Data Transfer Scenarios

Migration Online/Offline, Incremental

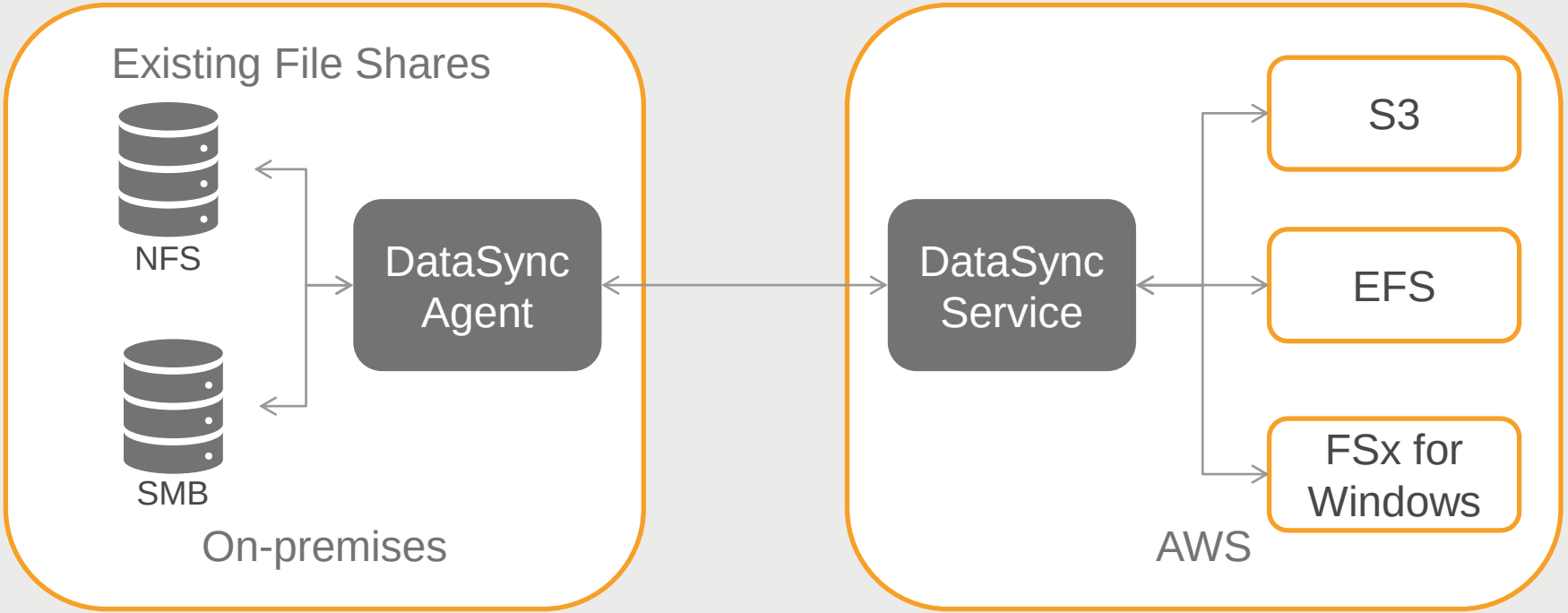
Chandra Lingam

Cloud Wave LLC

Several tools

- DataSync
- Storage Gateway
- Snowball
- S3 sync, cp
- S3 Batch Operations
- S3 Transfer Acceleration
- CloudFront PUT/POST
- Secure FTP

