Microsoft
Technology
Centers

Data Literacy Workshop Series

Dave Wentzel
   linkedin.com/in/dwentzel

Philadelphia MTC

Microsoft

Dave Wentzel

Decision Architect
Microsoft Technology Center
davew@microsoft.com

linkedin.com/in/dwentzel

Deloitte and Touche 7 years

Siemens Medical Solutions 5 years

Independent Consultant 4 years

CTO for data consultancy 2 years

Microsoft 4 years

Our People

The Architects of the MTC deliver immersive industry experiences and deep technical engagement focused on business outcomes. They help you understand the art of the possible and make it real by creating solutions to achieve business outcomes.

Architects

Todd Furst
Chief Technical Architect
Cross-workload specializing in *Azure-Infrastructure*
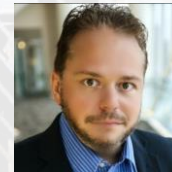Industry: Retail, Healthcare

Rich Ross
Principal Technical Architect
Cross-workload specializing in *Azure-App Dev*
Industry: Healthcare, Manufacturing

Dave Wentzel
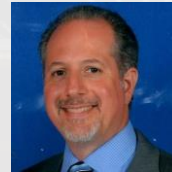*Principal Technical Architect*
Data and Analytics

Michael Mukalian
Principal Technical Architect
Cross-workload specializing in *Modern Workplace*
Industry: Financial Services, Healthcare

James Stento
Principal Technical Architect
Cross-workload specializing in *Biz Apps*
Industry: Media & Communications, Manufacturing

Understanding technology is less important than understanding data
- Pick the right tool for the user and use case

Self-service analytics initiatives are "underwhelming"

Your users' level of data literacy (the ability to find, work with, analyze, and "discuss" data is critical to building a self-service, *insights-driven* culture

These sessions (add'l 6-30 planned for 2022) are both tactical and strategic
- CoE → Prescriptive Analytics

It is my ambition to help you better integrate business analytics into the decision-making process, and brandish it for competitive advantage.

Our Process

Delivering the Right Experiences for our Customers

| Envision what's possible | Immersive engagement | Aligned solutions | Business outcomes |

Offerings

Design Thinking workshop
Explores the impact of digital transformation and innovation to help customers with vision-setting, strategy, roadmaps and organizational alignment.
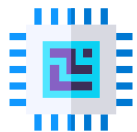
Strategy briefing
A strategic business and technical discussion to gain understanding of customer goals and challenges. Align Microsoft capabilities and solutions.

Architecture design session
Synthesizes the business and technical requirements for a solution including an initial scope and a high-level architecture to drive next steps.

Hackathon
A hands-on, intensely collaborative and inclusive sprint to determine the applicability of specific technologies against a set of business use cases.

Hands-on lab
A hands-on, immersive education experience to provide the skills and familiarity of a technology to enable solution development and adoption.

Rapid prototype
A tailored hands-on experience to demonstrate the key technical capabilities of a solution and address any challenges to accelerate decision making.

Customer Behavioral Analytics	Monday Feb 28 1-3pm EST	Without customers, your company is out-of-business. The most successful companies are using customer behavior data to make key business decisions. Many companies struggle with things like customer segmentation, loyalty, understanding Customer Lifetime Value, and "optimizing" churn. I'll bet your company has the data to tell which customers have churned and which might, but most struggle with "ok, now what do I do?" As data professionals, we are uncomfortable making opinionated recommendations on what we should do next based on what the data is telling us. In this session I'll show you how to use data and analytics processes to understand customer analytics issues and how to help your business leaders interpret their data to answer the question: "What should we do next?".
Marketing Analytics Using Demand Signal Repositories	Monday March 7 1-3pm EST	Most marketing analytics is still done in Excel. The reason is simple: Your marketing team wants to do complex analytics, but it's too cumbersome to do in your data warehouse. There's a better way. Data analysts need to be able to guide their marketing teams using Prescriptive Analytics techniques to show them "what do we do next". In this session we'll look at how to enable marketing teams to do "self-service analytics" (with the data team's guidance) by looking at real use cases like calculating Facebook campaign ROIs, performing RFM analytics, and understanding CAC (customer acquisition costs) in relation to churn, customer lifetime value, and product development. We want to show you how to build a Demand Signal Repository that goes viral inside your organization. This will be a hands-on session where we can create opinions using real data.

ETL is why Analytics Projects Fail: Here's a Better Way	Monday March 14 1-3pm EST	Traditional data projects spend a lot of time and resources copying data around the data ecosystem. This isn't fun for your data team, and it certainly doesn't add much value IF there is an alternative. In this session we'll show you some approaches to quickly acquire data sets, determine if they provide lift, and transform that data to add business value, quickly. You can leverage these techniques today to shorten time-to-value for any analytics project.
Latent Data Analytics	Monday March 21 1-3pm EST	It's likely that the most interesting data in your company is not being leveraged in your analytics. The fact is: the most valuable data your customer has is likely locked in unstructured data like Word docs and pdfs. In this session we'll show you how to crack that data and structure it to gain insights. We'll give you lots of use cases and ideas to spark your creativity.
Build the Corporate Decision Factory	Monday March 28 1-3pm EST	You've decided to be a more data-driven company. You are building data lakes, knowledge graphs, and data catalogs. You want to create better decision-making capabilities and focus less on the HOWs and more on the WHATs and WHYs but your team is spending a lot of time focusing on implementation bottlenecks that are derailing your transformation. At the Microsoft Technology Center we have concrete, repeatable processes that we've learned from helping our customers on their analytics journey. Simple questions like "how do I structure my analytics sandbox?" or "how can I store PII in a secure and compliant manner?" take months for many teams to implement. We want to share some patterns that you can use to shorten time-to-value and put your focus back on business outcomes.

