



DATASOURCE



B.I. IN THE CLOUD

BUSINESS INTELLIGENCE DATA INTEGRATION & ANALYSIS

Who Are We?

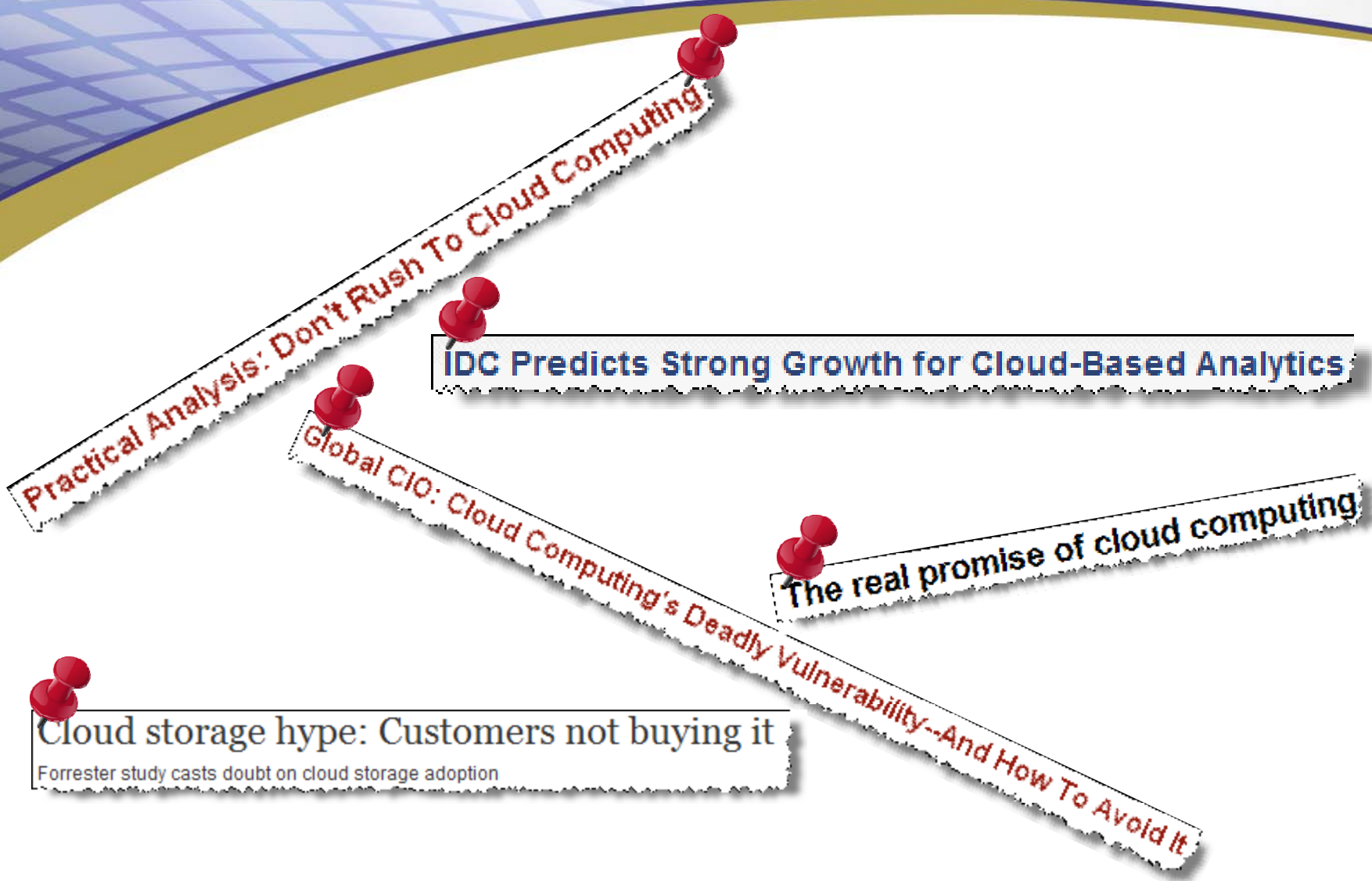


- Denver based Business Intelligence Consulting Company
- Specialize in delivering Lean, innovative, end-to-end business intelligence (BI) and data integration solutions that are successful, scalable and maintainable
- Partner with our customers by providing strategic guidance throughout their BI maturity lifecycle
- TDWI faculty & speakers, and judge for TDWI's Annual Best Practices Awards

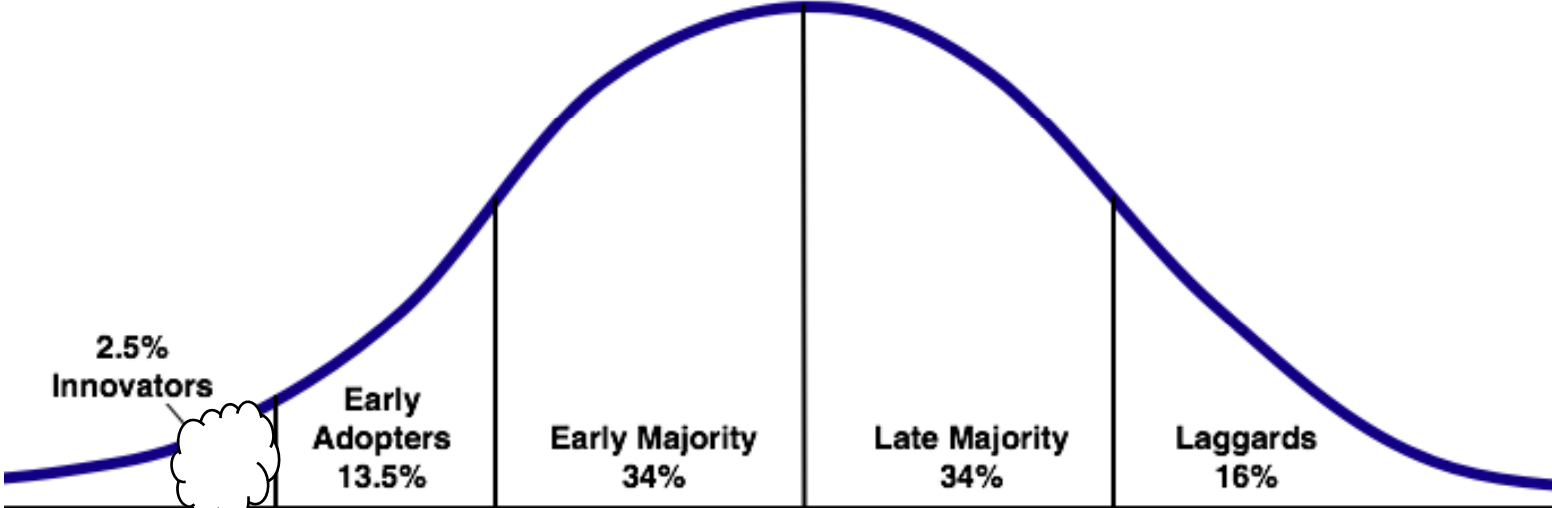
Agenda

- ☁ Cloud fundamentals & concepts
- ☁ Demo of BI in the EC2 Cloud
- ☁ Working in the Cloud
- ☁ Cloud challenges
- ☁ BI Vendors in the Cloud
- ☁ Case Study & Demo
- ☁ Summary

Navigating the Clouds



Consumer Adoption



Source: Everett Rogers, Diffusion of Innovations model



WHAT IS THE CLOUD?

Cloud Characteristics

**DYNAMICALLY
PROVISIONED**

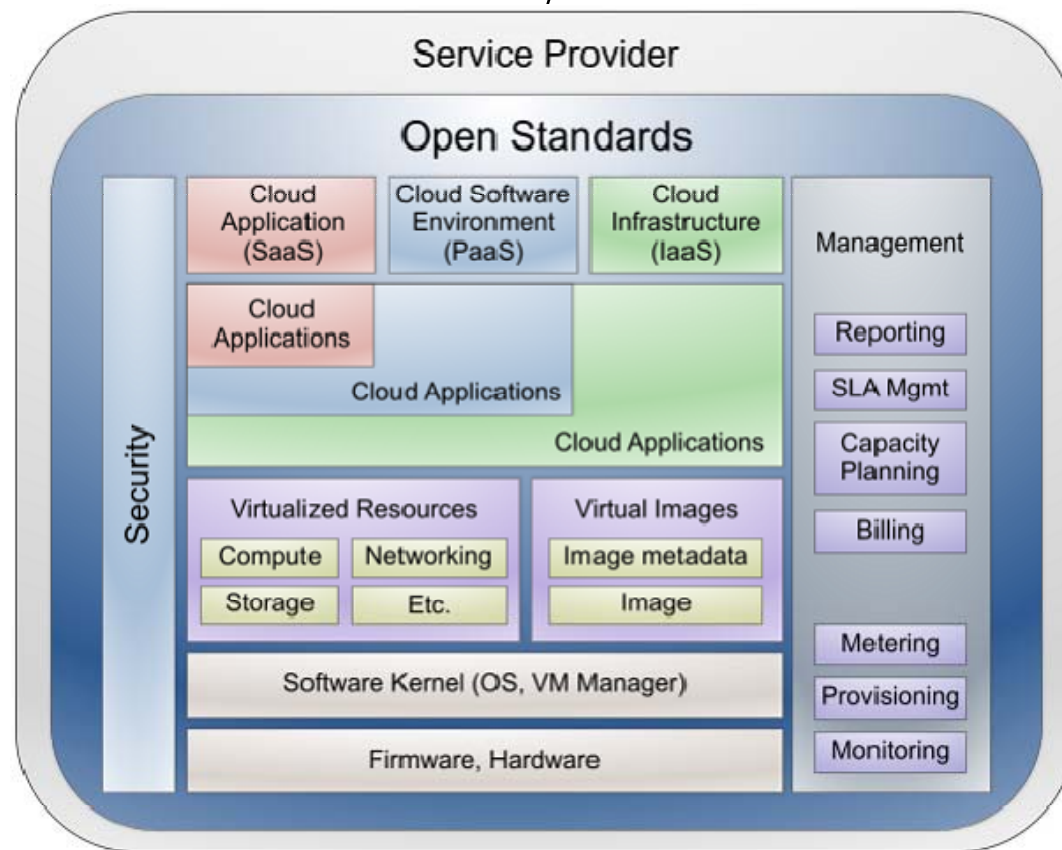
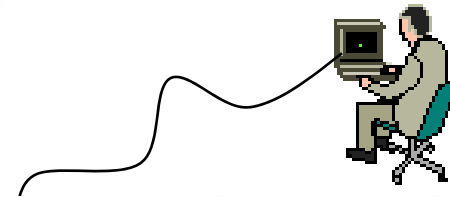
**UTILIZATION
BASED**

**SERVICE
ORIENTED**

**MULTI-
TENANT**

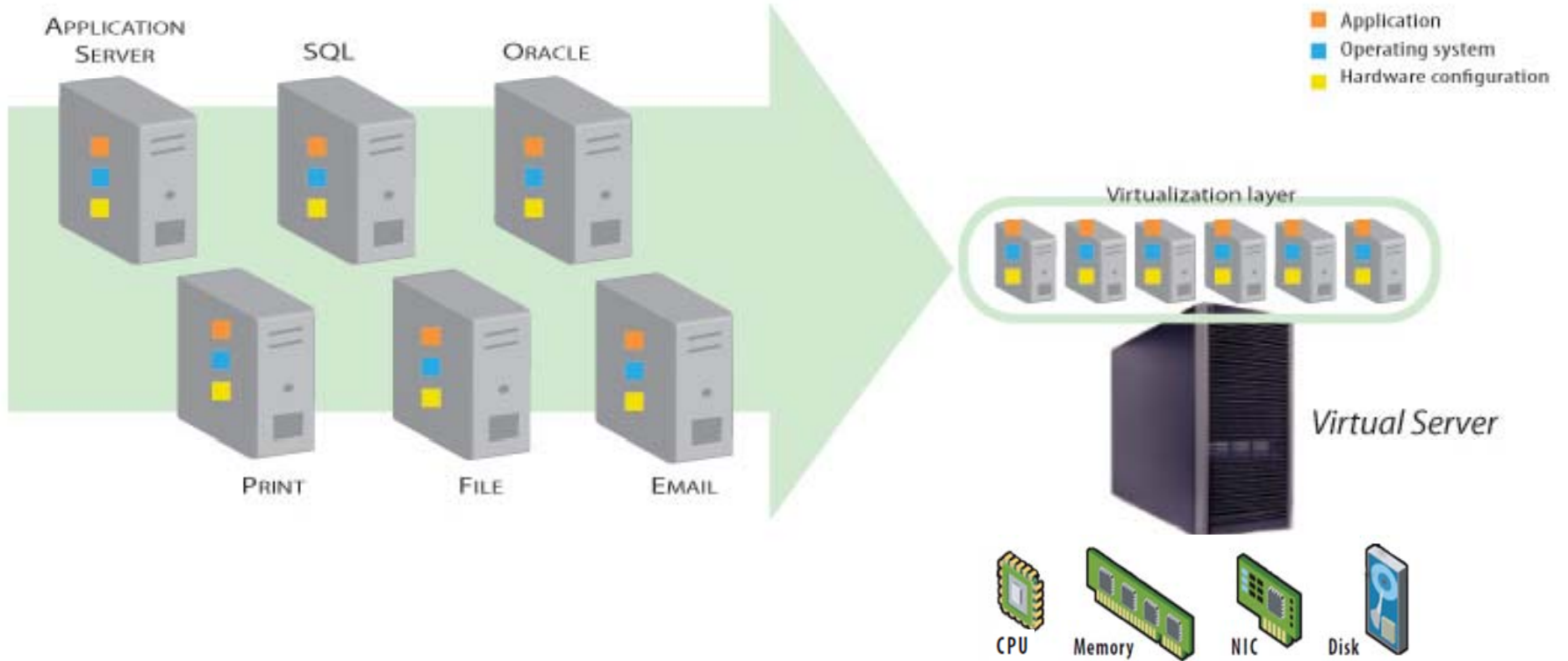
VIRTUALIZED

Example Cloud Architecture



http://cloud-computing-use-cases.googlegroups.com/web/Whitepaper_V2_Draft_3.pdf?hl=en&gsc=OV2YfQsAAAAAn_GLps52zkwO4Ra5Sw8jE

