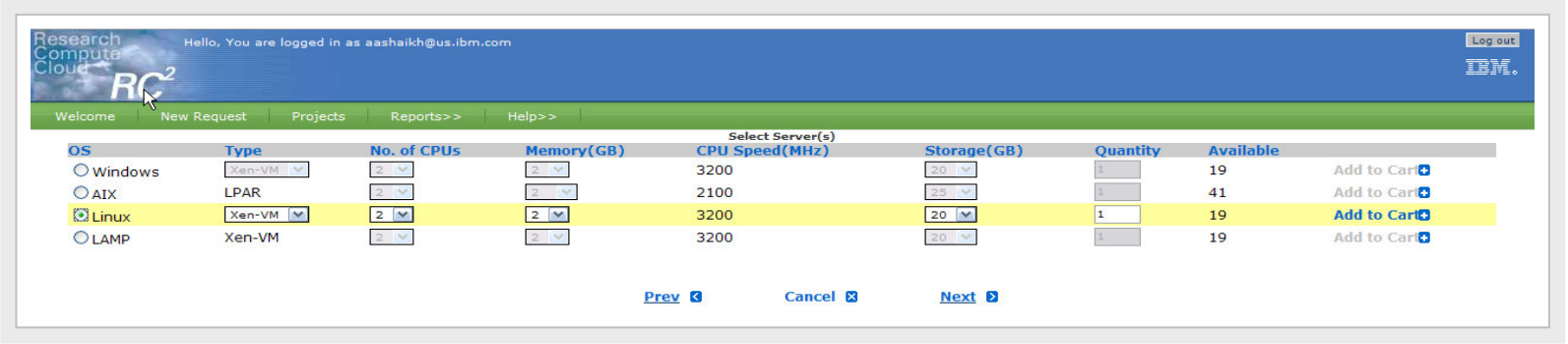
- Storage Cloud is an emerging marketplace opportunity
  - Technology is still behind, research needed
- The use of virtualization enables development of advanced solution
  - Meeting scalability, availability, cost challenges
  - Platform for storage services of many sorts
- VISION focus on use cases addressing telco/media/IT convergence
  - Synergy with Reservoir

© 2010 IBM Corporation

IBM Research Computing Cloud (RC2)
A living lab to advance Research projects

- Provides self service "on demand" delivery solution for research computing resources to cloud projects





© 2010 IBM Corporation

For More Information

- Dr. Yaron Wolfsthal, Sr Manager, System Technologies
- IBM Haifa Research Laboratory
- wolfstal at il dot ibm dot com

Credits

- Company/institute names, logos and trademarks are the property of their respective owners

© 2010 IBM Corporation