The Problem

What we hear from IT …

"Our people aren't ready (or capable) of doing advanced analytics"

"We are being asked to do advanced analytics by our business, but we don't know how"

"ETL is causing our analytics project to run late"

"Our data is a mess; we need to clean it up before we do anything else."

"We don't have data scientists that can really mine our data."

"Our last EDW project failed"

"We'd like to integrate social media and 3rd party data, but don't know how"

"We can't align our data strategy with our business strategy"

"We have a self-service analytics initiative, but it isn't working"

What we hear from business users …

"I want to control my own data"

"The data warehouse data is nightly; I need real-time data"

"We want to do analytics in *our* system."

"We don't know how to democratize data and make it an enterprise-wide capability"

"We have lots of data, but don't know what to do with it"

"I just want to get my job done"

"The data warehouse doesn't answer my questions"

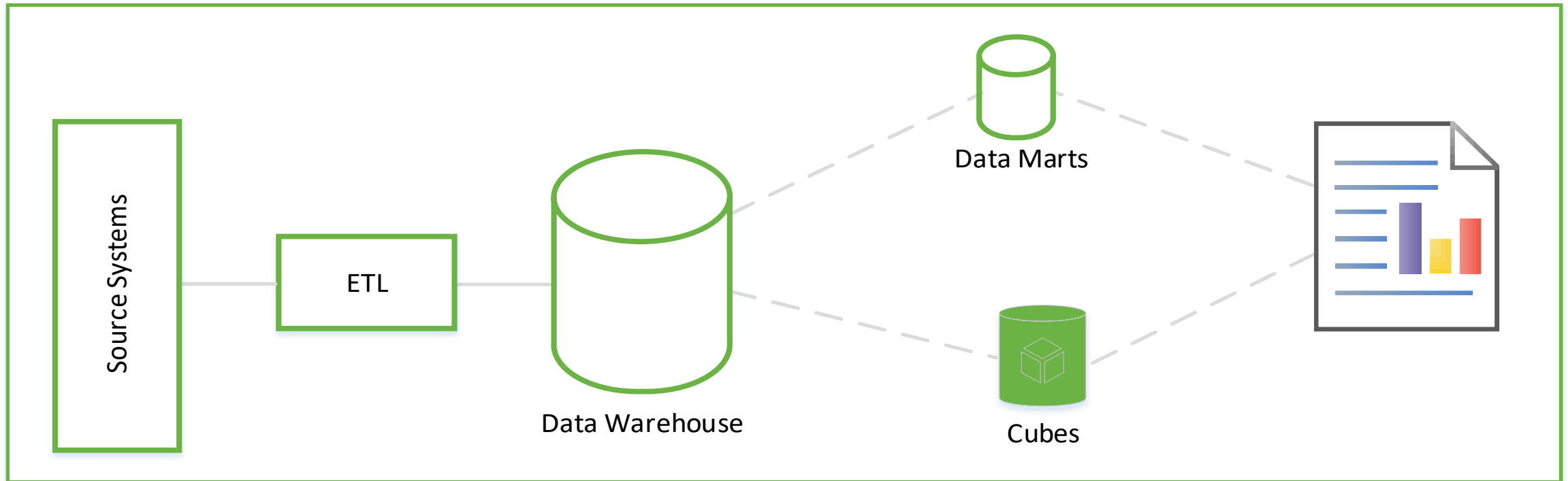"Even with the data warehouse and reports, I do analysis in spreadmarts"
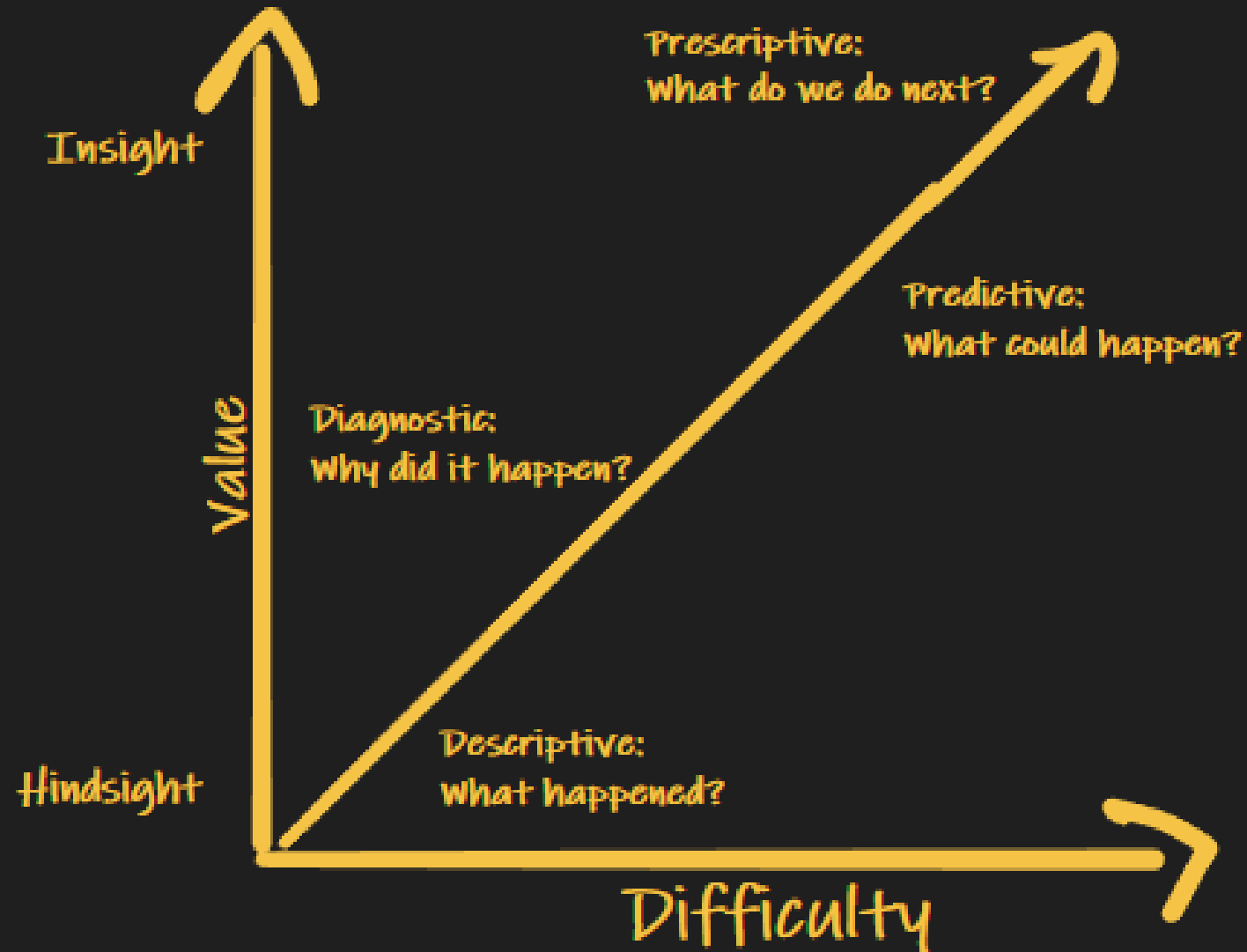
Is this how you do it?