AWS DataSync



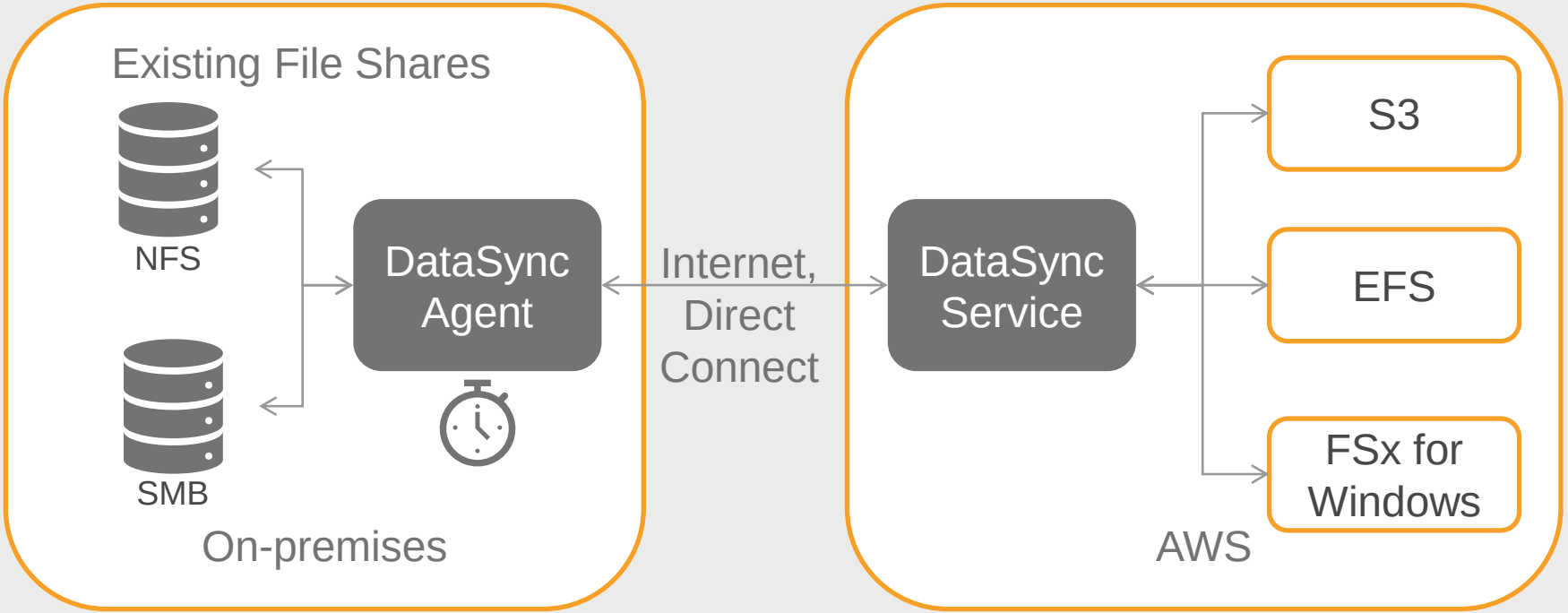
DataSync Agent

- On-premises VM Server
- SnowCone Appliance
- EC2 instance

Agent needs to be installed near data source

* don’t attempt to install in EC2 and read from on-premises

AWS DataSync



Server Configuration

On-premises VM:

- Four virtual processors
- 80 GB of disk space
- 32 GB RAM (for transfers up to 20 million files)
- 64 GB RAM (for transfers over 20 million files)

EC2:

- m5.2xlarge (for transfer up to 20 million files)
- m5.4xlarge (for transfer over 20 million files)

DataSync Features

- Data transfer between NFS, SMB, S3
- Optimized for moving large amounts of data
- Bandwidth throttle support
- One time or scheduled
- Store data directly in any of the S3 storage classes (including glacier and glacier deep archive)
- Copies metadata and permissions

Scenario 1

Customer needs to migrate 50 TB of data from on-premises systems to AWS cloud

Incremental data needs to be in sync until cutover to cloud

What cloud storage and data transfer service would you use for this?

Scenario 1 – Data Transfer Options

Service	Possible Solution
DataSync	Yes
S3 cp, sync	Yes
S3 Batch operations	No. It is used for S3 to S3
Secure FTP	Yes
Storage Gateway	No. due to cutover to cloud
Snowball	Yes. How to handle incremental transfer?
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

Which one to choose?

Network Connectivity

Transfer Speed (Mbps)	Amount of data transferred per day *	Transfer time for 50 TB
1,000	10 TB	5 days
500	5 TB	10 days
250	2.5 TB	20 days
100	1 TB	50 days
25	0.25 TB	200 days

* Approximate value. Easy to remember

Snowball turnaround time is around 5-7 days and you can send up to 80 TB in a single appliance (offline transfer)

Scenario 1 – Solution

Service	Possible Solution
DataSync	Yes, for initial transfer (if network is not a constraint) Yes, for incremental loading
S3 cp, sync	No. Too much effort to write logic
S3 Batch operations	No. It is used for S3 to S3
Secure FTP	No. Too much effort to write logic
Storage Gateway	No. due to cutover to cloud
Snowball	Yes, for initial transfer (if network constrained)
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