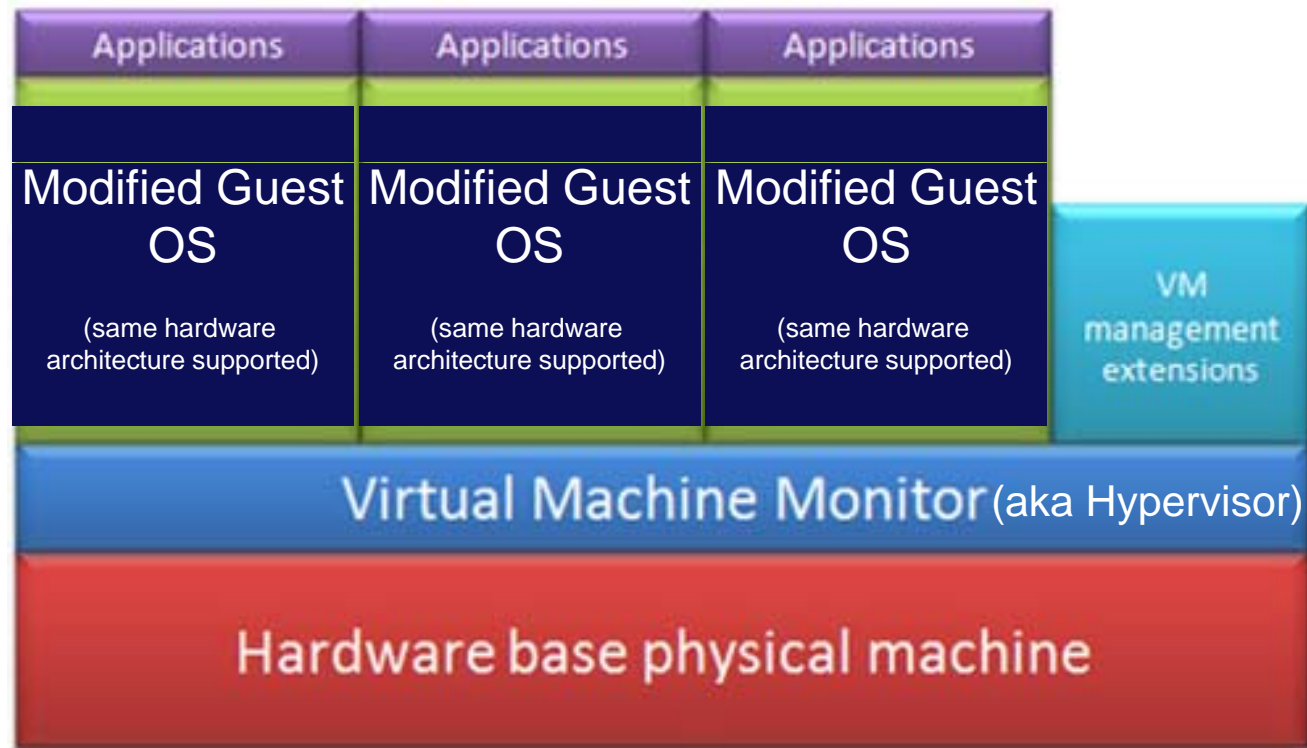
Server Virtualization



* <http://www.emdstorage.com/solutions/virtualization.asp>

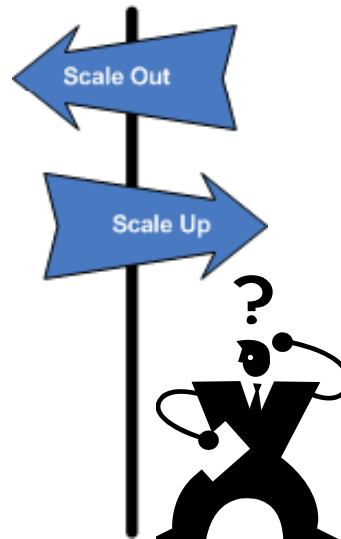
Copyright Datasource Consulting, LLC 2009

Paravirtualization



*<http://msdn.microsoft.com/en-us/library/dd430340.aspx>

Scale Up versus Scale Out

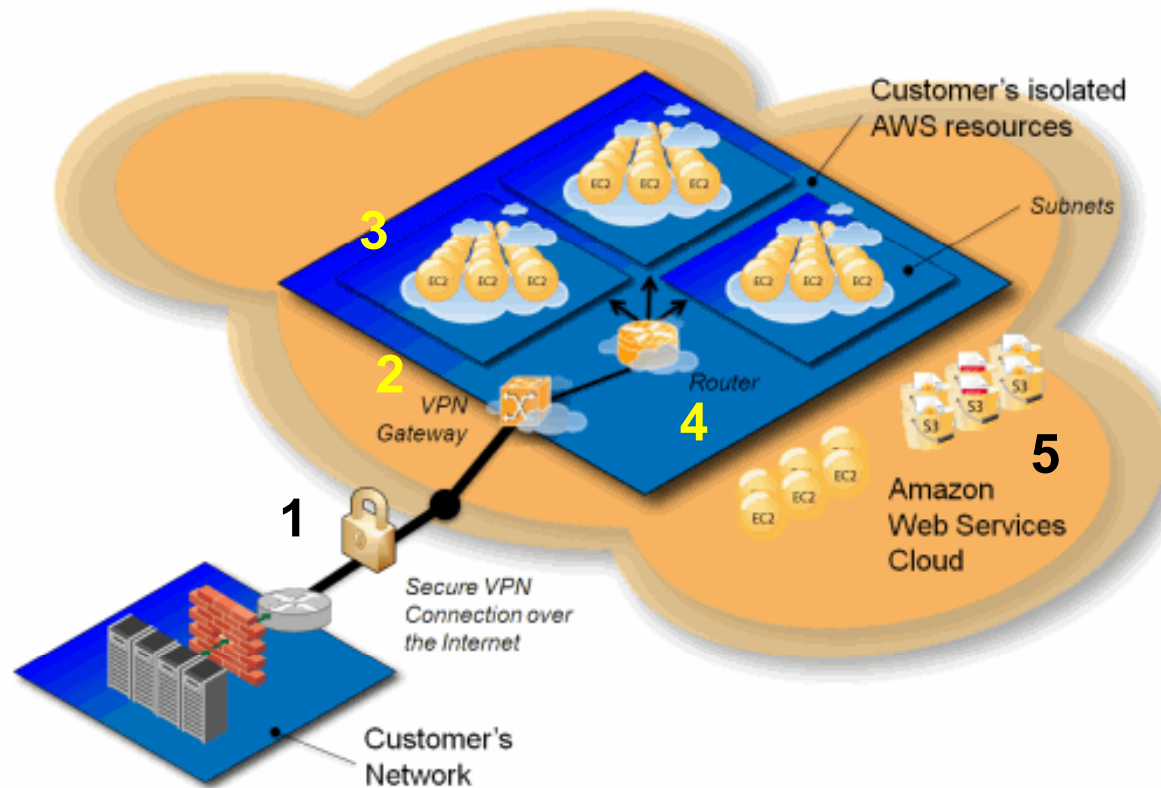


Types of Clouds

	Infrastructure Owned by	Infrastructure Location	Managed by	Access
Public	Third Party Provider	Off-Premise	Third Party Provider	Untrusted
Private	Organization	On-Premise	Organization	Trusted
	Third Party Provider	Off-Premise	Third Party Provider	
Managed	Third Party Provider	On-Premise	Third Party Provider	Trusted & Untrusted
Hybrid	Both Organization & Third Party Provider	Both On-Premise & Off-Premise	Both Organization & Third Party Provider	Trusted & Untrusted

* <http://www.rationalsurvivability.com/blog/?p=733>

Virtual Private Cloud

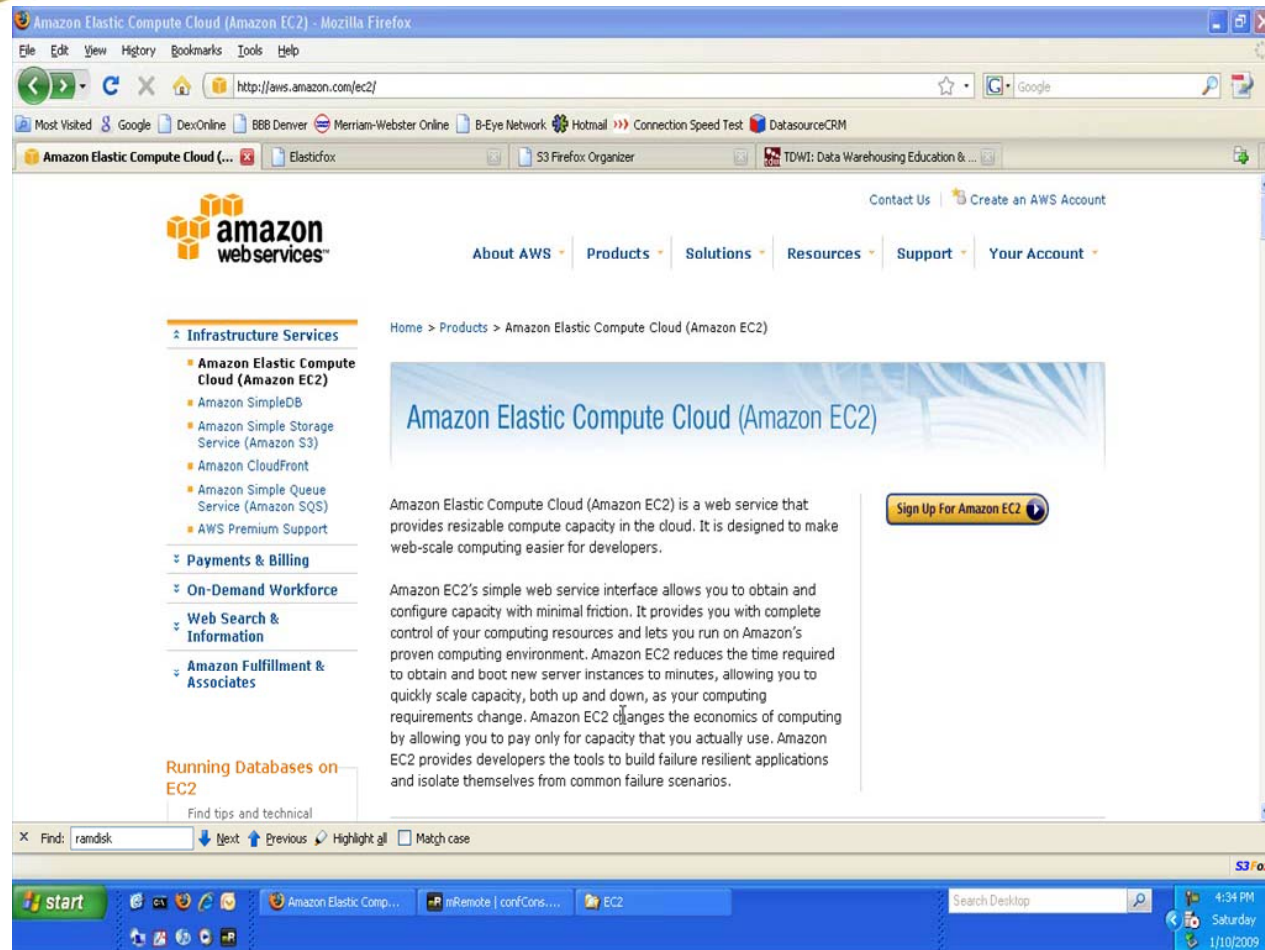


* http://news.cnet.com/8301-19413_3-10318114-240.html?tag=mncol;posts

The background features a blue sky with light clouds and a field of green grass at the bottom. A purple grid pattern is visible in the upper left corner, separated from the rest of the image by a curved yellow and blue border.

