

Azure and Integration Services

Integrating Data In the Cloud

Stacia Misner
blog.datainspirations.com
[@StaciaMisner](https://twitter.com/StaciaMisner)

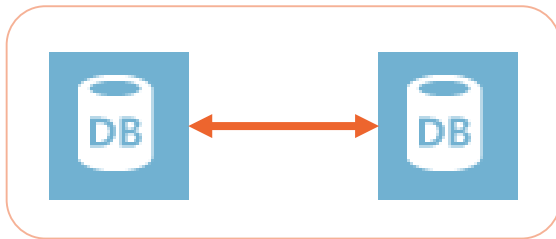


pluralsight 
hardcore dev and IT training

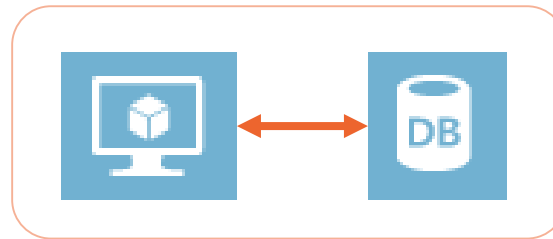
Overview

- **Architecture Options**
- **ETL Strategies**
- **Additional Considerations for Cloud Integration**

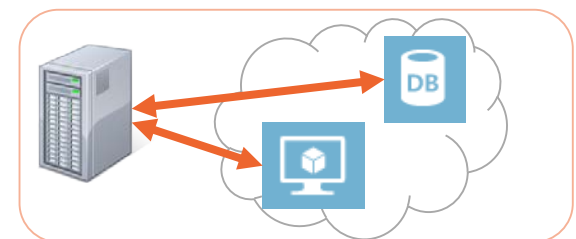
Architecture Options



SQL Databases

