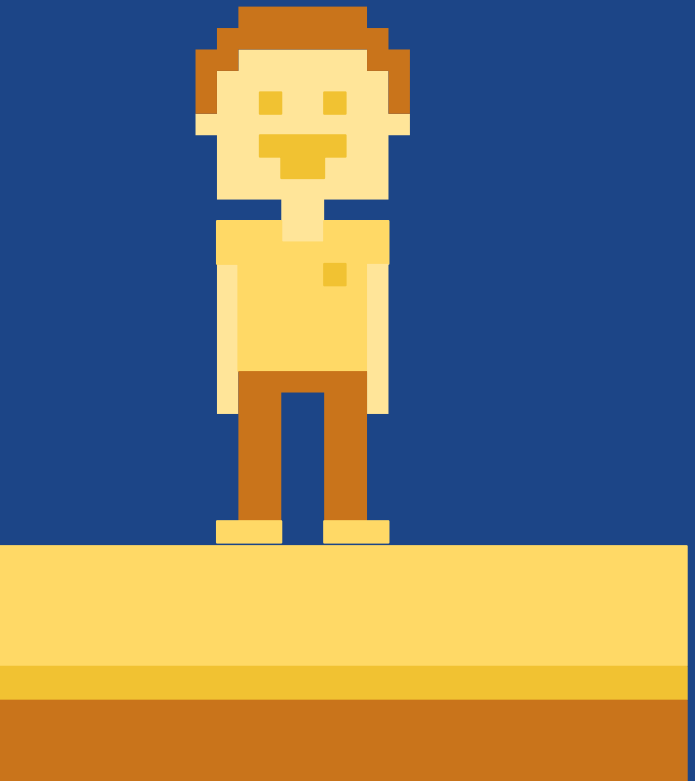
HTAPS

Are they the Future?

avanade

sqlbits

WELCOME!



Scott Bell

- Senior Consultant at Avanade and Previously Altius.
- Avanade UK&I Databricks SME
- Interested in Azure Data Platforms, Architecture and Design Pattens
- Masters Degree in Computer Science Focusing on Machine Learning in the Cloud
- Passionate about Beer, Rugby League and Most things that aren't Excel!

Agenda

01 What are they?

An Orgin Story

02 WHAT Challenges do they solve

Concrete Example

03 Maturity?

Are they still a Unicorn?

04 Q&A

01

WHAT Are They

WHAT Problem?

Traditionally, Transactional and Analytical Capabilities have been separated into discrete architectures

Traditional Data Processing

OLTP

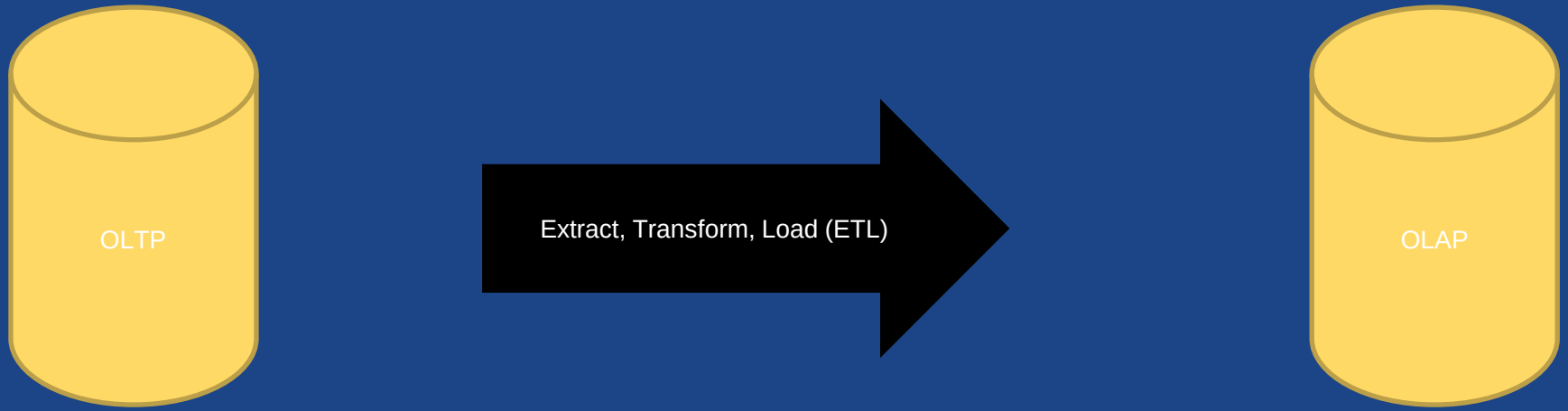
- Online transaction processing systems record business interactions as they occur in the day-to-day operation of the organization, and support querying of this data to make inferences.
- Focus on individual transactions
- low latency and high throughput with a mix of read and writes
- Need to be ACID Compliant
- Often Normalized to save space and increase efficiency
- Usually, store online banking, orders, messages

