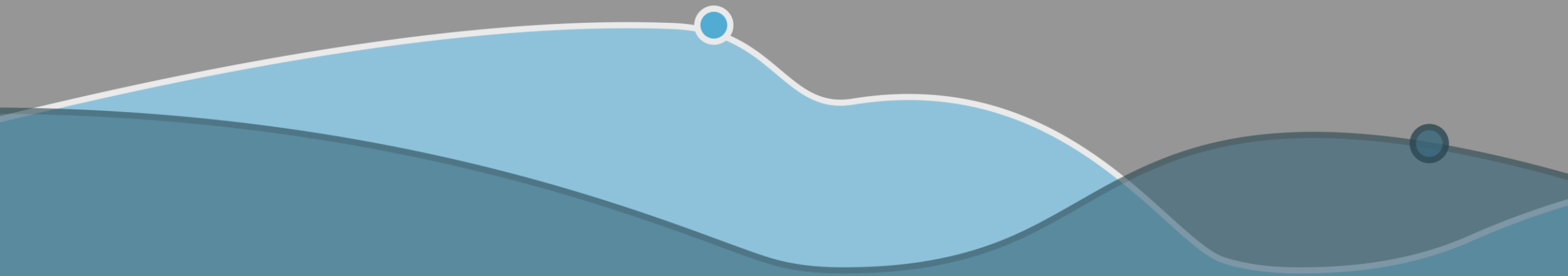




# Cortana Analytics Workshop

Sept 10 – 11, 2015 • MSCC



# *New Services in the Cortana Analytics Suite*

## *Azure Data Catalog*

Julie P. Strauss

Principal Group Program Manager

Information Management & Machine Learning Group

# Do You Recognize Any of These Challenges?

Users spend more time **looking** for data, than they do **analyzing** it

Data is sitting in **multiple sources**, but no insight into which data sits where

Many different **data ecosystems**, but no way to share data artifacts across them

Need data consumption in multiple **different tools**, but no common way of enabling discovery and access to data sources across them

Users are busy **re-producing** data assets that already exist



# What is Azure Data Catalog?

An *enterprise-wide* catalog in Azure that enables self-service discovery of data from any *source*

A metadata repository that allow users to *register*, *annotate*, *discover*, *understand*, and *consume* data sources