Legacy Thinking

The Philosophy: Model data » Transform data » Load data » *Understand* data

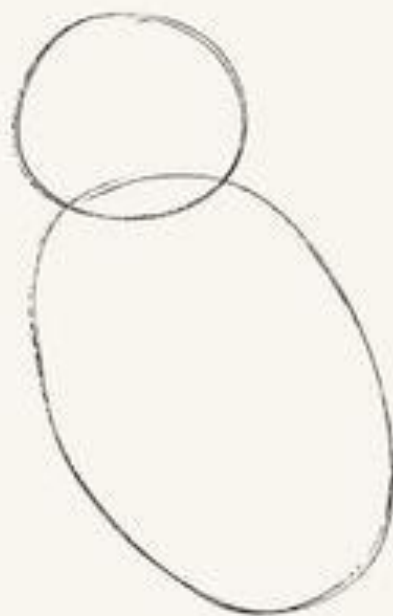# How to draw an Owl.

"A fun and creative guide for beginners"

Fig 1. Draw two circles

Fig 2. Draw the rest of the damn Owl
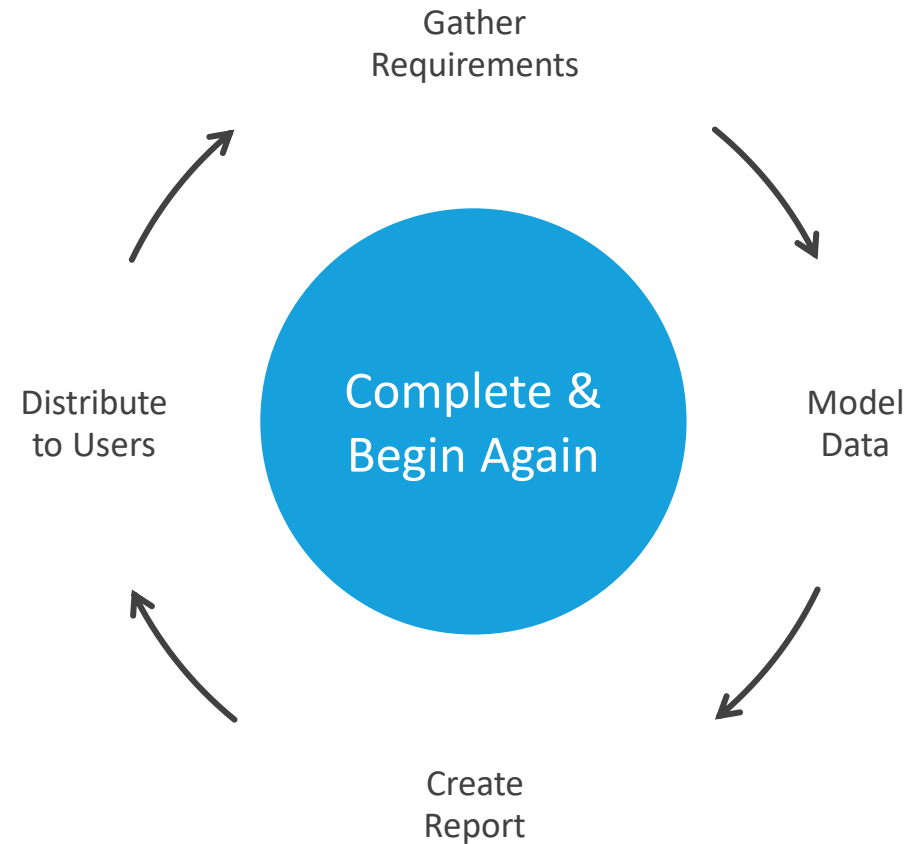
Data Projects have a high fail rate