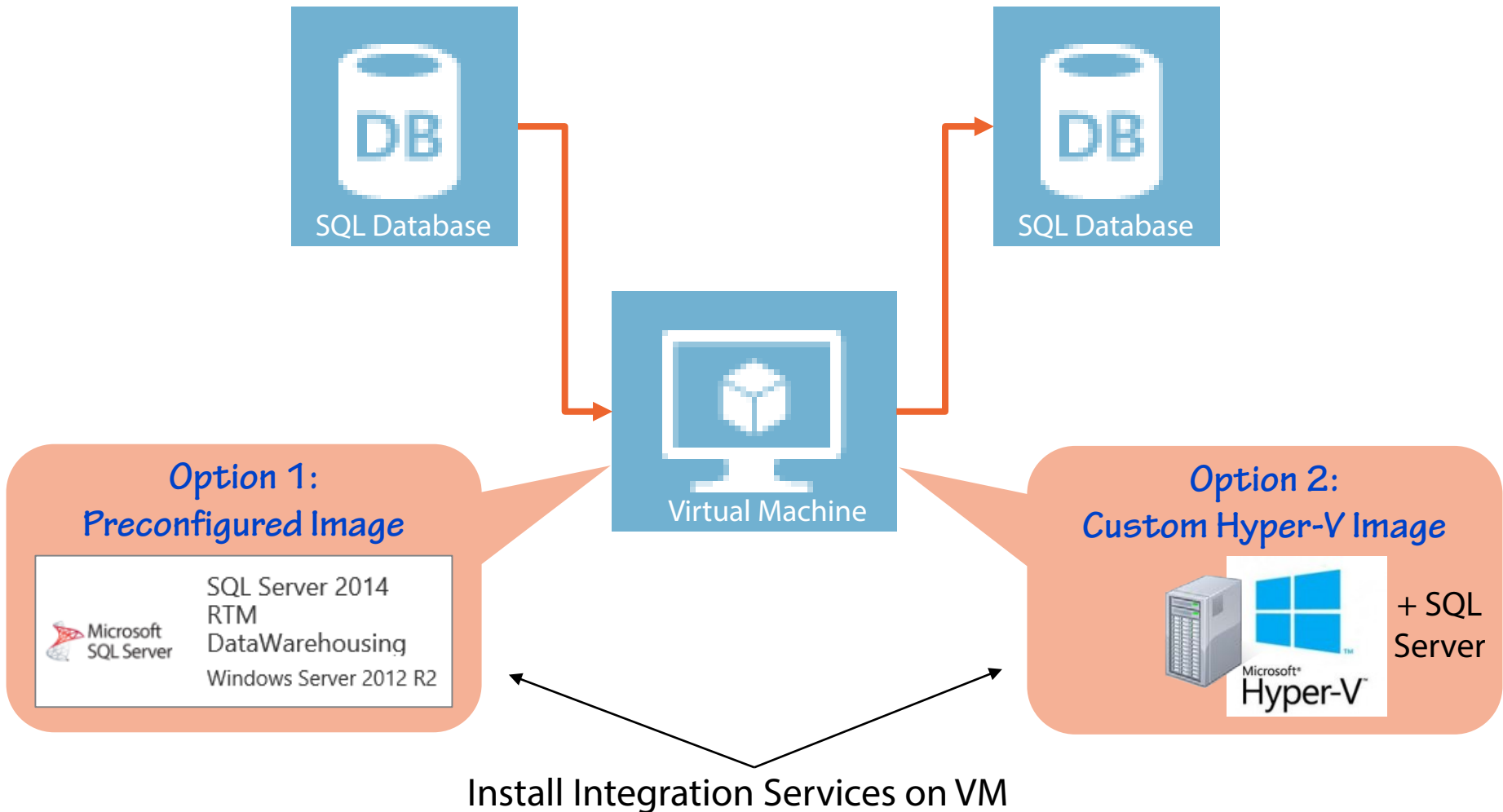


Azure Virtual Machine
and SQL Database

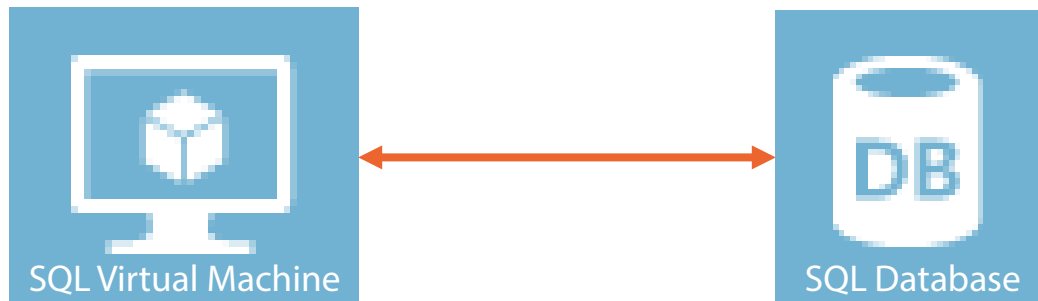


Hybrid:
On-Premises and
Azure

SQL Databases



Azure Virtual Machine and SQL Database



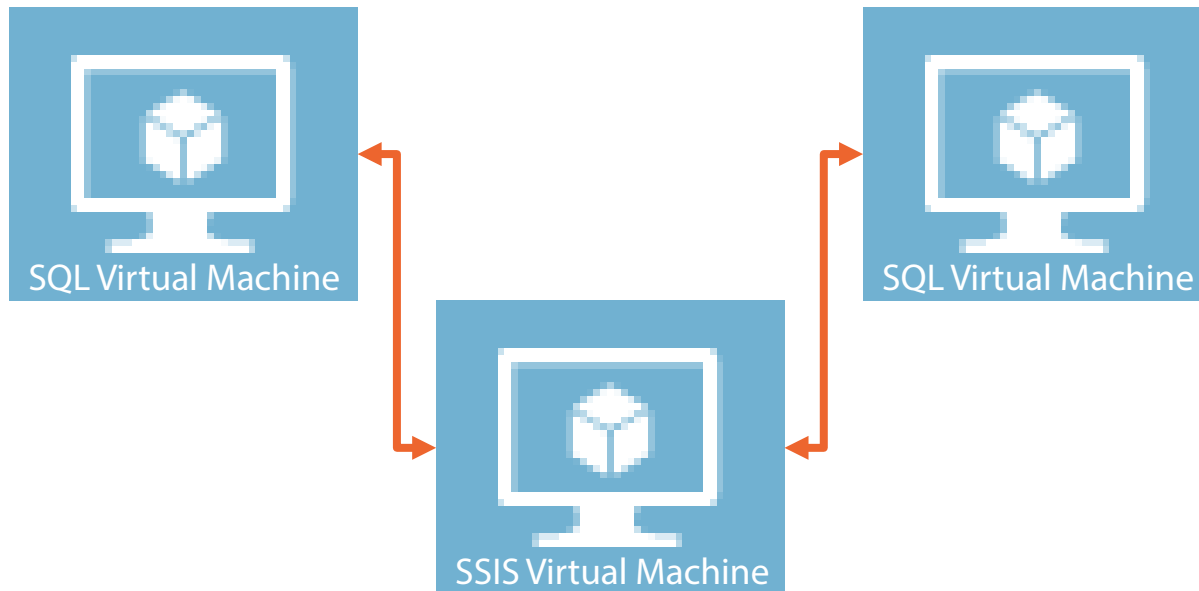
- Add Integration Services at minimum
- Optionally enable catalog for monitoring

Azure Virtual Machines



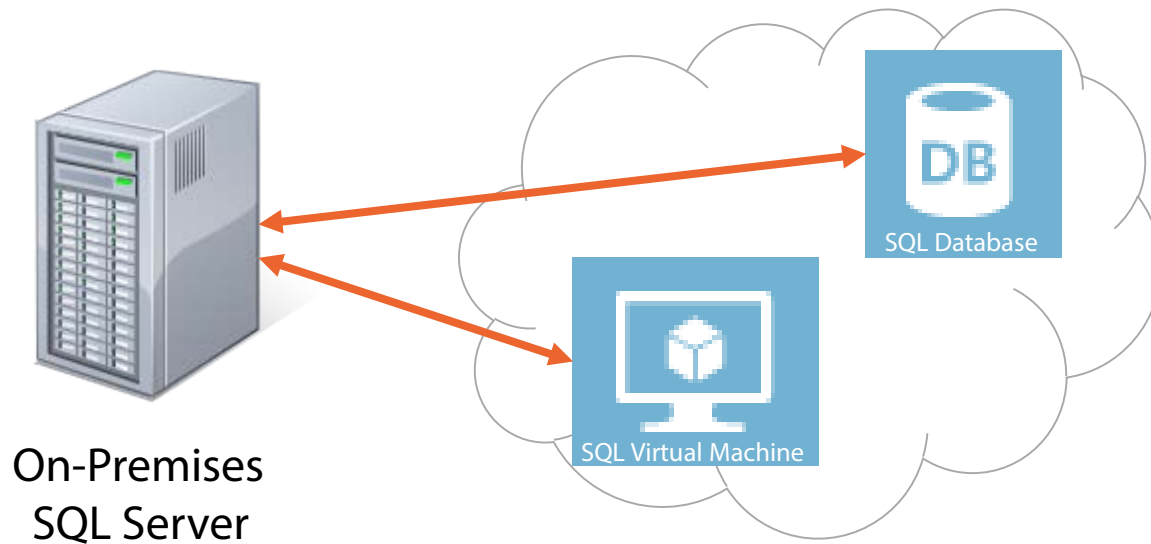
Run Integration Services on the virtual machine having more available resources