CLOUD DEMO

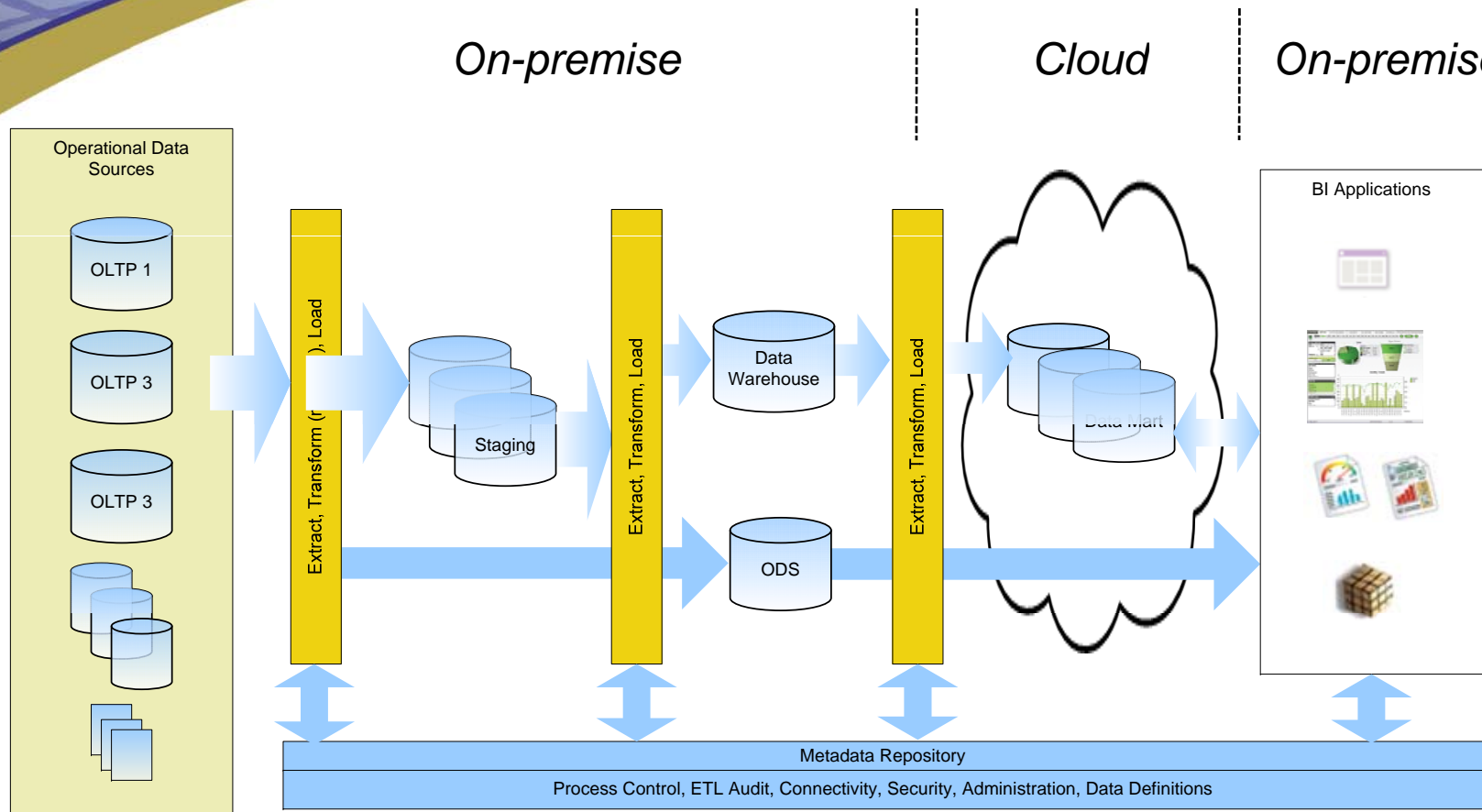
Demo of BI in the Cloud



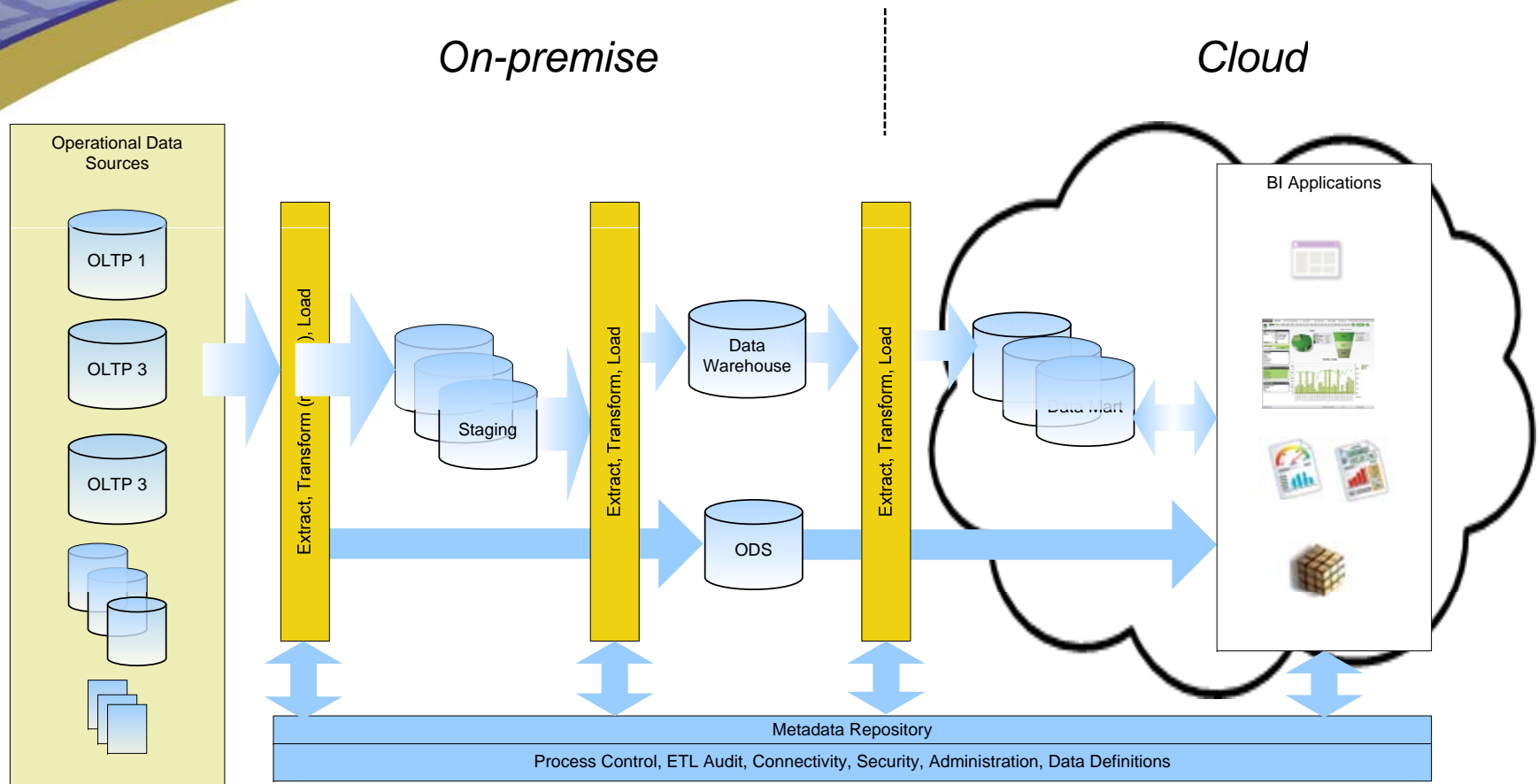
The background features a blue sky with light clouds and a field of green grass at the bottom. A purple grid pattern is visible in the upper left corner, partially obscured by a curved yellow and blue border.

WORKING IN THE CLOUD

Example Cloud BI Architecture



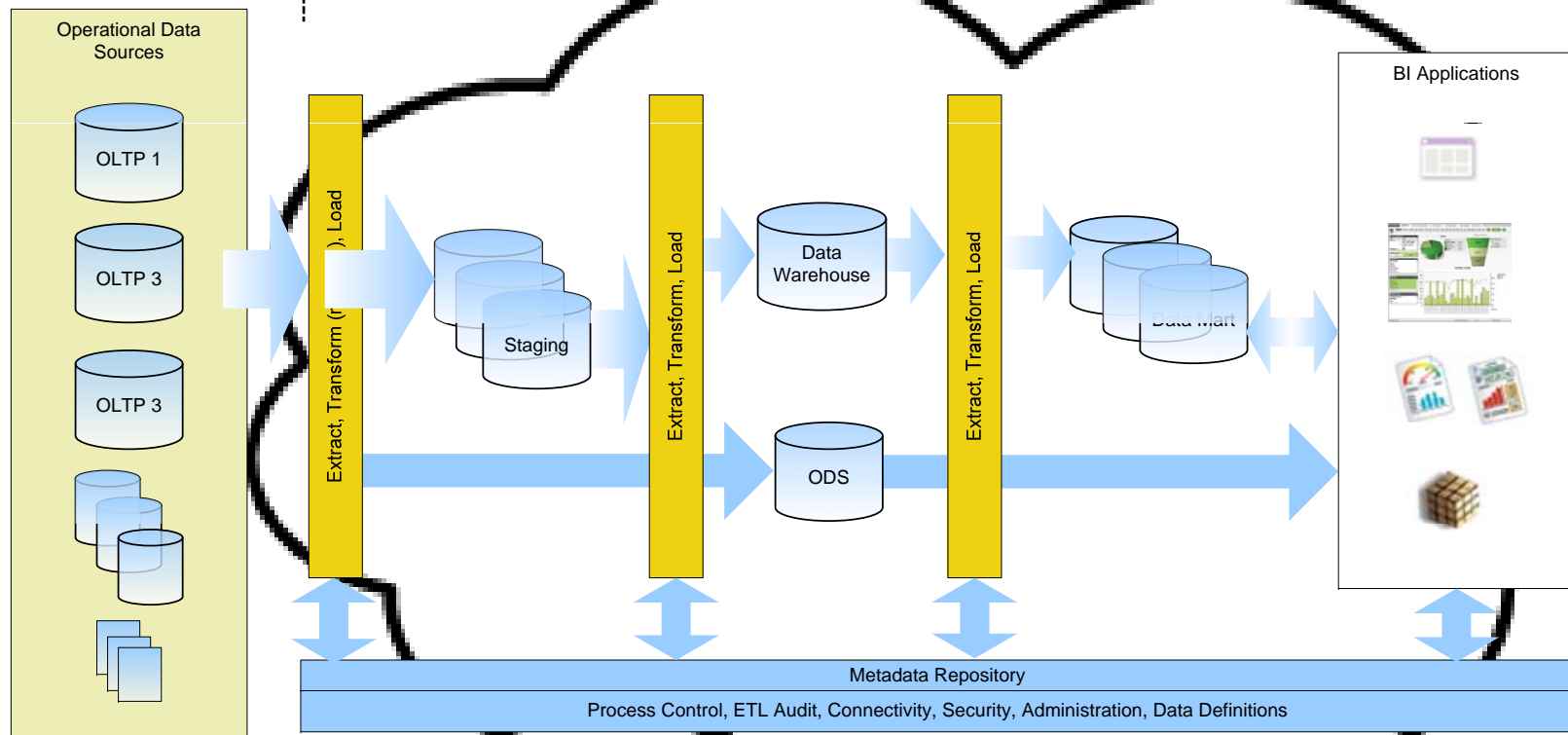
Example Cloud BI Architecture (cont)



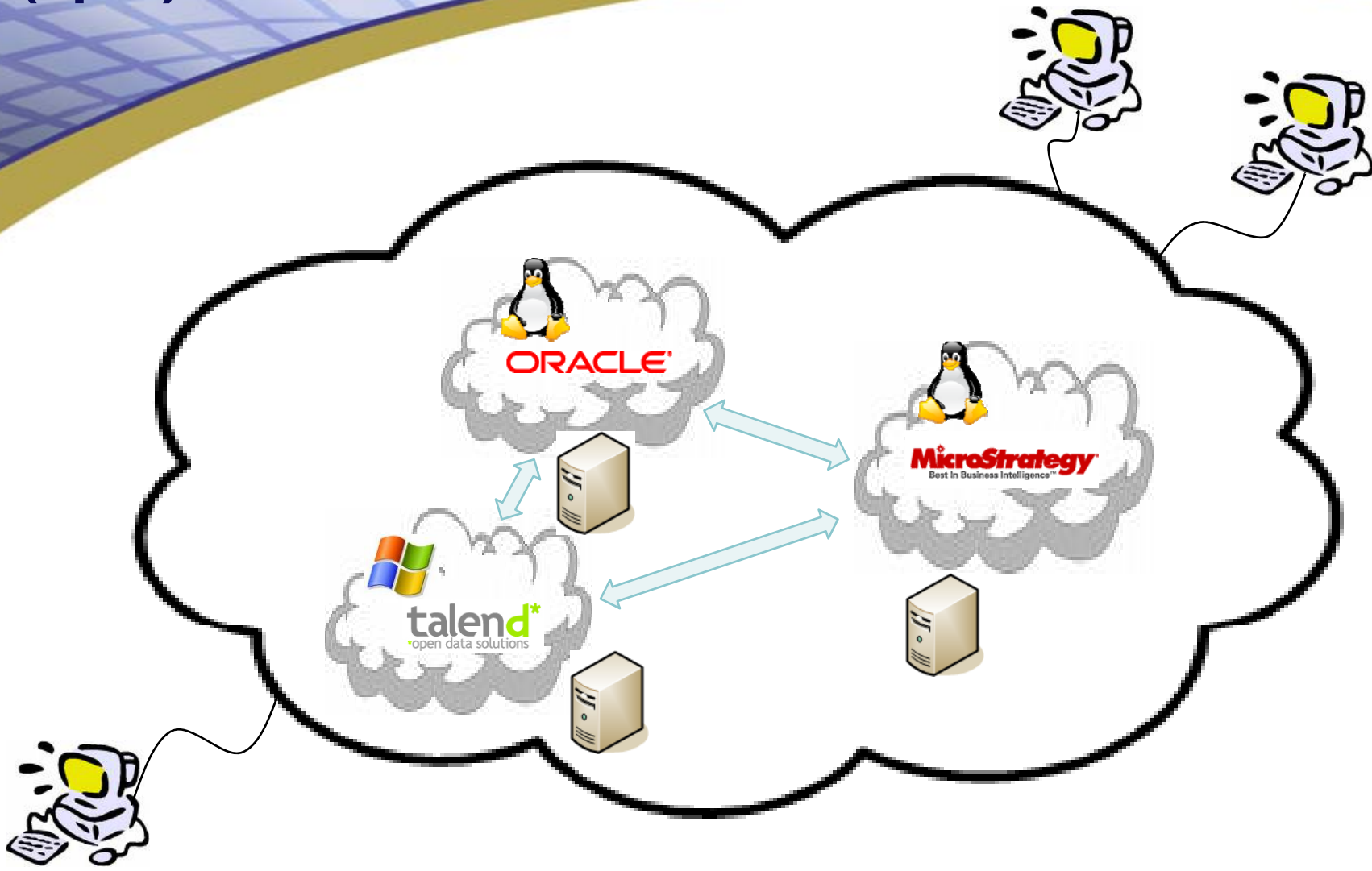
Example Cloud BI Architecture (cont)