# Cortana Analytics Suite

## Transform data into intelligent action



# Where do I use the Azure Data Catalog?

DATA SOURCES

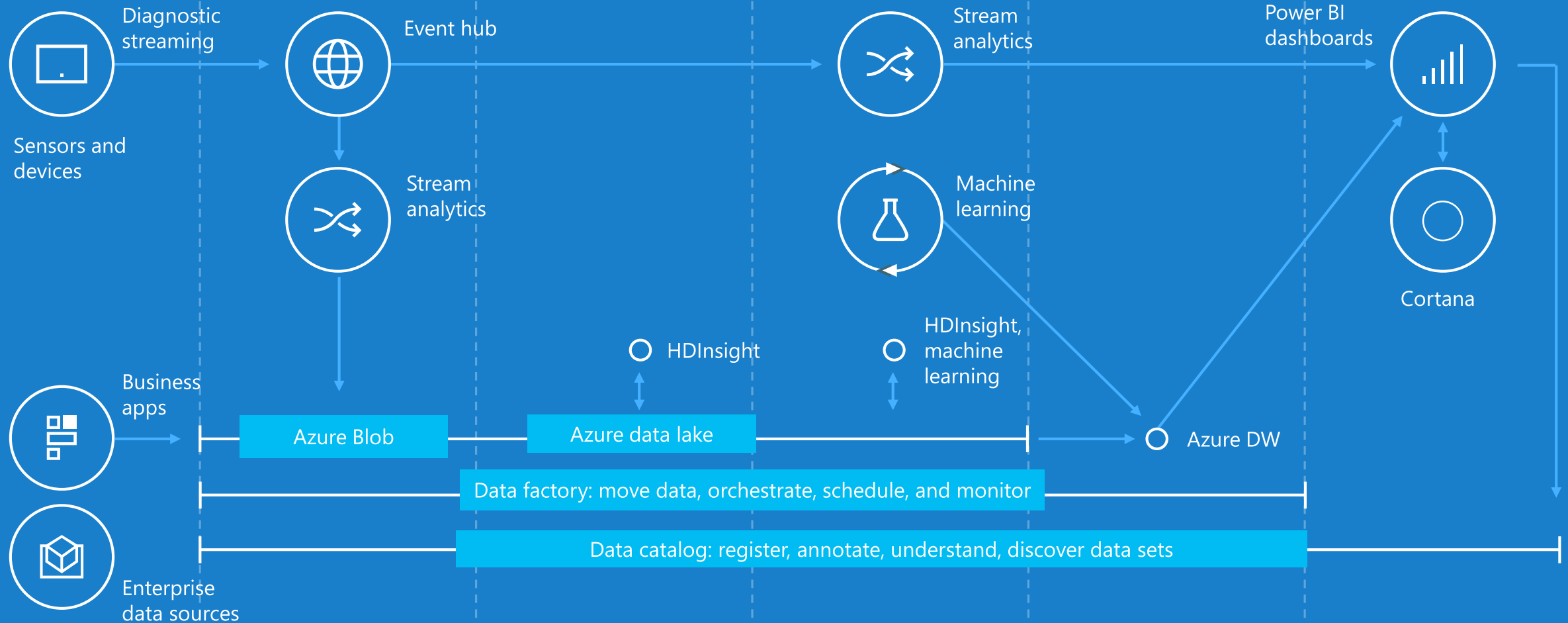
▶ INGEST

▶ PREPARE

▶ ANALYZE

▶ PUBLISH

▶ CONSUME



# Where do I use the Azure Data Catalog?

DATA SOURCES

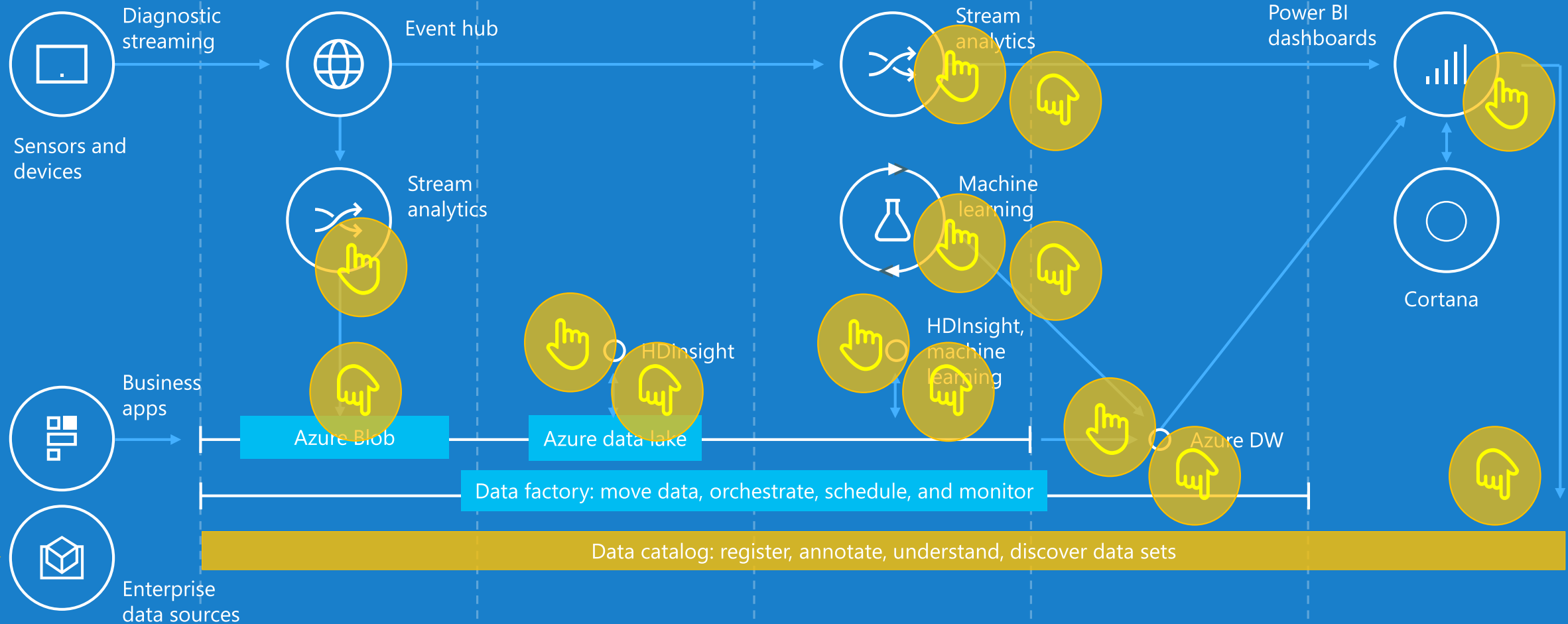
▶ INGEST

▶ PREPARE

▶ ANALYZE

▶ PUBLISH

▶ CONSUME



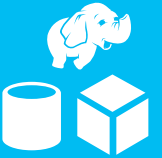
# How is Azure Data Catalog Different?

## Data **source** discovery



One-stop shop for all enterprise data sources  
No data movement or heavy up-front investment  
Time to value in minutes

## Data from **any** source



Cloud and on-premises data sources  
Structured and unstructured data sources  
Microsoft and non-Microsoft data sources

## Consumption through **any** tool



Enabling publishing, discovery and consumption  
of data sources through your tool of choice

## Annotation **crowdsourcing**



Empowering any user to capture and share  
their knowledge about data sources and  
their usage



# What Can I Do With It?

Publisher

Consumer

IT Admin

Publish

Register Data Sources

Discover

Browse - Search

Govern

Apply Policies - Control Access

Enrich

Categorize –  
Annotate

Understand

Get context – Identify Intent

Analyze

Track and monitor usage

← *Extend* →

Use open REST APIs to integrate data publishing and discovery into existing applications and processes

