What Makes SSIS Tick?
A Look at Internals and Performance

By Ravi Kumar
Lead Business Intelligence Developer
@SQLRavi
Kumar.ravi3@gmail.com

Why Present at a users group?

- Good networking
- Free knowledge
- Geek out

My Presentation Style

- Fun
- Engaging
- Talk about off topic things
- Make sure to Put positives and Negatives in Review.

Agenda

- Why Internals?
- Control Flow
- Data Flow
- Performance Tips
- Tips to make SSIS life easy (If time is left).
- New features in 2016 (if time is left).

Show of Hands on using SSIS!!

Why internals?

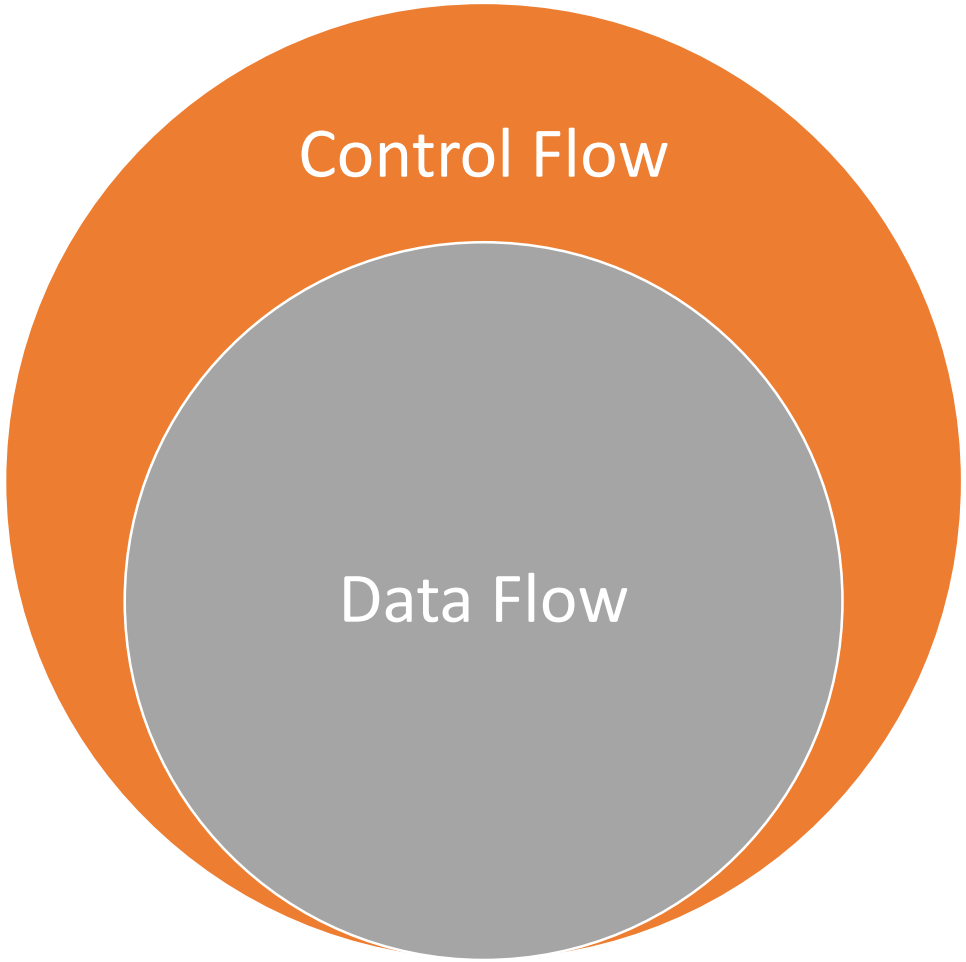
- Troubleshoot SSIS packages
- Build better SSIS packages

What is a package at Code level?

