

//build/

Data CodeLabs

Module 2 – High-Scale Data Processing in Azure

Romit Girdhar
Software Engineer – Developer Experience Group

#Build2016

Key Takeaways

Understand how to process Big data in Azure

- Analyze terabytes and petabytes of data to identify patterns, understand your audience better or perform complex computations to improve your business.

Learn about the tools available to process your data

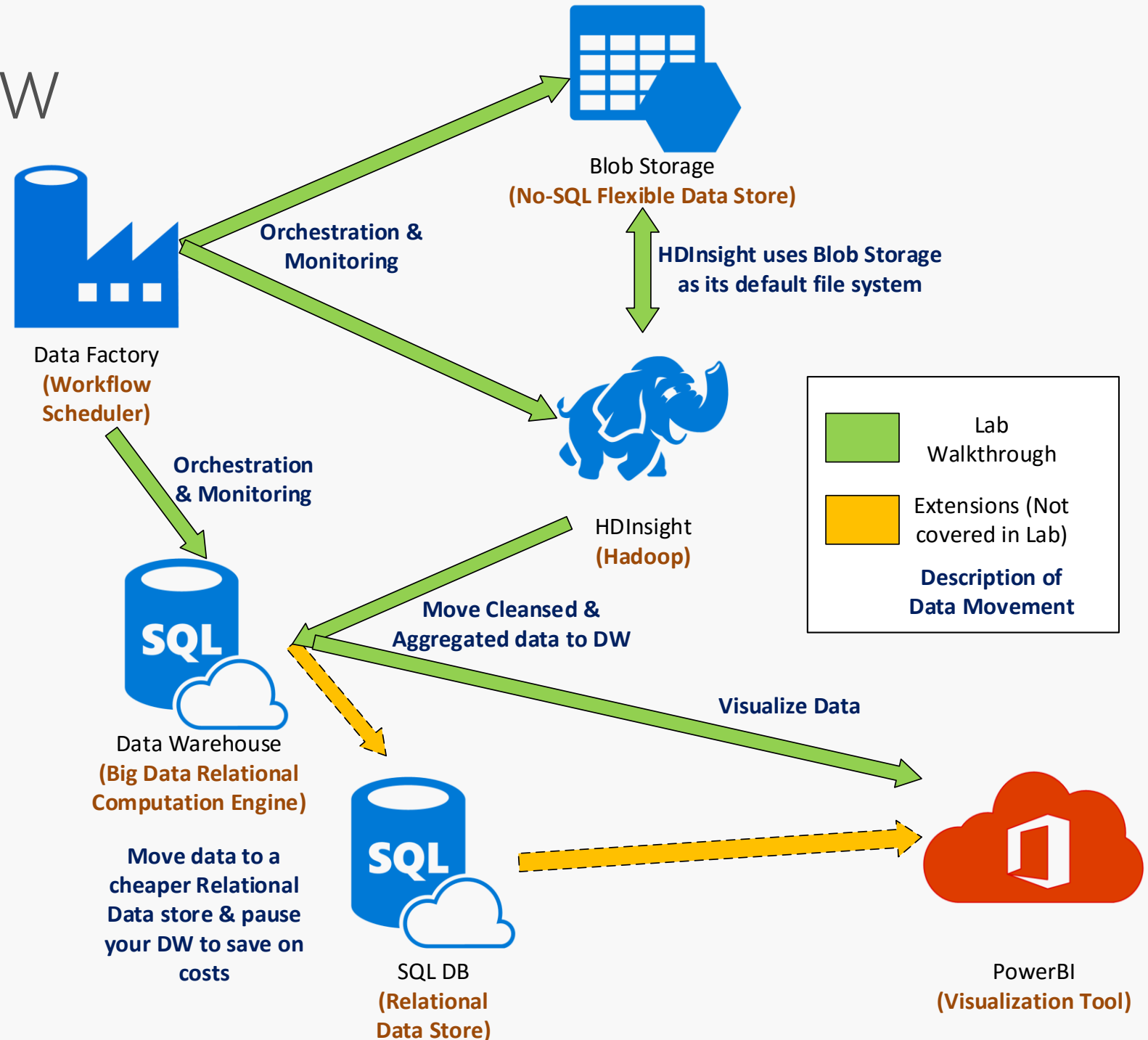
- From HDInsight (Apache Hadoop) to Azure SQL Data Warehouse, learn about the different tools available for you to process your data.

Orchestrate your Data Processing Workloads

- Using Azure Data Factory

Solution Overview

- Data stored in a cheap No-SQL storage gets picked up by your compute engine.
- The aggregated data gets moved to a "staging" data store
- PowerBI (or any other visualization tool) is used for visualization of data stored in the "staging" store.



Run the Setup.cmd

- Go through Steps 1, 2 & 3
- Open another Setup.cmd
- Go through Step #4
- 1st Three rows: East US
- 2nd Three rows: Central US
- 3rd Three rows: South Central US

(Alternatively) Create Azure HDInsight Cluster

- Go to the Azure Portal -> <https://portal.azure.com>
- Click on New -> Data & Analytics -> HDInsight
 - Cluster Name: **hdicluster-{SuffixNumber}**
 - Select Cluster Type -> OS: **Windows/Linux** & Version: **2.7.0**
 - Credentials -> Username: **admin** ; Password: **P@ssword123** (or something else...)
 - Data Source -> Create a New Storage Account
 - Location: **West US**
 - (Optional) Configure your Node VM sizes to optimize cost

(Alternatively) Create SQL DW Cluster

- From the Azure Portal
- Click New -> Data & Storage -> Data Warehouse
 - SQL DW Server Name: **buildlab-{SuffixNumber}**
 - SQL DW Name: **buildlab-dw**
 - Username: **dwadmin**
 - Password: **P@ssword123** (or anything else you'd like)
 - Location: **West US**

Azure Data Factory

- Create an Azure Data Factory
- Go to New -> Data & Analytics -> Data Factory
 - Name: **datalab-factory-{Suffix Number}**
 - Location: **West US**

Azure Data Factory

- Create Linked Services
 - Azure Storage Linked Service
 - HDInsight Linked Service