# Interested in Providing Feedback?

Contact: [Julie.Strauss@Microsoft.com](mailto:Julie.Strauss@Microsoft.com)

# Interested in Trying it out?

Go to: [AzureDataCatalog.Com](https://AzureDataCatalog.Com)



# Appendix



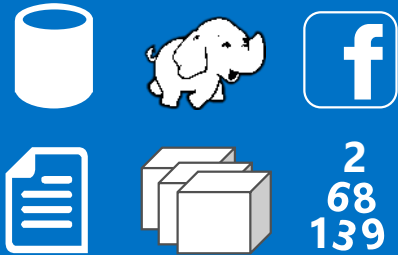
# Azure Data Catalog Programmability

- Programmability Scenarios
  - Incorporating data source discovery into client applications
  - Future: Defining custom data source and data asset types ("type extensibility")
- REST APIs
  - Register data asset operations: Register Data Asset, Delete Data Asset, and Search Data Asset
  - Annotate data asset operations: Annotate Data Asset, Update Annotation, Get Data Asset with Annotations, and Delete Annotation
  - Search API uses familiar, powerful, search syntax

# Enabling the Entire Enterprise Data Ecosystem

*Fostering a collaborative virtuous cycle for the Data Value Chain*

## The Enterprise Data



## The Public Data

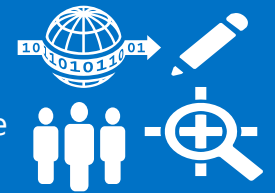


## DATA PRODUCERS



- Refine
- Combine
- Produce
- Annotate
- Share

## DATA STEWARDS



- Discover
- Combine
- Clean
- Annotate
- Publish/Share
- Certify
- Categorize
- View usage analytics and feedback

## INFORMATION CONSUMERS



- Discover
- Understand
- Connect
- Consume

## IT ADMIN

- Provision
- Configure



- Secure
- Monitor
- Operationalize