



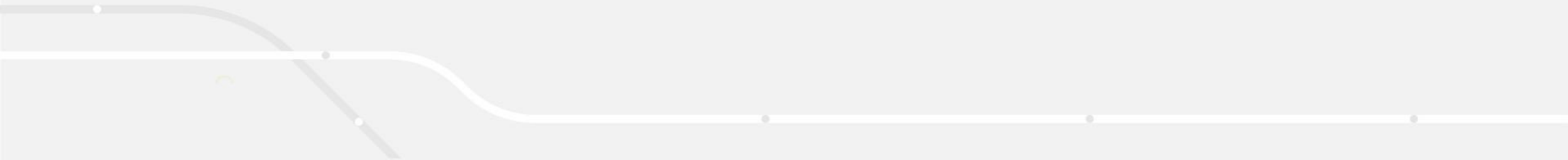
L'analyse des données IoT a portée de Synapse

CHRISTOPHER MANEU
@CMANEU

Presentation brief for MCT's

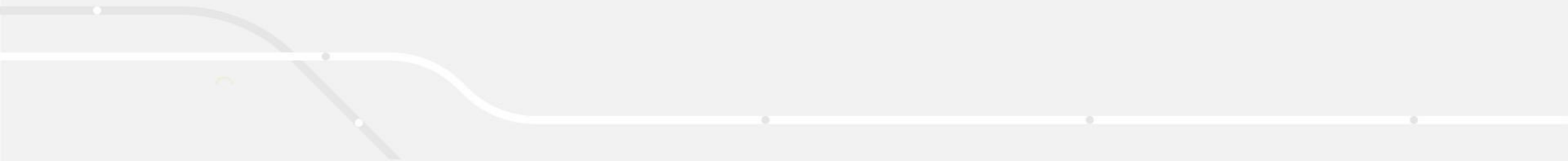
Title of presentation	Azure for the Data Engineer
Presentation\Lab length	Presentation: 40 - 45 minutes. Lab: 50 – 60 minutes
Audience	Data Professionals, Data Architects, Business Intelligence Professionals, Developers.
Purpose	1) Introduce data engineering from the perspective of the types of data, technologies, roles and responsibilities. 2) Understand the case study that will be used as the basis for all of the labs.
(think) Student key takeaways	1) The student should understand how data has evolved into many types of structured, semi-structured and non-structured. 2) The student should have an appreciation of the different data platform technologies at a high level. 3) The student should be able to understand the key roles and responsibilities of a data engineer.
(feel) Student desired mindset	1) The student should understand why data engineering extends the skillset of a DBA to new technologies and tasks. 2) The student may feel that this is overwhelming, but learn that this course provides a broad overview of key skills required.
(do) Student outcomes	1) The student should be able to give examples of how data has changed. 2) The student should be able to articulate a one sentence description of each data platform technology. 3) The students should be able to describe the Adventureworks case study.
More information	For MCT course feedback, visit the Microsoft Official Courseware forum

QUESTIONS FOR EDITORS

- In the M4 Story Framework, it seems that the hot path is presented in other modules. Which ones?
 - What definition we want to push for Big Data 2.0? (Insights?)
 - What is the grocery shopping / refrigerator analogy for Data Warehouse?
 - How we land the DWH into the Big Data 2.0 story?
 - Should we split “classic” datawarehouses and MPP dwhs?
- 

Internal presenter notes: *What attendees should take away*

- What is Big Data and its challenges
- What kind of data IoT produces and the challenges getting value out of them
- Why you should integrate your IoT Data



Internal presenter notes: What attendees should take away

Formatting Reminders:

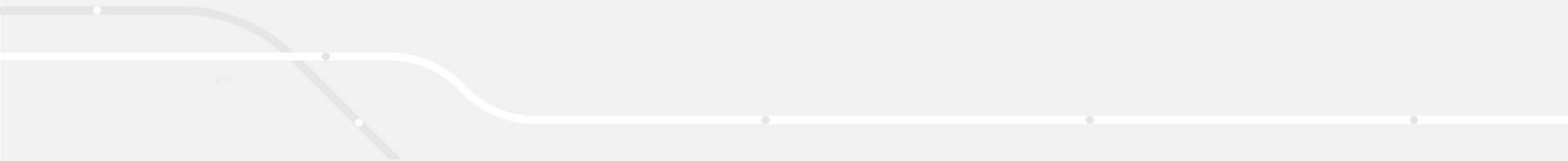
- 15-20 slides maximum
- Font: Segoe UI, 18pt minimum, 36pt maximum
- Code Font: Consolas, 20pt minimum

Include complete speaker notes on every slide

Use the slide options and styles as desired

Remember that we're building slides for mass public consumption, and any presenter should be able to use them

- Presenters are welcome to personalize the presenter slide or make minor adjustments throughout the deck, as long as they convey the main concepts to attendees.