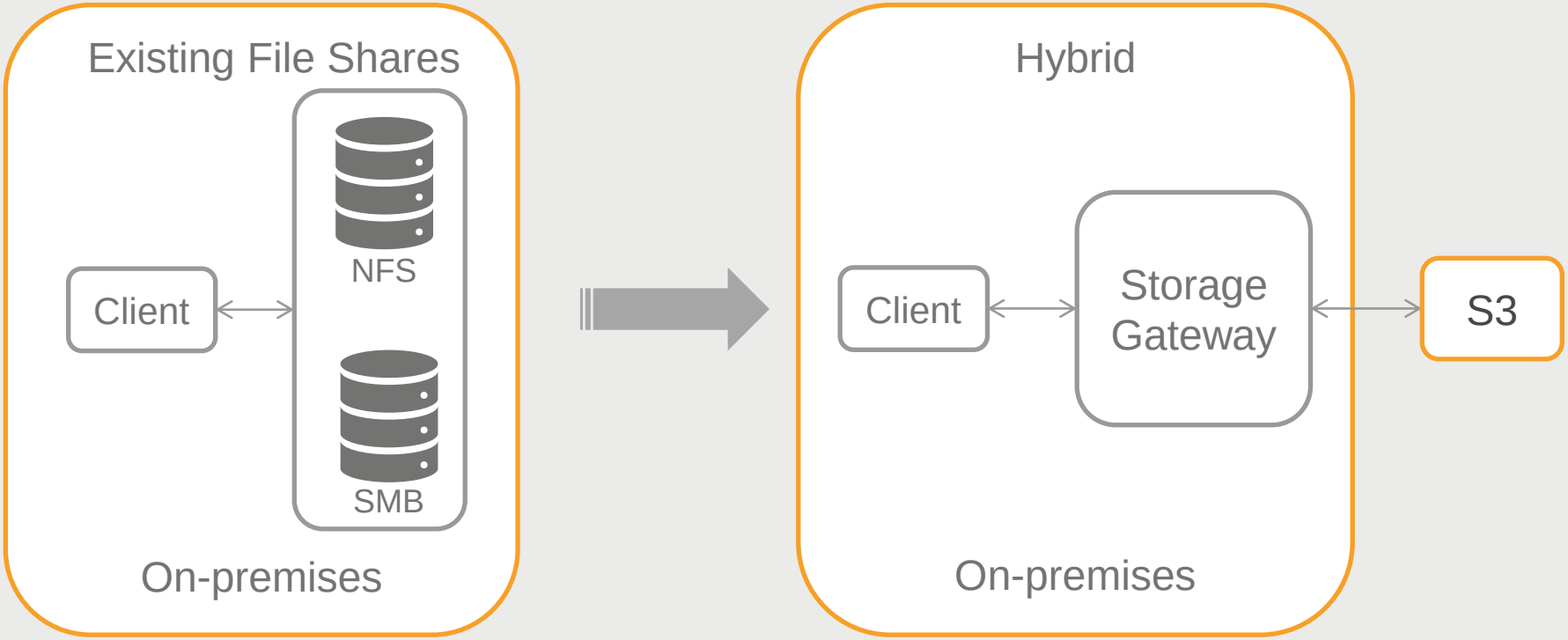
Scenario 2

Customer needs to migrate 50 TB from on-premises file shares to AWS cloud and make data available to on-premises applications.

Customer is interested in reducing the storage footprint on-premises.

What storage service and data transfer solution would you use?