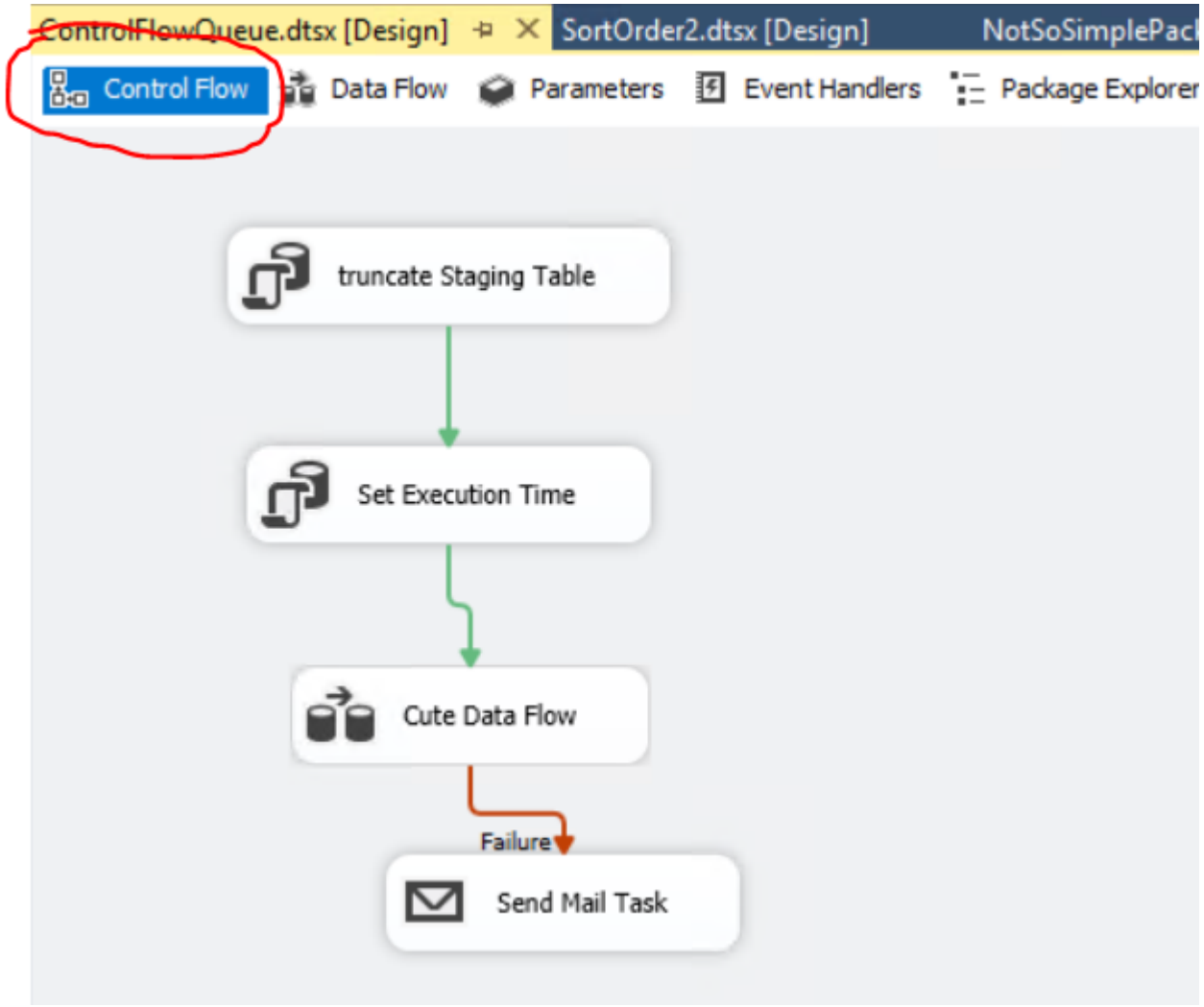
- Basic XML file

```
  DTS:Name="PackageFormatVersion">8</DTS:Property>
<DTS:Variables />
<DTS:Executables>
  <DTS:Executable
    DTS:refId="Package\Execute 1-Get_and_Load_Files"
    DTS:CreationName="Microsoft.ExecutePackageTask"
    DTS:DelayValidation="True"
    DTS:Description="Execute Package Task"
    DTS:DTSID="{806E1B6B-59FC-4D15-AF8A-D2C5B8412AAF}"
    DTS:ExecutableType="Microsoft.ExecutePackageTask"
    DTS:LocaleID="-1"
    DTS:ObjectName="Execute 1-Get_and_Load_Files"
    DTS:TaskContact="Microsoft Corporation; Microsoft SQL Server; Microsoft Corpor
    <DTS:Variables />
    <DTS:ObjectData>
      <ExecutePackageTask>
        <UseProjectReference>True</UseProjectReference>
        <PackageName>1-Get_and_Load_Files.dtsx</PackageName>
      </ExecutePackageTask>
```

Two Main parts of SSIS Engine

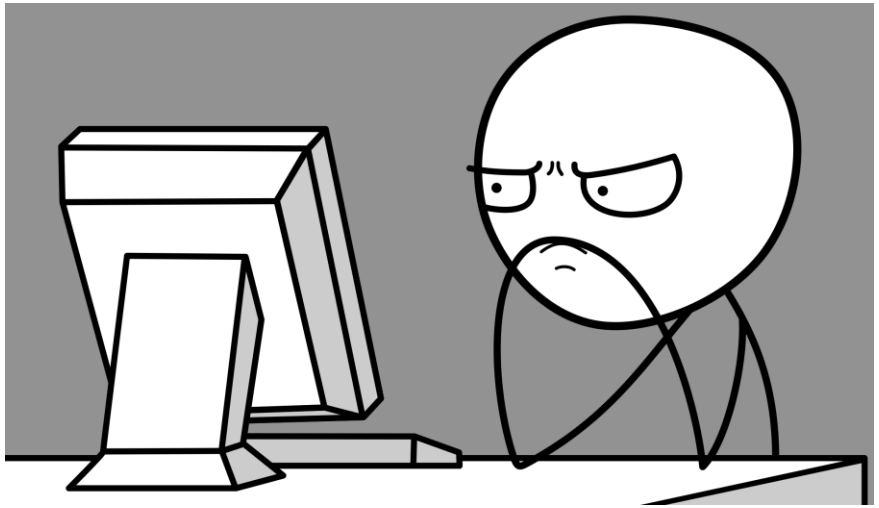


Control Flow



Control Flow

Load → Apply Parameters → Validate → Execute



Control Flow

Load → Apply Parameters → Validate → Execute

1. Read XML
2. Decrypt
3. Check Version
4. Load Objects
5. Apply configuration and expression

Control Flow

Load → **Apply Parameters** → Validate → Execute

- Used if Project Deployment model is used

Control Flow

Load → Apply Parameters → **Validate** → Execute

1. Validate the package in a tree like structure
2. Root > Containers > Sibling Nodes > Task
3. Every container and task validates itself
4. Delay validation comes in play here

Control Flow

Load → Apply Parameters → Validate → **Execute**

- Action time.
- Precedence Constrains come into play

- MaxConcurrentExecutables: Default = -1, cores + 2
  - How many tasks can run at one time

