

# Modern Data Warehouse

Ivan Kosyakov

MTC Technical Architect

<http://biz-excellence.com>

November 16, 2016

# Business Trends

Top Manager



IT



Big Data  
(3v)

Strategic  
Analytics

Business Unit  
Analytics

Analyst



Business User



End User  
Analytics

Advanced  
Analytics

Data Scientist



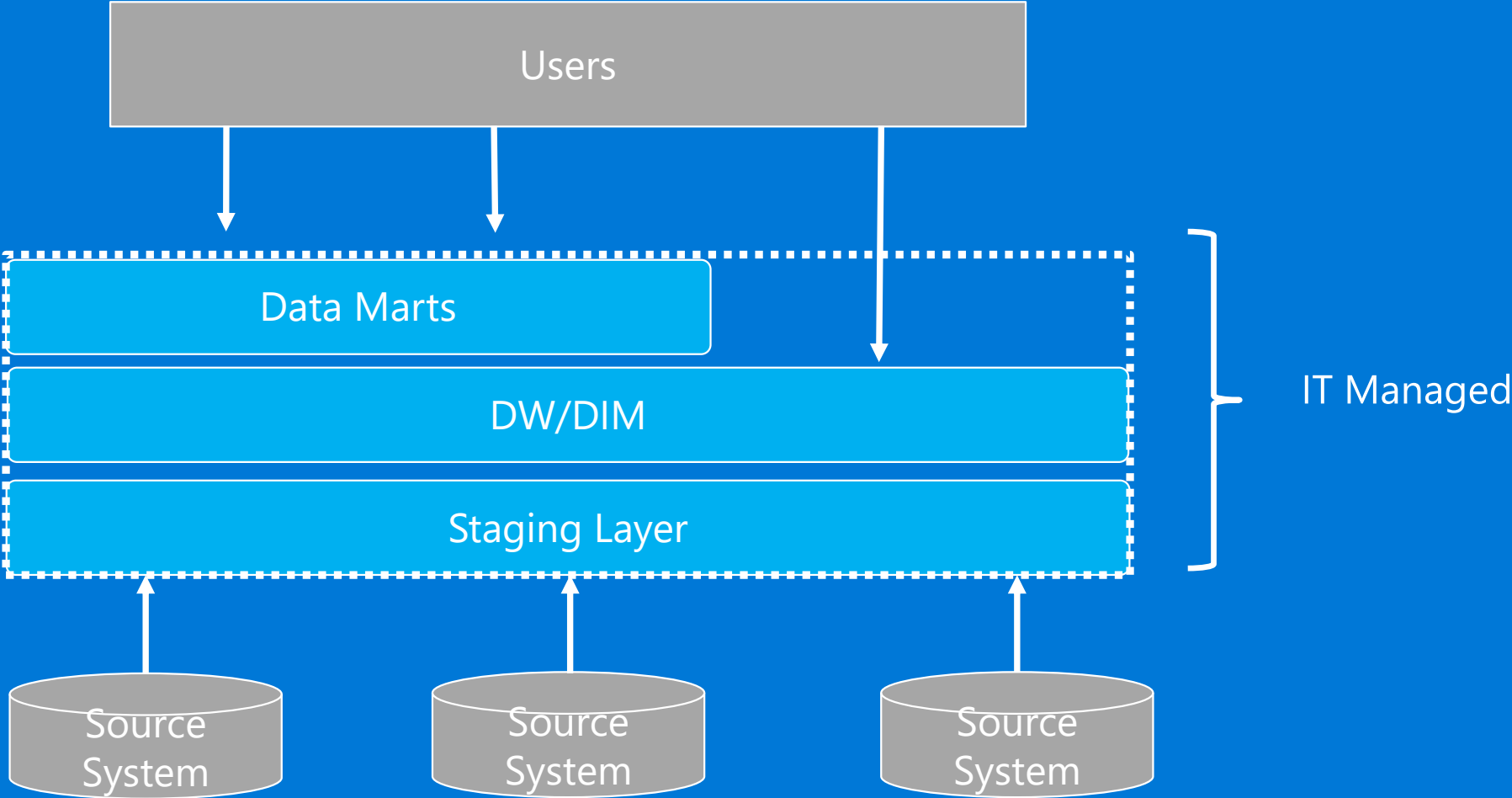
# Technology Trends

- The Cloud
- Scale-Out Repositories (e.g. MPP, Hadoop)
- NoSQL and Unstructured Data Repositories
- Columnar Databases
- In-Memory Analytics
- Data Virtualization/Query Federation
- Machine Learning and Advanced Analytics