Azure Virtual Machines

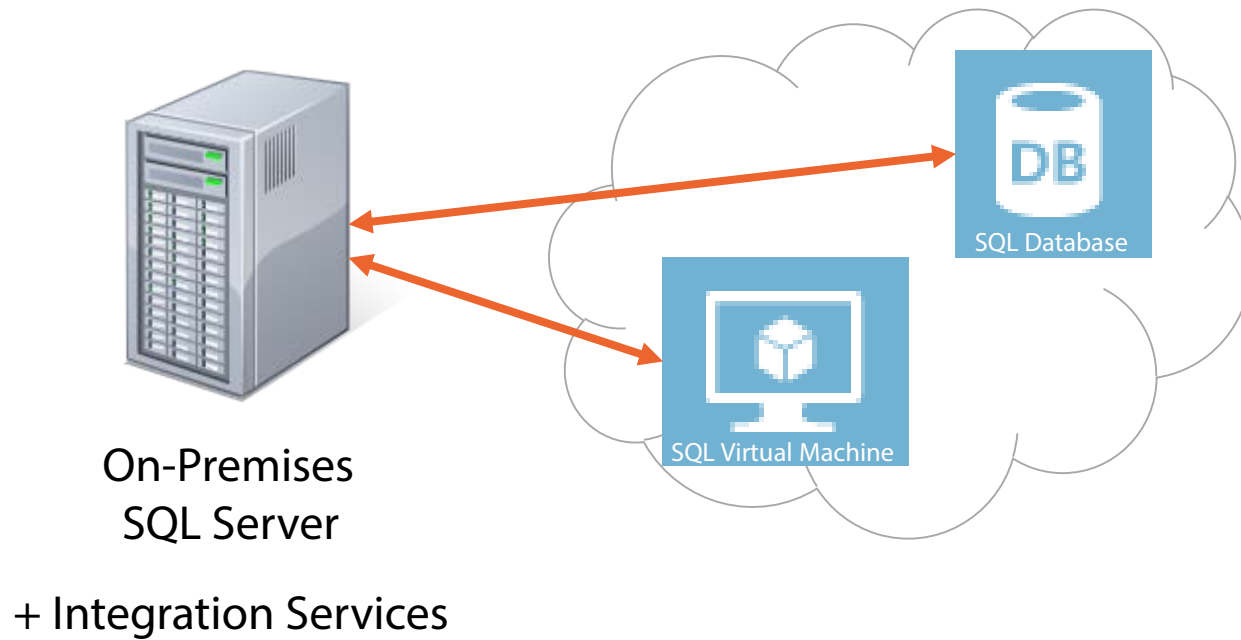


Run Integration Services on dedicated virtual machine

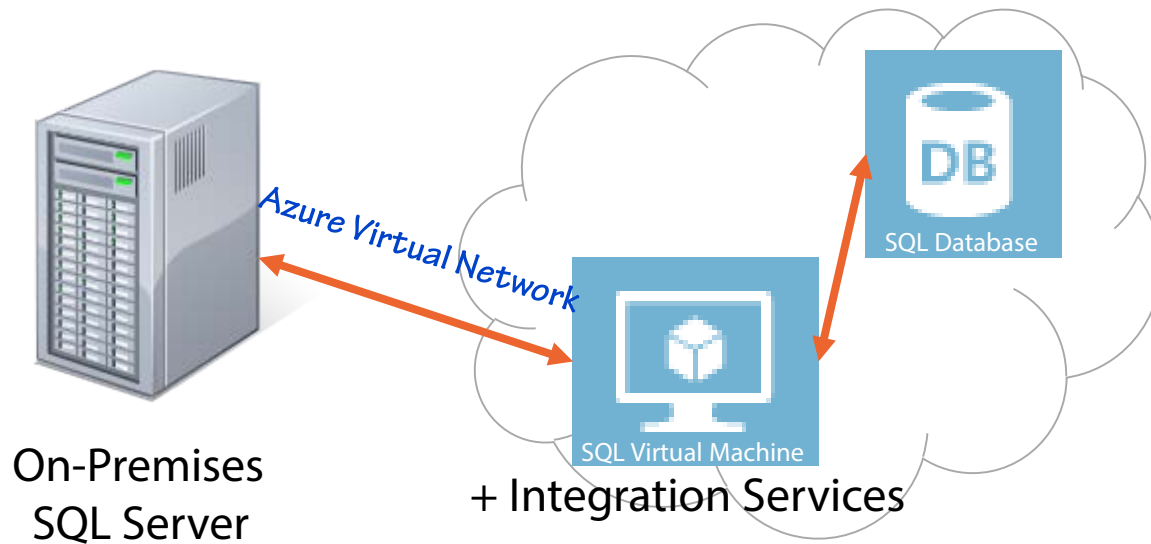
Hybrid Environment



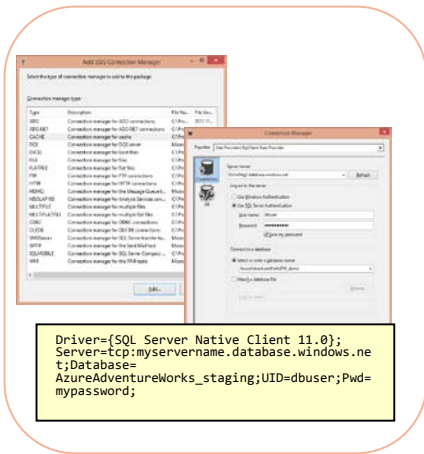
Hybrid Environment



Hybrid Environment



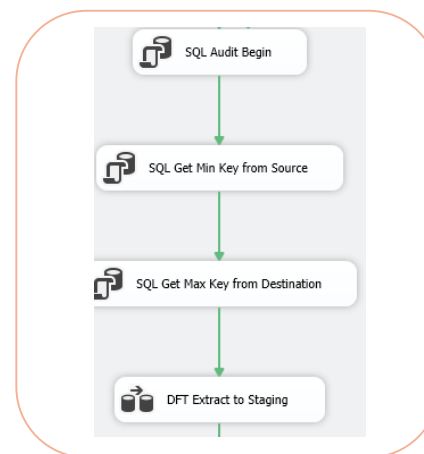
ETL Strategies



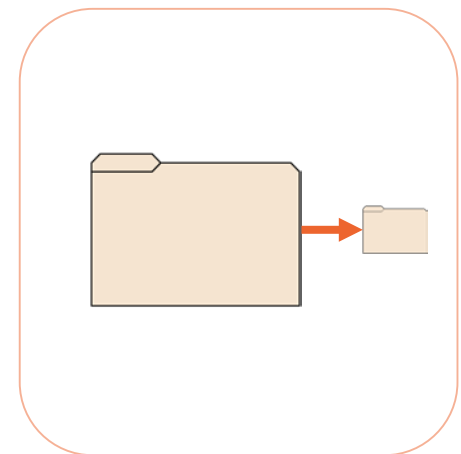
Data Connections



ETL Design Patterns
for the Cloud



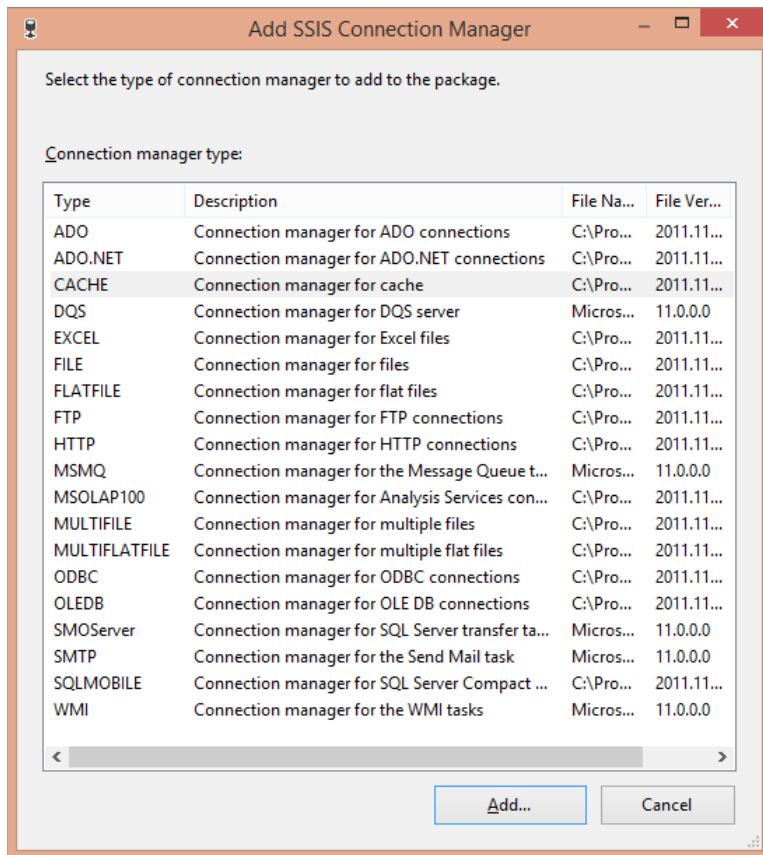
Data Subsets



File Compression

Data Connections

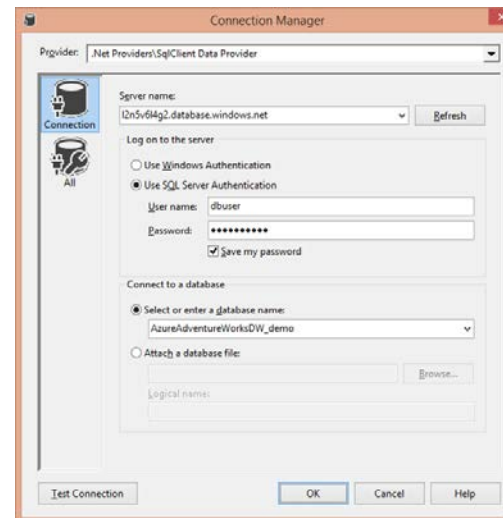
Virtual Machine or On-Premises



SQL Database

```
Driver={SQL Server Native Client 11.0};  
Server=tcp:myservername.database.windows.net;  
Database=AzureAdventureWorks_staging;  
UID=dbuser;  
Pwd=mypassword;
```

