Bailey Berry

Professor Posner

Systems and Infrastructures

Write Up #2

Storing data on the cloud means basically that instead of data being stored locally in your own hard drive, it is sent to remote data centers through the internet. These data centers are made up of series of network servers clustered together in racks and connected by routers. These systems are hooked up to a backbone, which connects the whole system to the internet. When your computer is also connected to the internet, you can send your data in the form of packages to the backbone and then into the data center, where it is stored in several different places.

When your data is stored in Dropbox, it is packaged into data packages, which can be sent via servers to Dropbox's storage racks and retrieved from them when the data is needed again. This all works via a system of networks, connectors, routers and switches. All these devices connect together to tell data where it needs to go and provide pathways for it.

Dropbox has data centers across the United States in Washington, California, Texas, Illinois, New York and Virginia. It has expanded its data centers internationally to Germany, Australia and Japan where more secure storage is available for some Dropbox Business users.

Within a data center, Dropbox has developed what it calls a Quad-Plane, 3-Tier fabric. This system was developed to address issues with storage capabilities and scalability. Originally, Dropbox centers worked using a "four-post architecture" scheme featuring a cluster made up of CC (Cluster connectors) and CR (Cluster Routers). At each data center, there were several Clusters, which are tiered network systems. Networks are simply systems of communication for sharing data. The Cluster Routers handled the data packages traveling within racks in each cluster and data traveling out of the cluster. Data traveling between clusters at a data center is delivered by Cluster Connectors. The system used eBGP between Cluster Connector and Router and iBGP between a rack and cluster router, the different languages used to package and communicate what data is and where it should go.

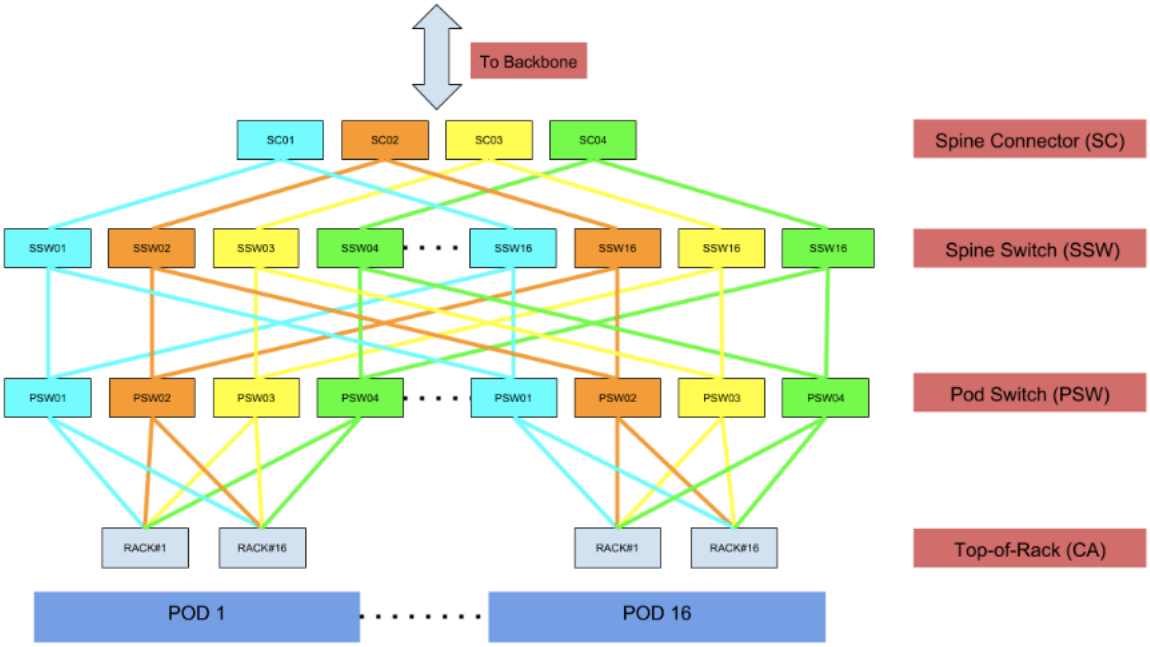


Figure 3: 256 rack fabric

[1]

Creating a "fabric" was meant to address design issues with limitations of stacking rigid hardware. Clusters were replaced with spine to reflect the more elegant horizontal connections between elements of the stack. Connecting hardware horizontally as opposed to vertically maximized the usage of ports on each device so that less ports went unused. This way, more racks could easily be gained to add capacity on demand. This new system unifies ASICs at every tier (Application-specific integrated circuit- the microchip which holds all of the potential logic gates, but created specifically to handle cloud computing/ data transfer operations), meaning that data can more easily communicate between different pieces of the stack. Uses ECMP (Equal-cost multi-path routing), which allows for several "best" routes for data to take to get to one destination. Typically, routers have to compute which particular route data must take to get to the proper destination. By offering more than one route, data transfers should occur more quickly. Individual Racks are organized into series of sixteen, called a "Pod". Instead of using both eBGP and iBGP, this system relies only on eBGP for ease of communication.