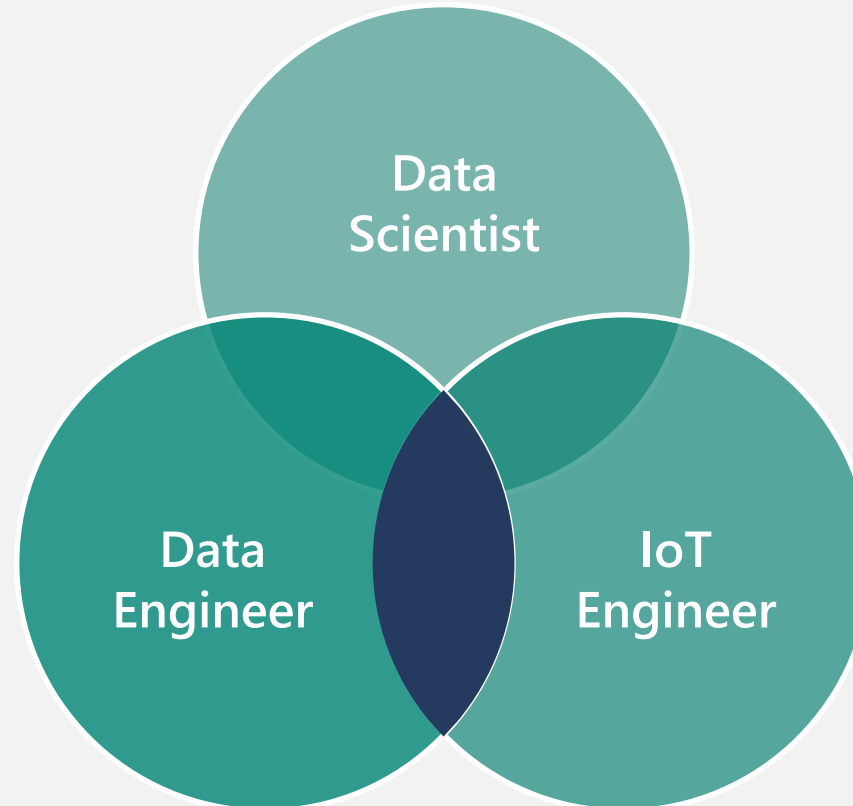
Agenda

A primer on Big Data Analytics

Integrating IoT and Big Data

Getting Insights & Actions from IoT Data

For whom is this session for?



C'est quoi le Big Data, et pourquoi l'IoT en a besoin?

What is Big Data?

Volume

La quantité de données à traiter

Variety

La qualité "destructurée" de la donnée (en opposition aux données relationnelles)

Velocity

La fréquence à laquelle la donnée est créée



Square Kilometer Array radio telescope

50 Tb

Volume of data analysé par jour



Signal &
image
data

7000 Tb

Donnée brute générée
par second

600Pb

Stocké par an

Square Kilometer Array radio telescope

50 Tb

Volume of data analysé par jour

Volumetry



Variety

Signal &
image
data

Velocity

7000 Tb

Donnée brute générée
par second

600Pb

Stocké par an

La nature des données IoT

JSON Documents

```
{
  "TimeStamp": "2020-0812T13:58:07.5203217Z",
  "Co": 20.6515621746199,
  "Co2": 48.47952265338019,
  "Ppm10": 257.44680750414,
  "Ozone": 163.747165663515,
  "EventProcessedUtcTime": "2020-08-12T13:58:13.2928373Z",
  "PartitionId": 3,
  "EventEnqueuedUtcTime": "2020-08-12T13:58:08.08500Z"
```

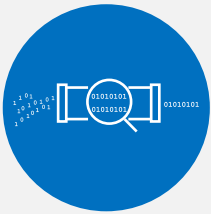
Metadata

DeviceId	14f8gq51gq
Customer	ContosoCity
LocationGps	43.5887 1.4522
DeviceType	AirQualityv2
Building	Main
BuildingType	School
Weather	Sunny
Wind (km/h)	28

Unstructured bin



4 challenges avec la donnée IoT



Manque de
consistance



Nécessite d'être
contextualisé



Donnée utilisée en
conjunction avec
d'autres sources



Souvent, avec une
retention infinie

Data lot recherche analyse...