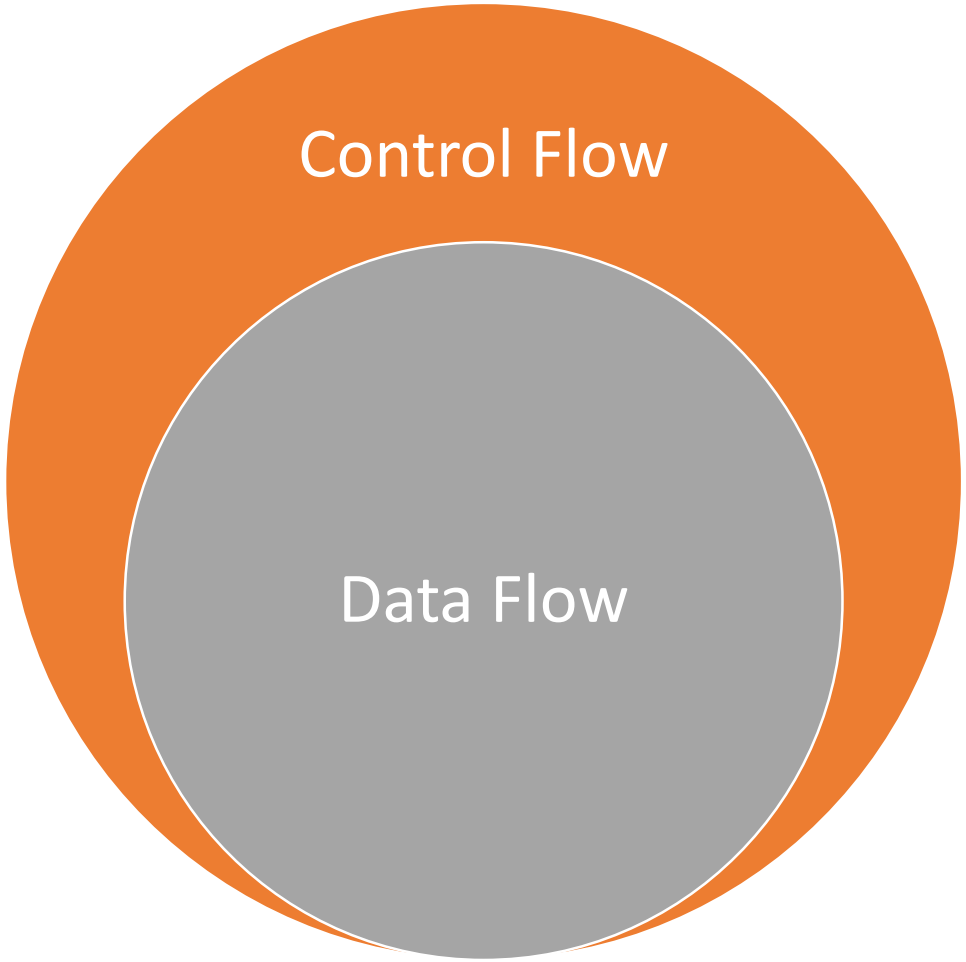
MaxConcurrentExecutables	-1

Questions on Control Flow…

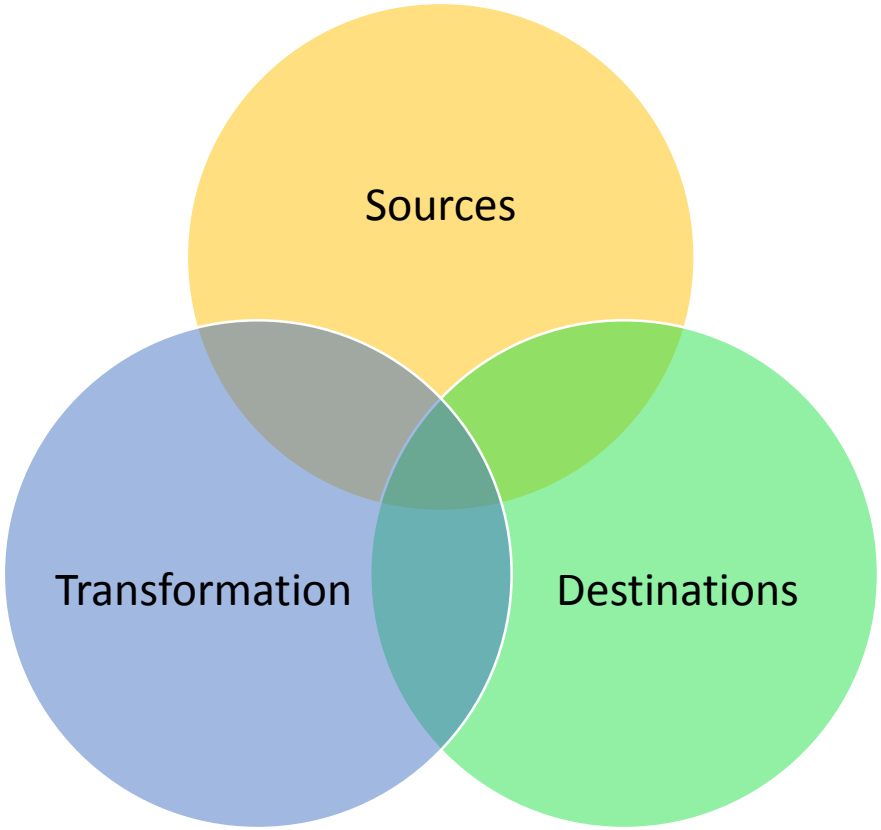
Demo

- A brief look at control flow.