OLAP

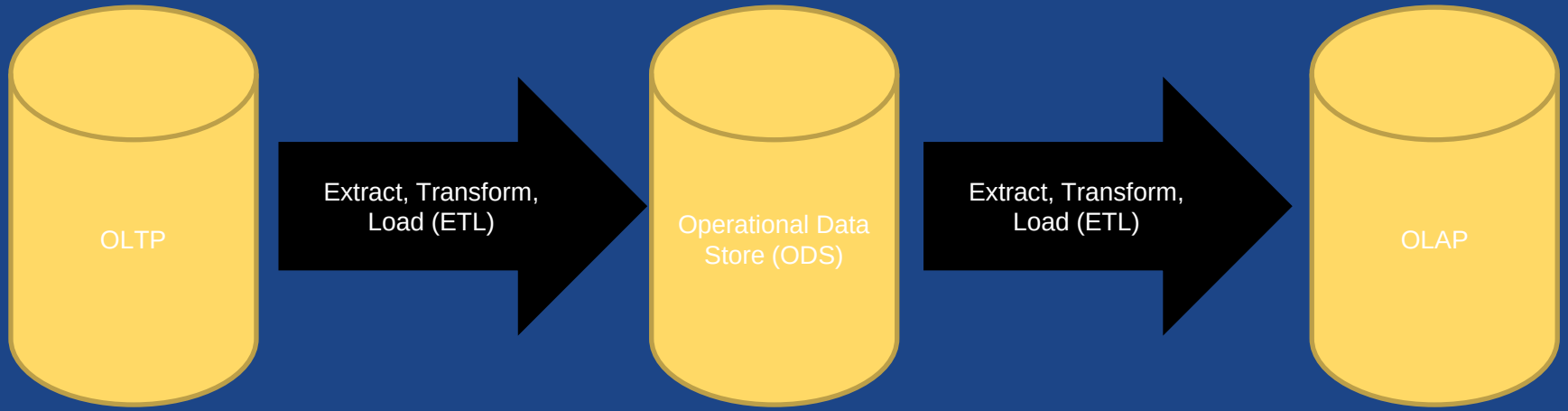
- Online analytical processing (OLAP) is a technology that organizes large business databases and supports complex analysis. It can be used to perform complex analytical queries without negatively affecting transactional systems.
- Models data in a business analytic friendly way
- Response time usually not real time, e.g. batch loads can be many minutes or even hours
- Don't modify data, usually read intensive
- Complex queries and modelling of data
- Larger in size