ODBC



ADO.NET

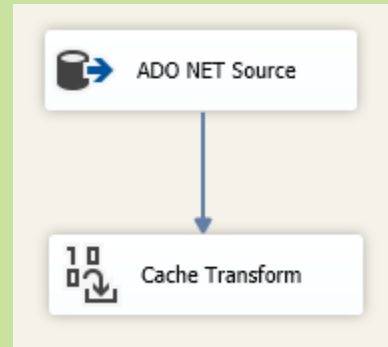
ETL Design Patterns for the Cloud

```
SELECT AuditKey from [dw].[DimAudit]
where TableName = @tableName and
ExecStartDT = @execStartDate
```

Use parameter names rather than ? for ADO.NET

Variable Name	Direction	Data Type	Parameter Name	Param
User::TableName	Input	String	@tableName	-1
System::PackageName	Input	String	@pkgName	-1

Execute SQL Task



Lookup

fx (project) Azure SQL DB DW

Project.params [Design] ➡ ✕

Name	Data type	Value	Sensitive	Required
AzureSQLBDW_ConnectionString	String	*****	True	True

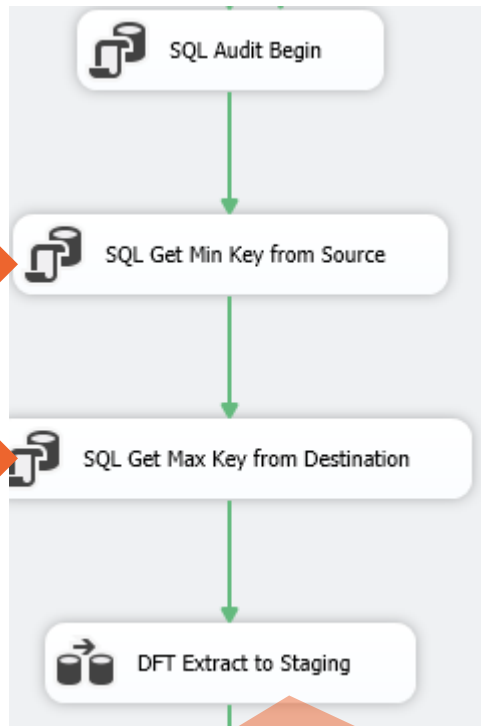
Sensitive Data

```
MERGE dw.DimProductCategory AS target
USING (SELECT
    Name,
    rowguid,
    ModifiedDate,
    ProductCategoryID,
    AuditKey
    FROM tmp.scdProductCategory) AS
SOURCE
```

MERGE

Use ? for ODBC query but start parameter mapping with 1

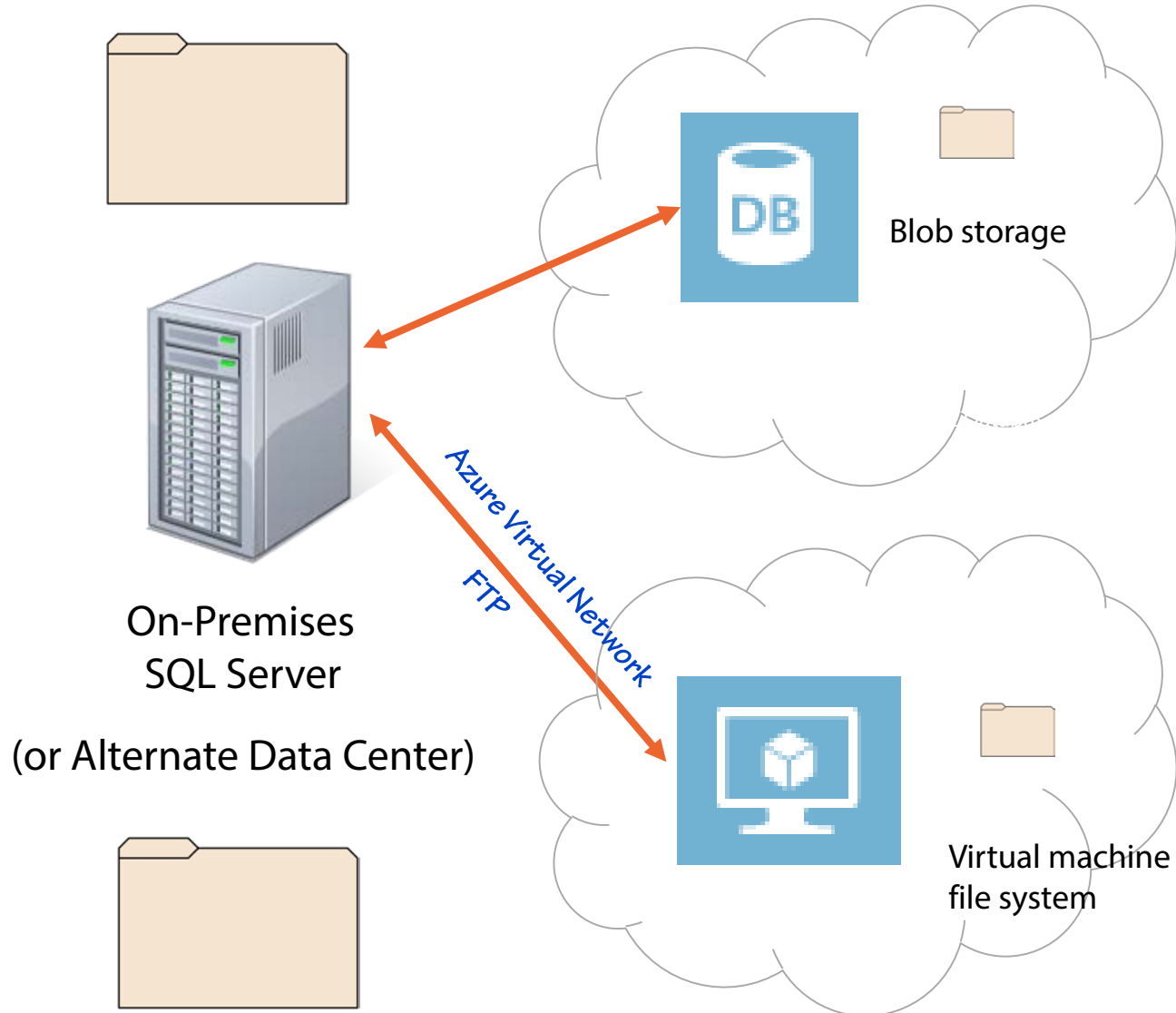
Data Subsets



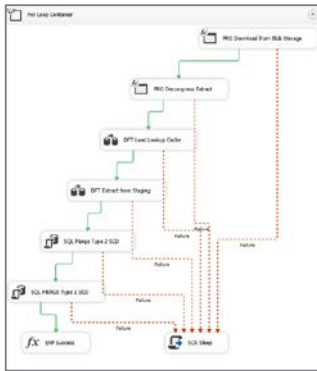
```
select  
IsNull(Max(ProductAlternateKey),  
@minKey) from dw.DimProduct
```

```
"SELECT  
[ProductID],[Name],[ProductNumber],[MakeFlag],[FinishedGoodsFlag]  
. . . FROM [Production].[Product] WHERE [ProductNumber] > '"' +  
@[User::MaxKey] + "'" ORDER by [ProductNumber]"
```

