

Data Factory 101

James Smith – Architect

April 27th 2019

Creating meaningful connections that help businesses run smarter.



Supply Chain Optimization

We help you invest smarter so you can manage today and transform the future.



Connected Workforce

We create a connected workplace so employees can work smarter.



Cloud & Data Center Transformation

We help you prepare for the future and align workloads to the right platforms.



Digital Innovation

We help you innovate smarter so you can make meaningful connections.

Keys to Digital Innovation

Engage Customers



Rethinking how companies connect with their customers

By **2020**, customers expect companies to **automatically personalize experiences**, as well as proactively address their current and future needs, not just predict them.

- Walker, Customers 2020: A Progress Report

Empower Workforce



Friction-free tools to maximize employee effectiveness

The **top three benefits** expected from **technology** transformation: **Productivity** (83%), **Innovation** (82%), **Cost Efficiencies** (82%).

- Accenture

Optimize Business



Operational optimization driven by information

90% of CEO's are changing the way they use **technology** to deliver on **new stakeholder expectations and demands**.

- PwC

Pillars of Digital Innovation

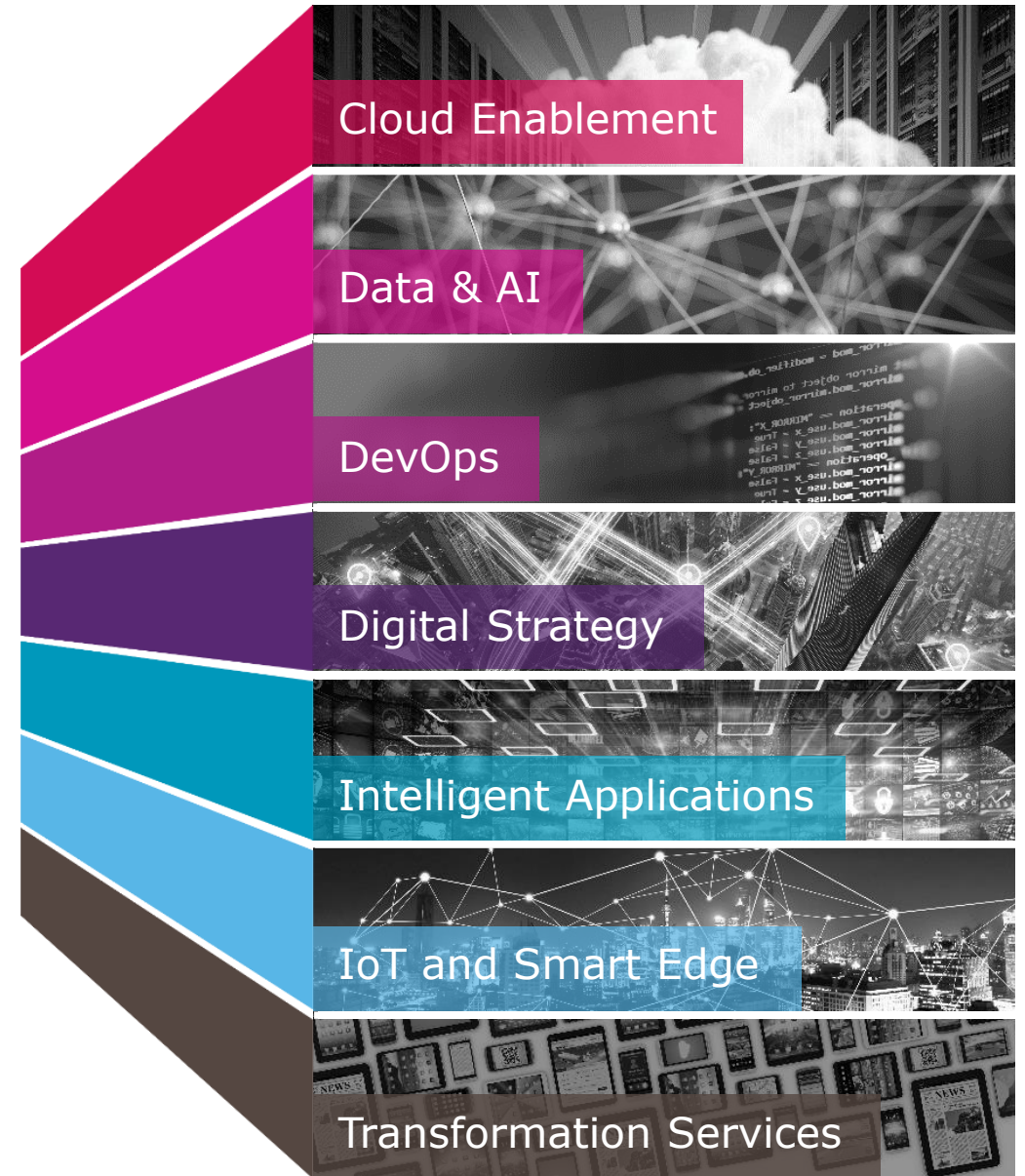
Digital Experience provides winning solutions that engage customers and employees in new, meaningful ways.