Azure Data Factory

- Create Datasets
 - Azure Storage Raw Dataset
 - Azure Storage Dummy Dataset

Azure Data Factory

- Create Pipeline
 - Adding Partitions for Hive

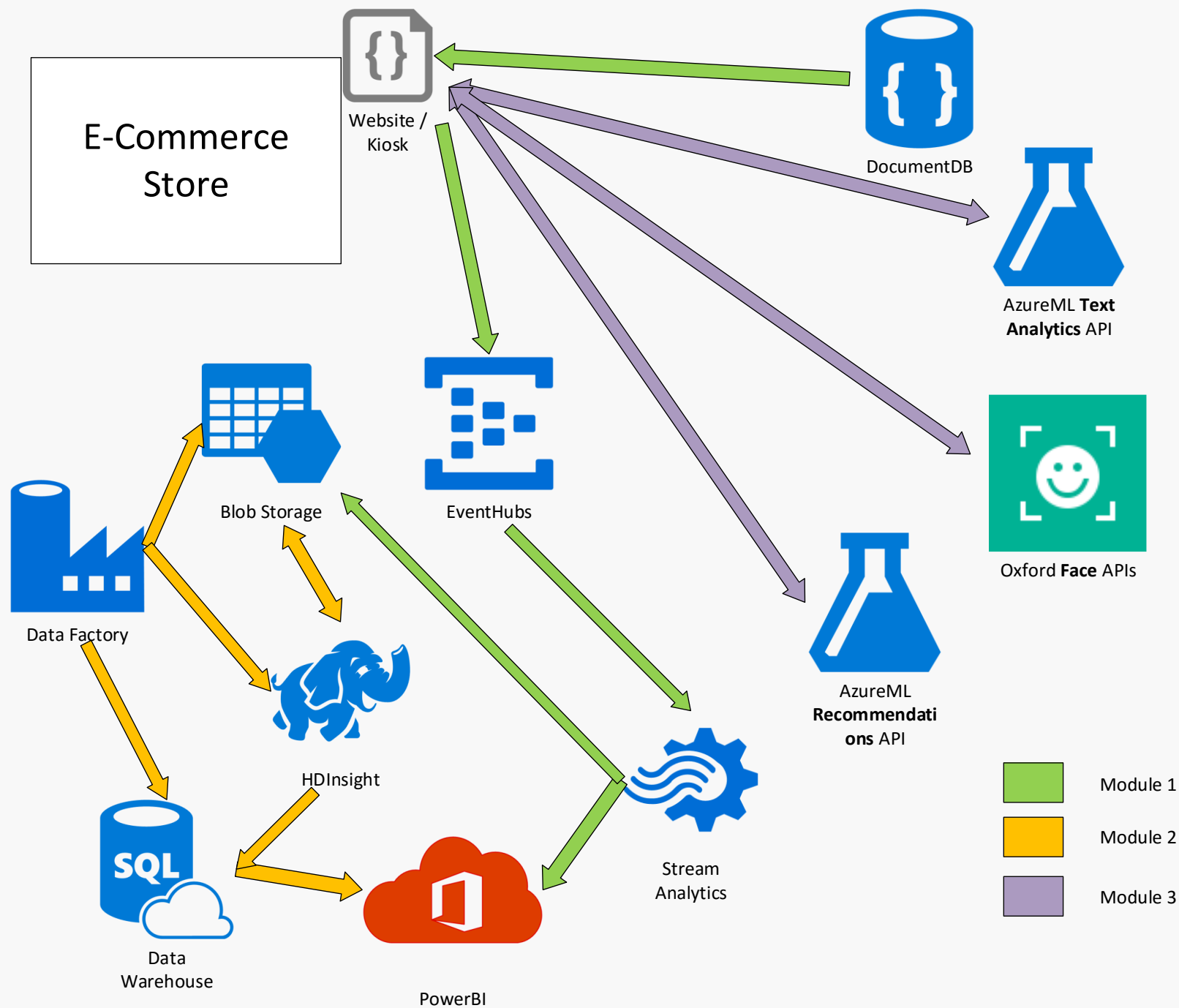
HDInsight

- Enable SSH on your Cluster through the Azure Portal
- Ssh into your cluster
- Putty is located under C:\CodeLabs-Data\putty.exe
- Enter the following Link in Putty:
`{Cluster Name}-ssh.azurehdinsight.net`
- Open the script 'createtables.hql' located in the 'Scripts' folder
- Replace the Hive variables with your storage account name and execute!

Data Factory Automation

- Open the ADFSetup program.
- Go to into the Project Properties
- Enter your Arguments
- Execute the Program!
- Follow the Prompts...

Architecture Across The Data Modules



Please Complete An Evaluation Form

Your input is important!

Required Slide

*delete this box when your slide is finalized

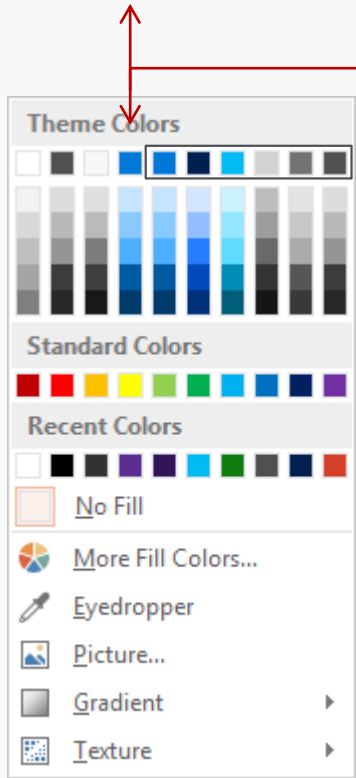


or



Slide palette info

The PowerPoint palette for this template has been built for you and is shown below. Avoid using too many colors in your presentation.



Theme Colors

Standard Colors

Recent Colors

No Fill

More Fill Colors...

Eyedropper

Picture...

Gradient

Texture

Select the 4th color from the left for subheads and 1st level non-bulleted text color, or wherever "color" text is preferred over the default black/white text

Accent colors 1-6 – (6 Theme Colors to the far right)