On-premise

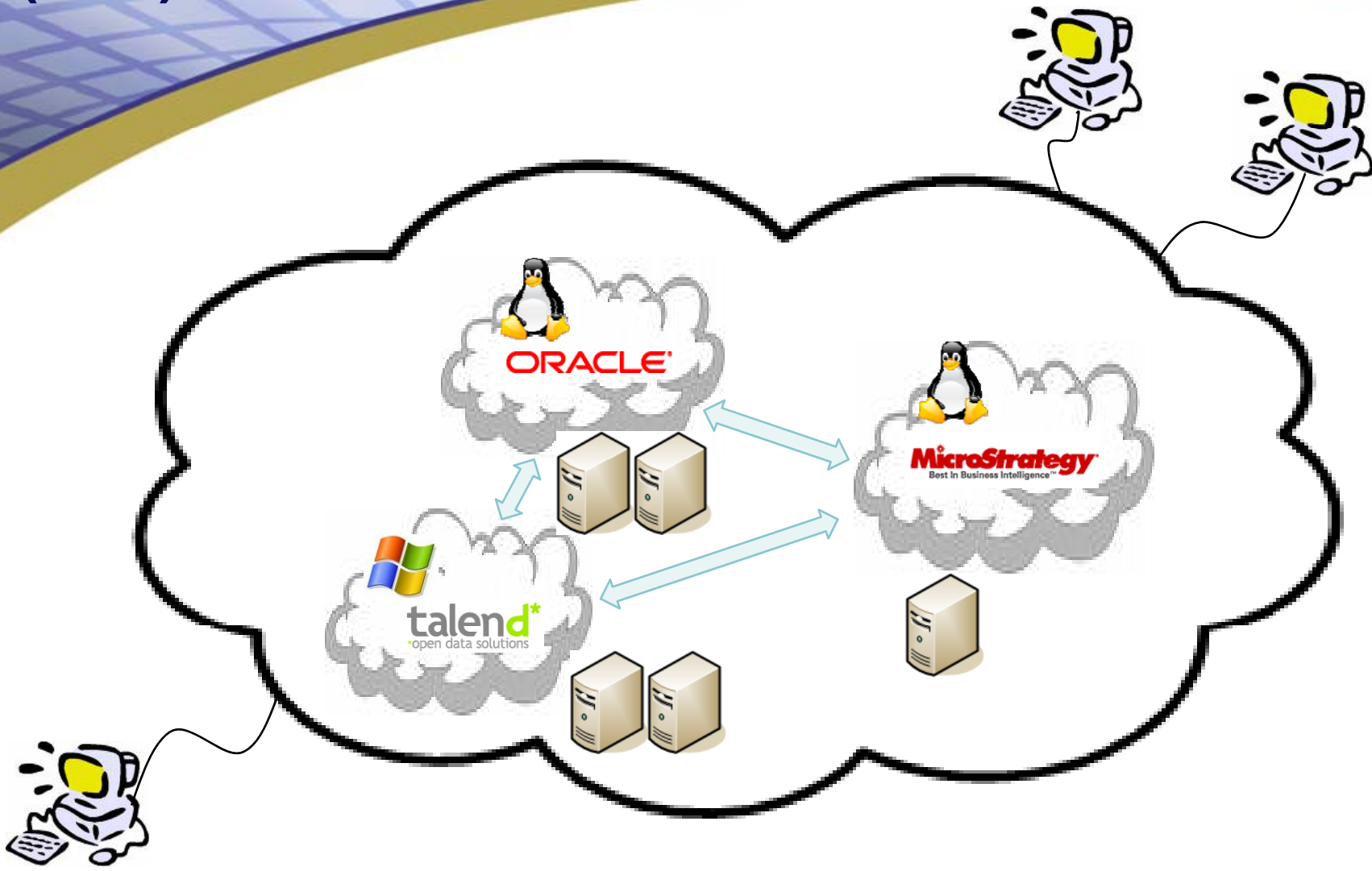
Cloud



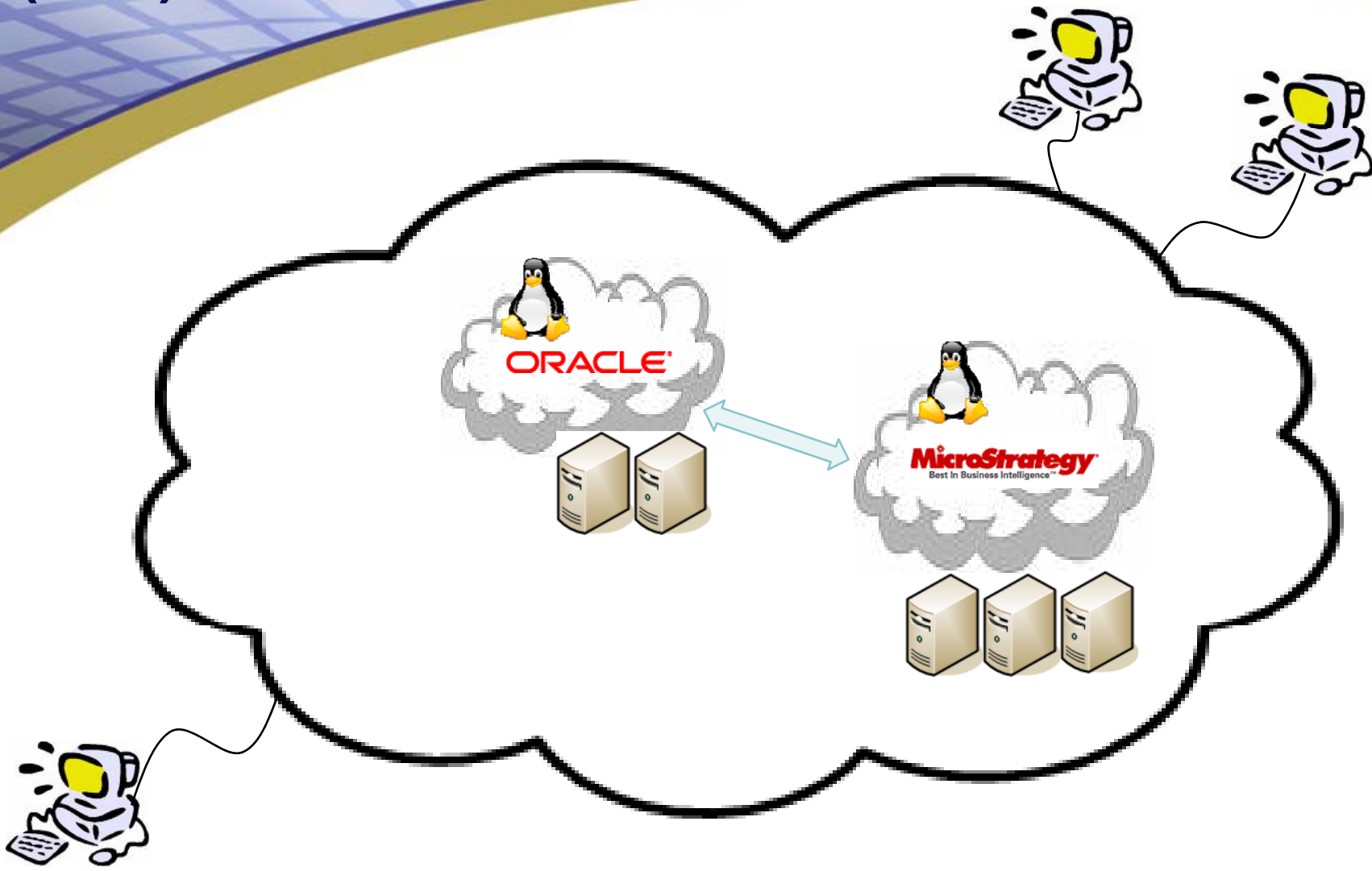
Example BI Server Architecture (9pm)



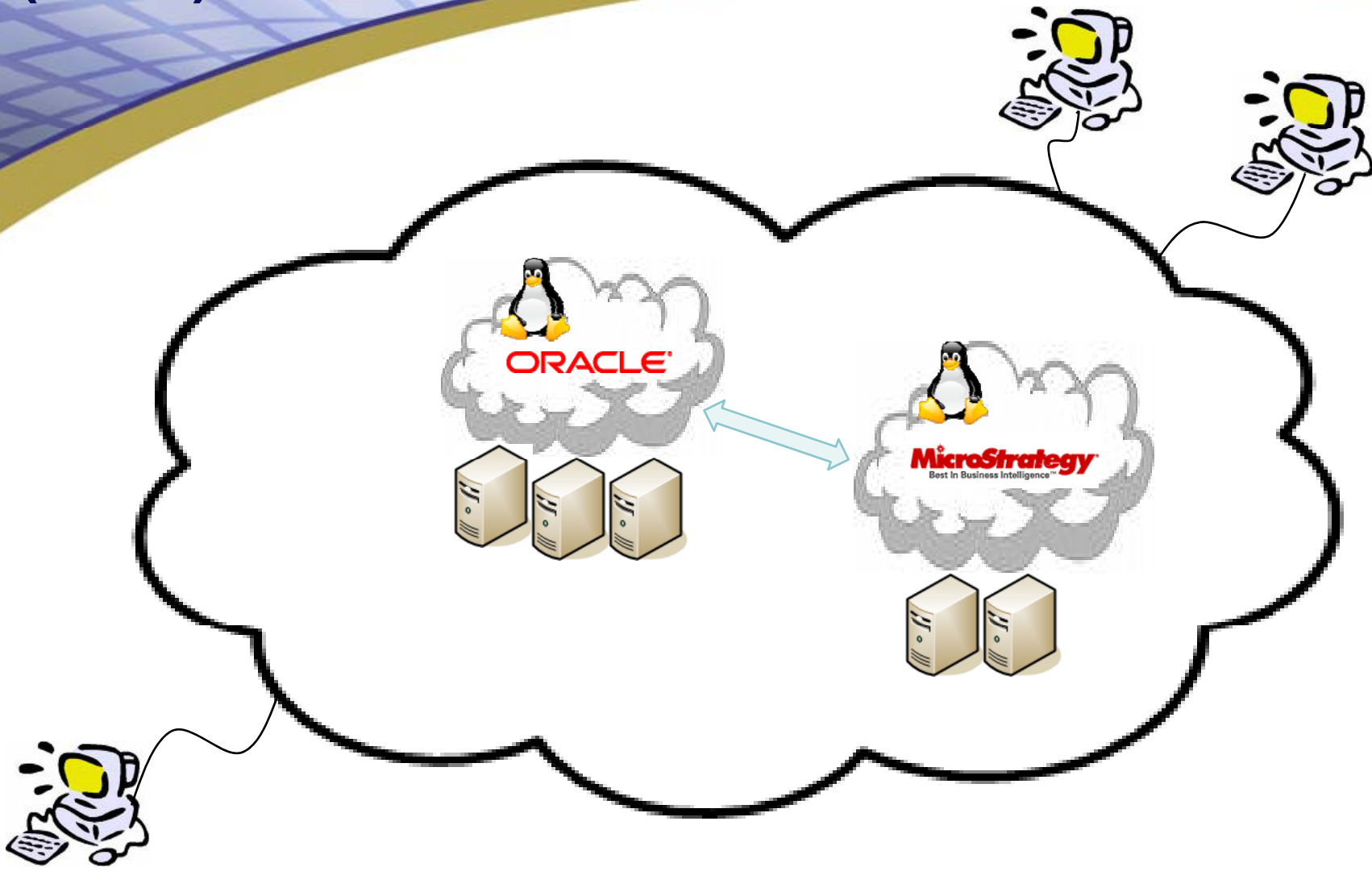
Example BI Server Architecture (2am)



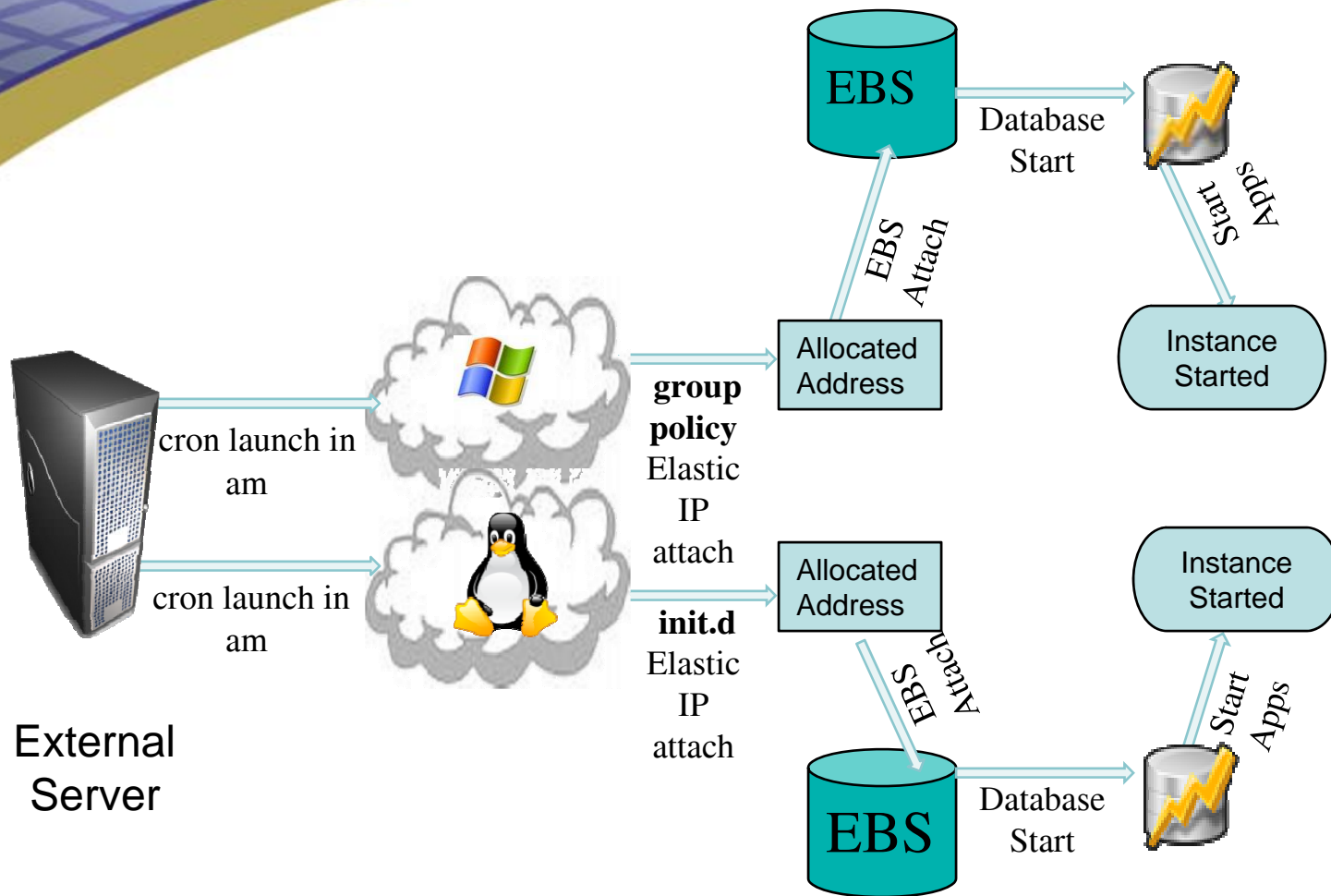
Example BI Server Architecture (4am)