Too much time is spent in:

- Requirements gathering
- Data modeling
- ETL

Users only see the fruits of the endeavor after the reports are created

In 2014, the Project Management Institute (PMI) released its Pulse of the Profession report. PMI found that "37 percent of all organizations reported inaccurate requirements as the primary reason for project failure."
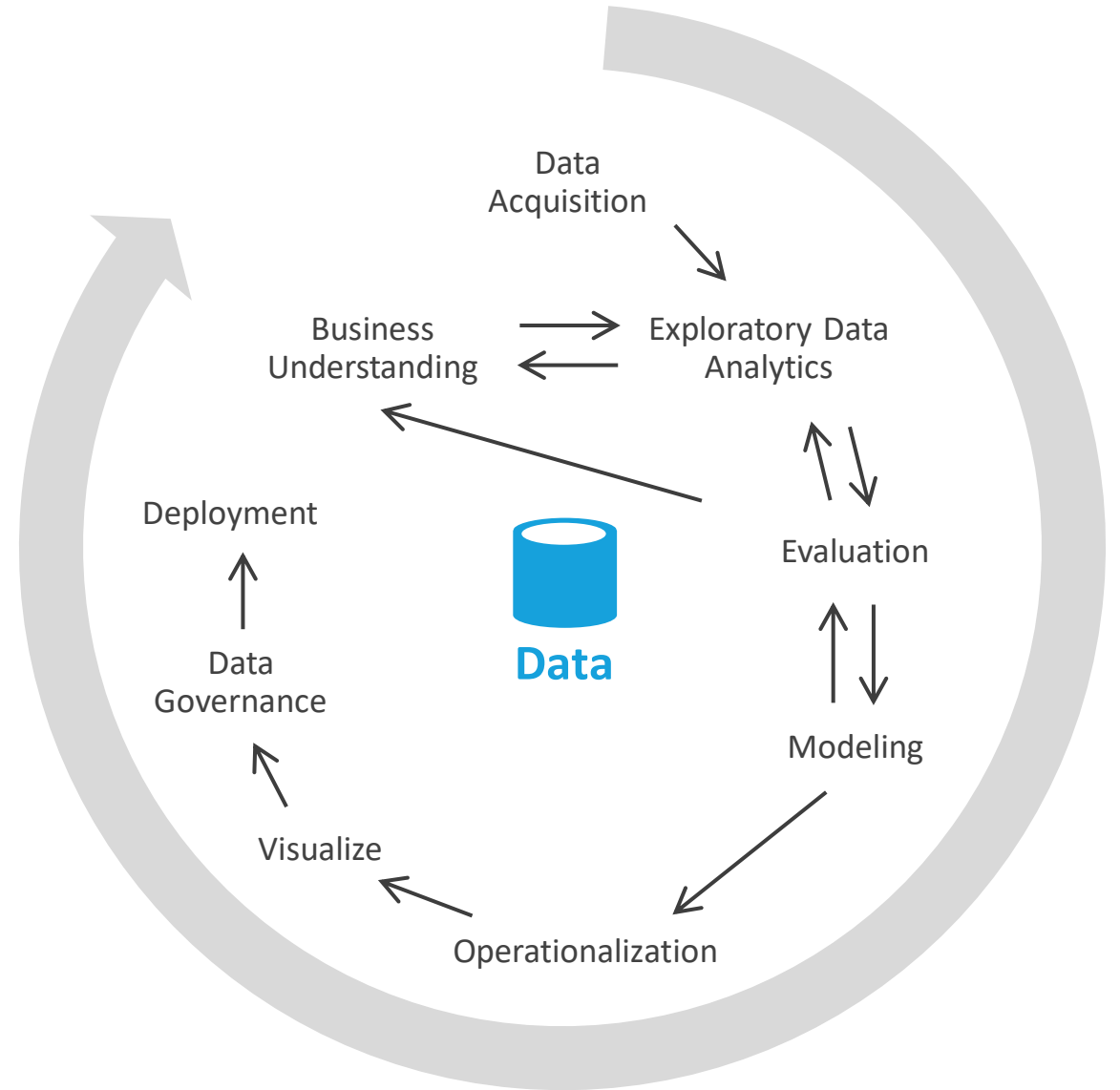https://www.pmi.org/learning/library/poor-requirements-management-source-failed-projects-9341

Gather Requirements

Model Data

Create Report

Distribute to Users

Complete & Begin Again

A Different Approach

Data Sandboxing/EDA

- A robust and well-proven methodology.

- Data science-like.

- Iterative.

- Stresses up-front understanding of data.

- Modeling is done later in the process (schema-on-read).