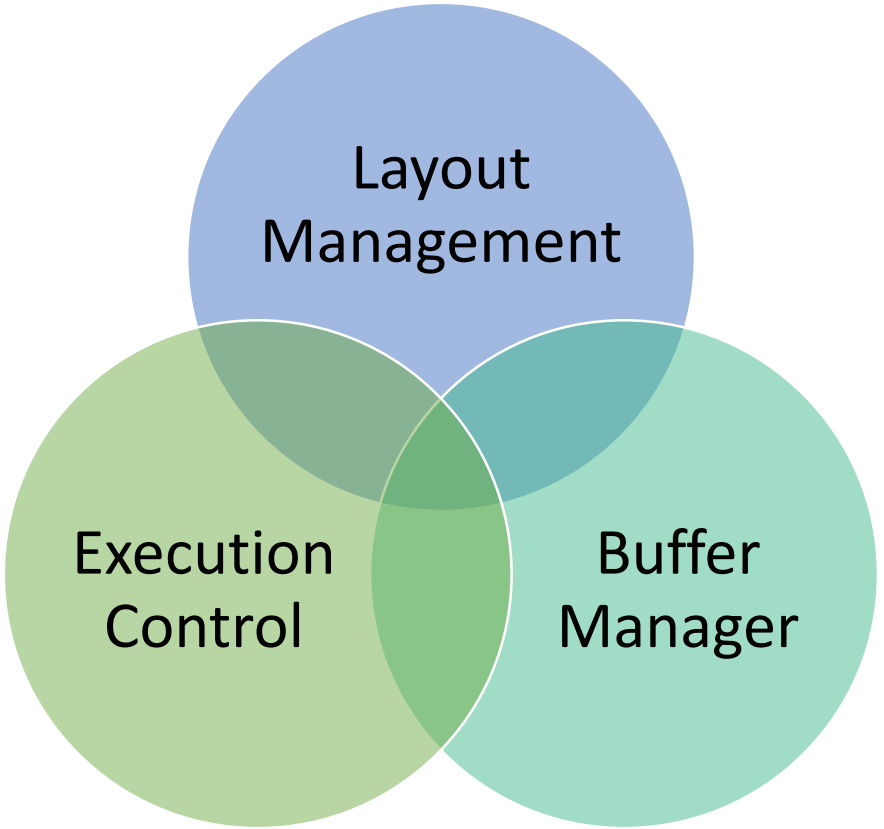
Two Main parts of SSIS Engine



Data Flow

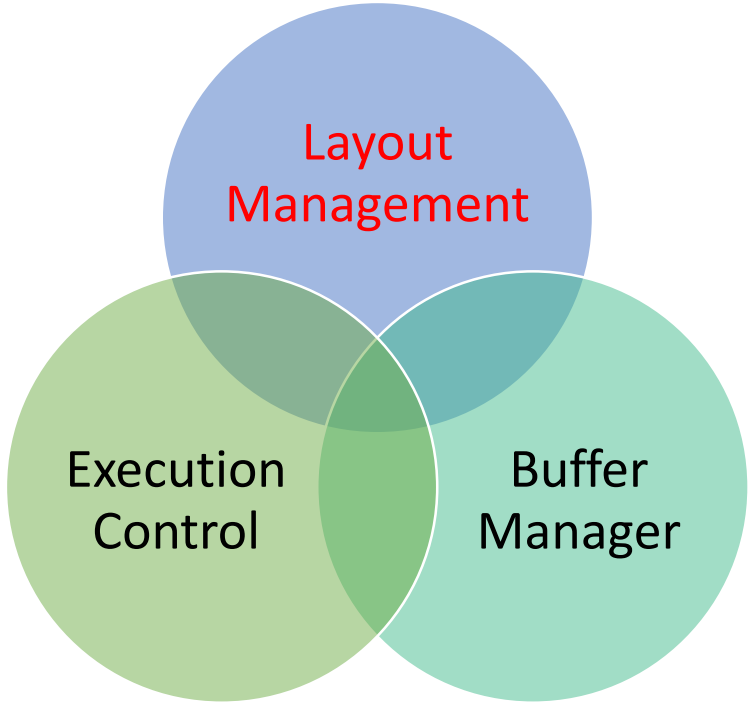


Data Flow Engine



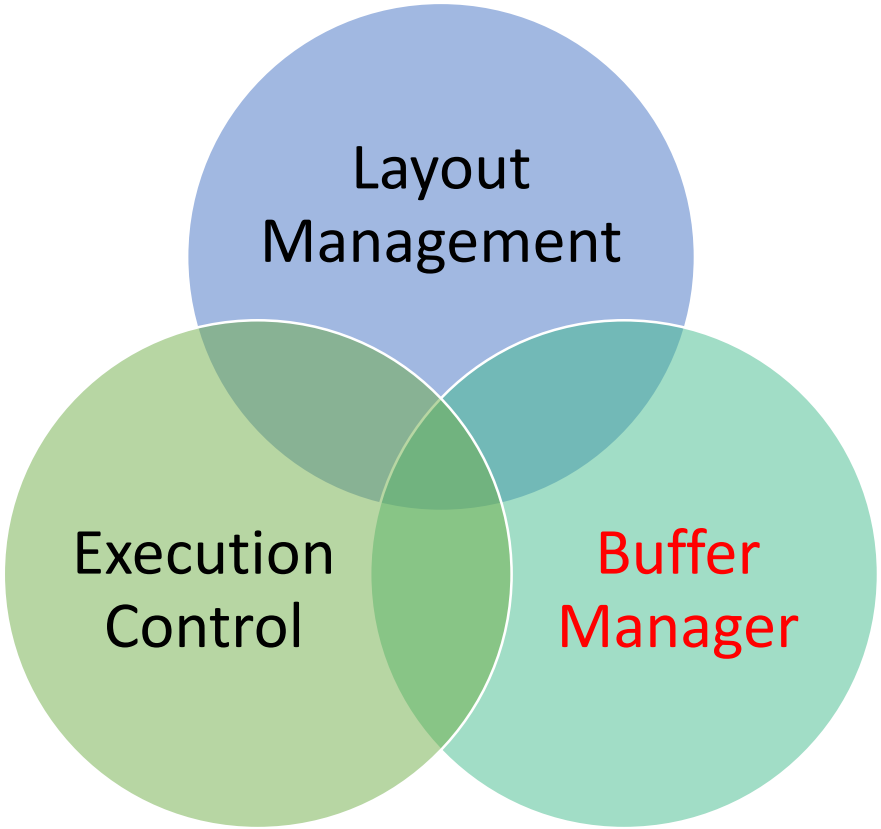
Data Flow Engine: Layout Management

- Relationship between data flow items/components
- It's stored in XML
  - <DTS:ObjectData> subnode