Text Dark 2

Accent 1

Accent 2

Accent 3

Accent 4

Accent 5

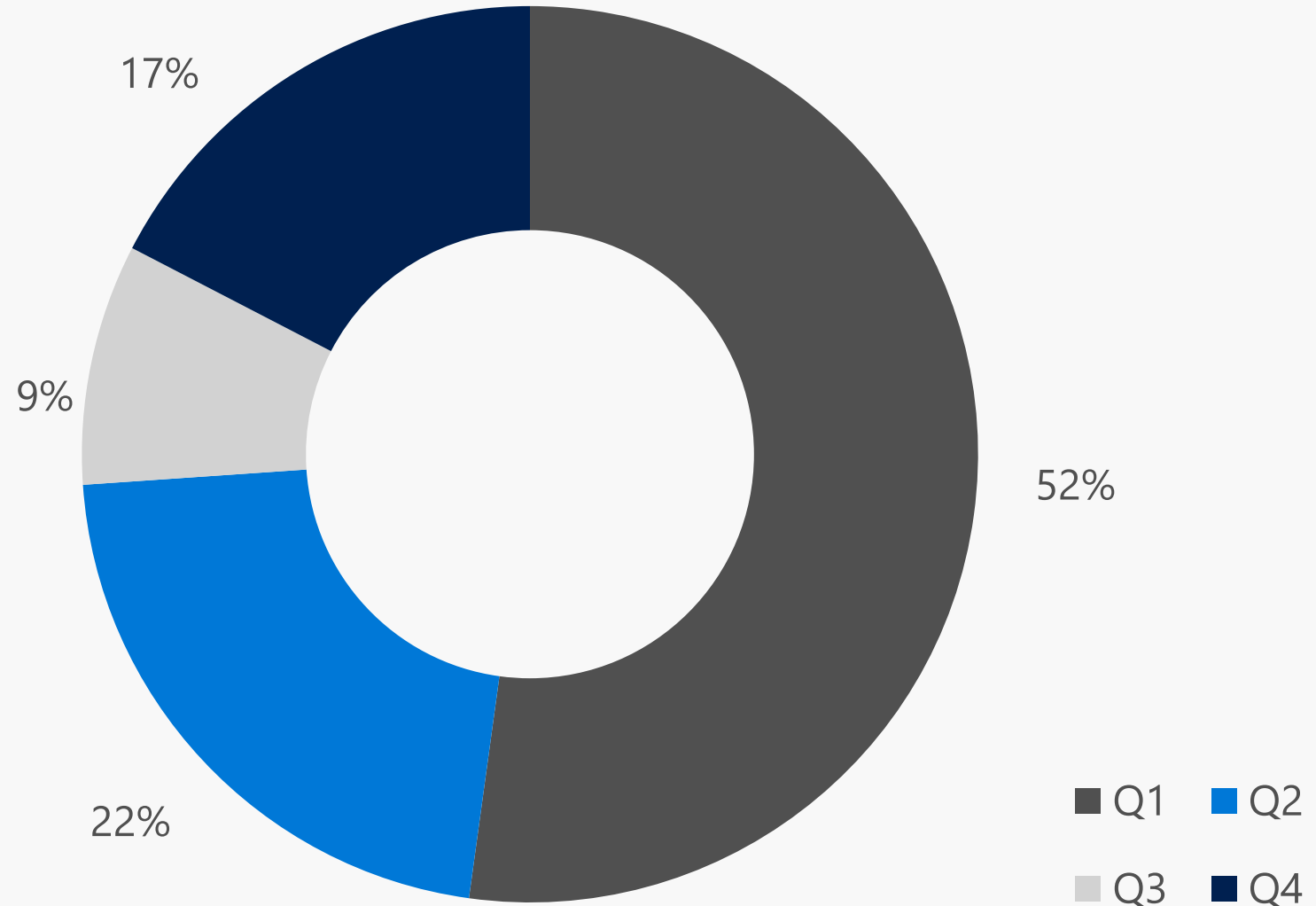
Accent 6

Use **Accent 1** as the main accent color. Use **Accent 2** and **Accent 3** when additional colors are needed.

Use **Accents 4-6** sparingly – only when more colors are necessary.

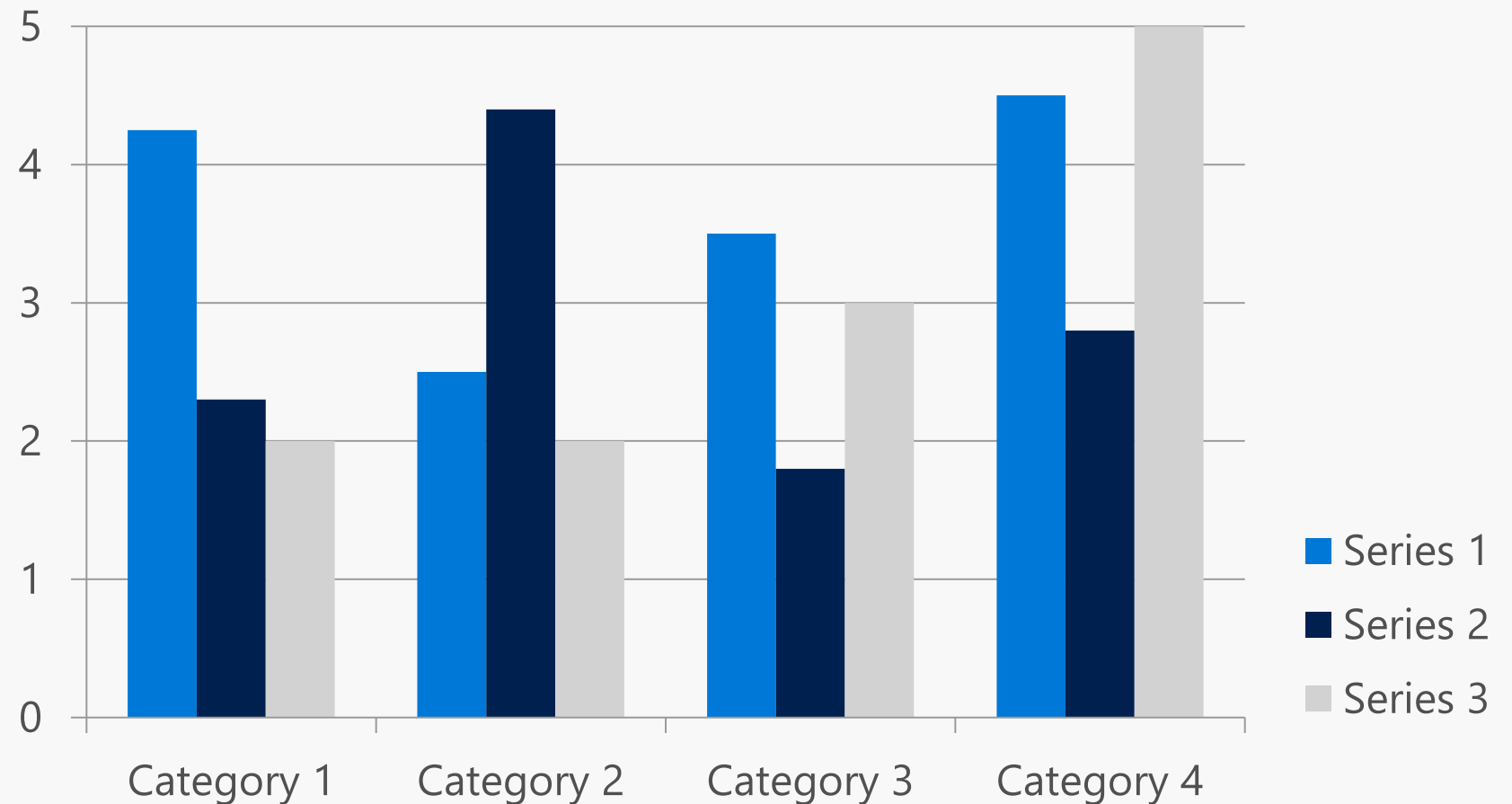
Charts

For content:
Type is what
articulates the
message. Text
doesn't need to
be bulleted. Text
size can reinforce
and impact.



Charts

For content:
Type is what
articulates the
message. Text
doesn't need to
be bulleted. Text
size can reinforce
and impact.



Table

This is a content
box, always
24pts and
sentence case.

[illegible]

Diagram with text

This is a content box, always 24pts and sentence case.