Volume

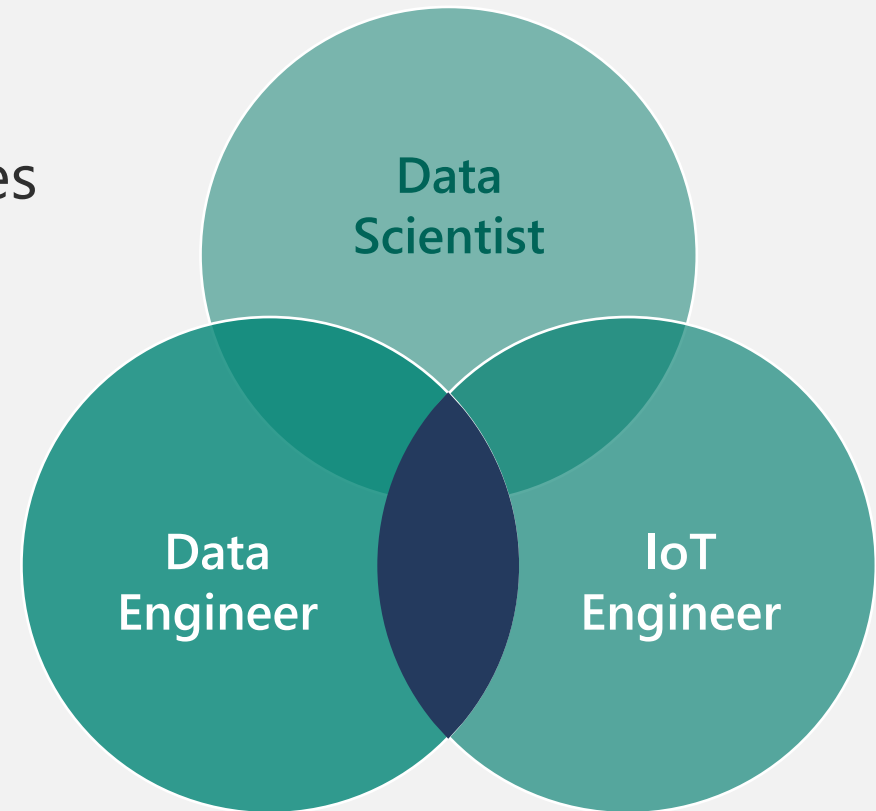
IoT génère un grand volume de données

Variety

IoT crée des données non structurées (images) ou semi-structure (json) complexes à intégrer dans du reporting