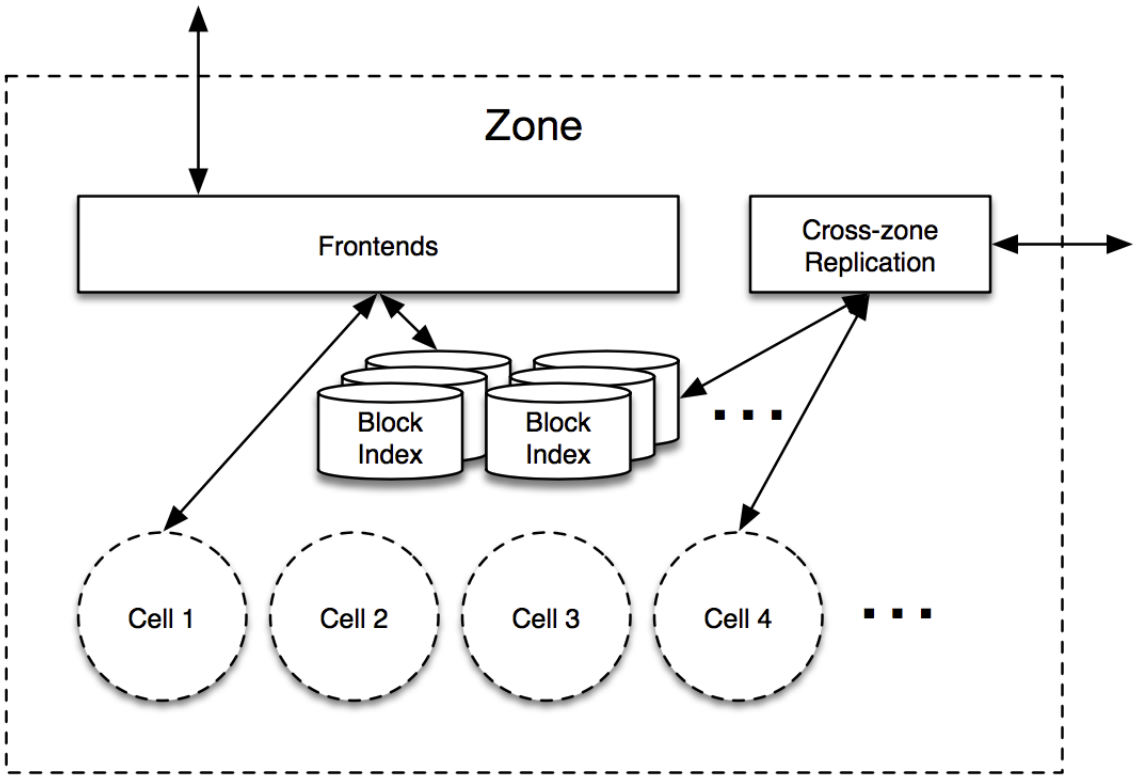
Connections within and between these building blocks in the fabric are made with Multimode fiber optic cables, the longest one being 140 feet.

Series of MDF (Main Distribution Frame), or cable rack, rows connect different pods to different power sources, meaning that in the event of any electrical anomaly, there are redundant electrical power distribution panels. Each MDF row "has diverse network pathways to each cabinet position".

Dropbox has something called a "Magic Pocket", which is the technology that manages the storage of data. Dropbox stores data using block technology. This means that files and the metadata about the files

[1] Sakpal, Vishal. "The scalable fabric behind our growing data center network". *Dropbox Blogs.* Dropbox: January 23, 2019. https://blogs.dropbox.com/tech/2019/01/the-scalable-fabric-behind-our-growing-data-center-network/ Accessed February 2, 2020.

are split into separate blocks of bits or data, encrypted, duplicated and stored separately for more secure storage. This is all done ultimately using Solid State Drives. Incoming data is accepted by front end nodes, which determine where to store the data. The data is split into blocks, which are identified by a unique hash and assigned a checksum for ensuring the authenticity of data. Blocks are copied several times and duplicates are then stored in different "zones", which refer to regions of the United States where data centers are located. Blocks are grouped together by a Block Index into buckets with other blocks, which are then grouped into volumes. The Block Index stores these locations by associating the hash with a cell, bucket and checksum. These volumes are written onto physical discs where they are stored in different cells.



[2]

Dropbox further categorizes data as "warm" storage or "cold" storage. It assumes that new data created will need to be accessed more often than older data. Dropbox's blocks are "immutable", so any changes to a piece of data will be stored in a separate system called "FileJournal". Data that is "warm" is kept at a higher level in the stack so that it can easily be sent out through the network. Data that is "cold" is the data that is written onto solid state disc drives.[3]

---

[2] Cowling, James. "Inside the Magic Pocket". *Dropbox Blog.* Dropbox: May 6, 2019. https://blogs.dropbox.com/tech/2016/05/inside-the-magic-pocket/. Accessed February 2, 2020.

[3] Le, Preslave. "How we optimized Magic Pocket for cold storage". *Dropbox Blog.* Dropbox: May 6, 2019. https://blogs.dropbox.com/tech/2019/05/how-we-optimized-magic-pocket-for-cold-storage/. Accessed February 2, 2020.

In order to use Dropbox, whether the website or the desktop app, your computer must be connected to the internet. This is the only way for your files and updates on your files to be sent to Drobox's data centers for storage. Dropbox does have options for offline access and even editing, but that information is only stored locally on your own computer until you are again connected to Wifi, meaning that it is vulnerable to any technological malfunctions that could destroy that information. Dropbox is very open about the ways in which it stores and transfers data. They seem to take pride in the methods they use to duplicate, spread and encrypt data to ensure secure storage and fast access.

https://help.dropbox.com/accounts-billing/security/physical-location-data-storage

https://computer.howstuffworks.com/cloud-computing/cloud-storage3.htm

https://blogs.dropbox.com/tech/2019/01/the-scalable-fabric-behind-our-growing-data-center-network/