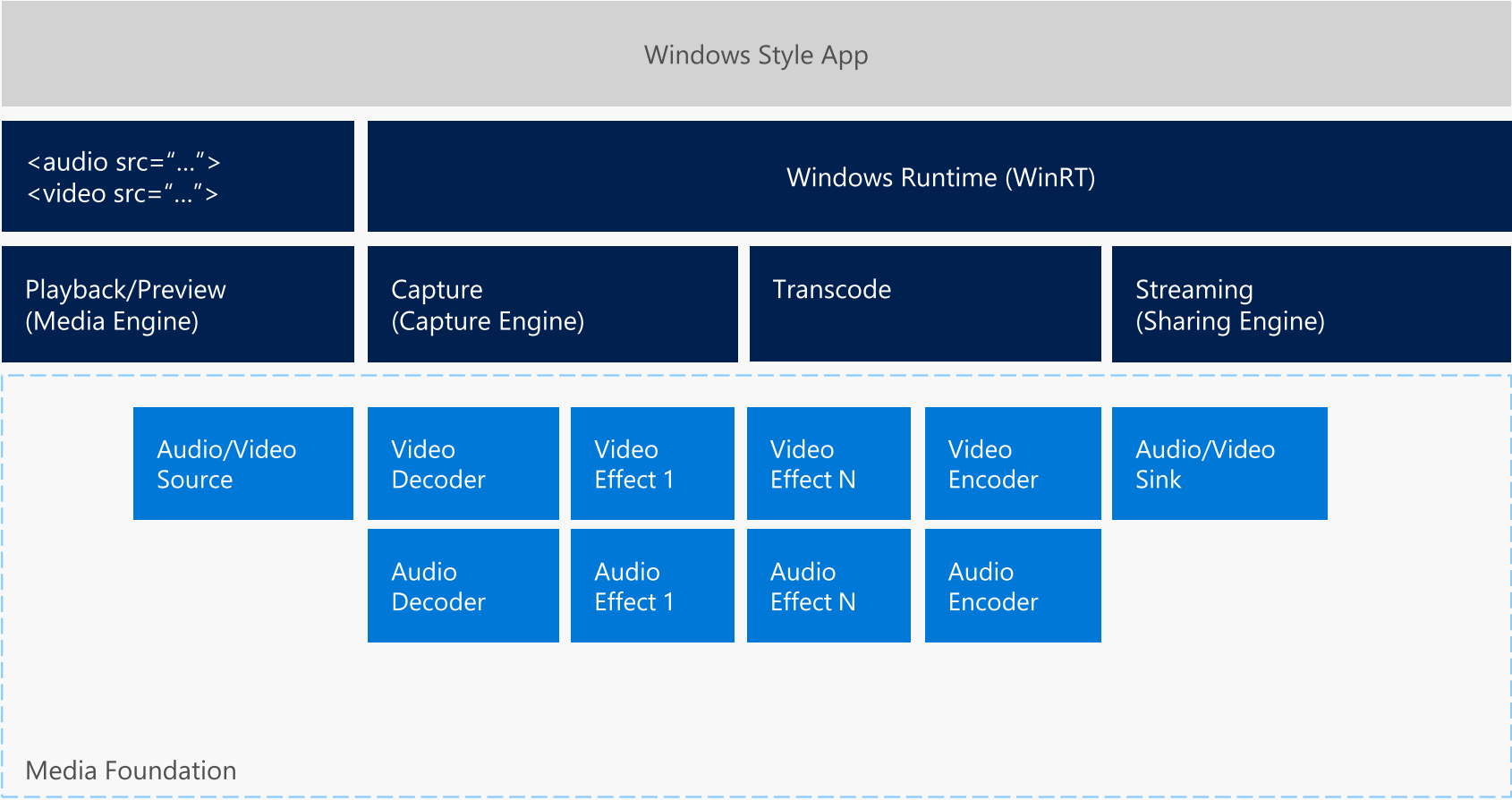
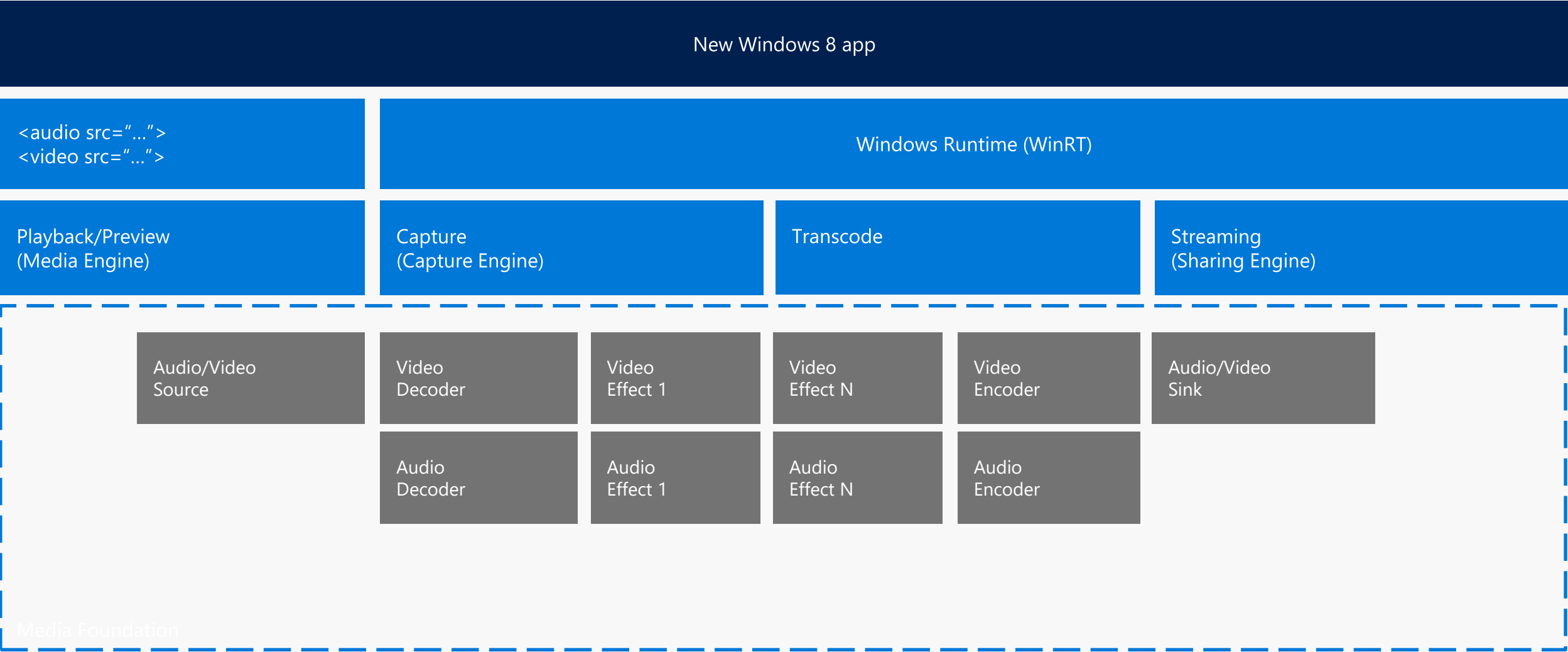


Diagram without text



Video

Demo

Speaker name

Photo layout 1



Microsoft brand guidelines

Looking for more slide resources?

Brand guidelines for PowerPoint templates is a separate slide deck that provides an overview of the Microsoft brand, guidelines, resources, tips and much more.