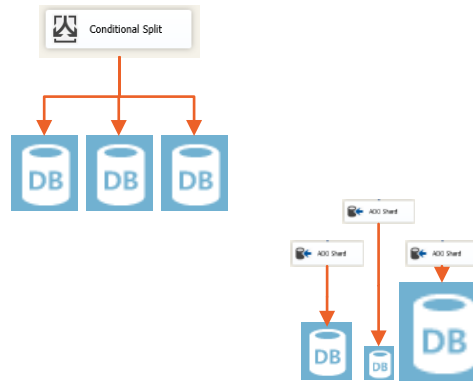
File Compression



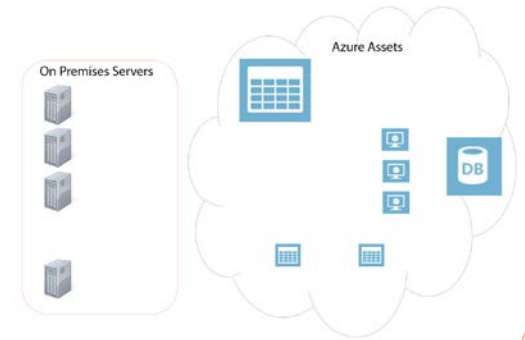
Additional Considerations for Cloud Integration



Planning for Failure



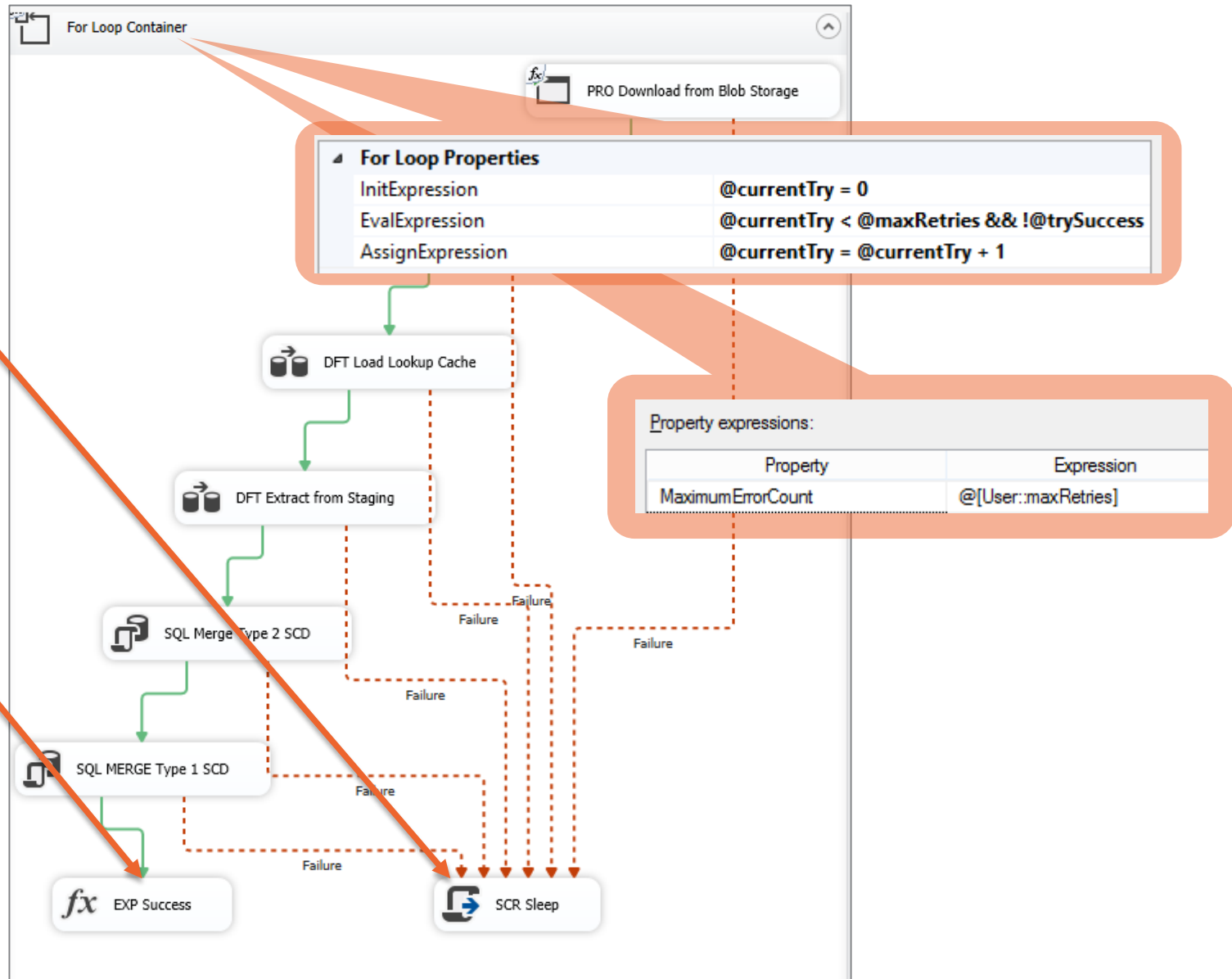
Sharding



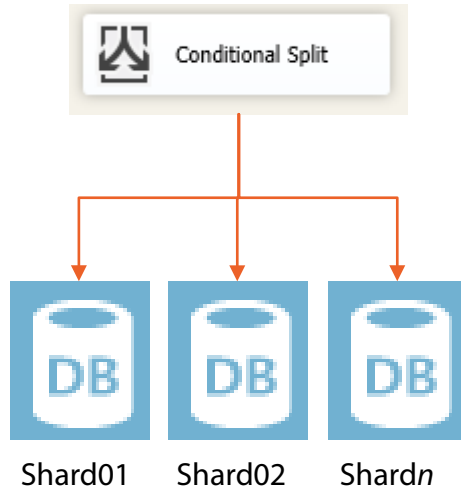
Scaling Processes

Planning for Failure

1. Define variables
2. Add FOR LOOP container
3. Add Script Task to sleep for short time
4. Add Expression task to force break out of loop on success
5. Add expression to control maximum error count