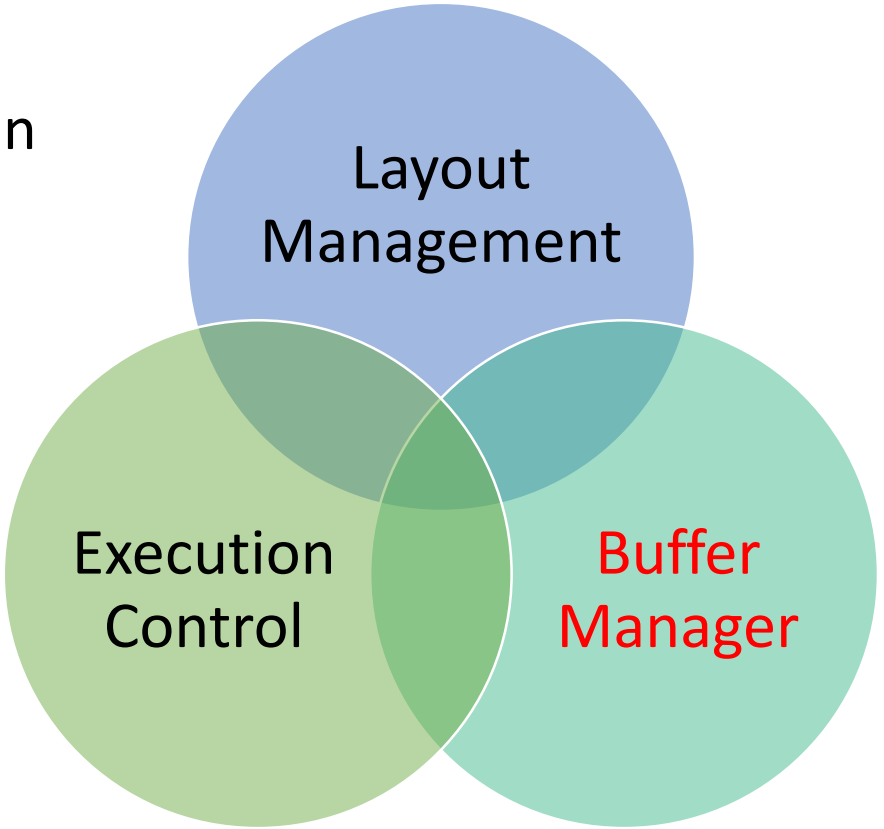
Data Flow engine: Buffers!!

- Types of buffers
  - Physical
  - Virtual
  - Private
  - Flat



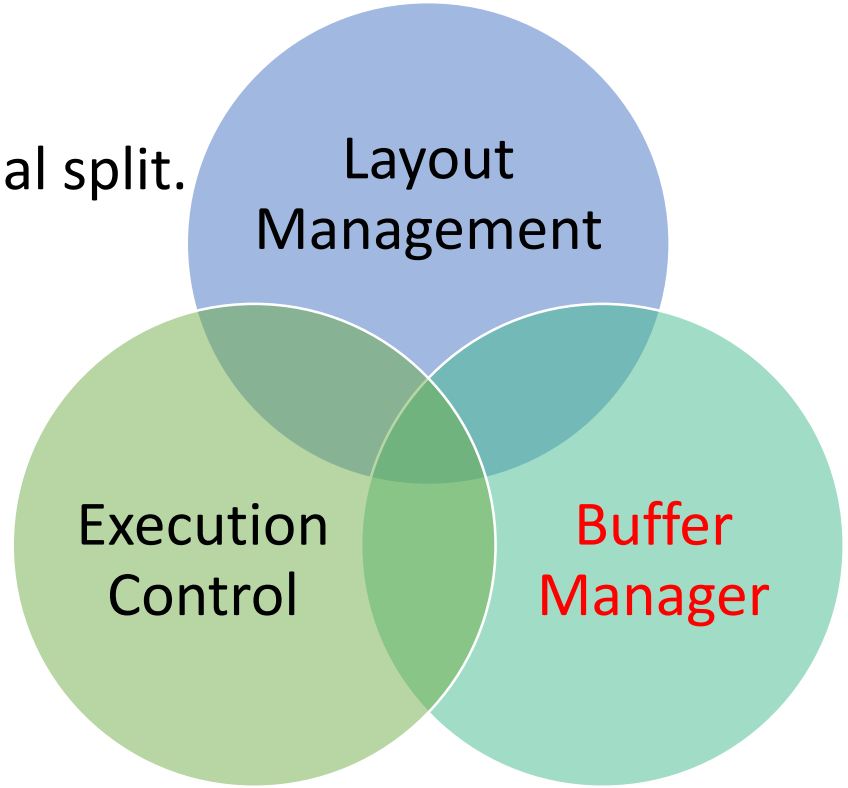
Data Flow engine: Buffers!!

- Physical Buffers:
  - Not same as SQL Server Engine buffers
  - Physical Memory with Rows and column dynamically size
  - Looks like real table
    - Class exercise: draw it out.



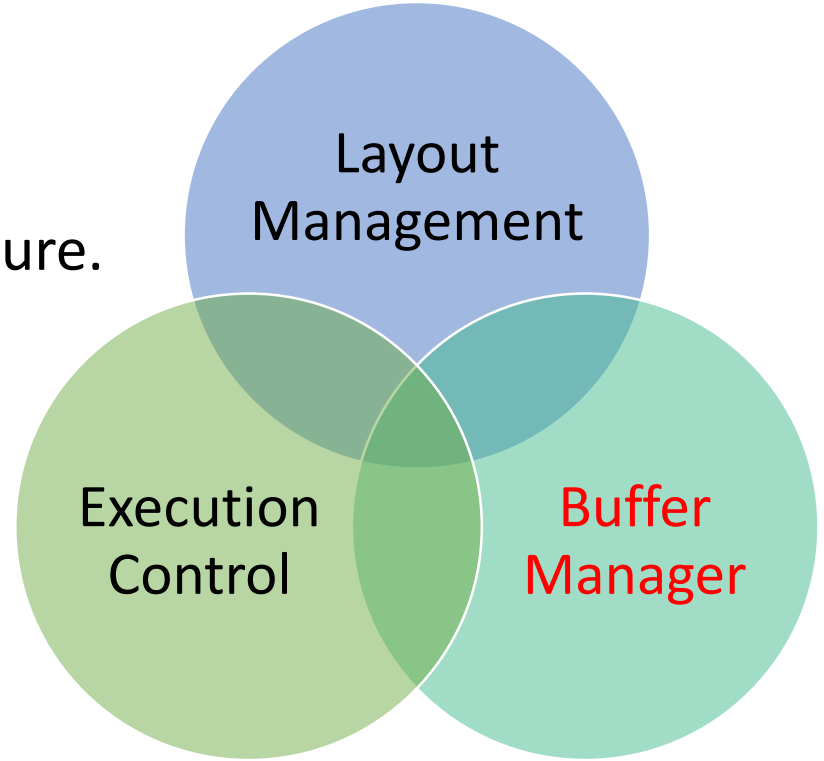
Data Flow engine: Buffers!!