A few of the slides are shown at right.

Download from:
<https://microsoft.sharepoint.com/teams/BrandCentral/Pages/Presentations.aspx>

Brand photography selections, 1 of 3

Below are thumbnails of a few of the many photos that you can find on Brand Tools. In show mode, click each image to download higher resolution copies.



Accessibility

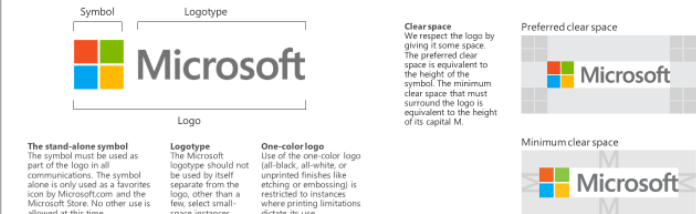
Type and background combinations must meet a minimum **4.5:1 contrast ratio or greater** for accessibility. Examples below demonstrate correct color use for on-screen applications. Colors shown with **notched corners** require black/dark text to be accessible.

Light Yellow #FFF176 #FFF176	Light Orange #FFCC80 #FF9800	Light Magenta #F48FB1 #E91E63	Light Purple #E1BEE7 #9575CD	Light Blue #BBDEFB #42A5F5	Light Teal #B2DFDB #80CBC4	Light Green #C8E6C9 #8BC34A	Light accent colors	
Yellow #FFF176 #FFEB3B	Orange #FF9800 #FF5722	Red #F44336 #E57373	Magenta #E91E63 #C2185B	Purple #9575CD #673AB7	Blue #42A5F5 #2196F3	Teal #80CBC4 #009688		
		Dark Red #C0392B #A52A2A	Dark Magenta #8E24AA #6A3D9A	Dark Purple #673AB7 #483D8B	Dark Blue #1976D2 #0056B3	Dark Teal #008080 #006D6C	Dark Green #556B2F #38761D	Dark accent colors (+ one mid accent color)
				Mid Blue #3498DB #2980B9				
White #FFF1F1 #FFFFFF				Light Gray #F5F5F5 #BDBDBD	Mid Gray #9E9E9E #808080	Dark Gray #696969 #404040	Rich Black #333333 #000000	White, black and gray accent colors
For more information on accessibility, visit: Microsoft.com/enable .								

For more information on accessibility, visit: microsoft.com/enable.

Microsoft logo

The logo has two components: the symbol with the multiple squares of color, which symbolize the company's diverse portfolio of products, and the logotype in the Segoe typeface, which is used in our products and marketing communications.



Device photography

Where to find

Microsoft device photos can be found on Brand Central...
<https://microsoft.sharepoint.com/teams/BrandCentral/>
(search on the product name and filter by asset type)
... and on News Center on Microsoft.com:
<http://news.microsoft.com/> (click on Our Products)

OEM & MS hardware (PC, tablet and phone) can be found on Featured Devices
<https://www.featureddevices.com/>

Hardware guidelines

Use the latest devices from the current year
Screen captures should be up-to-date

Remove OEM logos when possible

Drop shadows for devices only

Use drop shadows to provide subtle depth
Added drop shadows should appear as natural as possible



Section title

Section title

"If you could only sense how important you are to the lives of those you meet; how important you can be to the people you may never even dream of. There is something of yourself that you leave at every meeting with another person."

FRED ROGERS

Software code slide

This slide layout uses Consolas, a monotype font which is ideal for showing software code.

Call to Action

- Use this slide to list resources, white papers, videos and links.
- Let attendees know their next step after seeing this session.
- Re-visit Build on [Channel 9](#).
- Continue your education at [Microsoft Virtual Academy](#) online.

Please Complete An Evaluation Form

Your input is important!

Required Slide

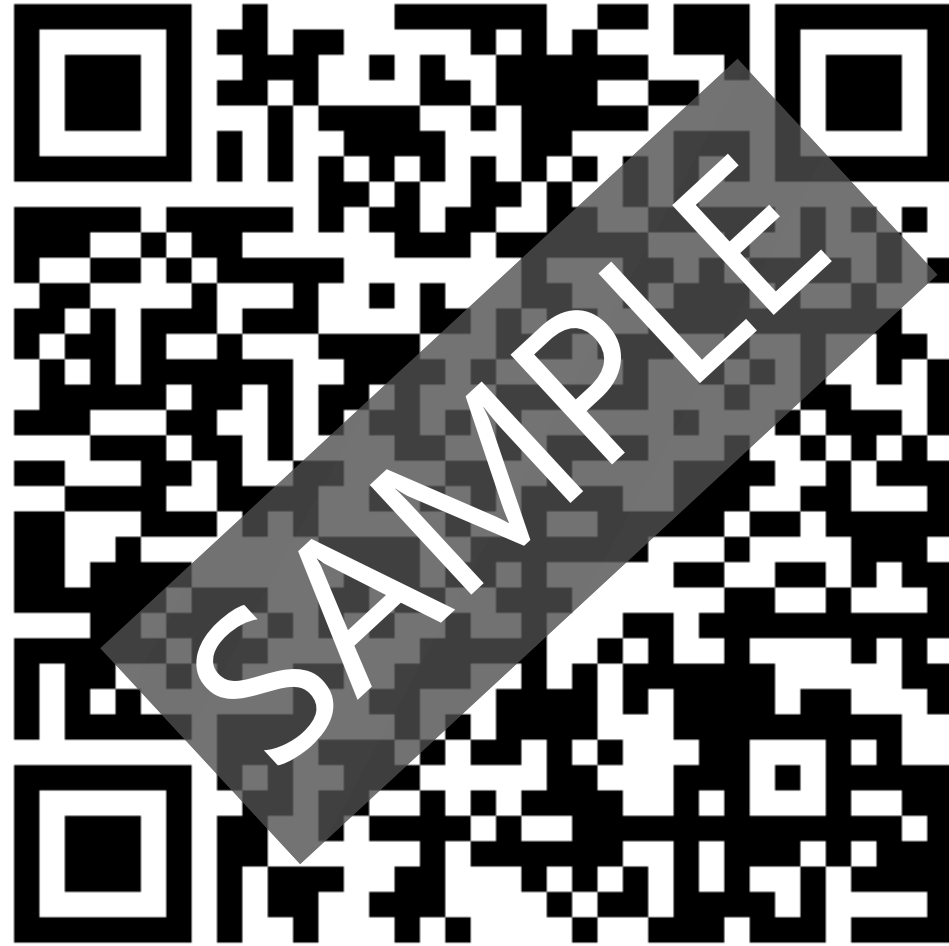
*delete this box when your slide is finalized



or



Evaluate this session



Required Slide

*delete this box when your slide is finalized

Speakers: This slide will be updated during the scrub process with a unique QR code. Attendees scan the QR code to access the eval for your session.



