Ebook:

FIVE CLOUD INFRASTRUCTURE DEPLOYMENT ARCHETYPES

ORACLE® + Dyn

🏠 dyn.com ☎ 603 668 4998 @dyn

Ebook:

Five Cloud Infrastructure Deployment Archetypes

Using DNS to Understand Cloud Adoption from the Edge

What does the edge say about your infrastructure? For most end users, the Internet begins and ends with names. Do they enter arbitrary IP addresses into their browsers? No. Instead of IP addresses, we use hostnames, which means the Domain Name System (DNS) is being used to translate these names into resources.

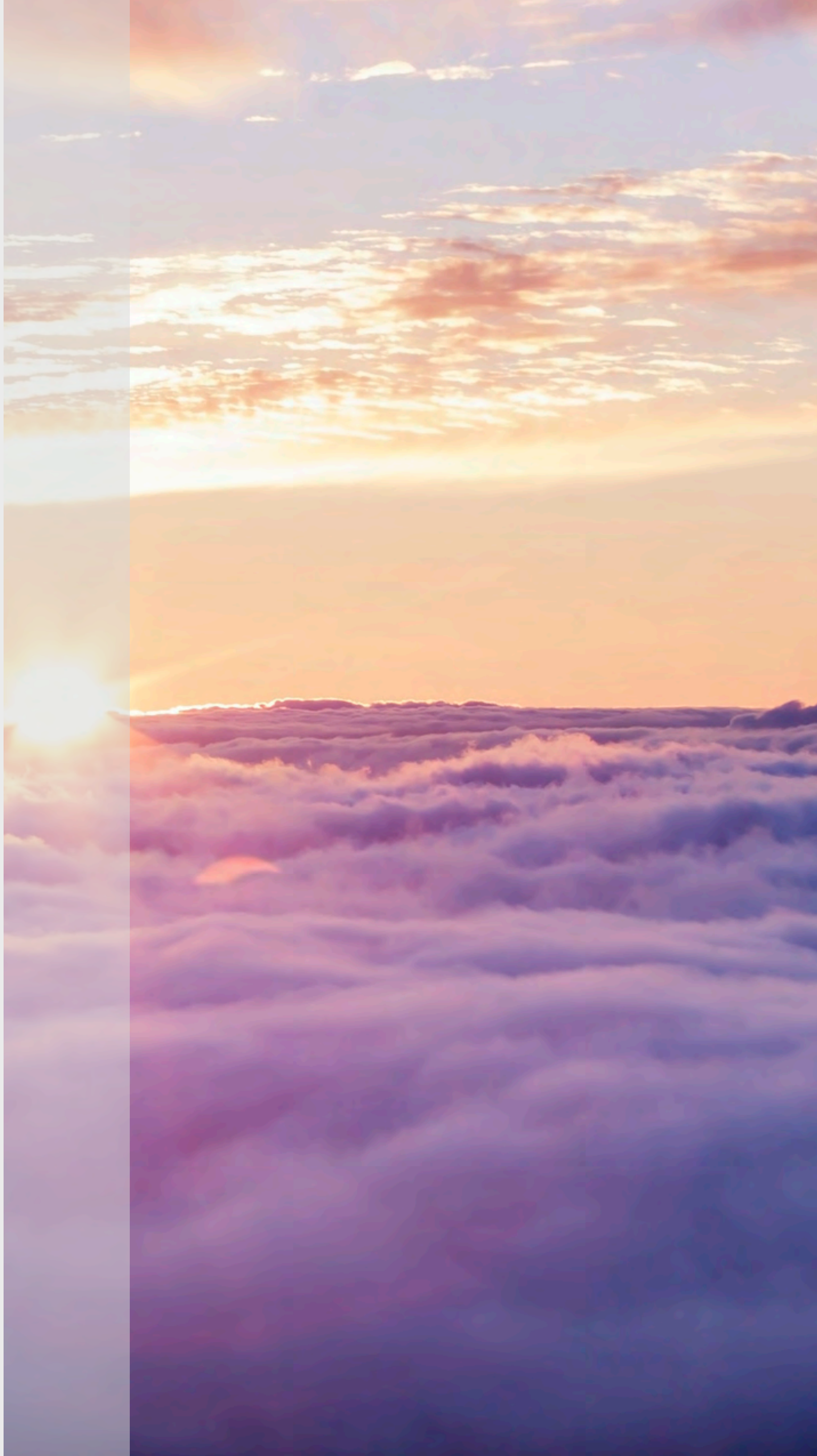
The DNS functionally defines the edge of your infrastructure, whether it is the A/AAAA records for appserver1.example.com, an MX record for mail.example.com, or a CNAME that implements load balancing for www.example.com. The resources defined in your domain namespace are functionally at your edge, abstractions that represent the infrastructure of your business.