- ETL might not be needed

Real-World Example – Customer 360

Data Marts

Source Systems → ETL → Data Warehouse

Cubes

Ingest all Data	Store all data	Do analysis	Operationalize
Extract and Load, NO Transform	In native format	Using almost any tool	Create schemas and pipelines

Social

LOB applications

Relational

External Data

Operational Databases

Data Lake (raw data)

Transformation
Experimentation
Sandboxing

Data Lake (enriched data)

Real-time dashboards

Machine Learning

Enterprise Data Warehouse

Data Catalog

Physical Structure

Analytics workflow

Trusted

Staging

RAW

Sandbox
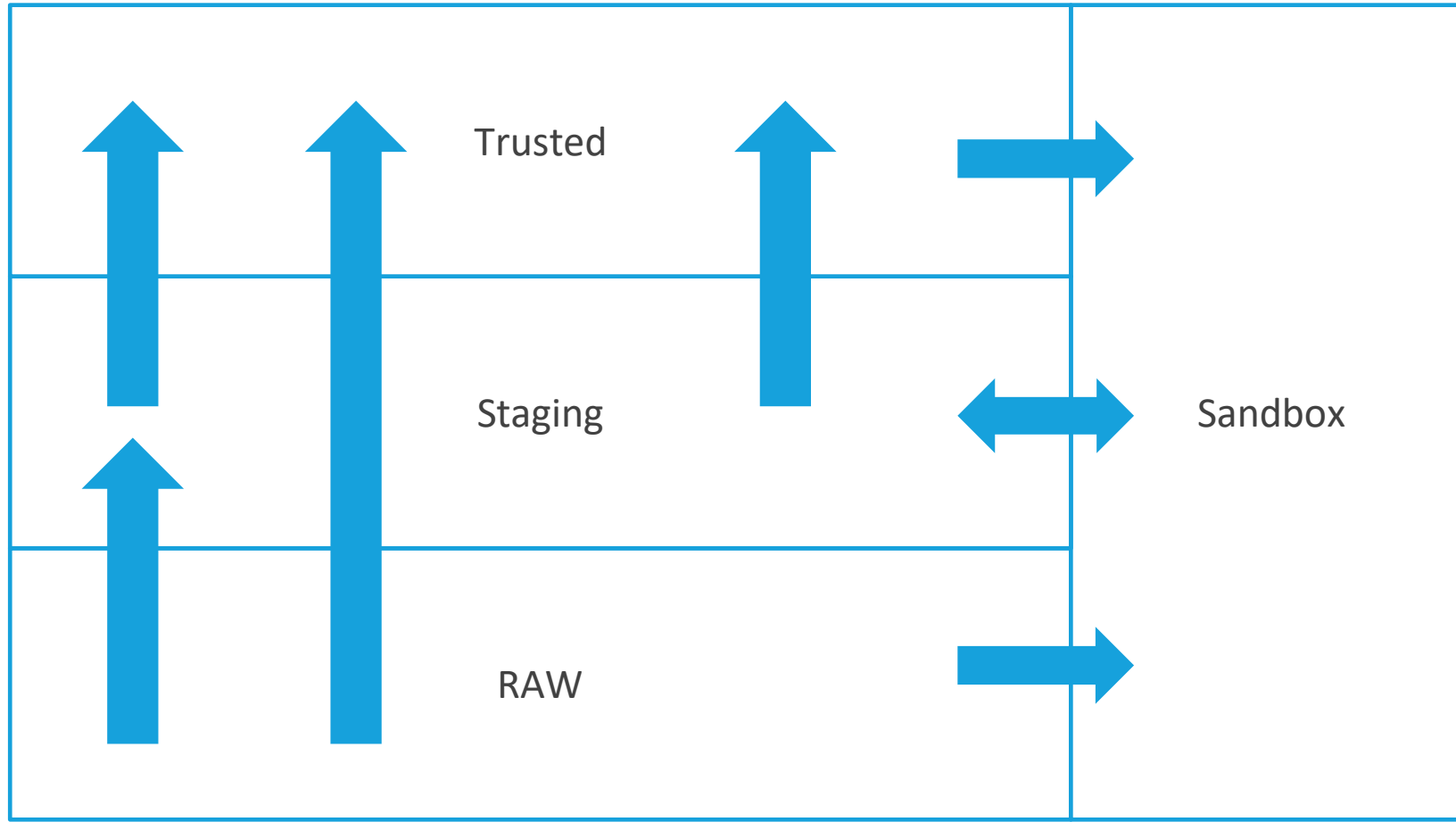
① Development / Data Ingestion

② Production

③ Data Scientists / Business Analysts/ Developers

④ Development

⑤ Production

Data Lake vs. Data Warehouse

Data Lake	Data Warehouse
Complementary to the EDW	Can be sourced from the Data Lake
Load first, understand later	Understand first, load later
Schema-on-read	Schema-on-write
System of Insight	System of Record
Detailed Data	Refined Data
Supports varied data formats	Structured data
Adapts to changing requirements	Difficult to change structure
Optimized for Cost	Optimized for Performance

Alternative Patterns to Traditional Data Warehouses

- Customer Data Platforms (CDP)
- PLM Software
- Demand Signal Repositories
- Behavioral Data Platforms
- traditional ERP/CRM software

- Let's talk about the analytics patterns that work best

Search   🔍

⊞ ☰ | Results Per Page: 10 ⌄ | Highlight ━● | Open In … ⌄   🗑 Delete

Searches

▼ Current Search

Save | Clear All

Source Type:
Azure Data Lake Store   ✕

▼ Filters

Tags

☐ earthquakes (4)
☐ datalaketest (3)
☐ Demo (2)
☐ datalaketest3 (2)

Object Type

☐ Directory (3)
☐ File (3)
☐ Data Lake (2)

Source Type

☐ SQL Server (4423)
☐ Filesystem (1551)
☐ SQL Server Analysis Services Tabular (396)
☐ SQL Server Analysis Services (140)

     see more

Daily Earthquake Records… ☑

Directory represents logical data set. Contains one file per day.