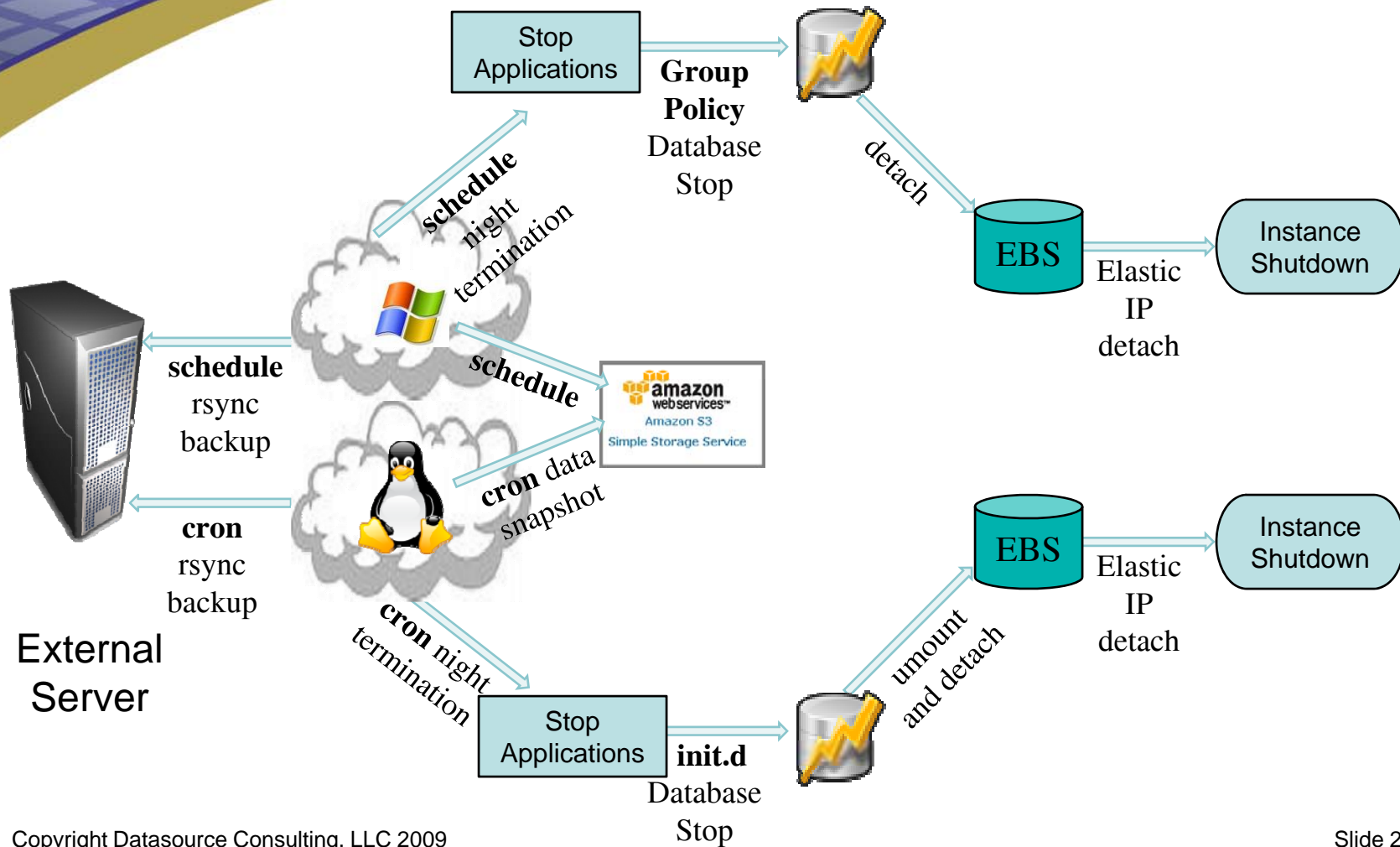
Example BI Server Architecture (10am)



EC2 Automated Server Start-up



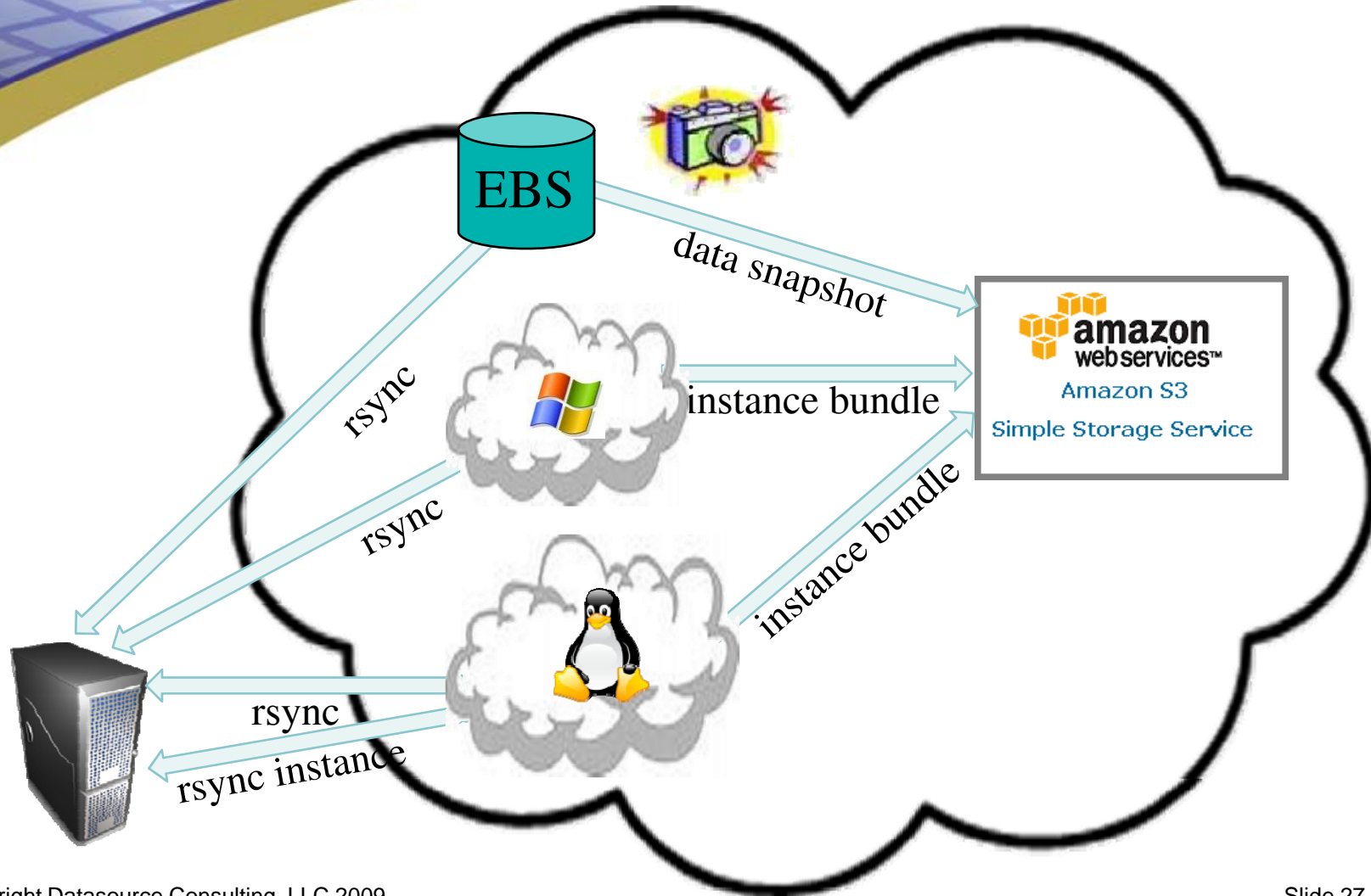
EC2 Automated Server Shutdown



Backup Requirements

- ☁ Hardware redundancy – no single point of failure
- ☁ Easily restored data by any user
- ☁ Small storage footprint and rapid backups
- ☁ Same strategy across instances - Linux and Windows Server instances
- ☁ Data availability without instance running
- ☁ Automated and unsupervised backups

Backup Strategy



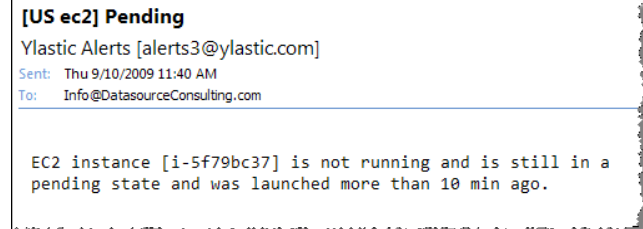
Rsync Backup Option

- Uses instance with cloud to internet speed and charges – solves S3 bottleneck
- Only copies and stores file deltas from previous backup – fast and small after first full backup
- Available on Linux and Windows
- Has live backups in familiar directory structure
- Uses ssh for host to host transfers
- Example command line:
 - `rsync -ave ssh --delete 196.255.255.196:/data /mnt/backups/backup.0/`

Cloud Services Management & Monitoring



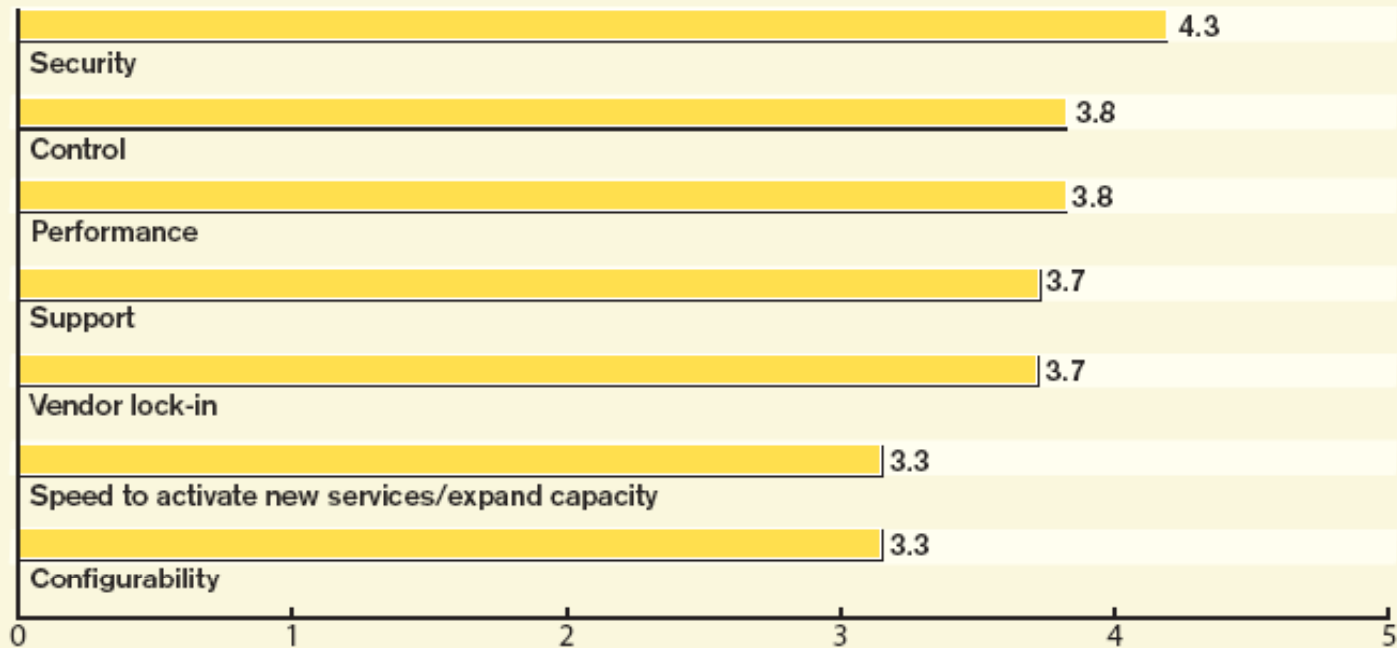
- Launching servers
- Server monitoring
- Performance monitoring
- Versioning
- Auto-scaling
- User management