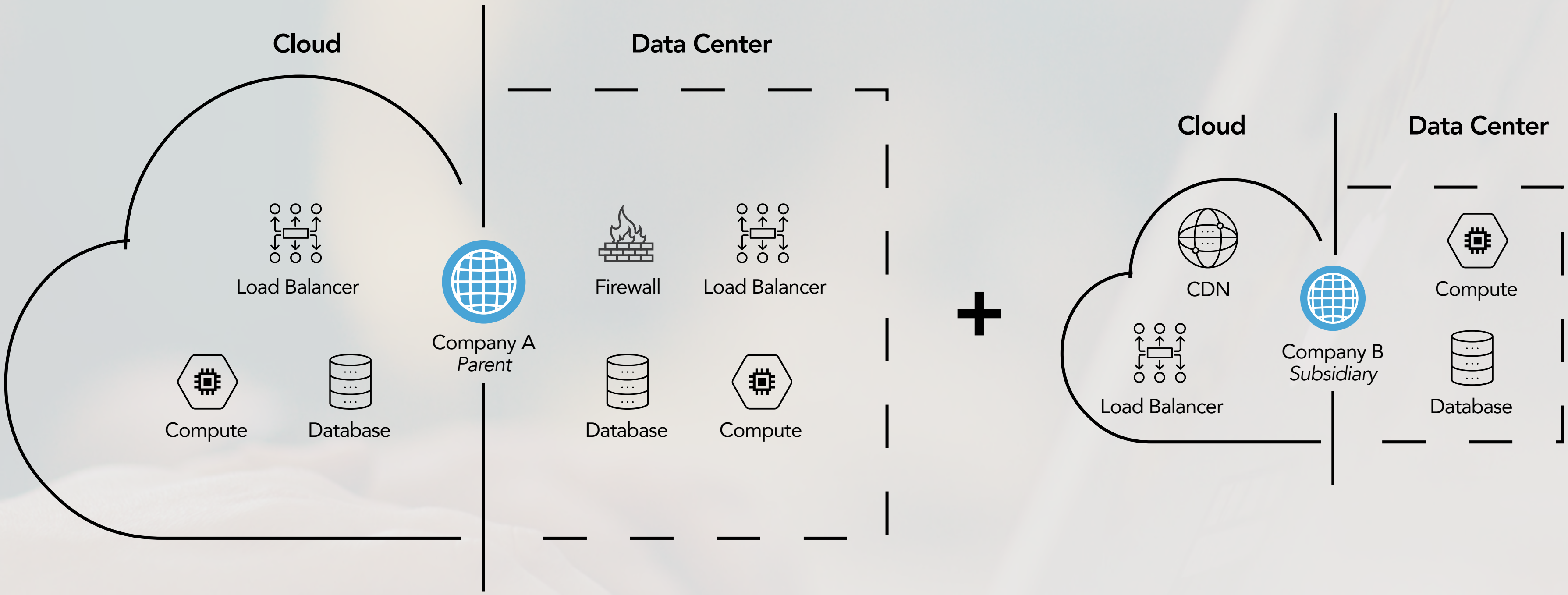
Through the analysis of the resource records within the DNS, we have identified five cloud infrastructure deployment archetypes:

1. Acquisitions Without Mergers
2. Hub and Spoke
3. Cloud Indigenous
4. Flux
5. Colony on Mars

These archetypes are reviewed below, followed by a discussion of the various types of DNS resource records that can help identify the deployment patterns.

.