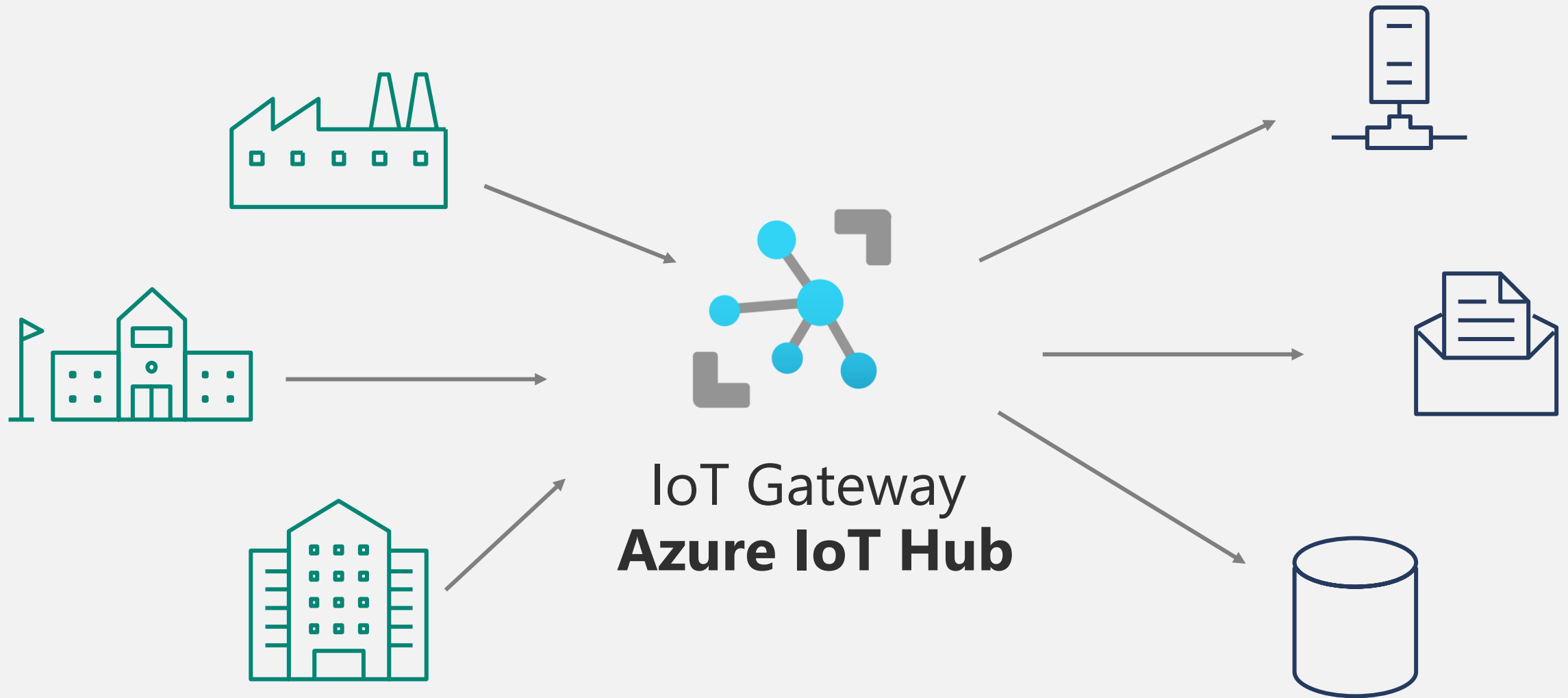
Velocity

IoT génère de la donnée rapidement, jusqu'à du temps réel (X events/secondes)



Synapse à la rescousse

D'où vient la donnée IoT ?



Où stocker la donnée ?



Azure IoT Hub

Storage



Storage
accounts



Data Lake



Time Series
Insights

Messaging



Service
bus



Event
Hubs

Open protocols

AMQP

Kafka

Processing



Stream Analytics



Azure Functions

Où stocker la donnée ?

Augment business apps

Relational
Databases

Leverage existing BI
capabilities

Data
warehouse

Unlock new scenarios

NoSQL
databases

Use existing Big Data &
Hadoop capabilities

Data lakes

Where you can land your IoT Data?

Augment business apps

Relational
Databases



Azure
SQL DB



PostgreSQL, Maria DB

Leverage existing BI
capabilities

Data
warehouse



Data warehouse



Azure Synapse
Analytics

Unlock new scenarios

NoSQL
databases



Azure
CosmosDB



Azure
Time Series Insights



Azure
Digital Twins

MongoDB, Cassandra

Use existing Big Data &
Hadoop capabilities

Data lakes



Azure Synapse
Analytics



Hadoop
(HD Insight)

How to ingest IoT Data depending on your target?

Relational databases

Azure Stream Analytics (ASA)
Specific features of Azure SQL (columnstore tables)

Data warehouse

Azure Stream Analytics

NoSQL & Specialized databases

Integrated with Time Series Insights,
Cosmos DB via ASA or Azure Databricks Spark