Disruptive Technology Innovations, both proven and emerging, captivate customers, empower employees and transform business operations.

Continuous Services sustains innovation with strategy, support, delivery, and data-driven insights.



DI



Awards



Our combined IT industry knowledge and technology expertise have earned us numerous Microsoft honors through the years.

2018

- Worldwide Artificial Intelligence Partner of the Year
- Worldwide Modern Workplace Partner of the Year
- U.S Partner Award for Apps and Infrastructure – DevOps
- U.S. Partner Award for Data & AI – Internet of Things
- U.S Partner Award for Apps & Infrastructure – Open Source Apps and Infrastructure on Azure

2017

- Worldwide Mobile Application Development Partner of the Year
- Worldwide Open Source on Azure Partner of the Year
- U.S. NSP Partner of the Year (and runner up)
- U.S. FED Partner of the Year
- U.S. EDU Partner of the Year

2016

- Worldwide Internet of Things Partner of the Year
- U.S. FED Partner of the Year
- OEM Device Partner of the Year

The Story So Far...

How did I get here? I just write code...

We have data, but...

I started on a new client project last summer. The team I joined delivered a brand new application by November and it went live in December. This application produces lots of data. User data, product data, application data, log data; you get the idea. We used Cosmos to store all this data and everything was great....until another team said they needed access to it.

However, they didn't really want access to the data. They wanted to create new reports and analyze it, but they wanted no part of the data in its Cosmos form. They wanted the data stored in a way they could understand it.

So!

The decision was made to create a relational model, move the data out of Cosmos and transform/store it in this new relational model hosted in Azure SQL.

Which leads us to Azure Data Factory...



What is Data Factory

Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

What does that mean?

In simple terms, Data Factory is a task scheduler primarily focused on data. It provides a set of tools to move, transform and analyze your data.

It also provides more advanced tools such as Machine Learning and Data Lake Analytics. We won't cover any of the advanced topics in this course, but keep them in mind if you decide to use Data Factory for your own projects.

Let's take a look at a Data Factory instance and go over the basic concepts.

Azure Data Factory

Dashboard

The dashboard is the primary management mechanism for a Data Factory instance. It provides pipeline authoring and managing capabilities, monitoring tools and links to documentation, samples and more.

The dashboard is a web application that can be accessed through the Azure portal. Simply navigate to the Data Factory instance and click the 'Author & Monitor' link.

Pipelines

Pipelines are a grouping of activities. They are the primary 'run' mechanism, or how you get work done.

For example, say you need to move data from Azure Blob Storage to SQL Server. You also need to transform some of that data after it moves. You can accomplish this with a Data Factory pipeline, utilizing the 'Copy' and 'Stored Procedure' activities.

Activities

Activities are pipeline processing steps. These are the individual building blocks that perform your work.

There are a slew of activities in addition to the 'Copy' and 'Stored Procedure' activities previously mentioned. For example, running SSIS packages, API lookups, conditionals and iterators.

There are more advanced activities such as Machine Learning, DataBricks and Data Lake Analytics. For a complete list and more details, please refer to the documentation.

Datasets

Datasets are data structures that represent your data. They are a view of the data that will be used in pipeline activities.

In a simple copy scenario, there would be two datasets; a source and destination.

Datasets are tied to Linked Services, which we will discuss in the next slide.

Connections

Connections or Linked Services are similar to connection strings. These are used in conjunction with Datasets to connect to and define the data needed by activities and pipelines.

There are dozens of connection types, including most Microsoft/Azure options and many third party connectors such as Oracle and DB2.

Triggers

Triggers determine when a pipeline is to be executed. There are simple manual triggers and more advanced options available.

For example, triggers run at a specific date and time or when a file is added to or removed from blob storage.

A photograph of four business professionals (three men and one woman) sitting around a conference table in a modern office setting. They are all smiling and engaged in conversation. The image is overlaid with a semi-transparent purple gradient. The text "To the Lab!" is centered over the image.

To the Lab!

And QnA.....

Lab – Move the data from Azure Blob Storage to our Azure SQL Database

<https://github.com/jamesmith119/datafactory101>



Thank you