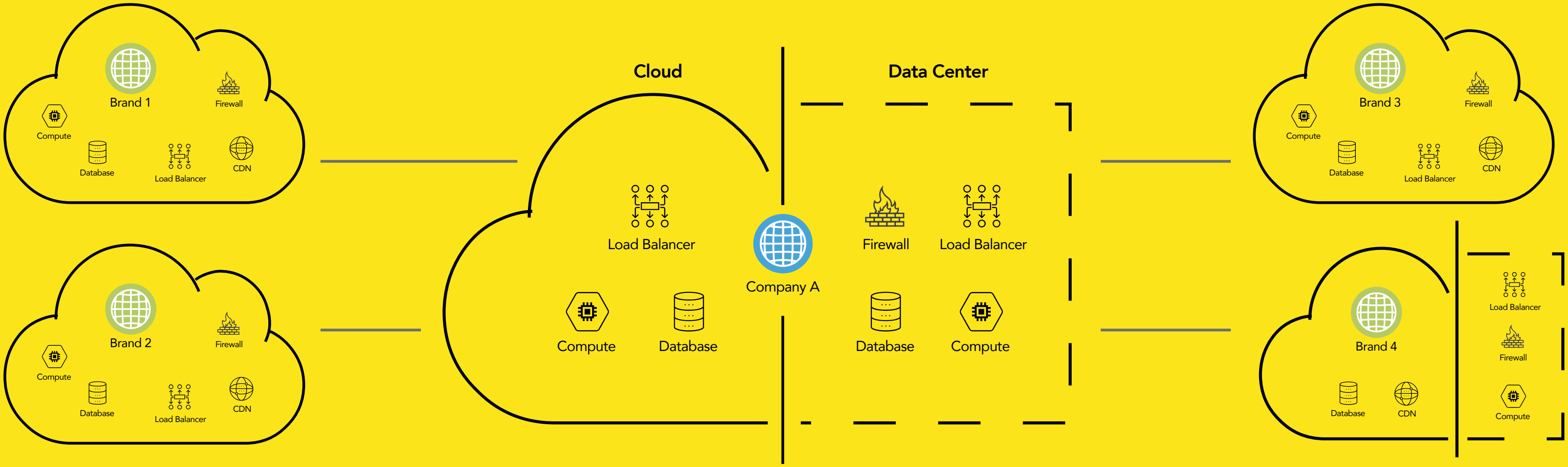
1. Acquisitions Without Mergers

A large pharmaceutical firm purchased a major competitor several years ago. After the closing of the merger, the two stock tickers were consolidated into one. In the DNS, however, both of their namespaces continued to exist. The parent company mainly used Amazon Web Services (AWS), and the company they acquired used SoftLayer. We were able to observe this based on: the namespaces of the two companies, some of the namespaces associated with their popular products, and the resource records they had configured.

As we were able to associate additional namespaces to the entities, we gained an increased understanding of the company's edge. After evaluating the namespaces associated with each pharmaceutical company and their products, it was shown that they were also hosting resources in Linode, OVH, and Unified Layer. We also saw different brands from these companies using various Content Delivery Networks (CDNs) to deliver content for their campaigns.

From the CFO's perspective, the company is missing out on being able to negotiate better rates from providers, as multiple contracts issued to individual brands diminish purchasing power.

From the CSO and CISO perspective, there are now multiple sets of credentials for a number of services that need to be monitored and audited.

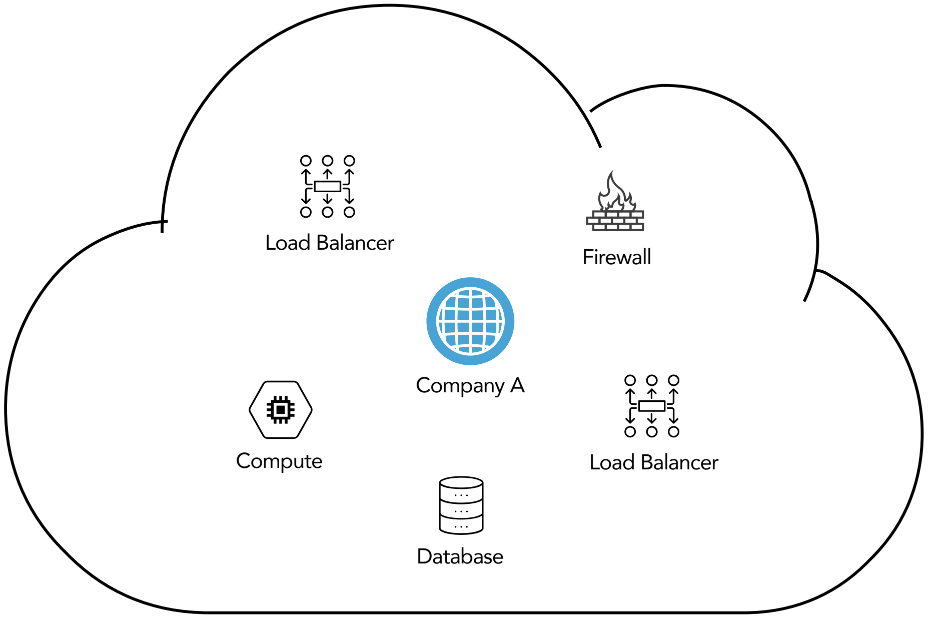


2. Hub and Spoke

In the quest to maximize shareholder value, some corporations grant autonomy to individual business units to avoid shackling them with bureaucracy. We have seen the hub and spoke pattern in a few verticals, but the strongest is in media, in which a large media parent company has a number of individual brands each operating their own infrastructure. In this scenario, each brand is making independent infrastructure decisions, including cloud hosting providers and CDNs.