https://docs.microsoft.com/en-us/azure/architecture/data-guide/

Data Processing Flow



People Tried Solving the problem



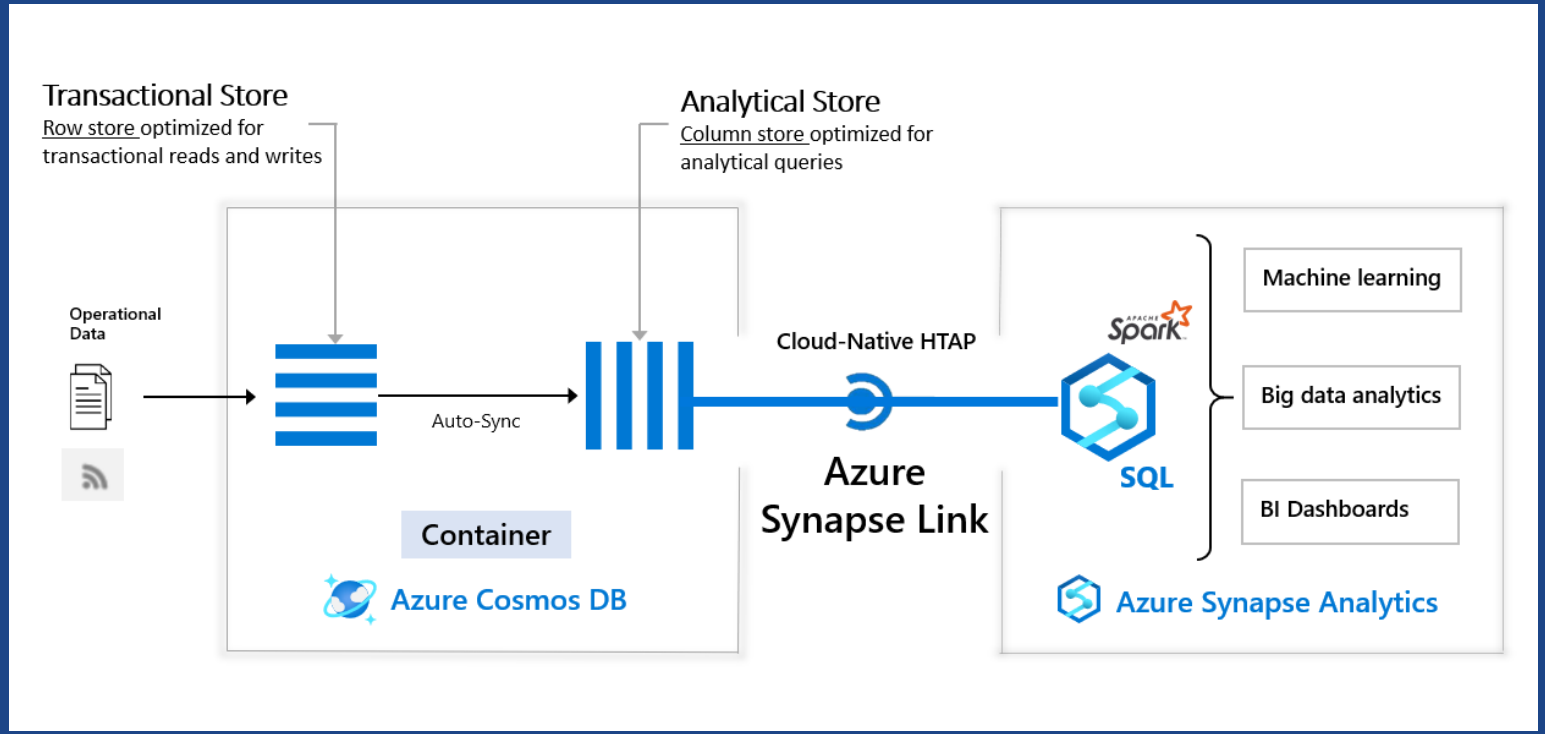
WHAT IS IT?

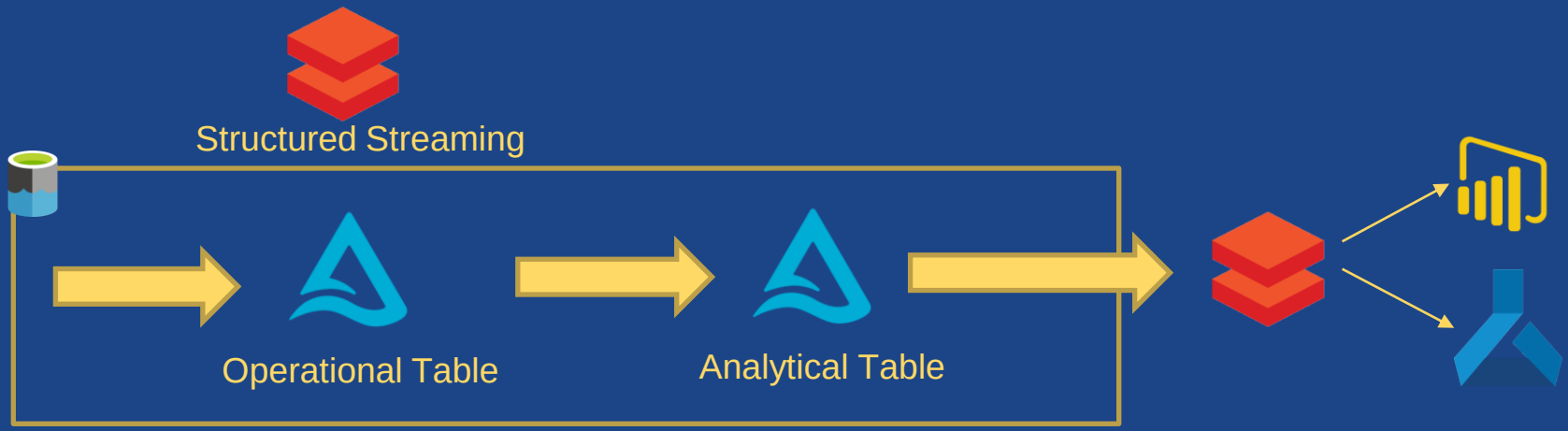
Bridges the gap into a single Architecture, enabling near real time analytical capabilities on transactional data