- Virtual Buffers:
  - Subset of physical buffers
  - Class Drawing (Draw it).
  - Examples: Derived column and conditional split.

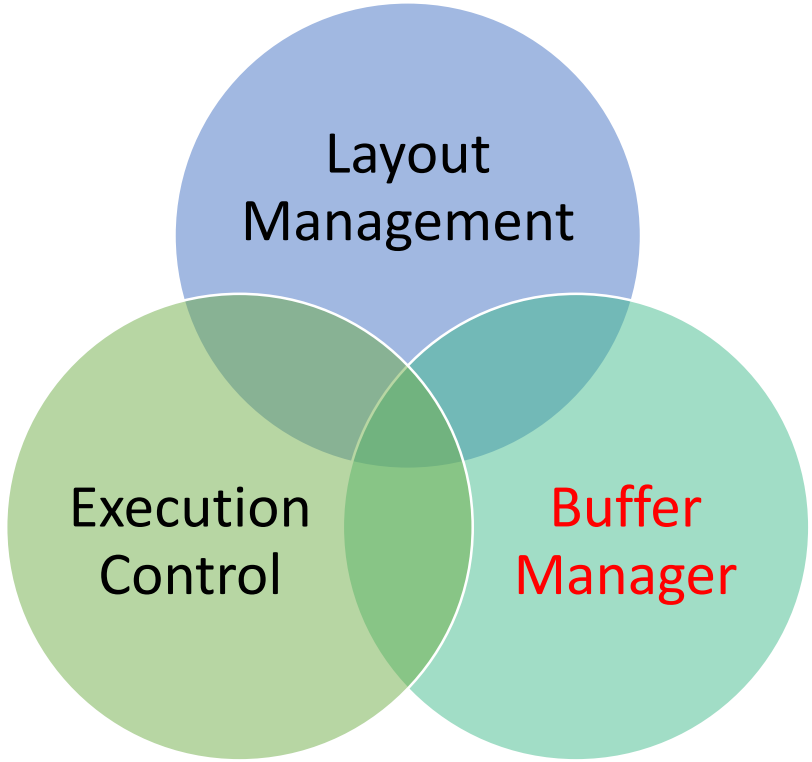


Data Flow engine: Buffers!!

- Private Buffers:
  - Physical structure but owned by a component in Data Flow
  - Example: Sort component.
- Flat Buffers:
  - Physical memory but without any structure.
  - Example: Lookup transformation.



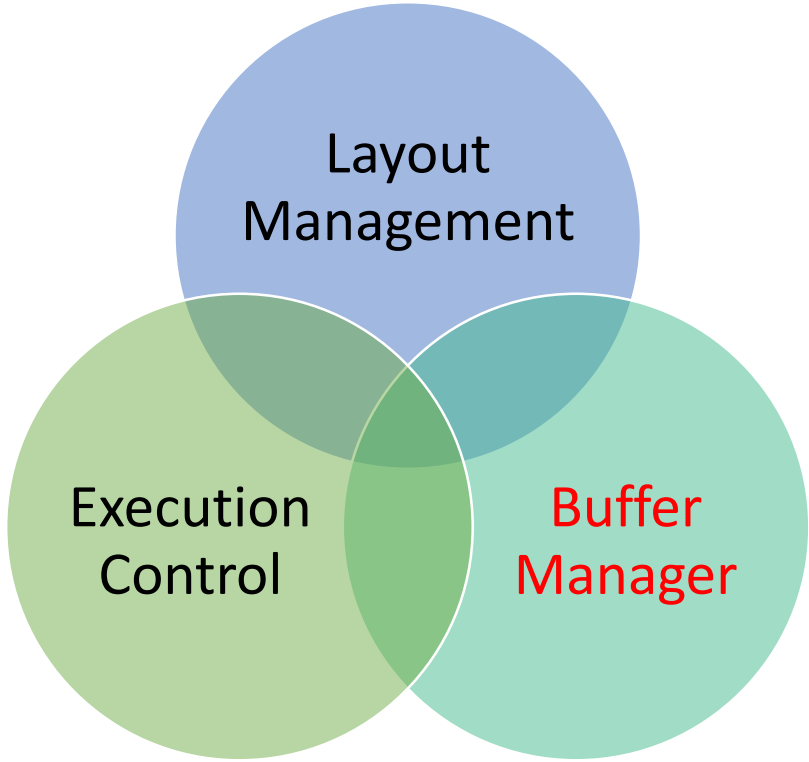
Data Flow engine: Buffers!!



- During ETL, less data is more performance
- Use only columns needed
- Use small/narrow data types
  - Char instead of varchar
  - Varchar instead of nvarchar if possible.
- Try to do data conversion at the source, not add columns
- RunInOptimizedMode – not available in SSDT.