//build/

Presentation title goes here

Speaker name
Title

Preferred text layout (no bullets)

Main topic 1: size 40pt

Size 20pt for the subtopics

Size 20pt for the subtopics

Main topic 2: size 40pt

Size 20pt for the subtopics

Size 20pt for the subtopics

Main topic 3: size 40pt

Size 20pt for the subtopics

Size 20pt for the subtopics

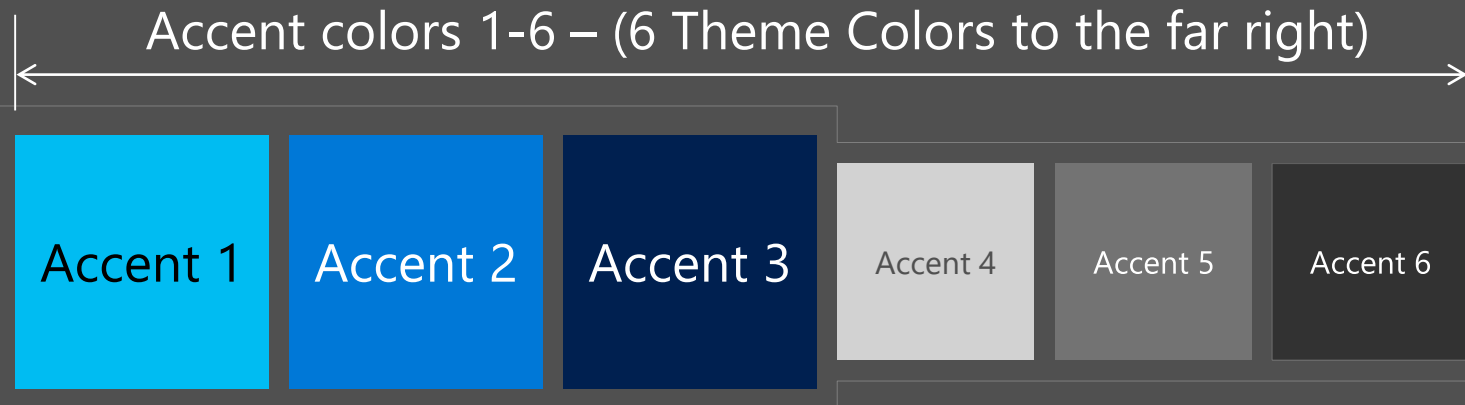
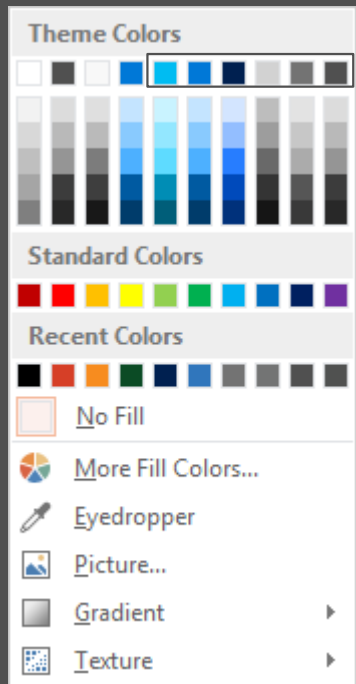
Bullet points layout with subtitle

Subtitle

- Example of a bulleted slide with a subhead
 - Set the slide title to "Sentence case"
 - Set subheads to "Sentence case"
- Hyperlink style
 - www.microsoft.com

Slide palette info

The PowerPoint palette for this template has been built for you and is shown below. Avoid using too many colors in your presentation.

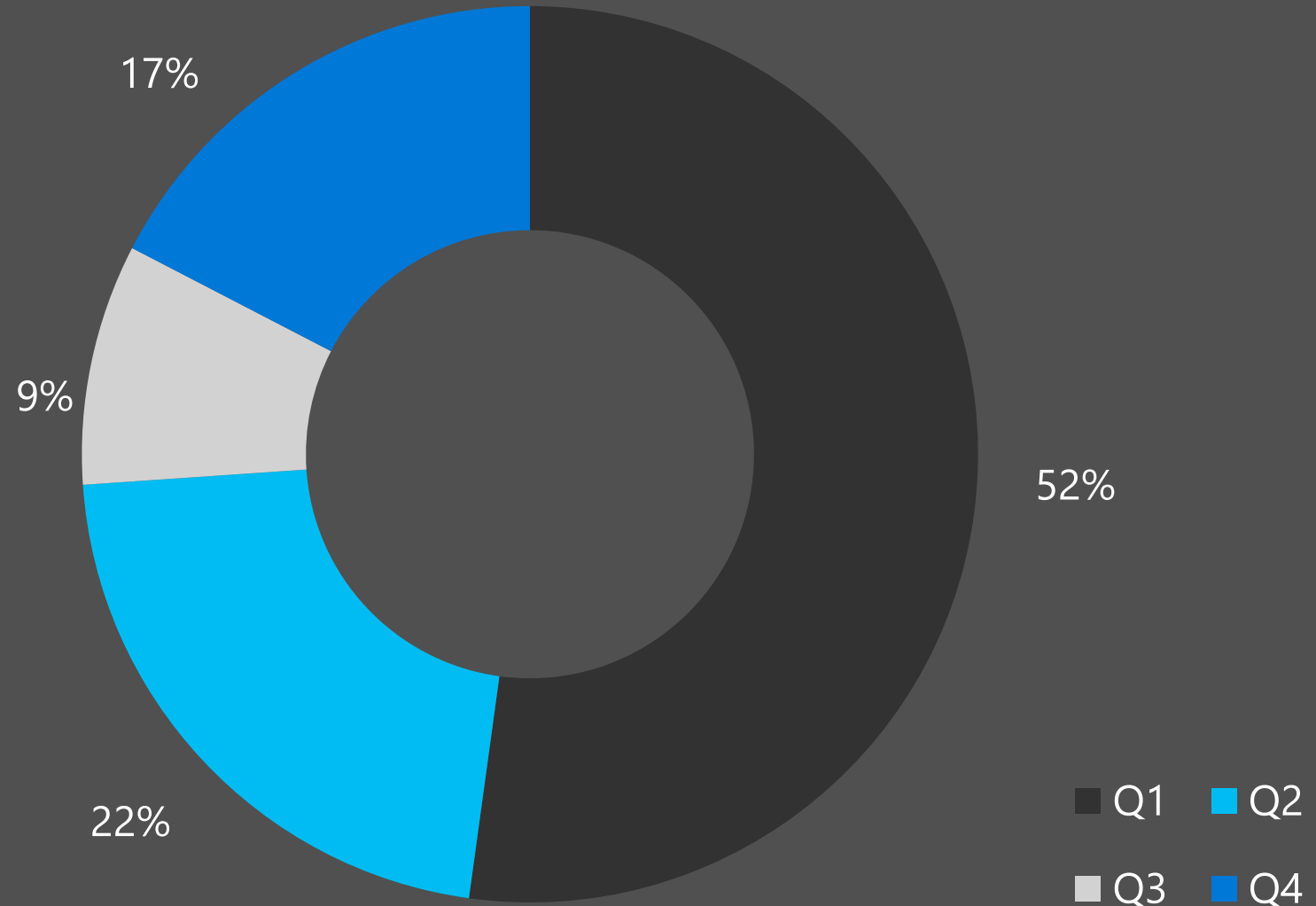


Use **Accent 1** as the main accent color. Use **Accent 2** and **Accent 3** when additional colors are needed.