Experts:
amitkul@microsoft.com

earthquakes

Contained In Data Lake:
adlsadc.azuredatalakes…

📁 ⚡

📂 **DATA LAKE DIRECTORY**

Open In …   ⌄   Explore Data Lake ❯

region_names.asc

click tile to add a description…

Experts:
amitkul@microsoft.com

earthquakes

Contained In Data Lake:
adlsadc.azuredatalakes…

📁 ⚡

📄 **DATA LAKE FILE**

Open In …   ⌄   Explore Data Lake ❯

AzureDataCatalogErrorEve…

click tile to add a description…

Experts:
linm@test.com

Demo   datalaketest

test1

click tile to add a description…

Experts:

datalaketest3

❯ Properties

▦ Preview   ⊞ Columns   📖 Docs

Name:
DailyEarthquakes

Friendly Name:

Daily Earthquake Records from USGS

Description:

Directory represents logical data set. Contains one file per day.

Experts:

amitkul@microsoft.com

Add…

Tags:

earthquakes

Preview   Columns   Documentation

We have a winning idea, help us operationalize it
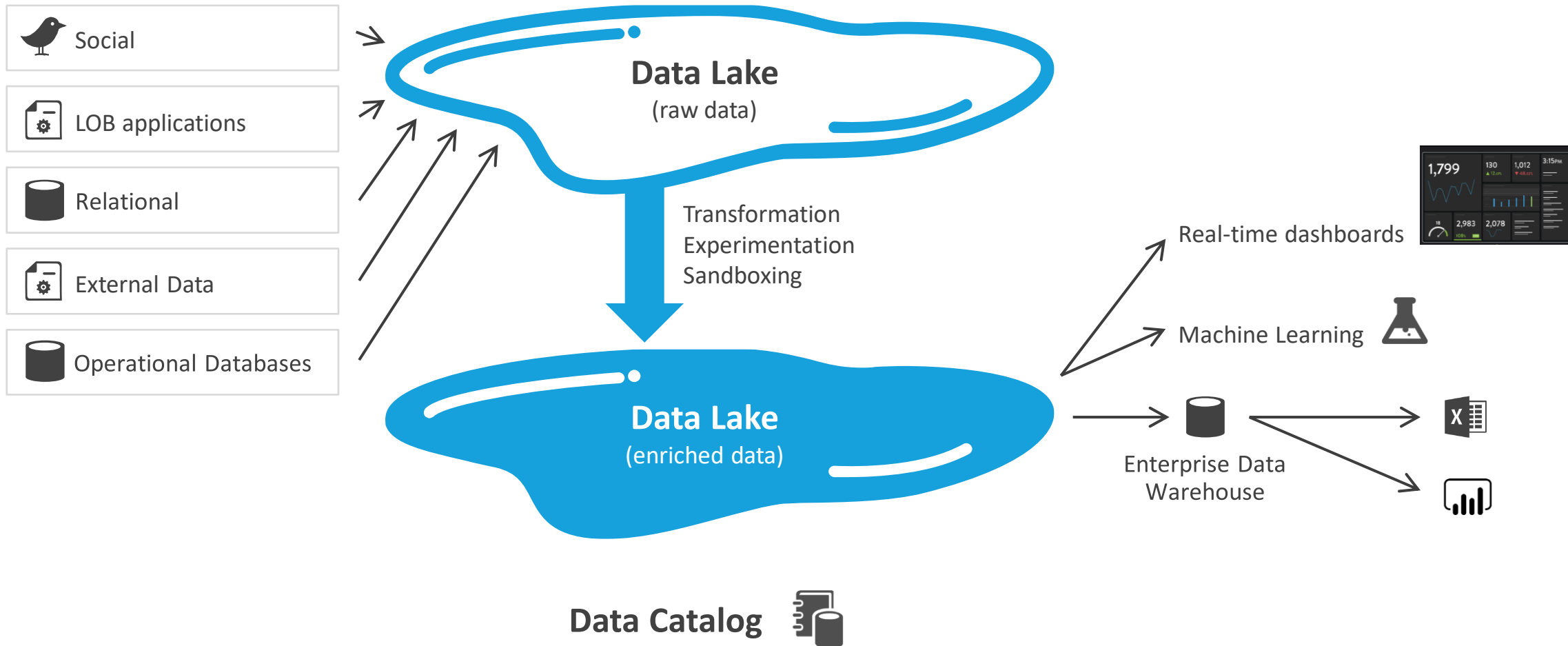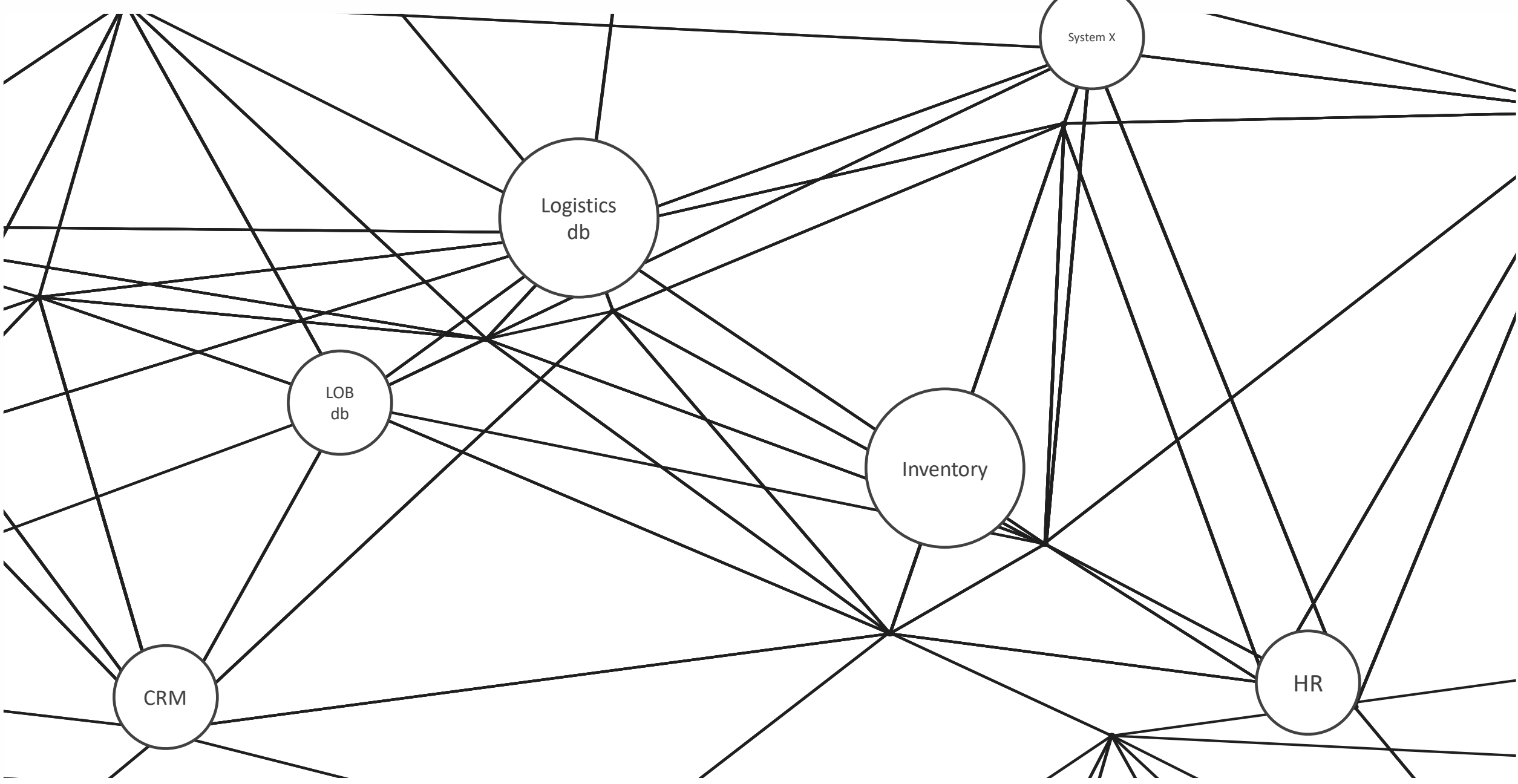
Data Ingestion Factory
(Kappa Architecture)
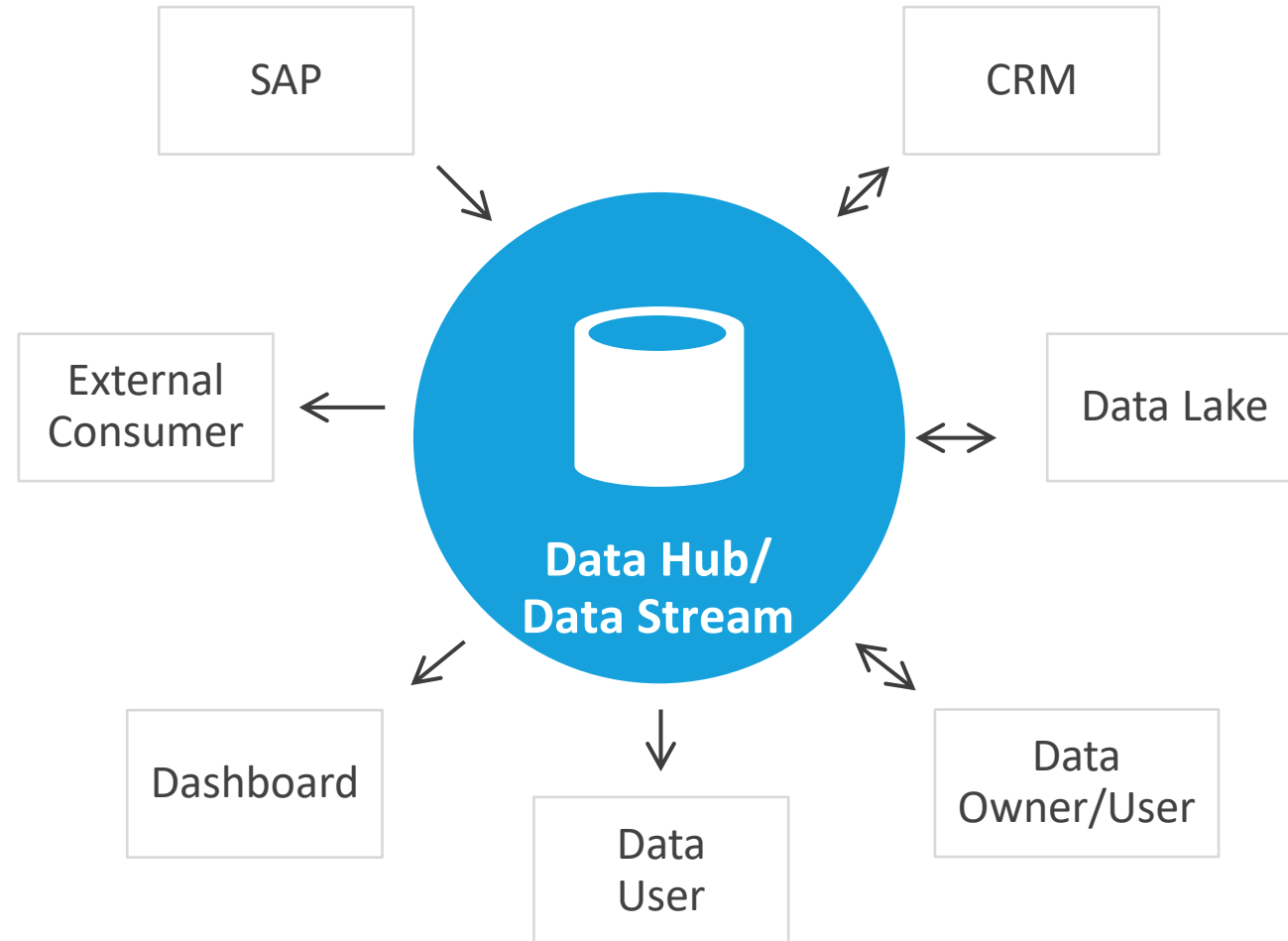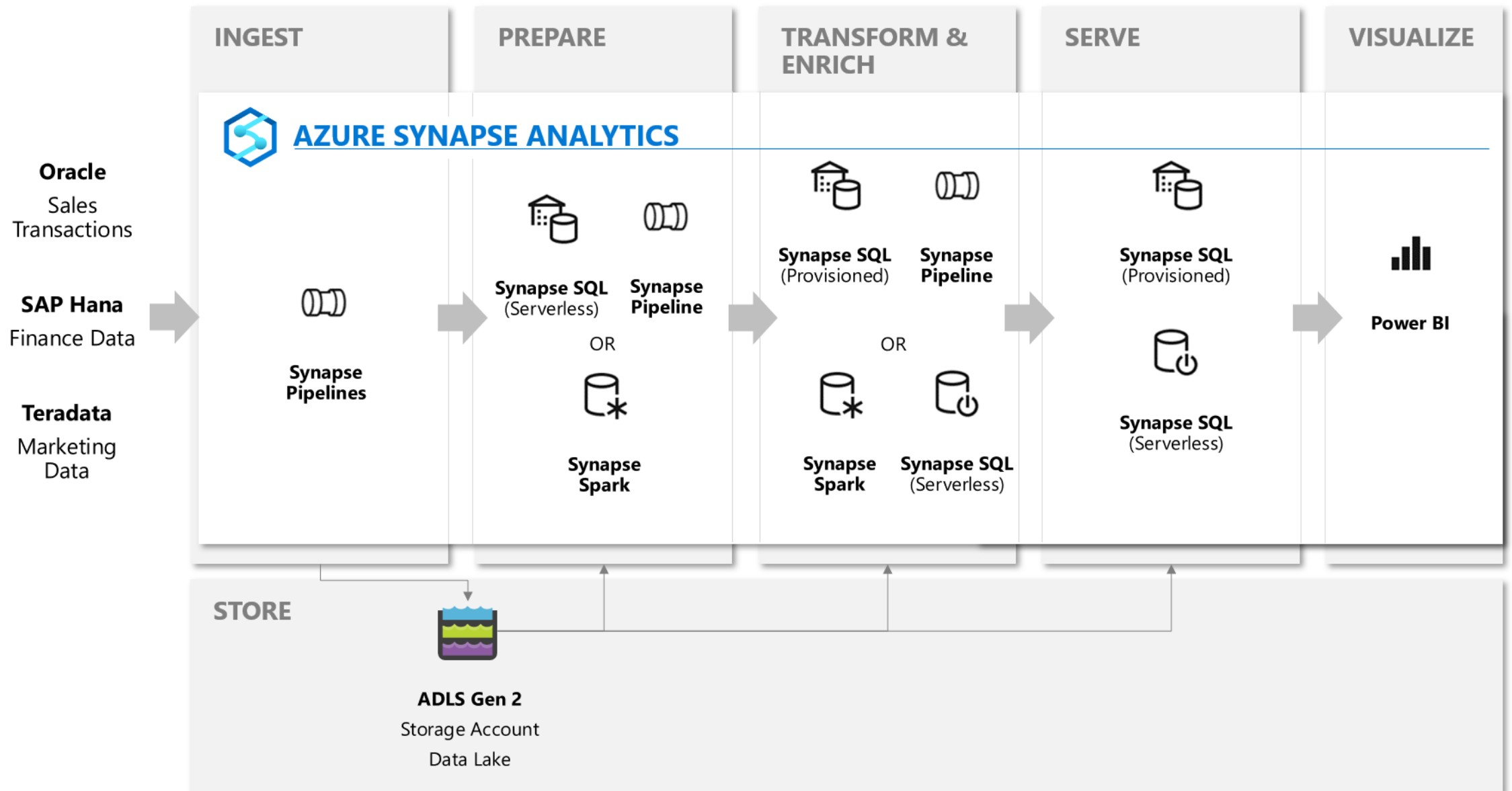
Handling Streaming Data

Twitter
Tweets

WebJob
Twitter Feed
Consumer

Event Hubs
Streaming
Data

**Stream
Analytics**
Process Data

ADLS Gen 2
Storage Account
Data Lake

Synapse SQL
(Provisioned)

Power BI
Dashboards
& Reports

IoT Sensors
In-store Traffic
Data

IoT Hub
Streaming
Data