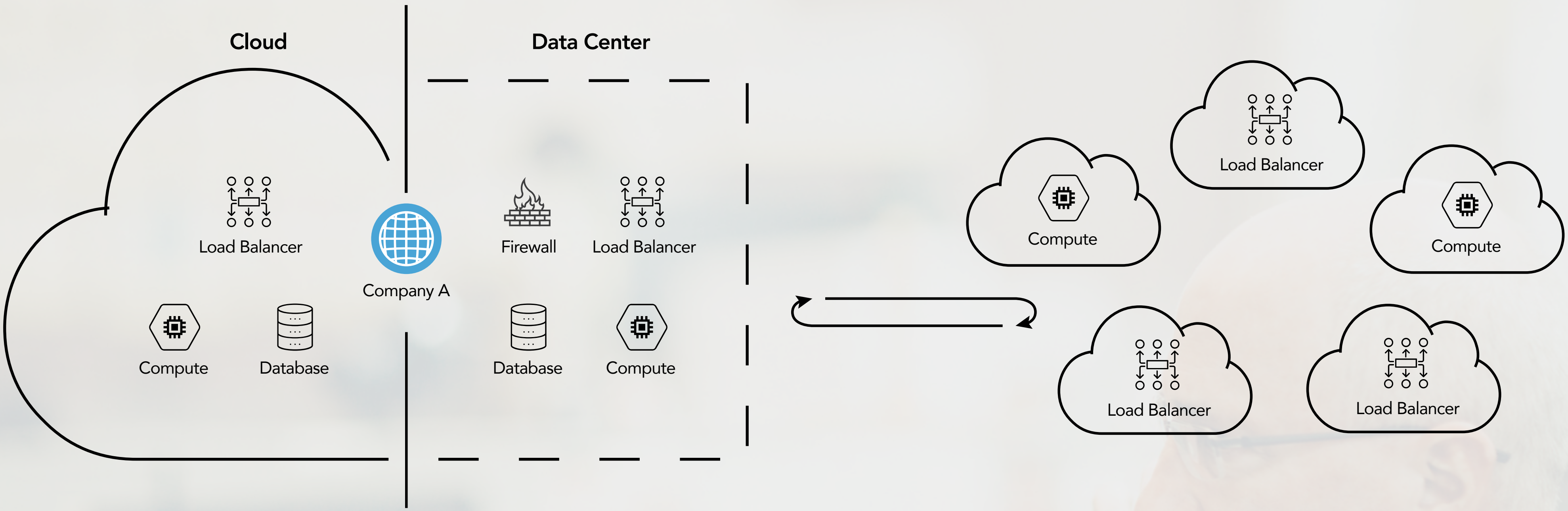
The distributed responsibility might enable individual brands to cut down on their time to market and select the vendors with the features sets that matter most to them, but the cost to the parent business is a lack of centralized purchasing power and the associated favorable pricing terms that can often be negotiated. A subpattern of this is seen as larger companies offer independence in infrastructure to advertising campaigns and special promotions. They often operate out of their own namespace which incorporates the theme of the campaign: example-kids.com, example-contest.com, and so forth.

The challenge in observing this pattern is establishing the metadata that ties a number of namespaces together to form a notion of a larger corporate entity. Through visibility into the DNS, we build out the resources associated with the individual namespaces, but awareness of the corporate entity that owns the namespace is what makes viewing the corporate hub and spoke configuration possible. To gain that awareness, you can use the Bloomberg company overview to identify subbrands or Crunchbase to find recent acquisitions.