Sharding

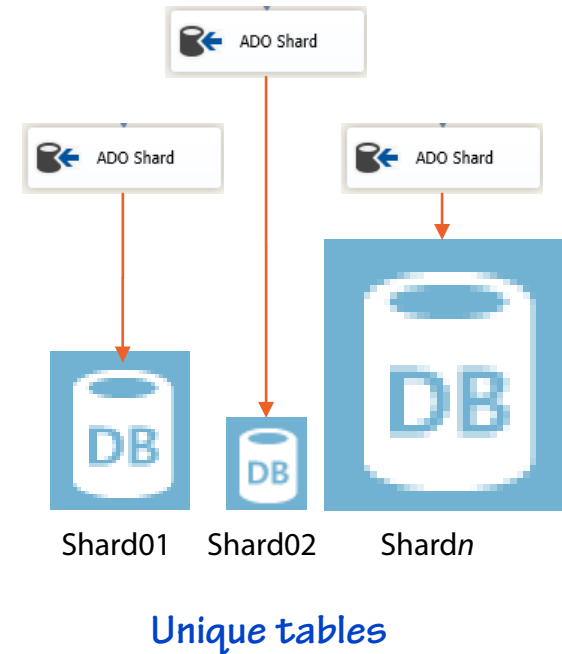


Identical tables
Unique data in each table

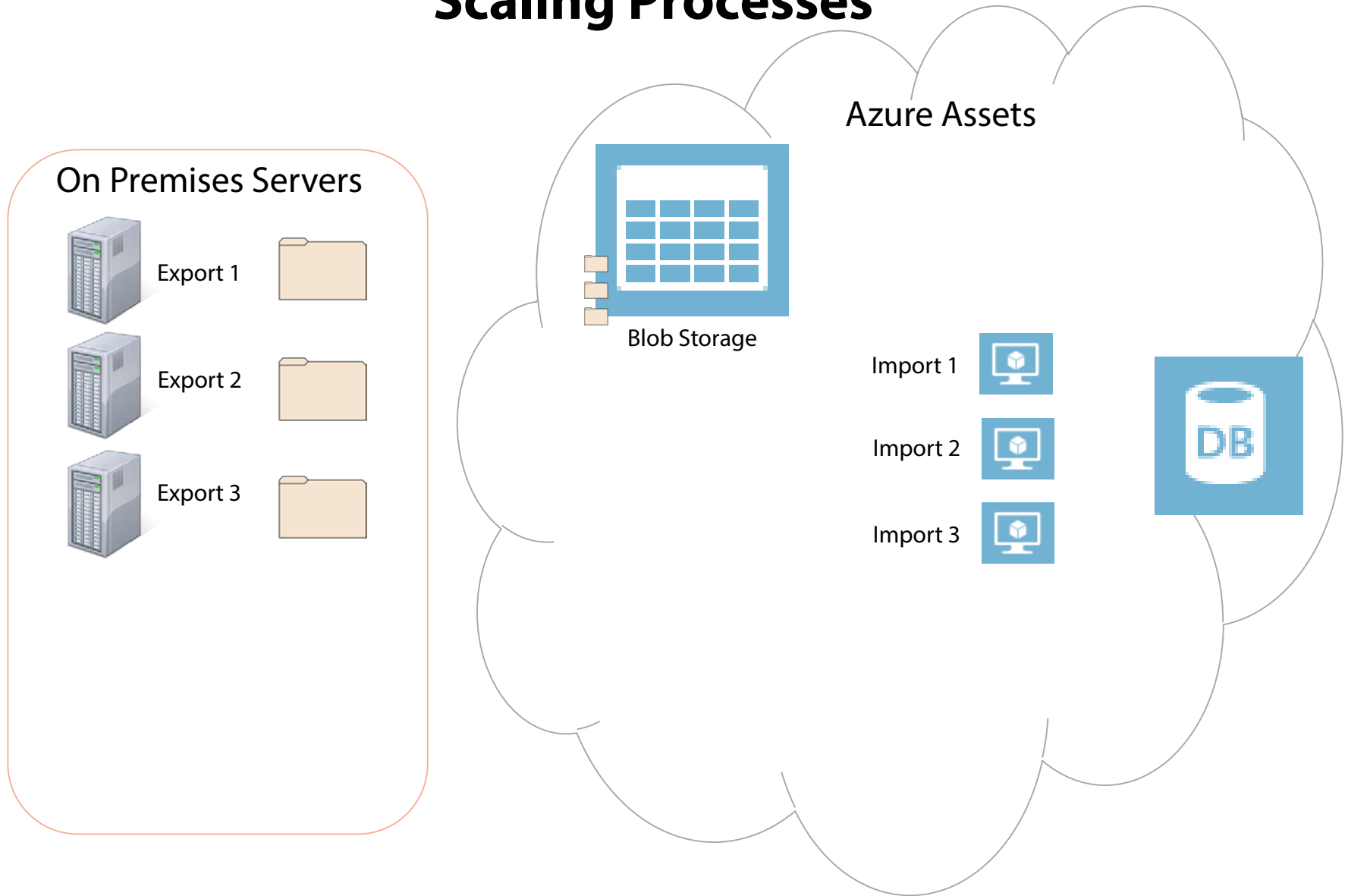
- Manage capacity with *horizontal sharding*
- Use Conditional Split in ETL

Sharding

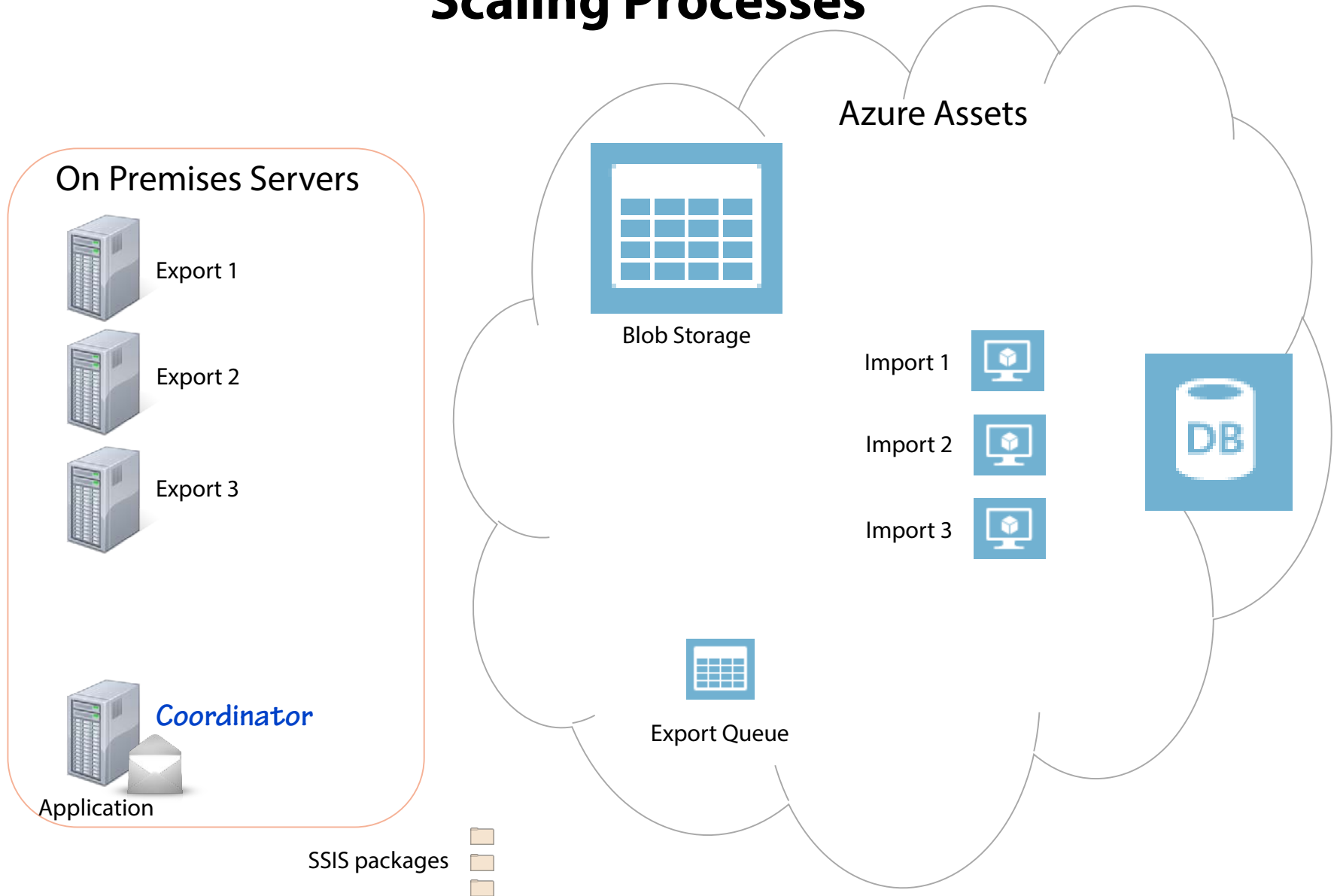
- Manage performance with *vertical sharding*
- Configure Destination in ETL