Use **Accents 4-6** sparingly – only when more colors are necessary.

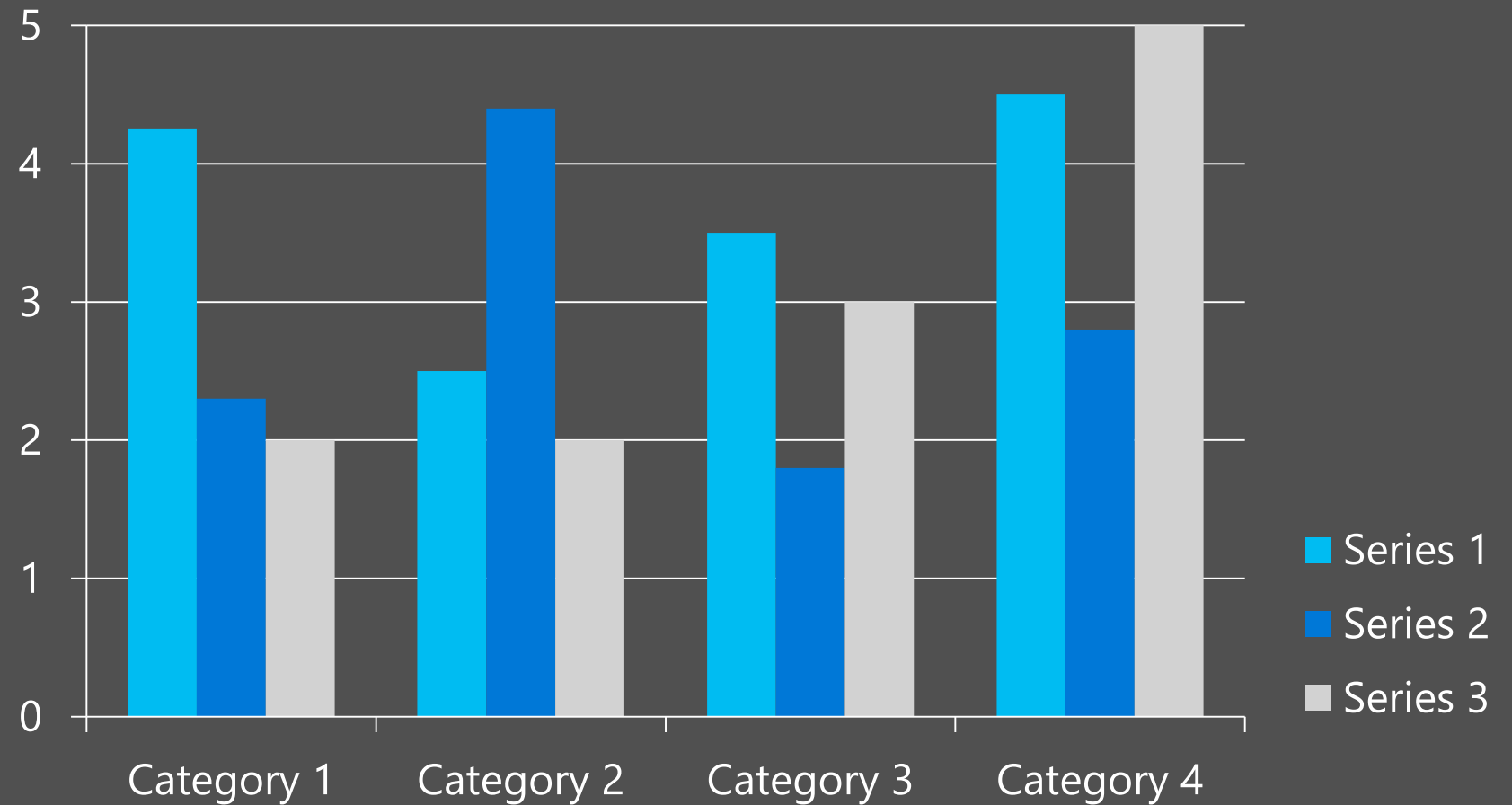
Charts

For content:
Type is what
articulates the
message. Text
doesn't need to
be bulleted. Text
size can reinforce
and impact.



Charts

For content:
Type is what
articulates the
message. Text
doesn't need to
be bulleted. Text
size can reinforce
and impact.



Table

This is a content
box, always
24pts and
sentence case.

[illegible]

Diagram with text

This is a content box, always 24pts and sentence case.