WHAT IS IT?

- Usually, a hybrid of large in memory datastores and datalakes.
- Can support complex analytical queries while low latency real time transactional updates.
- Highly scalable to handle demand.
- Is it a unicorn?

Microsoft Reference Version




Structured Streaming
Operational Table
Analytical Table

RECAP

Transactionally based

Handles analytical capabilities performantly on transactional systems

Bridge the Gap

Bridge the gap between Operational and Analytical Architectures

---

HIGH THROUGHPUT

Need to be able to handle high transactional throughput

HIGHLY SCALABLE

Need to be highly scalable independent of each capability

02

WHAT Challenges do they solve

A real example

- Client is a FTSE 100 company that owns a diverse business within a sector
- They own many brands each with their own systems, loyalty schemes and marketing.
- Wish to centralise all data into a common semi structured schema from their disparate sources to support analytics
- Need to use fuzzy matching algorithms to merge customers from different brands and information
- Want a real time view of every customer query able and updated via APIs (Customer and Business facing)
- Secure and restrict access to personally Identifiable information while providing masked data for analytics
- Need to be able to have complex analytical queries executed
- Need to deliver daily batch outputs to their marketing systems
- Need to support real time outputs to their Pricing Engines for personalising offers
- Want to develop a feature store to support machine learning and data science opportunities
- Needs to be highly available, scalable and have low latency that is geo redundant