3. Cloud Indigenous

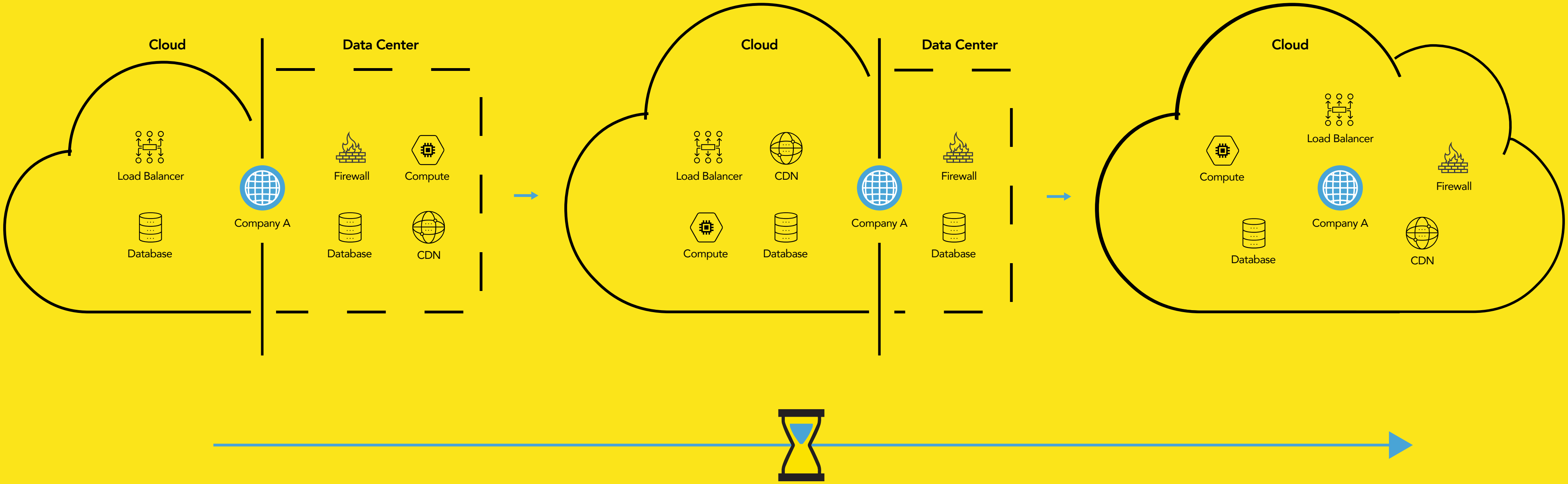
The "cloud indigenous" have built their services to leverage all the advantages of the "as-a-service" revolution. Their balance sheets hold no IPv4/IPv6 assets, autonomous systems, or Internet-facing hardware. Instead, they have invested in developing an "infrastructure-as-code" environment. Younger companies tend to be single-homed. Their namespace is filled with CNAMEs pointing to cloud based load balancers, customer service platforms, and marketing tools. Their TXT records contain a manifest of proof of domain ownership entries for different providers and SPF records for their transactional email providers—all of which will be explained in further detail later in this piece.



4. Flux

The “flux” pattern is a signature of a cloud-native company, or a company moving to the cloud. There is a central core of fixed infrastructure in a data center, or provider cloud and CDN, that supports the front door/landing page of the business. However, the rest of the infrastructure used to deliver service moves between providers, potentially driven by cost optimization, customer endpoint colocation/fast connect, or other factors. Similar to the business metadata driving visibility of the connection between the hub and spoke, time series analysis of a company’s namespace reveals this continuous churn of cloud provider CNAMEs and IP addresses. The flux pattern can be a source of operational information leakage. A provider may deploy test infrastructure for a demo or trial, which the customer accesses via a designated subdomain or hostname.

Over time, one can watch the gradual increase of subdomains/hostnames to witness growth, or a gradual decline signaling some detail about their user base. Similarly, you can get a sense of specific end-user requirements as the subdomain resolves to a new cloud infrastructure provider for the first time. If, for example, the CNAMEs or A records shift from all being in one cloud provider to include a second cloud provider after CompanyM is brought on board. We can look at the DNS zone data for CompanyM and see that most of their infrastructure falls into the new providers space. This might be a hint that supporting their existing infrastructure provider was required to acquire the new business.



5. Colony on Mars

The entity in question owns some of their own IP space and appears to keep most of their core services hosted in their own IP address space. However, they have moved their customer-interfacing edge into cloud provider infrastructure. Their landing page is behind a web application firewall (WAF) and their web server farm has moved to the cloud for autoscaling. Additionally, they are distributing assets related to the web experience from a CDN. They are maintaining their corporate data center, possibly because they have an enterprise workload that they haven't found the right cloud provider for, or maybe because they have a longer depreciation schedule for the last round of hardware they purchased. Or perhaps they are starting their migration to the cloud, starting with customer-facing assets.

Some entities own and operate infrastructure from their own IP space, or they have a long-running relationship with a hosting provider, but have built out an expanding presence in a cloud provider over time. The implemented services all remain in the classic environment, but all new development appears in the cloud provider. Observing this requires a time series of the namespace of the company and shows a steady increase of resource records which resolve to the enterprise cloud provider, observed across the changes in the number of CNAMEs and owner of A/AAAA record related IP space.

So how, exactly, did we discover these five cloud infrastructure deployment archetypes? Let's take a peek behind the curtain.

Understanding DNS Resource Records

The DNS provides a map of a business. The resource records outlined below, provide a list of relationships and dependencies of the business itself, starting with the name servers defined in NS records, and ending with domain ownership verification in TXT records and SPF records.

- NS records specify the nameservers that are associated with a given domain name, and establish which infrastructure is making sure the rest of the names in the namespace are available for resolution. Observing whether a single provider or multiple providers are used can give insight into corporate policy and/or organizational risk tolerance—avoiding single points of failure by requiring multiple vendors.
- A and AAAA records provide a map of which infrastructure is being used to deliver services. They map hostnames to IP addresses. The IP addresses associated with these A/AAAA records ARE the edge; they are the resource that will receive the packets. Does the owner of the namespace own the IP address space it is operating in? If not, who does? The answers can provide insight into the type of cloud archetype an organization may be associated with.
- CNAME records are essentially a pointer to another resource. CNAMEs are commonly used to implement CDN services, advanced traffic steering features, or to integrate third-party services. For example, if a CNAME is defined for newsletters.example.com, and it is a pointer to c90210bae.emailmarketing.com, it is clear which vendor is providing email newsletter distribution services.
- MX records clarify if the company is operating its own mail infrastructure or if they have outsourced it to a third-party provider.
- TXT records contain SPF (Sender Policy Framework) records, which are a map of which IP addresses and third parties are approved to send mail on behalf of the domain. TXT records also contain site and domain verification hashes used by third-party providers to clarify domain ownership.