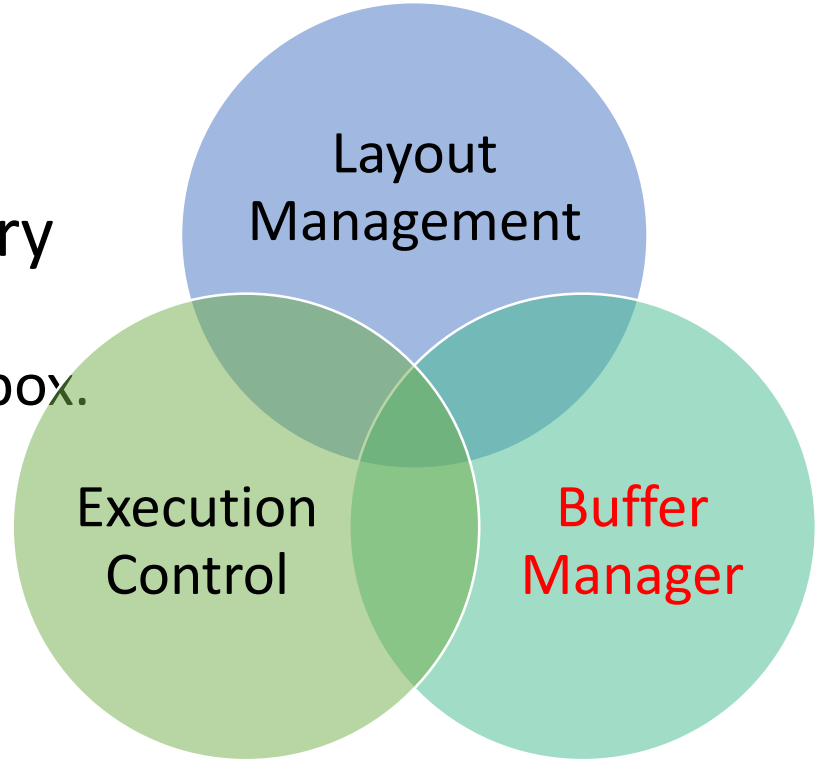
Data Flow engine: Buffers!!

- The fruit example..
- Buffer Sizes
  - DefautBufferSize: Default 10MB, Max 100MB
  - DefaultBufferMaxRows: Default 10000
  - Auto Adjust Buffer Size: New for 2016
  - Calculation
    - RowSize = Data*1024/NoOfRows
    - DefaultBufferMaxRows= DefaultBufferSize(Bytes)/RowSize
  - Turn on BufferSizeTuning Logging

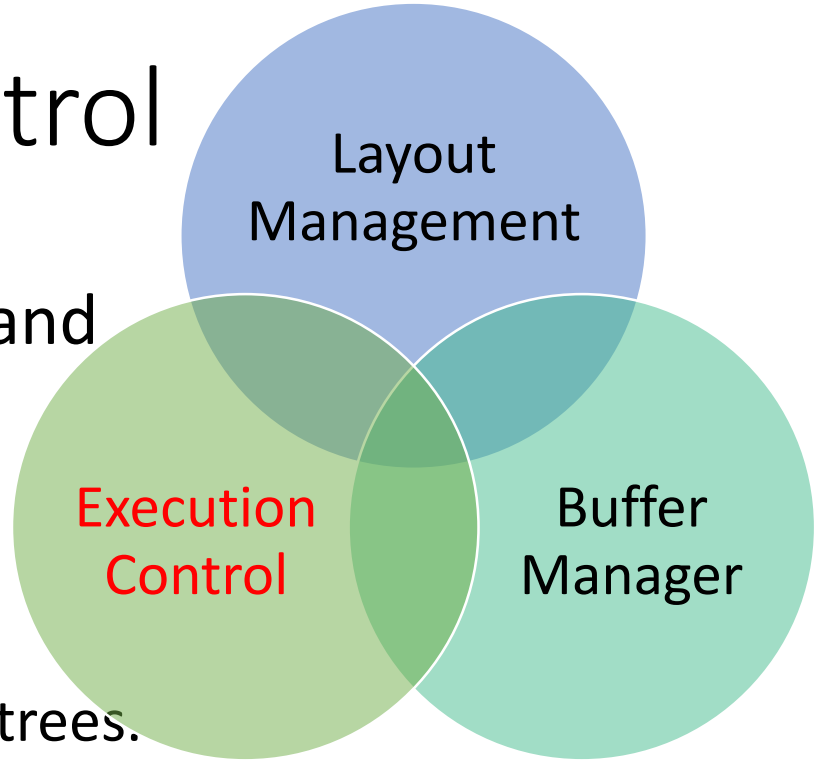


Data Flow engine: Buffers!!

- Blob Data – The love-hate relationship
  - Varchar(max)/Nvarchar(max)/Varbinary(max)
  - DT_TEXT, DT_NTEXT, DT_IMAGE
  - Share buffers in half
- Beware of data writing to disk due to memory constraints
  - Maybe because SQL Server is running on same box.
- BLOBTempStoragePath
- BufferTempStoragePath
  - https://msdn.microsoft.com/en-us/library/ms137622(SQL.110).aspx