CLOUD CHALLENGES

Top Cloud Concerns

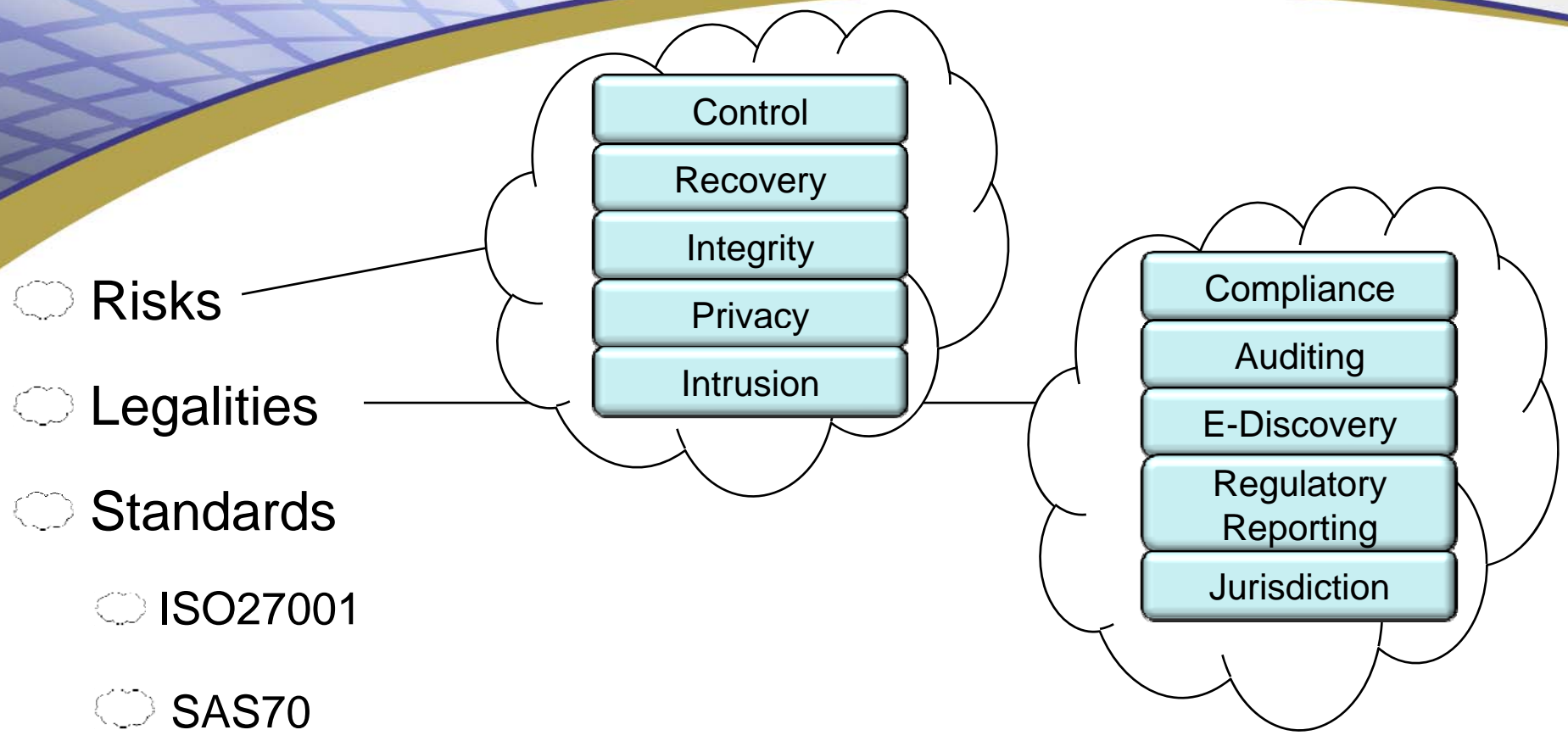


Note: Mean average ratings based on a five-point scale, Where 1 is "not at all concerned" and 5 is "very concerned"

Base: 172 respondents at organizations currently receiving or considering services from a cloud provider

Data: *InformationWeek* Analytics Cloud Computing Survey of 456 business technology professionals

Security Concerns



Security Mitigation



- Physical and Logical Security

- Hidden Data Locations with Secure and Limited Hardware Access
- 24x7 Security/Advanced Alarm System/Cameras

- Internal Network Security

- Firewall Defaults to No Port Authorization – Each Port Must Be Opened by the Customer and Customer Can Only Connect with Assigned IP Address
- Traffic monitoring to Prevent Distributed Denial-of-Service Attack (DDoS) and Port Scanning
- Man-in-the-Middle Attack Prevention by Secure Shell Channel (SSH)

Security Mitigation (Cont)

- External Network
 - Generate Own Login Key Pairs
 - Network Key Encryption
 - Security Groups which can Limit Open Ports
- Operating System
 - Customer only access to instance
 - Login options such as token authentication
 - Privilege escalation such as 'sudo'
- Client Desktop
 - Multiple access keys

Benchmarking Performance

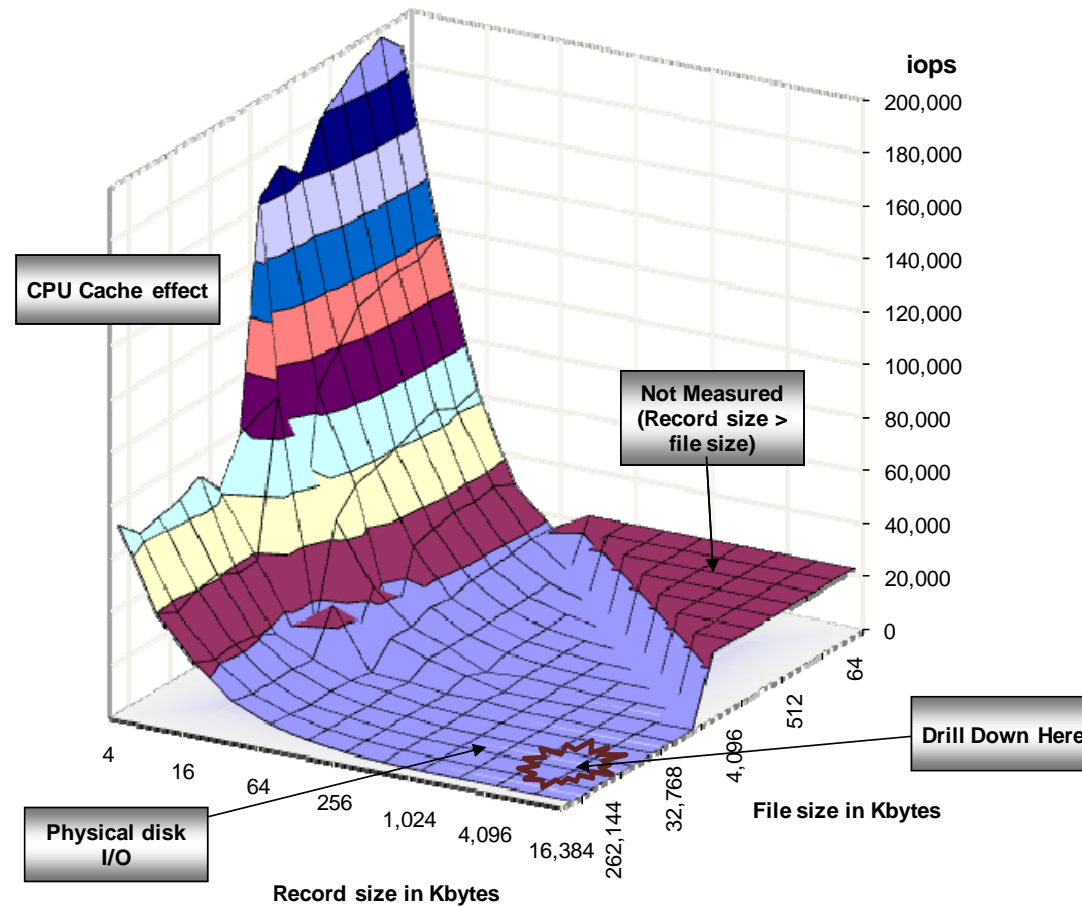
- A benchmark is a computer program that can be run on different architectures in order to objectively compare the performance
- The benchmark tool, iозone3, was chosen because it:
 - Runs on Linux and MS Windows Server 2003
 - Is an updated open source program
 - Provides multiple input parameters
 - Provides random read/write in input operations per second (iops) as an output parameter

Benchmarking Goals

- To provide some metrics as to expected rates of I/O performance using a cloud architecture
- To compare read and write rates of different Amazon EC2 architectures
- To compare read and write rates of Amazon EC2 architectures with traditional hardware configurations

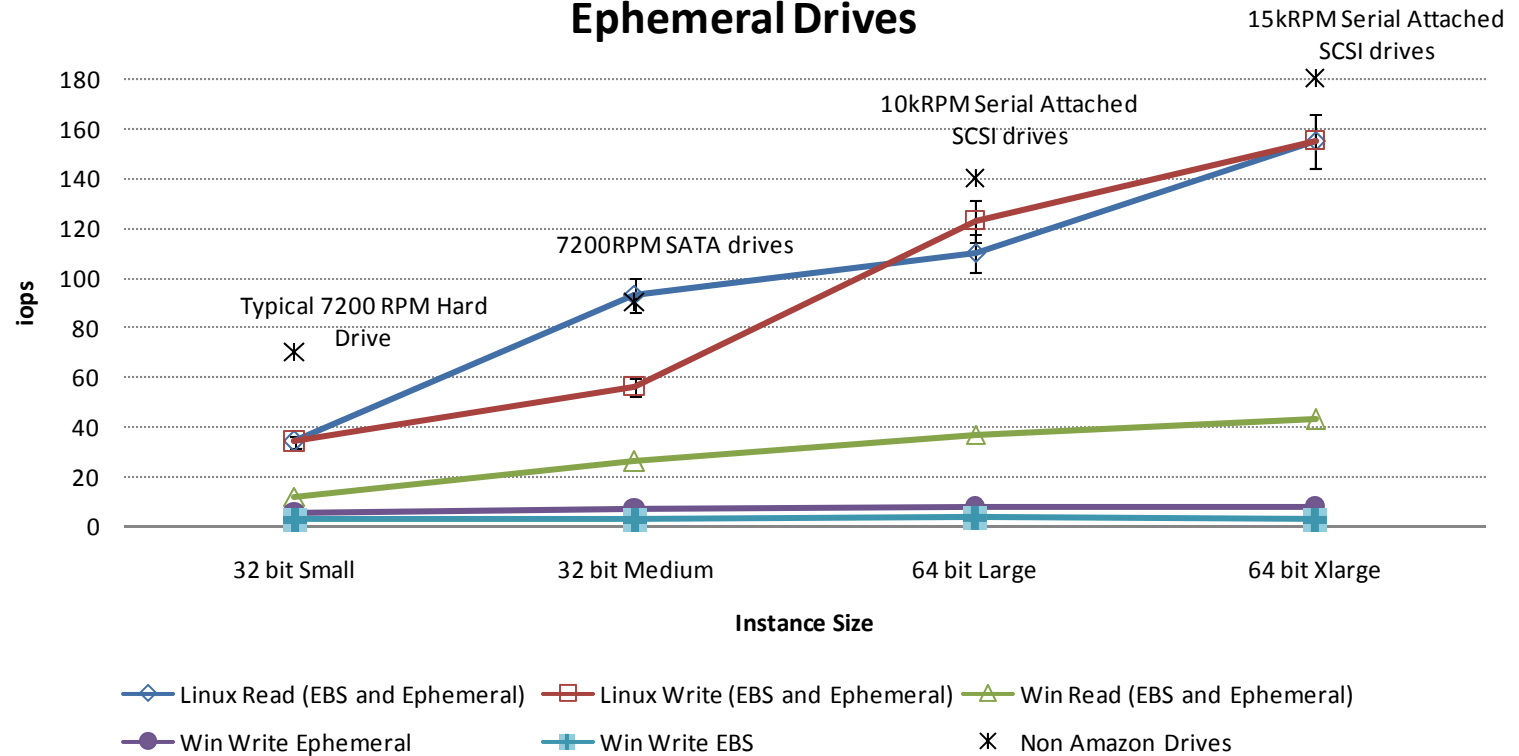
Benchmarking Performance

Random Read Ephemeral Performance Linux 32 bit Small Instance



Benchmark Results

Amazon EC2 iops Benchmark iozone3 Results for EBS and Ephemeral Drives



Benchmark Conclusions



- Linux does indeed scale up with I/O performance
- Linux is magnitudes faster than Windows Server
- EBS and Ephemeral (D: and /mnt) have comparable performance
- Amazon EC2 I/O speeds are comparable to traditional hardware
- Amazon EC2 has rapid i/o cache rates for up to 6 Meg record and file sizes

The background features a blue sky with light clouds and a field of green grass at the bottom. A purple grid pattern is visible in the upper left corner, separated from the rest of the image by a curved, gold-colored border.