A DNS-based view of the world can help you understand trends in edge services. Analysis can be done on a corpus of DNS data, which can come from zone files, passive DNS observation, web crawl logs, and so forth. The core attributes of the resulting data set are the names and associated resource records. To illustrate the production use of these records, we will use the results of a dig dyn.com ANY for some examples. (Note that the results of the command have been trimmed down to only include one record of each type.)

```
dyn.com.   21599 IN     NS     ns3.p01.dynect.net.
dyn.com.   29    IN     A      204.13.248.106
dyn.com.   599   IN     AAAA   2600:2001:0:3::106
dyn.com.   21599 IN     MX     10 aspmx.l.google.com.
dyn.com.   3599  IN     TXT    “v=spf1
ip4:204.14.232.0/21 ip4:216.6.202.0/24
ip4:195.160.236.246/32 ip4:195.160.237.248/32
ip4:52.38.191.241 include:mktomail.com
ip4:182.50.76.0/22 include:_spf.google.
com “ “ip4:216.146.45.0/24 ip4:96.43.144.0/20
ip4:104.209.130.71 ip4:104.209.133.54
ip4:104.209.134.151 ip4:104.209.133.195
ip4:104.209.130.35 ip4:104.209.132.14
ip4:111.221.94.91 ip4:23.98.64.25 “ “ip4:138.91.37.12
ip4:23.101.16.102 ip4:23.101.17.118 ip4:23.101.18.6
ip4:40.123.47.107 ip4:13.68.100.116 include:spf.
dynect.net include:stspg-customer.com include:_spf.
salesforce.com ~all”

dyn.com.   3599   IN    TXT    “status-page-domain-
verification=n00jg490wdwp”

dyn.com.
3599 IN    TXT    “google-site-verification:
VRIuHYFo0s9aWRVZxlAPyReiMYiKsMfWduASFCCyDlI”
```

Let’s start with the NS record: seeing a dynect.net hostname in the associated value tells us that dyn.com is (unsurprisingly) using Oracle Dyn’s Managed DNS service. Keywords present within the target value(s) of the NS record can provide insight into the third-party providers. Examples include:

- *dynect* >> Oracle Dyn Managed DNS
- *domaincontrol* >> GoDaddy
- *ultradns* >> Neustar UltraDNS
- *verisigndns* >> Verisign
- *cloudflare* >> Cloudflare

The target value(s) may also include the company’s own domain name or one associated with the company. This may indicate that they are doing their own authoritative DNS, or it could indicate that they are using vanity nameserver records. In the latter case, the NS records are in the company’s namespace or a namespace associated with the company, but a managed DNS provider is actually being used as the authoritative. In these cases, we can take the NS records and look up all their associated A / AAAA records and associate the IP address(es) to a provider.

We are about to dive into a nuanced problem that plagues a number of researchers and organizations who are trying to combine and attribute disparate data sources. The data starts with observations in the DNS, but then quickly expands into data from the Regional Internet Registries, RIR, which track which organization is responsible for IP space and autonomous systems, as well as domain registration data which exposes on whose behalf a domain was registered. This process of working from a name and a resource record value to the owner of a domain, an IP block and autonomous will be referenced again when looking into MX and CNAME records.

This provides a transition to looking at classifying A / AAAA records. Whether it’s dyn.com. IN A 204.13.248.106 or ns1.example.com IN A 192.168.10.10, the goal is to associate the namespace with an IP address space owner. The first step is to take the IP address and associate it with the data held by the Regional Internet Registries (RIRs), which is available via a WHOIS query. Below is an example of a slightly truncated WHOIS query - it includes a good bit of data.

```
NetRange:       162.88.0.0 - 162.88.255.255
CIDR:           162.88.0.0/16
NetName:        DNSINC-4
NetHandle:      NET-162-88-0-0-1
Parent:         NET162 (NET-162-0-0-0-0)
NetType:        Direct Allocation
OriginAS:       AS33517
Organization:   Dynamic Network Services, Inc. (DNS-33)
RegDate:        2013-11-27
Updated:        2016-10-20
Ref:            https://whois.arin.net/rest/net/NET-162-
88-0-0-1

OrgName:        Dynamic Network Services, Inc.
OrgId:          DNS-33
Address:        150 Dow St.
City:           Manchester
StateProv:      NH
PostalCode:     03101
Country:        US
RegDate:        2004-01-06
Updated:        2017-01-28
Ref:            https://whois.arin.net/rest/org/DNS-33
```