Data Flow Engine: Execution Control

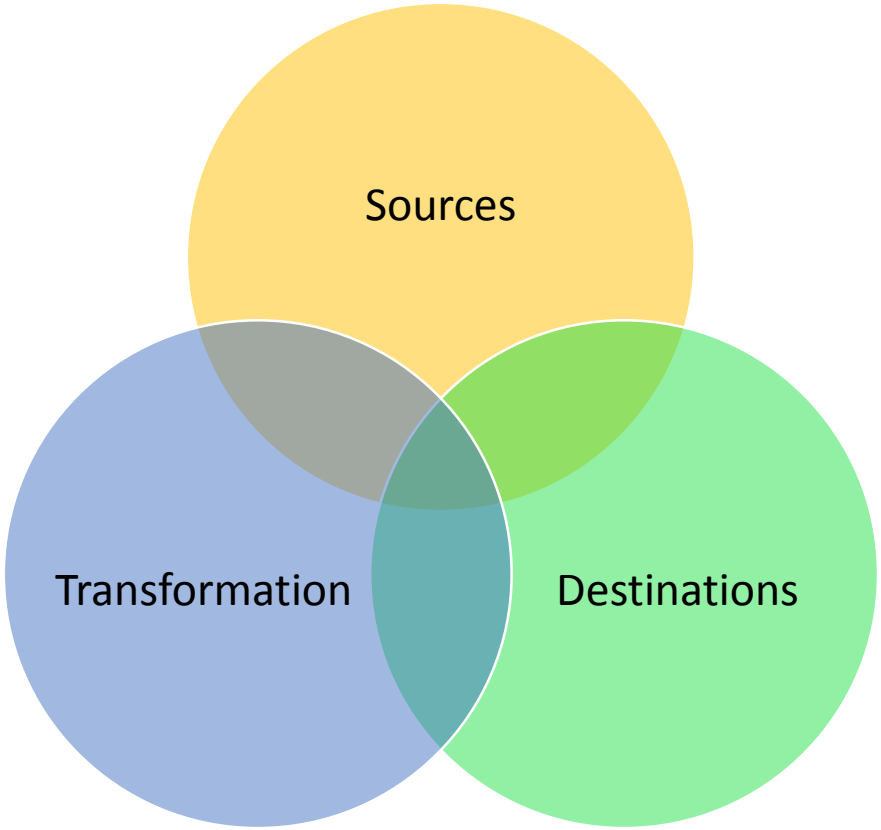


- Controls how and when sources, transformation and destinations are executed.
- Execution Trees –
  - Start and end with buffers
  - Thread allotment
  - Blocking transformation start and end new execution trees.
- Buffers are reused when no longer used.
- Max 5 buffers per execution tree
- More than one thread can be assigned to a execution tree.

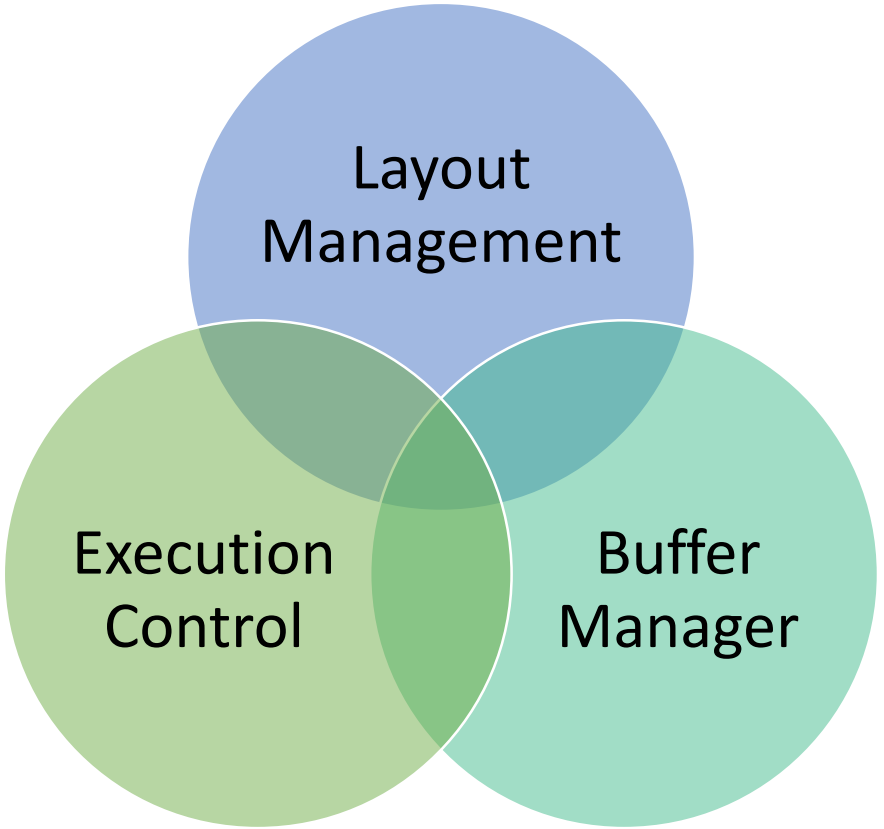
Data Flow Engine: Execution Control

- Backpressure
- Enhanced backpressure (introduced in 2012)
  - Example, Merge and one input is slow

Data Flow



Data Flow Engine



Data Flow: Transformations

- Blocking
- Semi-Blocking
- Non-Blocking

Note: buffers don’t move.

Data Flow: Transformations

Non-Blocking transformations	Semi-blocking transformations	Blocking transformations
Audit	Data Mining Query	Aggregate
Character Map	Merge	Fuzzy Grouping
Conditional Split	Merge Join	Fuzzy Lookup
Copy Column	Pivot	Row Sampling
Data Conversion	Unpivot	Sort
Derived Column	Term Lookup	Term Extraction
Lookup	Union All	
Multicast		
Percent Sampling		
Row Count		
Script Component		
Export Column		
Import Column		
Slowly Changing Dimension		
OLE DB Command		

Data Flow: Transformations

	Non-blocking	Semi-blocking	Fully-blocking
Synchronous or asynchronous	Synchronous	Asynchronous	Asynchronous
Number of rows in == number of rows out	True	Usually False	Usually False
Must read all input before they can output	False	False	True
New buffer created?	False	True	True
New thread created?	False	Usually True	True

Data Flow: Output Types

- Synchronous – buffers are pushed downstream quickly.
  - Non blocking transformation.
  - All input types (destinations).
- Asynchronous – buffers are held for processing.
  - Semi-Blocking and blocking transformations.
  - Different buffers between input and output.

Data Flow: Class exercise

- Class Exercise
- Demo

SSIS 2016 Features

- AutoAdjustBufferSize
  - If **AutoAdjustBufferSize** is set to true, the engine data flow engine uses the calculated value as the buffer size, and the value of**DefaultBufferSize** is ignored.
- Column Names in errors in data flow
- Custom logging level and RuntimeLineage logging level
- Incremental package deployment
- Azure Feature pack for SSIS
- Hadoop (HDFS Support)
- Support for always on availability group.

Performance Tips

- Try not doing Merge and Sorts
- Keep events handlers to a minimum
- Leave the logging to SSIS Catalog
- Try Performance logging in SSIS Catalog rather than normal.
- Don’t do Select *

Tips to make SSIS life easy

- Buy standard extended package for SSIS
- Implement Standards
- Centralize ETL Packages
- Plan them out on whiteboard.