On-premises



Scenario 2 – Solution

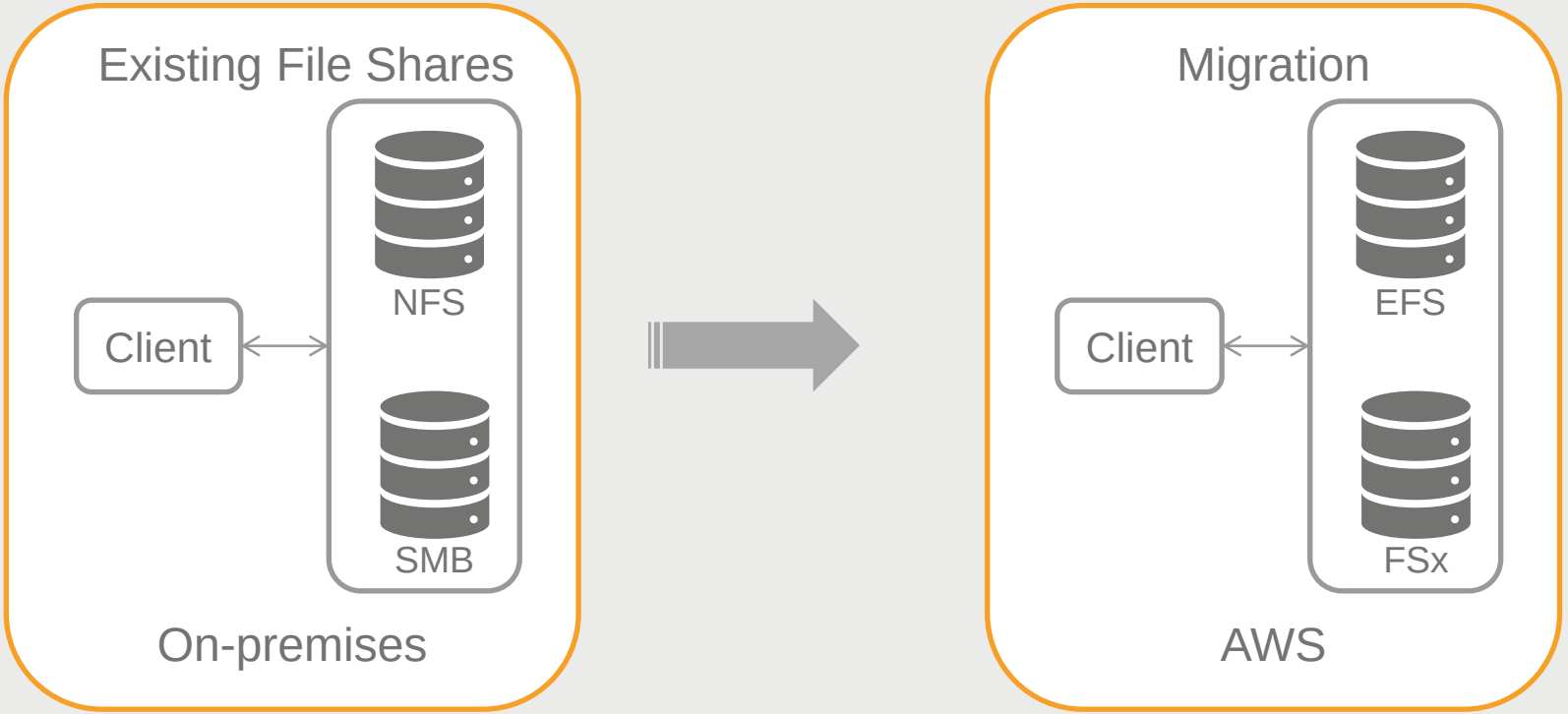
Service	Possible Solution
DataSync	Yes, for initial transfer (if network is not a constraint)
S3 cp, sync	No. Too much effort to write logic
S3 Batch operations	No. It is used for S3 to S3
Secure FTP	No. Too much effort to write logic
Storage Gateway	Yes, for incremental loading and for replacing on-premises storage (block, file share, tape)
Snowball	Yes, for initial transfer (if network constrained)
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

Scenario 3

Customer has on-premises Linux and windows applications that use NFS, SMB file shares. Customer is migrating the application to cloud.

What storage service and data transfer service would you use that minimizes changes to the application?

Scenario 3 – Lift and Shift



Scenario 3 – Solution

Service	Possible Solution
DataSync	Yes, for initial transfer (if network is not a constraint) Yes, for incremental transfer
S3 cp, sync	No
S3 Batch operations	No. It is used for S3 to S3
Secure FTP	No
Storage Gateway	No
Snowball	Yes, for initial transfer (if network constrained). Copy to S3 first and then copy to EFS or FSx
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

Scenario 4

For disaster recovery, customer needs to replicate the content of S3 bucket to another bucket in a different region.

```
Source: S3 bucket
Destination: Another S3 bucket
```