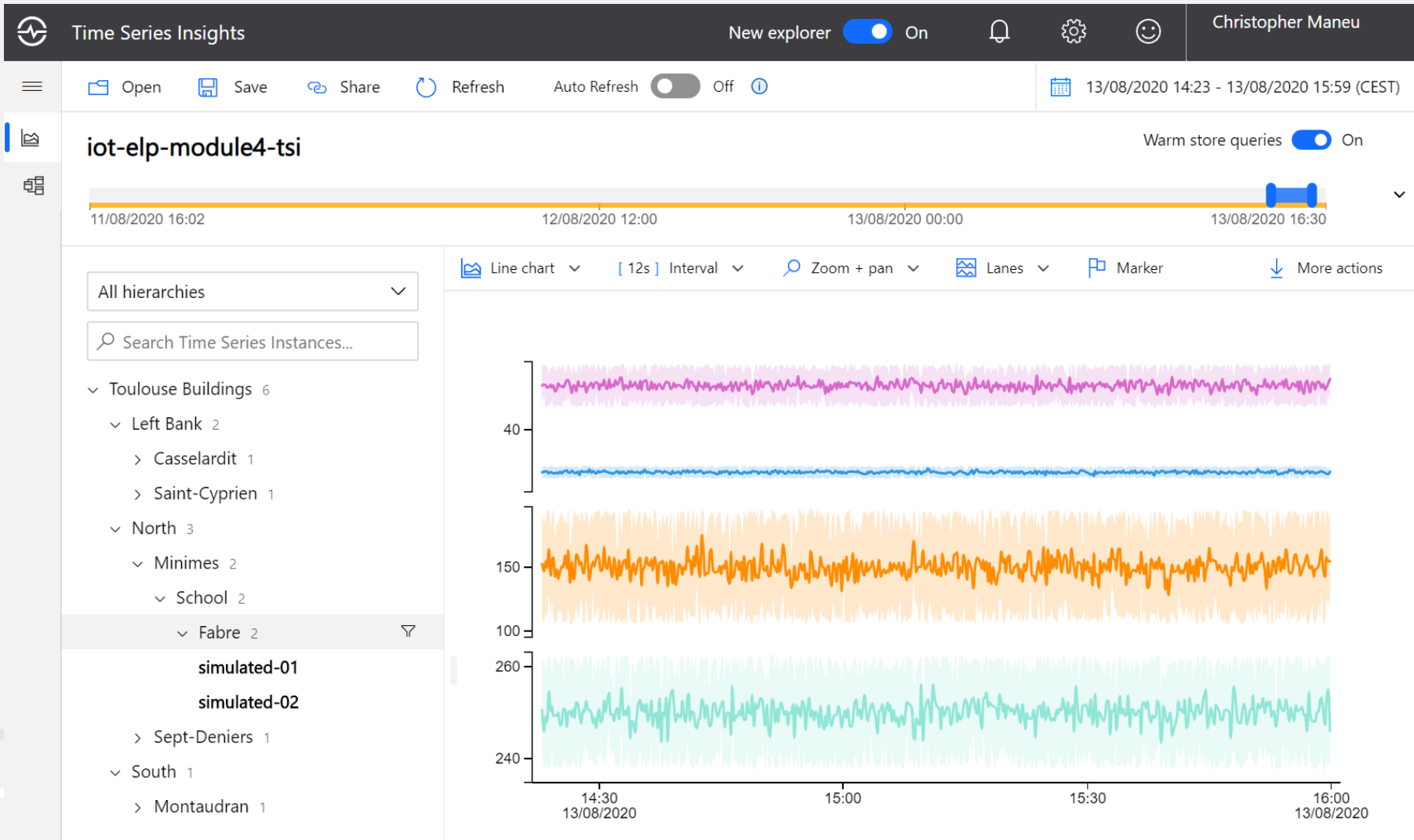
Data Lake

Integrated with IoT Hub

Azure Synapse

Integrated with Stream Analytics or directly query Data Lake

So, where you *should* land your IoT Data?



So, where you *should* land your IoT Data?

Big Data Data Lake

Experimentation
Fast exploration
Semi-structured data

OR

Relational Data Data Warehouse

Proven security & privacy
Dependable performance
Operational data

Introducing Azure Synapse Analytics



A **limitless** analytics service with **unmatched time to insight** that delivers insights from all your data, **across data warehouses and big data analytics systems** with **blazing speed**

Simply put, **Azure Synapse is Azure SQL Data Warehouse evolved**

We have taken the same industry leading data warehouse and elevated it to a whole new level of performance and capabilities

Demo

aka.ms/iot40/watch



Resources



Session Resources Hub

aka.ms/iot40



Internet of Things Learning Path

aka.ms/iot10/learn



Envie de découvrir comment créer une solution IoT?

aka.ms/iotlp

/learn alert

Formation rapide et interactive sur l'analyse de données IoT.

aka.ms/iot40/learn

 Microsoft | Docs Documentation Learn Q&A Code Samples Search 

IoT Learning Path Module 4 - Big Data 2.0

2 hr 19 min • 1 Learning Path • 2 Modules • Created by  Christopher (Microsoft)  

100% completed

Items in this collection



Introduction to implementing lambda architecture for IoT solutions

Module • 28 min

★★★★☆ 4.6 (318)

100% completed



Explore and analyze time-stamped data with Time Series Insights

Module • 37 min

★★★★★ 4.7 (268)

100% completed

En voir plus?

Découvrez comment
analyser vos données IoT
avec Synapse et d'autres
services Azure en 40 minutes

aka.ms/iot40/watch



Thanks to these open resources



Photo by [Donald Giannatti](#) on [Unsplash](#)