>>

```
OrgTechHandle: IAA29-ARIN
OrgTechName:   IP Address Administrator
OrgTechPhone:  +1-603-296-1598
OrgTechEmail:  ip-admin@dyn.com
OrgTechRef:    https://whois.arin.net/rest/poc/IAA29-
ARIN

OrgAbuseHandle: ABUSE514-ARIN
OrgAbuseName:   Abuse Department
OrgAbusePhone:  +1-603-668-4998
OrgAbuseEmail:  abuse@dyndns.com
OrgAbuseRef:    https://whois.arin.net/rest/poc/
ABUSE514-ARIN

OrgNOCHandle: NOC1473-ARIN
OrgNOCName:   Network Operations Center
OrgNOCPhone:  +1-603-668-4998
OrgNOCEmail:  noc@dyndns.com
OrgNOCRef:    https://whois.arin.net/rest/poc/NOC1473-
ARIN
```

What does this information tell us?

- The CIDR (Classless Inter-Domain Routing) is going to tells us the range of addresses associated with this entity.
- The OriginAS (Origin Autonomous System) provides another source of organizational information, along with email addresses to confirm when we look up the associated autonomous system details.
- The Organization and OrgName provide sets of strings which can be used later on to try to assign identity or infer relationships.
- The various email addresses expose namespaces associated with the IP address that was queried. In the example above, we now know that the domains dyndns.com and dyn.com are associated with this range. (However, make sure you don't overassociate things based on addresses from free email providers such as Gmail, Yahoo, or Hotmail).

For each A / AAAA record we now have a dataset that contains the following details:

> domain name, IPv4 or IPv6 address, CIDR, origin autonomous system, organization, organization name, and email addresses.

We can then take the origin autonomous system and see what other details are available:

```
Number            33517
Name              DYNDNS
Handle            AS33517
Organization      Dynamic Network Services, Inc. (DNS-33)
Registration Date 2005-01-11
Last Updated      2012-03-02
Comments
RESTful Link      https://whois.arin.net/rest/asn/AS33517
See AlsoRelated POC records.
See AlsoOrganization's POC records.
```

From this set of information, we are mainly interested in the Name and Organization. Depending on the scope of your research, you may want to focus on a specific component. For example, if you are looking to identify cloud provider usage within a namespace, instead of considering ALL of the ASes, you can focus on specific sets. For example, if the namespace has an A or a AAAA for which the OriginAS falls into any of the groups below, chances are they are using that provider's cloud service:

- Oracle = [31898]
- Softlayer = [36351, 6461]
- Rackspace = [15395, 19994, 27357, 33070, 44716, 45187, 58683]
- Azure = [8075, 8068]
- Amazon = [14618, 16509, 38895, 4230, 8987]
- Google = [15169]

Similar to the processing of the NS records, MX records are interpreted by the domain in the value field. In the example, we have an MX record that tells a sender the following: if you are looking to send email to anyone with an @dyn.com email address, the servers responsible are defined by resource records associated with aspmx.l.google.com.:

```
dyn.com.   21599  IN     MX     10 aspmx.l.google.com.
```

There are large clusters of customers of a few hosted email providers, which makes 80 percent of the work of identifying email providers very easy. Dealing with the last 20 percent depends on your requirements. If your goal is to identify who is using one of the major providers and who is not, then mapping the domain of the MX record will do the trick. If not, then following the process outlined above for attributing vanity nameservers and looking for the A or the AAAA associated with the domain in the MX record and using the IP ownership data can help you make an informed guess.

Next up, we have TXT records. These records provide the ability to associate some arbitrary and unformatted text with a host or domain name. For our purposes here, we will look at Sender Policy Framework (SPF) records and “other” records, both of which can appear as the value of a TXT record. The SPF record contains a manifest of domains and IP addresses that are allowed to send mail on an organization’s behalf. This represents the perfect blend of domain matching and IP classification.