Scenario 4 – Options

Service	Possible Solution
DataSync	Yes, for initial transfer of existing objects
S3 cp, sync	Yes
S3 Batch operations	Yes, for initial transfer of existing objects
S3 Replication	Yes, for incremental changes – automatic
Secure FTP	No. Too much effort to write logic
Storage Gateway	No. commonly used for hybrid setup
Snowball	No. Used for offline transfer
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

Scenario 4 – Solution

Service	Possible Solution
DataSync	Possible for initial transfer of existing objects (server based)
S3 cp, sync	Possible. Too much effort (server based)
S3 Batch operations	Yes, for initial transfer of existing objects (serverless)
S3 Replication	Yes, for incremental changes – automatic (serverless)
Secure FTP	No. Too much effort to write logic
Storage Gateway	No. commonly used for hybrid setup
Snowball	No. Used for offline transfer
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

Scenario 5

An enterprise needs to replicate the content of the S3 bucket to buckets in multiple regions. The application reads from the closest bucket

```
Source: S3 bucket
Destination: Multiple S3 buckets
```

Scenario 5 – Solution

Service	Possible Solution
DataSync	Yes for initial transfer and incremental transfer. Automatically tracks changes (server based)
S3 cp, sync	Possible. Too much effort (server based)
S3 Batch operations	Yes, for initial transfer of existing objects. Incremental transfer requires you to build an inventory of changed object using S3 Events + Lambda
S3 Replication	No. Works only for one source and destination
Secure FTP	No. Too much effort to write logic
Storage Gateway	No. commonly used for hybrid setup
Snowball	No. Used for offline transfer
S3 Transfer acceleration	No. Used for global users
CloudFront	No. Used for global users

Scenario 6

Globally distributed customers routinely upload files to your S3 bucket

```
Source: Customer location (global)
Destination: S3 bucket
```

Scenario 6 – Solution

Service	Possible Solution
DataSync	No
S3 cp, sync	Yes with transfer acceleration
S3 Batch operations	No
S3 Replication	No
Secure FTP	Possible
Storage Gateway	No. commonly used for hybrid setup
Snowball	No. Used for offline transfer
S3 Transfer acceleration	Yes, uses edge location (for routine transfer of files greater than 1 GB)
CloudFront	Yes, uses edge location (for routine transfer of files less than 1 GB). PUT/POST

Summary

Type	Source	Destination	Online/Offline	Limits	Use-case	Frequency
DataSync	NFS,SMB,S3	NFS,SMB,S3	Online	Network Bandwidth	Online data transfer service. agent (server based) for one time migration and on-going incremental data transfer, built-in bandwidth throttle. S3 destination can be any storage class including glacier and deep archive. Flexible source and destination.	On-demand or scheduled
S3 cp, sync	NFS,SMB,S3	S3	Online	Network Bandwidth	Onetime backup or on going backup	On-demand or scheduled
S3 Batch Operations	S3	S3	Online	No limit. Uses AWS network	Region to region copy, bulk processing of objects and invoke Lambda on existing objects	On-demand or scheduled
Secure FTP	FTP Client	S3	Online	Network Bandwidth	Client requires support for FTP based transfer	On-demand or scheduled
Storage Gateway	Files, Blocks, Tapes	S3	Online	Network Bandwidth	Hybrid storage, continous, minimizes or eliminates need for custom storage on-premises	Automatic
SnowCone	On-premises	S3	Offline, Online	up to 8 TB per appliance	Offline data migration to cloud, tactical edge compute, bandwidth constrained or transferring data from remote locations	Physically ship appliance
Snowball	On-premises	S3	Offline, Online	upto 80 TB per appliance	Offline data migration to cloud, tactical edge compute, bandwidth constrained or transferring data from remote locations	Physically ship appliance
Snowmobile	On-premises	S3	Offline	upto 100 PB per container	Offline data migration to cloud, bandwidth constrained or transferring data from remote locations	Physically ship appliance
S3 Transfer Acceleration	Global	S3	Online	Multi-GB to Multi-TB	Global users uploading to central S3	Client initiated
CloudFront PUT/POST	Global	S3	Online	< 1 GB	Global users uploading to central S3	Client initiated
Directconnect	On-premises	AWS region	Online		Private network connection, dedicated bandwidth, low latency	
VPN	On-premises	AWS region	Online		Secure channel either over the internet or inside direct connect	

Chandra Lingam

57,000+ Students

For AWS self-paced video courses, visit:

https://www.cloudwavetraining.com/

Cloud Wave LLC