The Vision

Brand A

Brand B

Brand C

Brand D

Brand E

Brand F

Social Media

Web

Loyalty Accounts

Loyalty Accounts

Business Accounts

3rd Parties

Operational Database

Communication Engine

Data Platform

Customer Relations

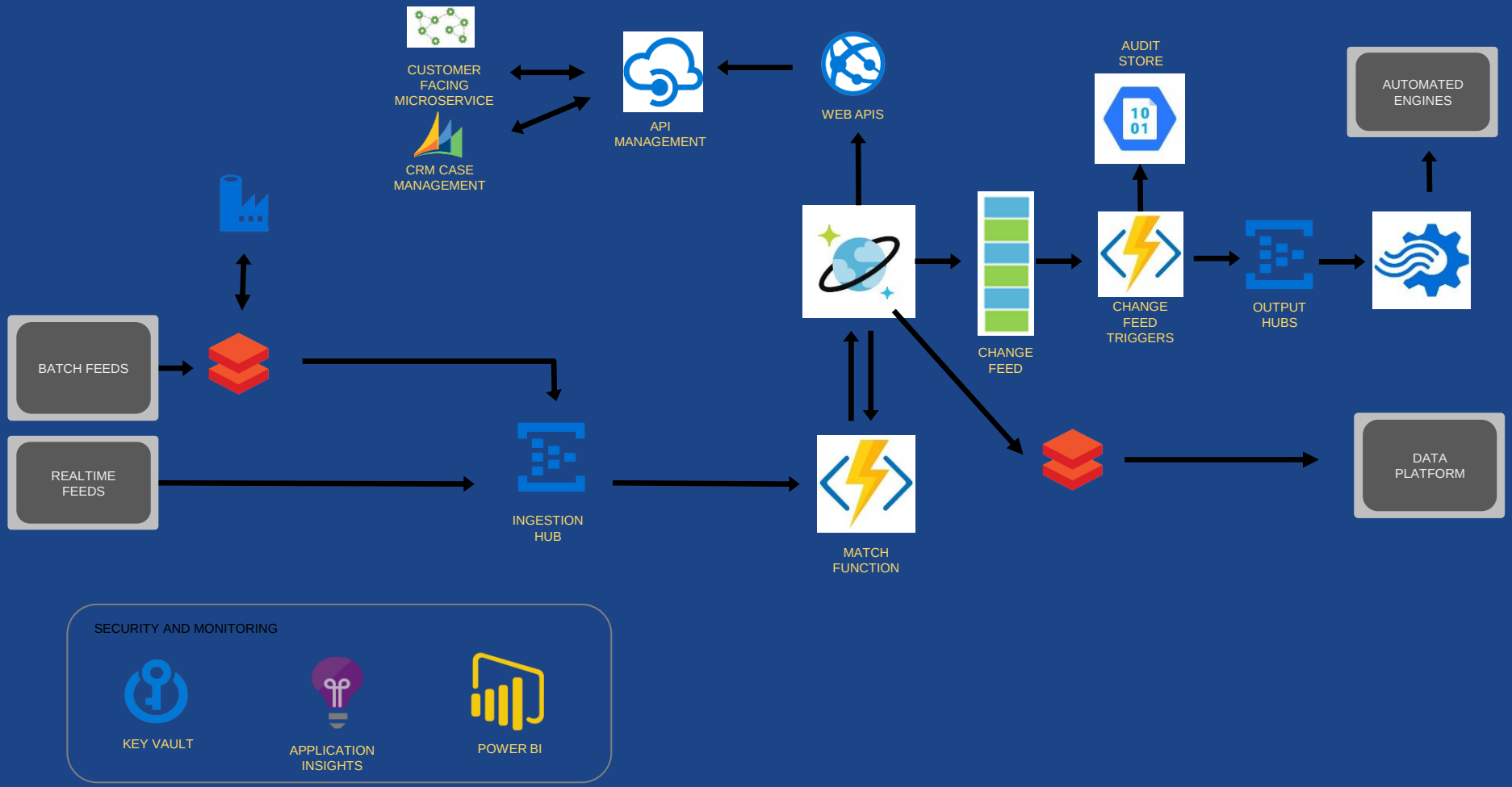
Social Media

Email Campaigns

App Notifications

Smart Displays

Tailored Offers


CUSTOMER FACING MICROSERVICE
CRM CASE MANAGEMENT
API MANAGEMENT
WEB APIS
AUDIT STORE
AUTOMATED ENGINES
BATCH FEEDS
REALTIME FEEDS
INGESTION HUB
CHANGE FEED
CHANGE FEED TRIGGERS
OUTPUT HUBS
MATCH FUNCTION
DATA PLATFORM
SECURITY AND MONITORING
KEY VAULT
APPLICATION INSIGHTS
POWER BI

03

Maturity?

Do Mature HTAPs Exist?

Monolithic by disguise?

- Centralising capabilities into a single product(s)
- Often expensive and proprietary?

Data Mesh?

- A decentralised, distributed domain specific data ownership and consumption architecture
- Enable Self Service
- Data As a Product

04 MORE INFO

MORE INFO

- **How to build HTAP workloads using Azure Cosmos DB and Azure Synapse Link -** https://arcade.sqlbits.com/sessions/?_gl=1*1fx7wnh*_gcl_aw*R0NMLjE2NDQzNDA0MjMuRUFJ4UIRb2JDaE1JcTIEcXM4M3c5UUIWNW9CUUJoMTJwUU1nRUFBWUFTQUFFZ0wtemZEX0J3RQ..#
- **COSMOS 101 -** https://arcade.sqlbits.com/sessions/?_gl=1*1fx7wnh*_gcl_aw*R0NMLjE2NDQzNDA0MjMuRUFJ4UIRb2JDaE1JcTIEcXM4M3c5UUIWNW9CUUJoMTJwUU1nRUFBWUFTQUFFZ0wtemZEX0J3RQ..
- Microsoft Learn HTAP Module -https://docs.microsoft.com/en-us/learn/modules/design-hybrid-transactional-analytical-processing-using-azure-synapse-analytics/
- Data mesh - https://martinfowler.com/articles/data-mesh-principles.html

THANKS!

Do you have any questions?
me@scottjamesbell.com
@fusionet24
www.myyearindata.com

Feedback Link

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.