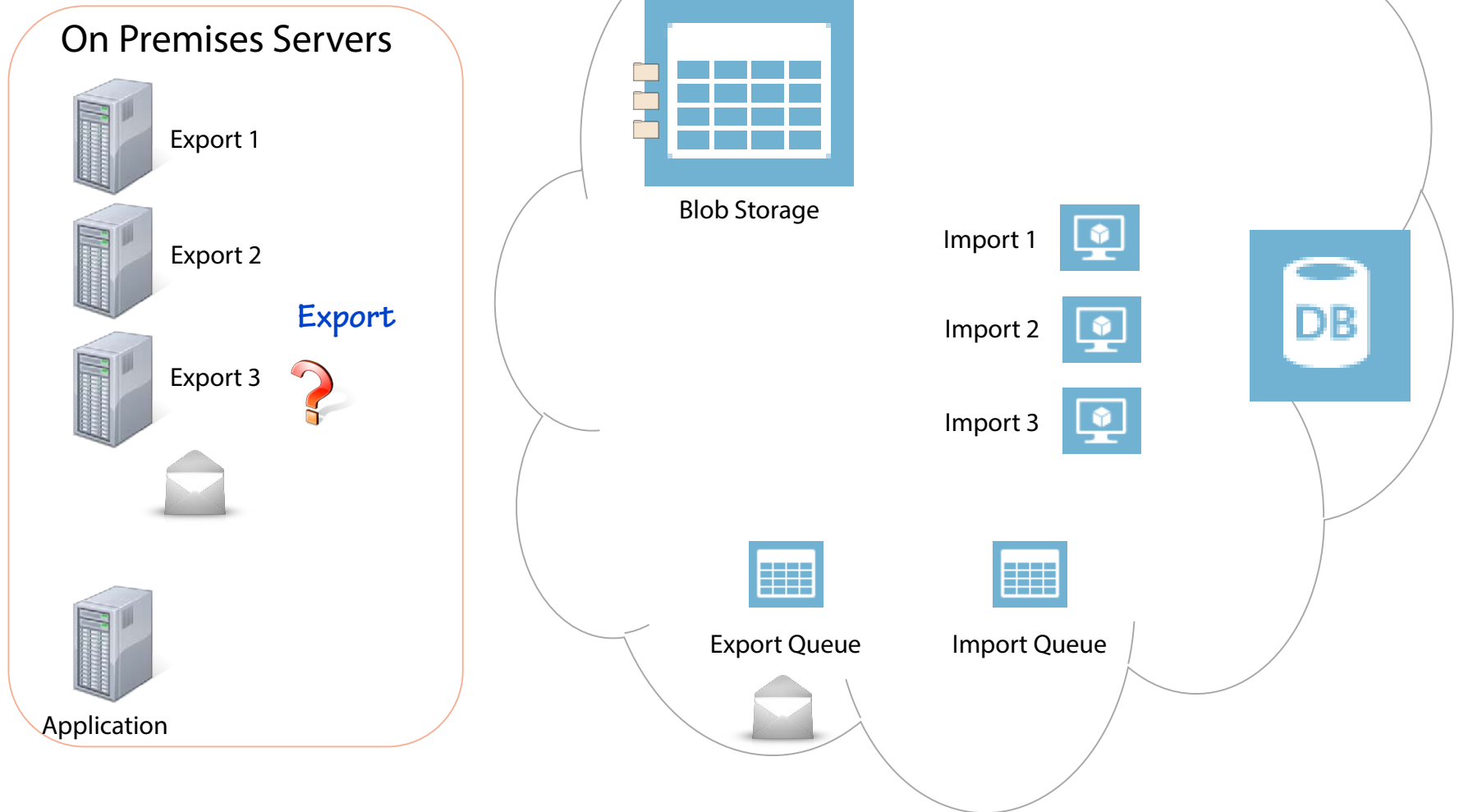
Scaling Processes



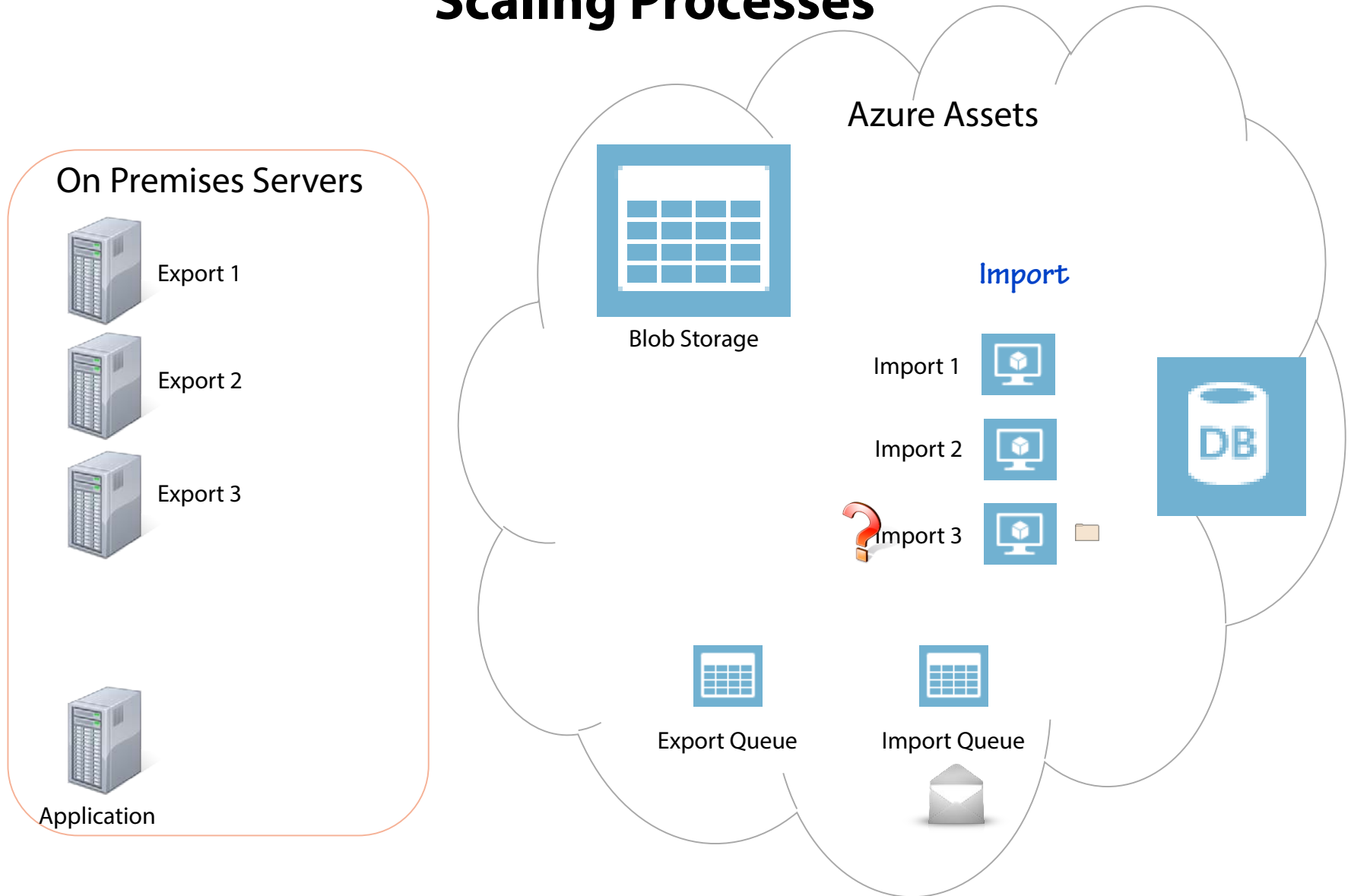
Scaling Processes



Scaling Processes



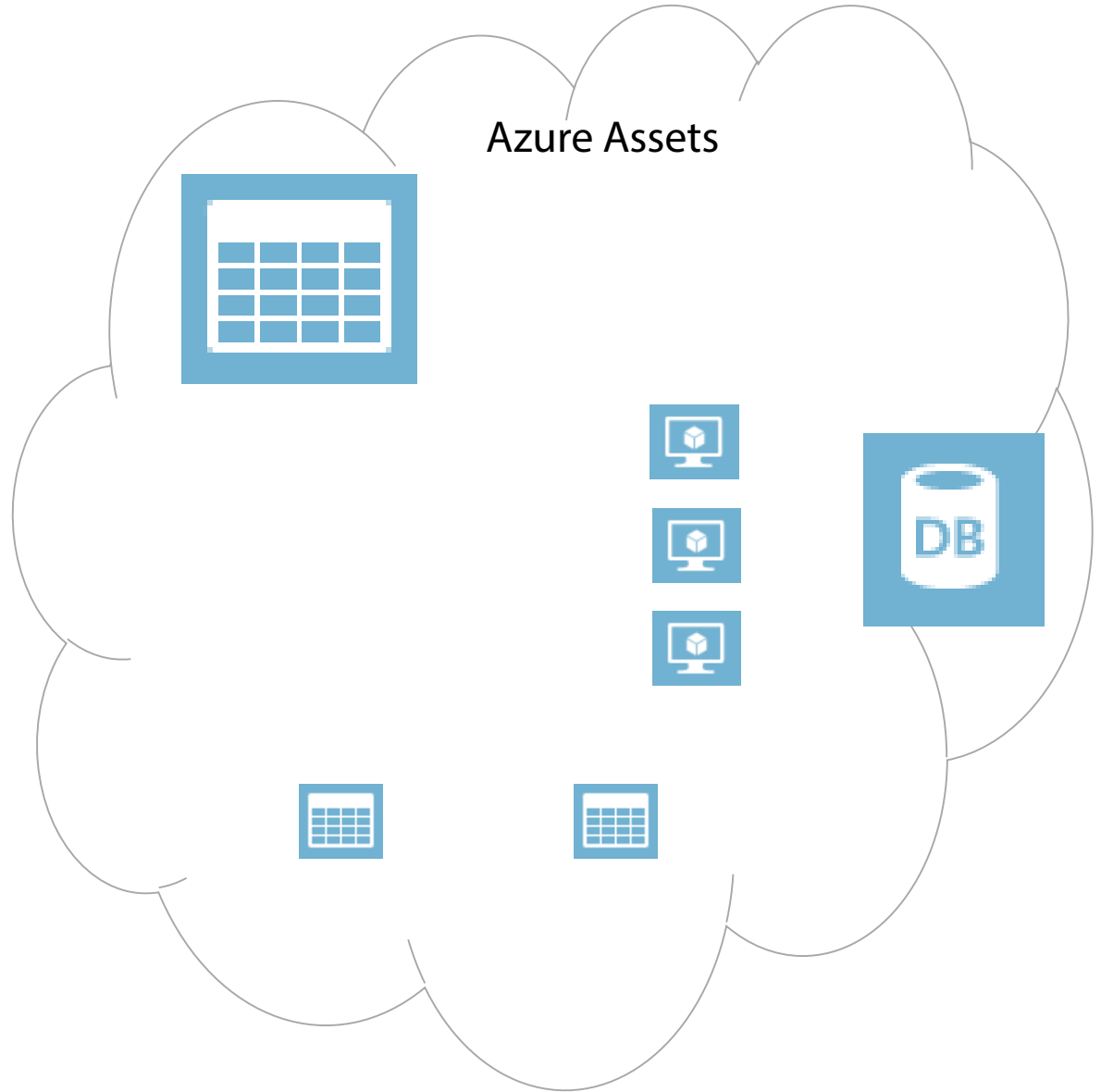
Scaling Processes



On Premises Servers



Azure Assets



Summary

- **Architecture Options**

- SQL Databases, Azure Virtual Machines, Hybrid

- **ETL Strategies**

- ADO.NET or ODBC data connections
- Revised ETL design patterns (parameters, lookups, MERGE, sensitive data)
- Incremental loads using data subsets
- File compression

- **Additional Considerations for Cloud Integration**

- Retry logic with FOR LOOP container
- Performance or capacity management with sharding
- Scale out with custom code to manage parallel processes

Resources

- **Pluralsight Courses**

- Windows Azure: The Big Picture – <http://tinyurl.com/kn6yyl2>
- Introduction to Windows Azure - <http://tinyurl.com/n8lq32n>
- SQL Azure - <http://tinyurl.com/n5t95y2>

- **Getting Started with Microsoft Azure SQL Database**

- <http://tinyurl.com/p9x3wgl>

- **Getting Started with SQL Server in Azure Virtual Machines**

- <http://tinyurl.com/qaeohc3>

- **SSIS Operational and Tuning Guide**

- <http://tinyurl.com/l2ndown>

- **SSIS for Azure and Hybrid Data Movement**

- <http://tinyurl.com/mlqnnj4>

Resources

- **Point-to-Site Certificate Creation**
 - <http://tinyurl.com/mkn2ee7>