Below is an example of one of the dyn.com TXT records which details who can send mail originated from the dyn.com namespace.

```
dyn.com.          3599   IN
TXT “v=spf1 ip4:204.14.232.0/21 ip4:216.6.202.0/24
ip4:195.160.236.246/32 ip4:195.160.237.248/32
ip4:52.38.191.241 include:mktomail.com ip4:182.50.76.0/22
include:_spf.google.com “ “ip4:216.146.45.0/24
ip4:96.43.144.0/20 ip4:104.209.130.71 ip4:104.209.133.54
ip4:104.209.134.151 ip4:104.209.133.195 ip4:104.209.130.35
ip4:104.209.132.14 ip4:111.221.94.91 ip4:23.98.64.25 “
“ip4:138.91.37.12 ip4:23.101.16.102 ip4:23.101.17.118
ip4:23.101.18.6 ip4:40.123.47.107 ip4:13.68.100.116”
```

You can look at the SPF record as a blob of domains and IPs or another web of business relationships. For example, a number of the IP ranges in the SPF record belong to Oracle Dyn, however others belong to Microsoft, Amazon, Salesforce, and others. This insight helps contextualize how different SaaS / PaaS platforms are used and integrated into service offerings in relation to where mail-related resources are located. As previously mentioned, in the context of nameservers you often need to dig a bit deeper for contextual understanding to decipher if the IP is related to the cloud provider or a service operating from the cloud provider’s platform. In this case we can look at the RIR WhoIS data to see that the IP referenced ip4:52.38.191.241 belongs to a subdelegation labeled (Message Systems AWS-MESSAGESYSTEMS-MD (NET-52-38-191-224-1) 52.38.191.224 - 52.38.191.255). From this point on, we can look at passive DNS to see that the IP has been associated to some names related to mail transfer agents that fall into the sparkpostmail.com namespace as well as email.influitive.com.

In the “other” category, it’s often a function of knowing what you’re looking for. Both of the examples below are related to domain verification.

In these cases, a service provider requires the customer to create a TXT record with a specific string or unique hash to prove they own a domain.

```
dyn.com.          3599   IN     TXT    “status-page-domain-
verification=n00jg490wdwp”

dyn.com.          3599   IN     TXT    “google-site-
verification: VRIuHYFo0s9aWRVZxlAPyReiMYiKsMfWduASFCCyDlI”
```

Last, but not least, we will cover CNAMEs. CNAME records are the primary means of integrating cloud or software / platforms as a service. It is common to see www.example.com configured as a CNAME to a cloud load balancer or web application firewall. CNAMEs are mappings between domain names, so the same mapping process applied to NS and MX records applies yet again. Due to the number of products and services implemented via CNAMEs, the key/value mapping gets large quick. Different domains or subdomains connote different services, which can have an impact on how you scope your analysis: do you just want to map a CNAME back to a provider? or do you also want to know what service was implemented? Examples of a provider/service mapping is below:

- *akamai.net, edgesuite.net* >> Akamai standard (HTTP) content delivery
- *edgekey.net, akamaiedge.net* >> Akamai secure (HTTPS) content delivery
- *akadns.net* >> Akamai global traffic management service
- *cloudfront.net* >> Amazon Web Services CloudFront content delivery
- *s3.amazonaws.com* >> Amazon Web Services S3 object storage
- *compute.amazonaws.com* >> Amazon Web Services EC2 cloud compute platform
- *elb.amazonaws.com* >> Amazon Web Services Elastic Load Balancer

The end result of this data processing and tagging is a key value of namespace and company, or in some cases namespace and company—product pairs. This dataset forms the foundation from which analysis can be done to identify cloud archetypes. However, it is missing one key component: the notion of an entity, such as a company, which operates using one or more namespaces. Example *Industries*, for instance, supports a few different lines of business; each has its own namespace such as: example.com, examplecloud.net, or exampledata.com. Each of these namespaces can define a function or business unit of the company's public-facing Internet infrastructure.

Conclusion

The DNS provides a service for mapping abstractions to resources. It can be used to steer traffic based on the geolocation of the requestor or based on which endpoint will provide the optimal end user experience. Studying the relationships defined in the DNS in the context of the business entities that own and operate the corresponding namespaces and IP address ranges paints a picture of "who is using what." If we look at this data with a dimension of time, we can start to understand and contextualize the change by understanding "who was using what, when." This time series allows the analyst to explore rates of adoption, conversion, and churn, or to cluster by common providers, time of change, and so forth. Individuals and organizations define their edge resources in the DNS; when Internet-accessible resources change, the change is implemented in the DNS.

The future is here. **Visit dyn.com to learn more.**

Rethink DNS.

Oracle Dyn is global business unit (GBU) focused on critical cloud infrastructure. Dyn is a pioneer in DNS and a leader in cloud-based infrastructure that connects users with digital content and experiences across a global internet. Dyn's solution is powered by a global network that drives 40 billion traffic optimization decisions daily for more than 3,500 enterprise customers, including preeminent digital brands such as Netflix, Twitter, LinkedIn and CNBC. Adding Dyn's best-in-class DNS and email services extend the Oracle cloud computing platform and provides enterprise customers with a one-stop shop for infrastructure as a service (IaaS) and platform as a service (PaaS).

ORACLE® + Dyn

dyn.com ☎ 603 668 4998 @dyn