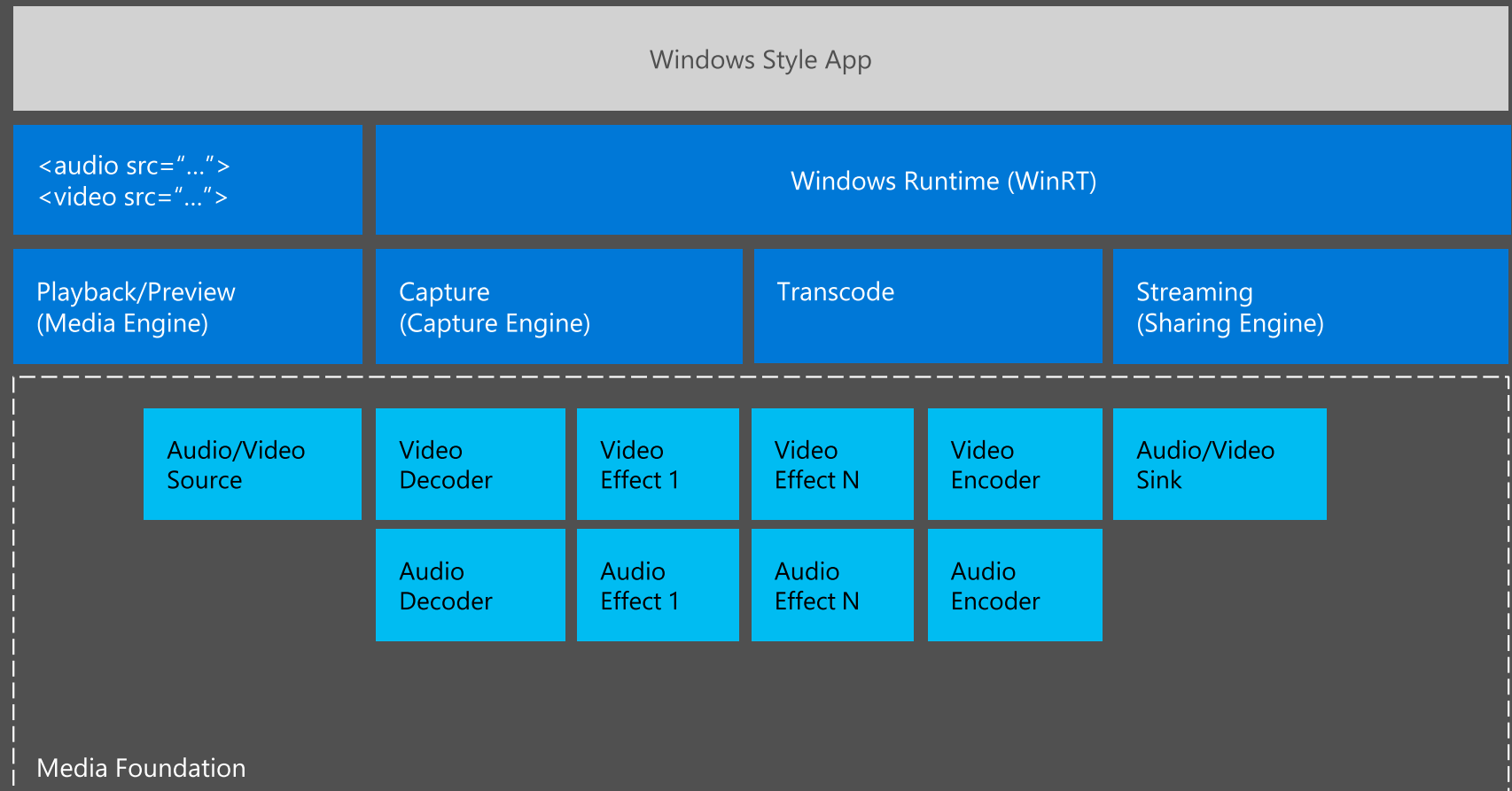
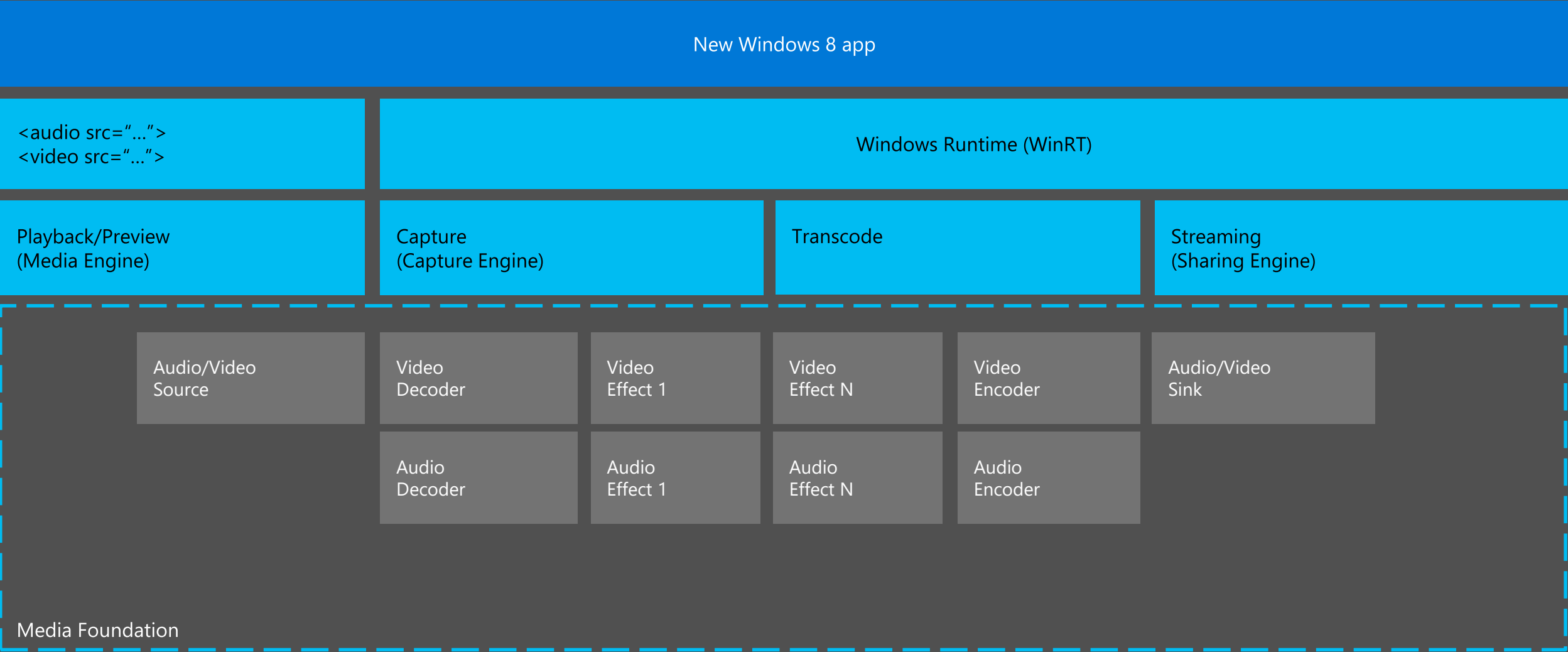


Diagram without text



Video

Demo

Speaker name

Photo layout 1



"If you could only sense how important you are to the lives of those you meet; how important you can be to the people you may never even dream of. There is something of yourself that you leave at every meeting with another person."

FRED ROGERS

Section title

Section title

Software code slide

This slide layout uses Consolas, a monotype font which is ideal for showing software code.

Call to Action

- Use this slide to list resources, white papers, videos and links.
- Let attendees know their next step after seeing this session.
- Re-visit Build on [Channel 9](#).
- Continue your education at [Microsoft Virtual Academy](#) online.

Please Complete An Evaluation Form

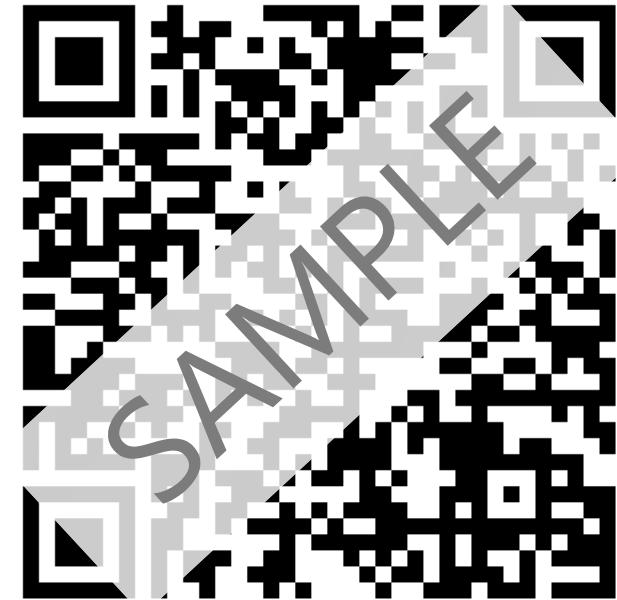
Your input is important!

Required Slide

*delete this box when your slide is finalized



or



Evaluate this session



Required Slide

*delete this box when your slide is finalized

Speakers: This slide will be updated during the scrub process with a unique QR code. Attendees scan the QR code to access the eval for your session.