CLOUD VENDORS

Public Cloud Vendor Options



☁ Amazon **Elastic** Compute Cloud: (EC2)

Pricing
by
Size

- ☁ Monthly by Storage, Data In, Data Out, and Requests (PUT/LIST, Other)
- ☁ STANDARD: small, large, extra large or HIGH-CPU: medium, extra large

☁ **Flexiscale**



Charged
for 5 Key
Components

- ☁ Customer base in the UK & Europe
- ☁ Cloud Server Instance Hour, MS Windows OS licensing, Storage, Data Transfer, Option Services (Firewalling)

Public Cloud Vendor Options (cont)

☁ **Rackspace Cloud** by Rackspace Hosting



- ☁ Managed vs. Private solutions
- ☁ 1.5 cents per hour of use; 15 cents per Gig of Storage
- ☁ Prides themselves on customer support

“Fanatical Support” →

☁ **GoGrid** a division of ServePath (dedicated hosting)

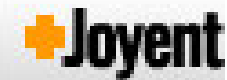
- ☁ Charge per GB Ram hour (as opposed to CPU hour), Outbound Data Transfer (Unlimited Inbound), Storage (after 10 GB)
- ☁ Pre Paid Plan Options


Pricing Options →



Public Cloud Vendor Options (cont)


Joyent



-  Claim to be faster than Amazon and other cloud providers through their "Accelerators"

Offerings



-  Uses Cloud Control™ via a web interface and APIs to manage full life cycle of complete application deployment architectures across multiple servers

-  Only offers fixed price by month or year

Comparison of Public Cloud Pricing



	Monthly Example		Amazon EC2		Rackspace		FlexiScale ³		Joyent ⁴		GoGrid ⁵	
	Usage		Unit \$	Total	Unit \$	Total	Unit \$	Total	Unit \$	Total	Unit \$	Total
Linux 1.7 G RAM 1 ECU	175	run hours	\$ 0.09	\$ 15	\$ 0.08	\$ 14	\$ 0.12	\$ 21	\$ 250	\$ 250	\$ 0.38	\$ 67
Linux 7.5 G RAM 4 ECUs	175	run hours	\$ 0.34	\$ 60	\$ 0.40	\$ 70	\$ 0.61	\$ 106	\$ 1,000	\$ 1,000	\$ 0.80	\$ 140
Windows 1.7 G RAM 1 ECU	175	run hours	\$ 0.12	\$ 21		\$ 100 ¹	\$ 0.17	\$ 29		N/A	\$ 0.20	\$ 35
Windows 1.7 G RAM 5 ECUs	175	run hours	\$ 0.29	\$ 51		\$ 529 ²	\$ 0.32	\$ 56		N/A	\$ 0.48	\$ 83
GB Data Transfer in	10	Gbytes	\$ 0.10	\$ 1	\$ 0.08	\$ 1	\$ 0.11	\$ 1			\$ -	\$ -
GB Data Transfer Out	1	Gbytes	\$ 0.17	\$ 0	\$ 0.22	\$ 0	\$ 0.16	\$ 0			\$ 0.29	\$ 0
Elastic IP Hours	1,054	off hours	\$ 0.01	\$ 11	\$ -	\$ -	\$ -	\$ -			\$ -	\$ -
Storage	500	Gbytes	\$ 0.15	\$ 75	\$ 0.50	\$ 225	\$ 0.33	\$ 166	\$ 0.15	\$ 0.08	\$ 0.15	\$ 74
Silver Support			\$100.00	\$ 100	\$100.00	\$ 100	\$ -	\$ -		\$ -	\$ -	\$ -
TOTAL				\$ 333		\$ 1,039		\$ 380		\$ 1,250		\$ 399
Two Rackspace Dedicated Servers						\$ 838						
Differentiators							United Kingdom					

* Prices as of 1/15/2010 (see notes for fine print)

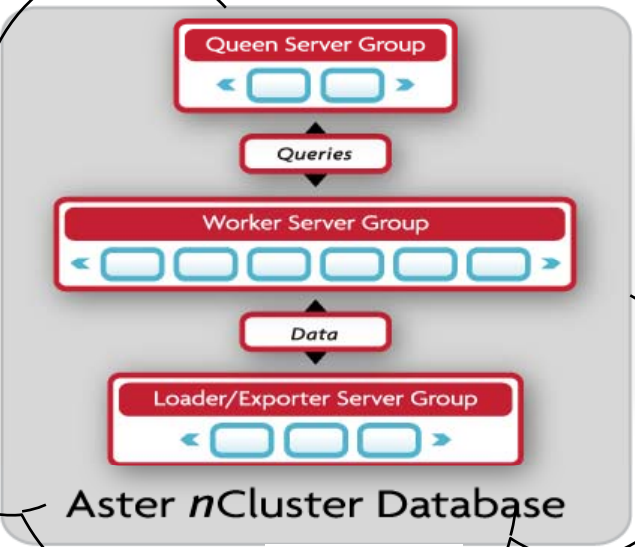
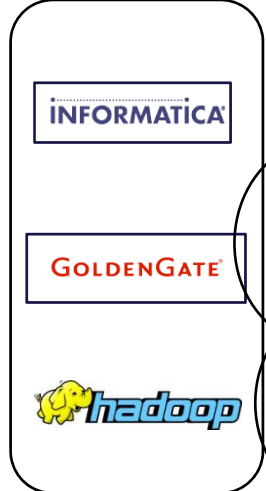
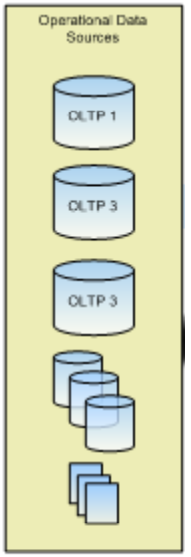
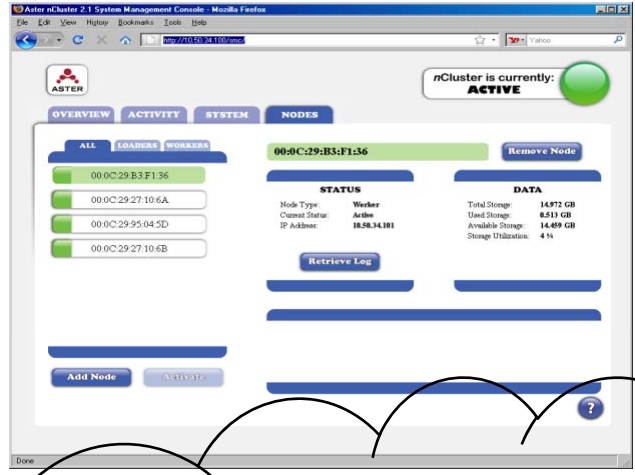
Pricing Complexities

	Rate	Usage	Totals
Amazon Elastic Compute Cloud			
View/Edit Service			
Amazon EC2 running Windows			
	\$0.125 per Small Windows Instance (m1.small) instance-hour (or partial hour)	120 Hrs	15.00
	\$0.30 per High-CPU Medium Windows Instance (c1.medium) instance-hour (or partial hour)	53 Hrs	15.90
	\$0.100 per GB Internet Data Transfer - all data transfer into Amazon EC2	3,004 GB	0.30
	\$0.170 per GB Internet Data Transfer - first 10 TB / month data transfer out of Amazon EC2	0.361 GB	0.06
	\$0.00 per Elastic IP address remap - first 100 remaps / month	1 Count	0.00
	\$0.10 per GB-month of provisioned storage	11.626 GB-Mo	1.16
	\$0.10 per 1 million I/O requests	283,810 IOs	0.03
		View Usage Report	32.45
Amazon Simple Storage Service			
View/Edit Service			
	\$0.170 per GB - first 10 TB / month data transfer out	0.000850 GB	0.01
	\$0.01 per 1,000 PUT, COPY, POST, or LIST requests	785 Requests	0.02
	\$0.01 per 10,000 GET and all other requests	381 Requests	0.01
	\$0.150 per GB - first 50 TB / month of storage used	0.233 GB-Mo	0.03
		View Usage Report	0.07
Taxes			
Estimated Taxes			
(Due February 1, 2009)			
Charges due on February 1, 2009+			
			32.52

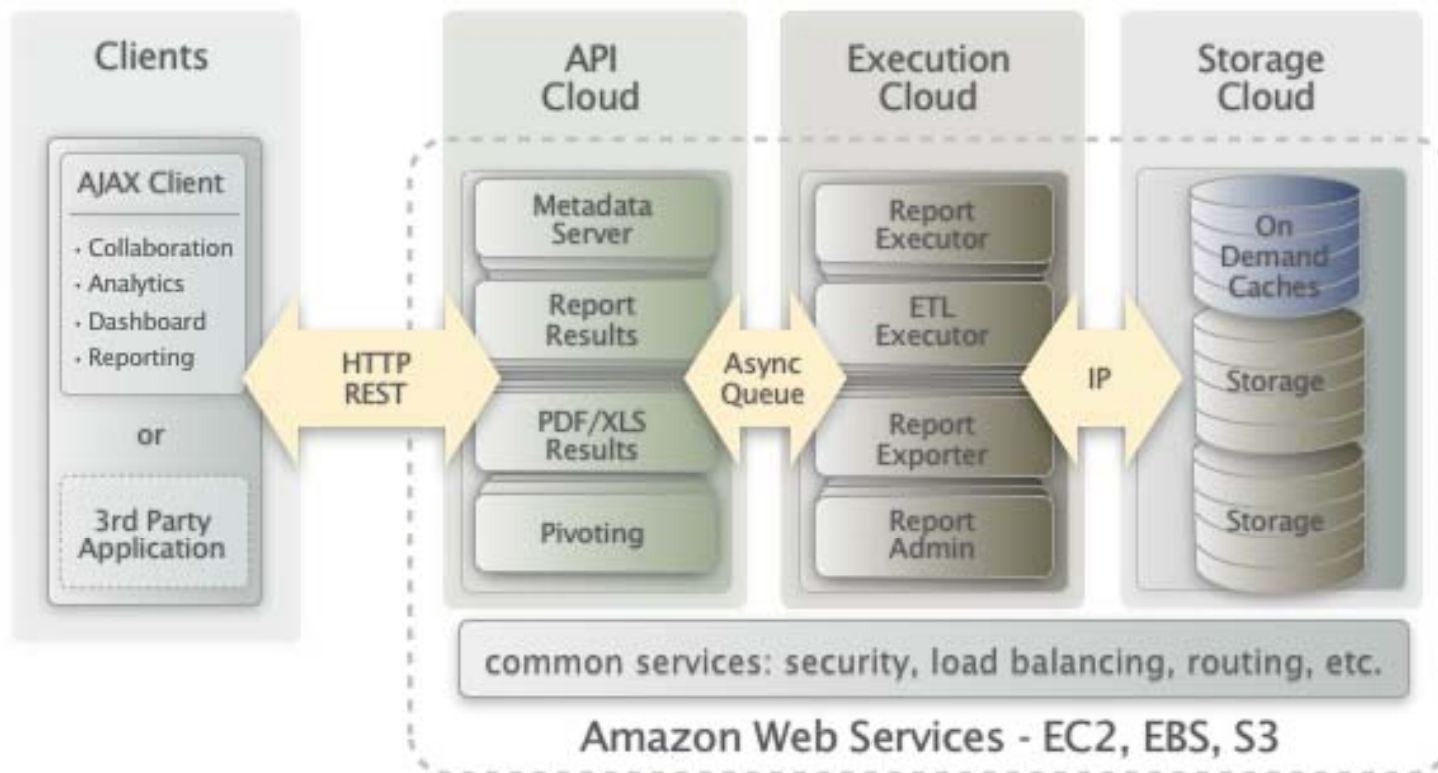


BI VENDORS IN THE CLOUD

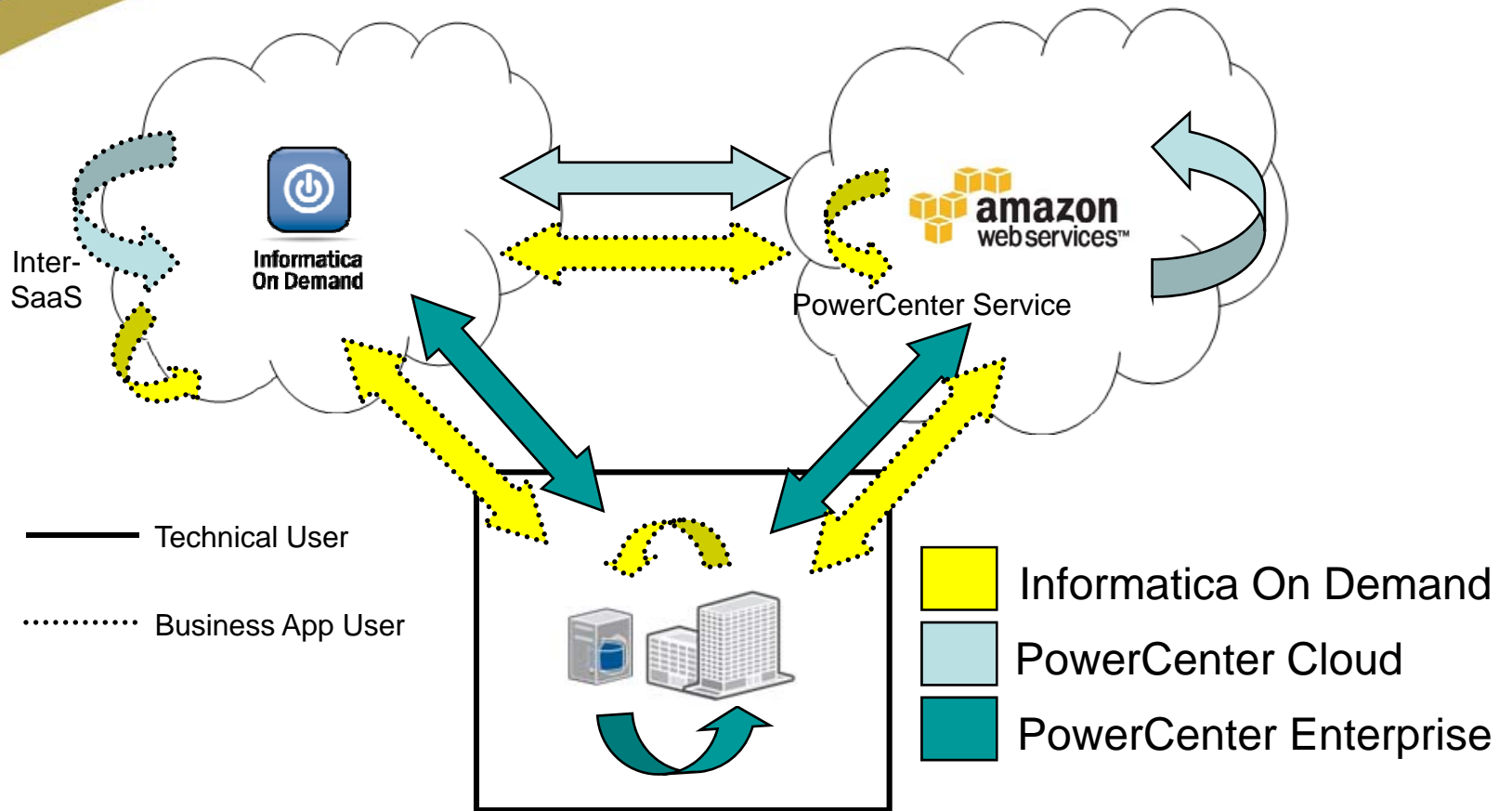
Aster Data Systems



GoodData - SaaS in the Cloud



Informatica Data Integration Cloud



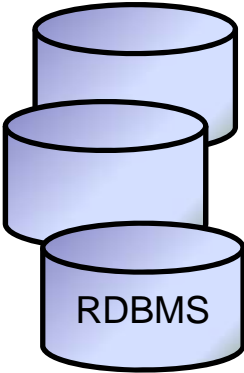
JasperSoft Cloud Solution



RIGHT SCALE®

- AUTOMATION ARCHITECTURE
- CLOUD-READY SOLUTIONS
- EXPERTISE & SUPPORT

Sources:



talend*
open data solutions

DW / Data Integration Developer

VERTICA

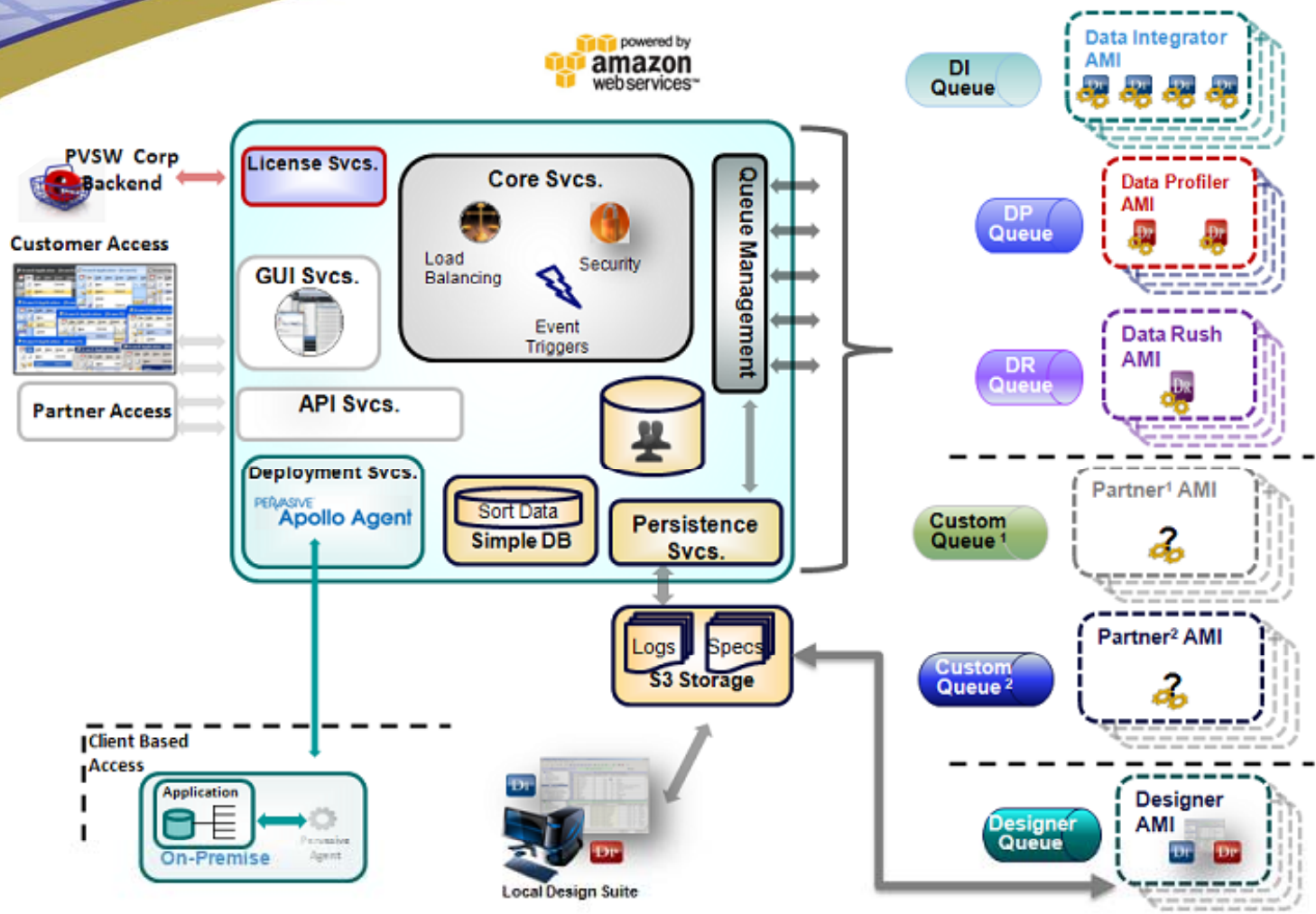
THE VERTICA DATABASE™
SIMPLY FAST.™

DW Professional

JASPERSOFT

JasperAnalysis

JasperServer



BI in the Cloud Case Study

The screenshot displays the BeyeNETWORK website interface. At the top left is the logo for BeyeNETWORK with the tagline "Global coverage of the business intelligence ecosystem" and "US EDITION |". To the right are links for "Login" and "Become a member!" and a search bar. A left-hand navigation menu lists categories such as Home, Channels by Industry, Channels by Topic, Channels by Expert, News, Articles, White Papers, Events, Web Seminars, Blogs, Spotlights, Podcasts, BeyePERSPECTIVE, Videos, BeyeRESEARCH, BeyeCONNECT, BeyeBLOGS, BeyeSEARCH, Resources, and Stay Informed. The main content area is divided into several sections: "THE LATEST WORD" with a "NEWS" sub-section listing articles like "Birst Announces Comprehensive Cloud BI Suite"; "ARTICLES" with titles like "Predictive Modeling at the Transaction Level"; and "WHITE PAPERS". Other sections include "Spotlights" featuring "John Doe" from "ACME Company", a "Video" section with a video player titled "Video Blog: Ask the Experts", "Of Note" with a link "Interested in Optimizing Performance?", and "Special Coverage". On the right side, there are links for "RSS FEEDS", "PODCASTS", "BLOGS", "NEWSLETTERS", and "WEBSITES", as well as an "Events" section listing dates like "November 5, 2009" and "November 12, 2009" with associated topics.

BeyeDW Project Requirements

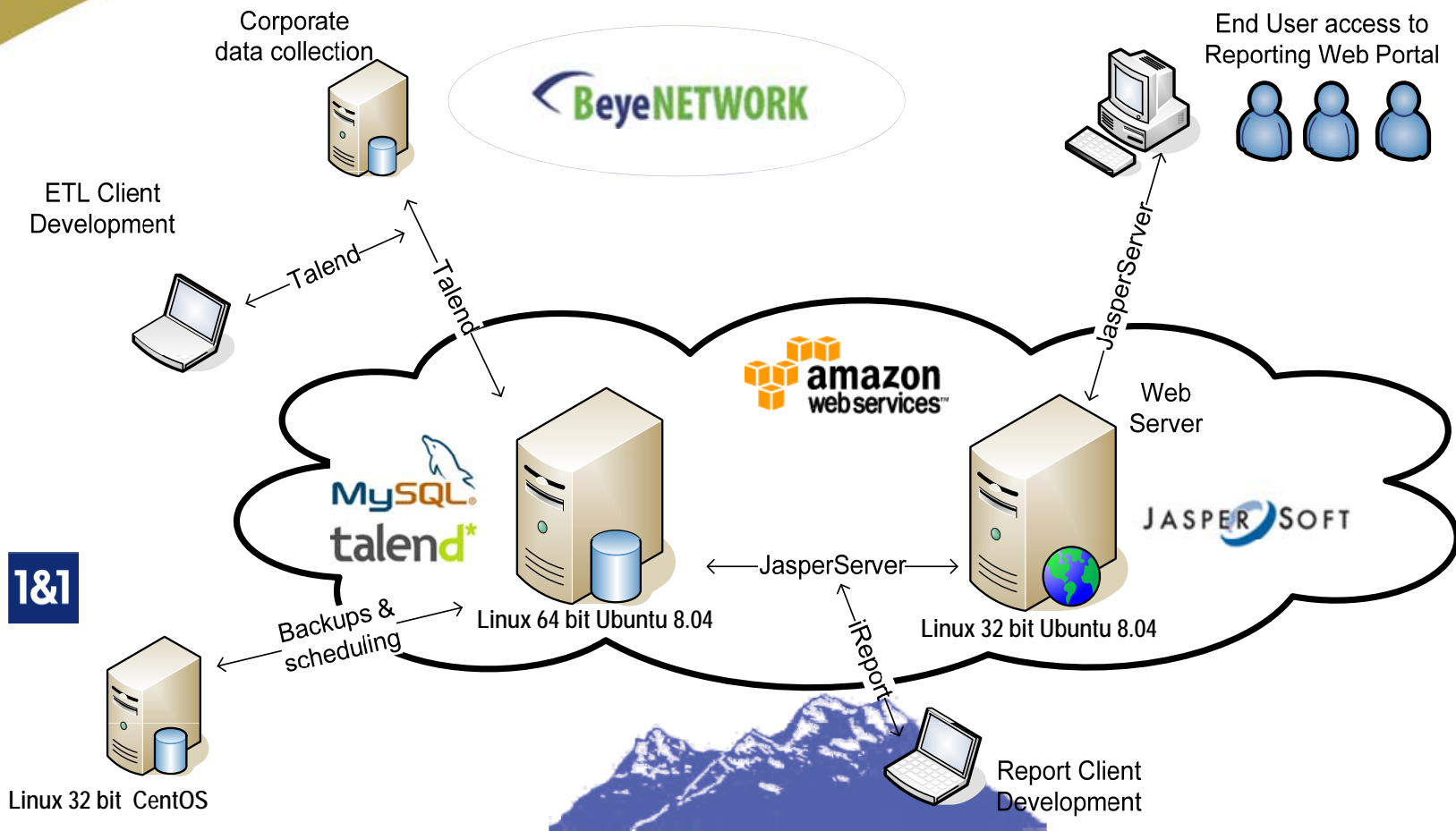


- ✓ Utilize open source software and LAMP stack
- ✓ No on-premise server hardware
- ✓ Ability to access BI software remotely
- ✓ Enablement of self-service BI
- ✓ Ability to load data 3 to 4 times per day

BeyeDW Reporting Requirements



BeyeDW Architecture



BeyeDW Demo

The screenshot shows the RazerSQL interface with a SQL query editor and a results grid. The query is as follows:

```
139 order by count(*) desc;
140
141 select max(bas_insrt_ts) from content_view;
142
143 select count(*) from content_view
144 where bas_insrt_ts > '2010-02-00'
145 and etl_job_run_id = 3694;
146
147 select * from content_view
148 where visitor_dwid = 7657746
149 order by session_id, view_datetime;
150
151 select session_id, ip_address, view_datetime, count(*)
152 from content_view
153 where view_datetime > '2009-11-30 00:00:00'
154 group by session_id, ip_address, view_datetime
155 having count(*) > 1;
156
157 select * from content_view
158 where view_datetime between '2010-01-29 00:00:00' and '2010-01-29 23:59:59'
```

The results grid displays the following data:

	content_dwid	content_type_dwid	channel_dwid	visitor_dwid	member_dwid	site_dwid	country_ip_dwid	sor_dtbs_nm	sor_tbl_nm	acq_ro
1	9592	4	74	7657746	20214	2	225	ben_ben	memberscontentviews	831097
2	10677	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831097
3	10678	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
4	10720	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
5	10721	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
6	10721	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
7	10175	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831097
8	10677	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
9	10809	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
10	10811	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
11	10809	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
12	10809	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	831098
13	9594	4	-2	7657746	20214	2	225	ben_ben	memberscontentviews	119029



BI IN THE CLOUD SUMMARY

Benefits to BI in the Cloud

- Flexibility to scale computing resources with few barriers
- Ability to shorten BI implementation windows
- Reduced cost for BI programs
- Ability to add environments for testing, proof-of-concepts and upgrades
- Geographic scalability

Challenges to BI in the Cloud

- The ability to scale-up is limited
- Difficult to quell security concerns
- Viability of moving large amounts of data
- Scalability of physical data access
- Reliability of service concerns
- Pricing is variable and complex

Lessons Learned (thus far...)

- Shared development can be a challenge
- Dynamically provisioning servers requires configuration scripts upon startup
- Keep your ear to the ground as technology and landscape changes rapidly
- Persistence is an important architectural consideration
- Consistent image naming standards are critical
- Beware of server clocks and time zones

Conclusions

- Deploying BI in the Cloud can help programs become more flexible, scalable and agile
- Cloud service architectures and vendors that provide them are still immature and is best suited for sandbox, development and test environments
- It can be challenging to configure databases and BI tools to run in the Cloud
- The Cloud holds a great deal of potential to change the way that we deliver BI to the masses

Special Thanks to:



- Ray Light, Director of Product Marketing, GoodData (www.gooddata.com)
- Dave Menninger, VP Marketing & Product Management, Vertica (www.vertica.com)
- Todd Freemon & Michael Kuhl, Pervasive Software (www.pervasive.com)
- John Thompson, CEO, North America, Kognitio (www.kognitio.com)

Contact Info



- If you have further questions or comments:

Steve Dine
President, Datasource Consulting, LLC
sdine@datasourceconsulting.com
888-453-2624