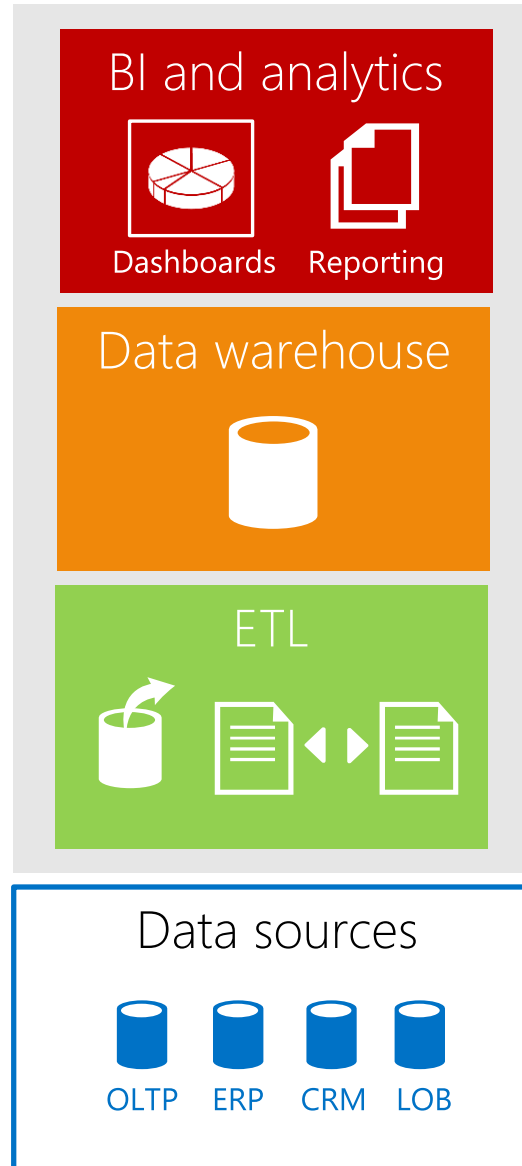
# Traditional Enterprise Data Warehouse

- Vision: Single Source of Truth for The Enterprise
  - All repositories are Relational
  - All data transformation is ETL
  - All users access *only* cleaned and modeled data

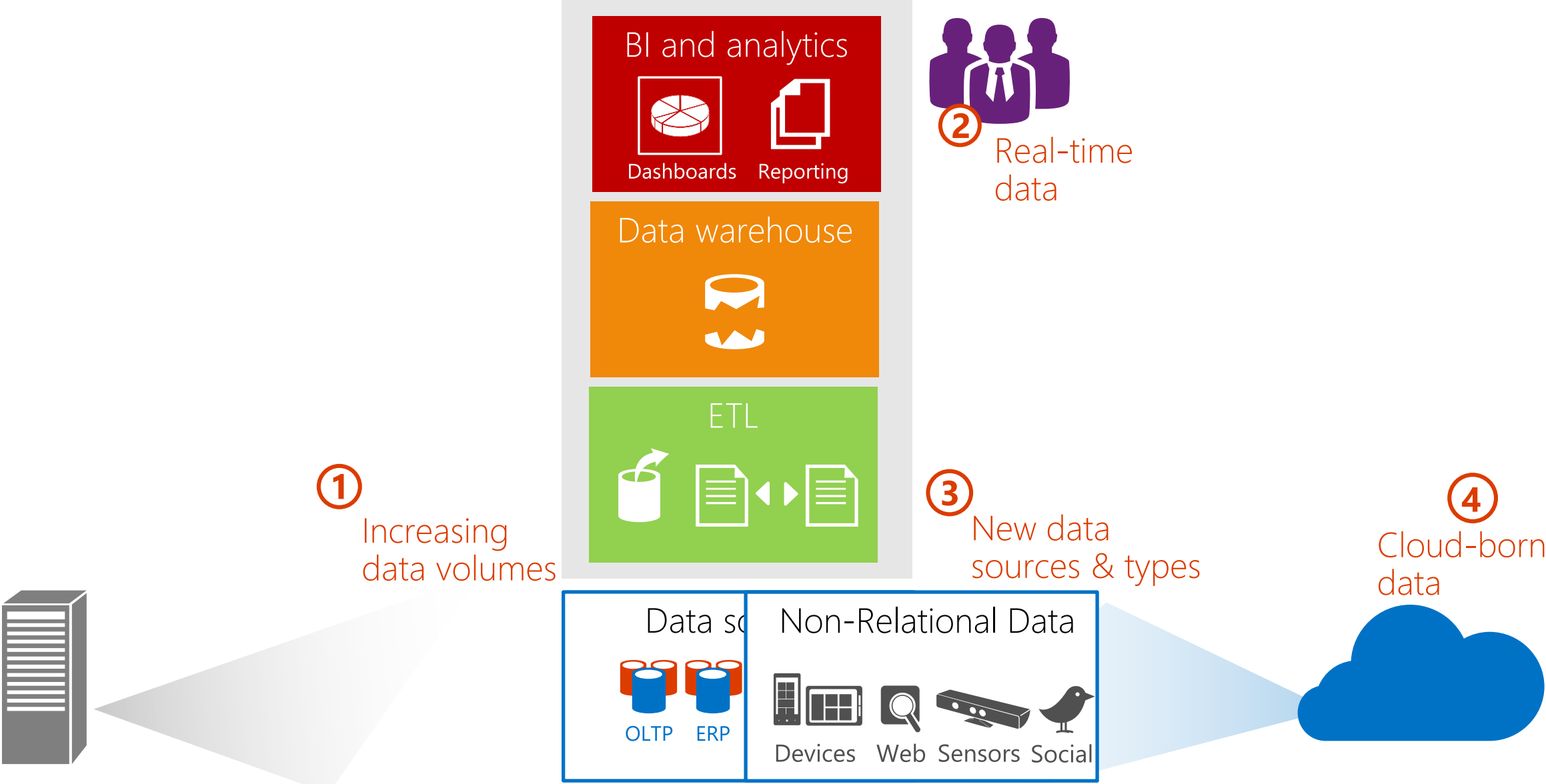
# Traditional Enterprise Data Warehouse Logical Architecture



# Traditional Data Warehouse



# Traditional Data Warehouse



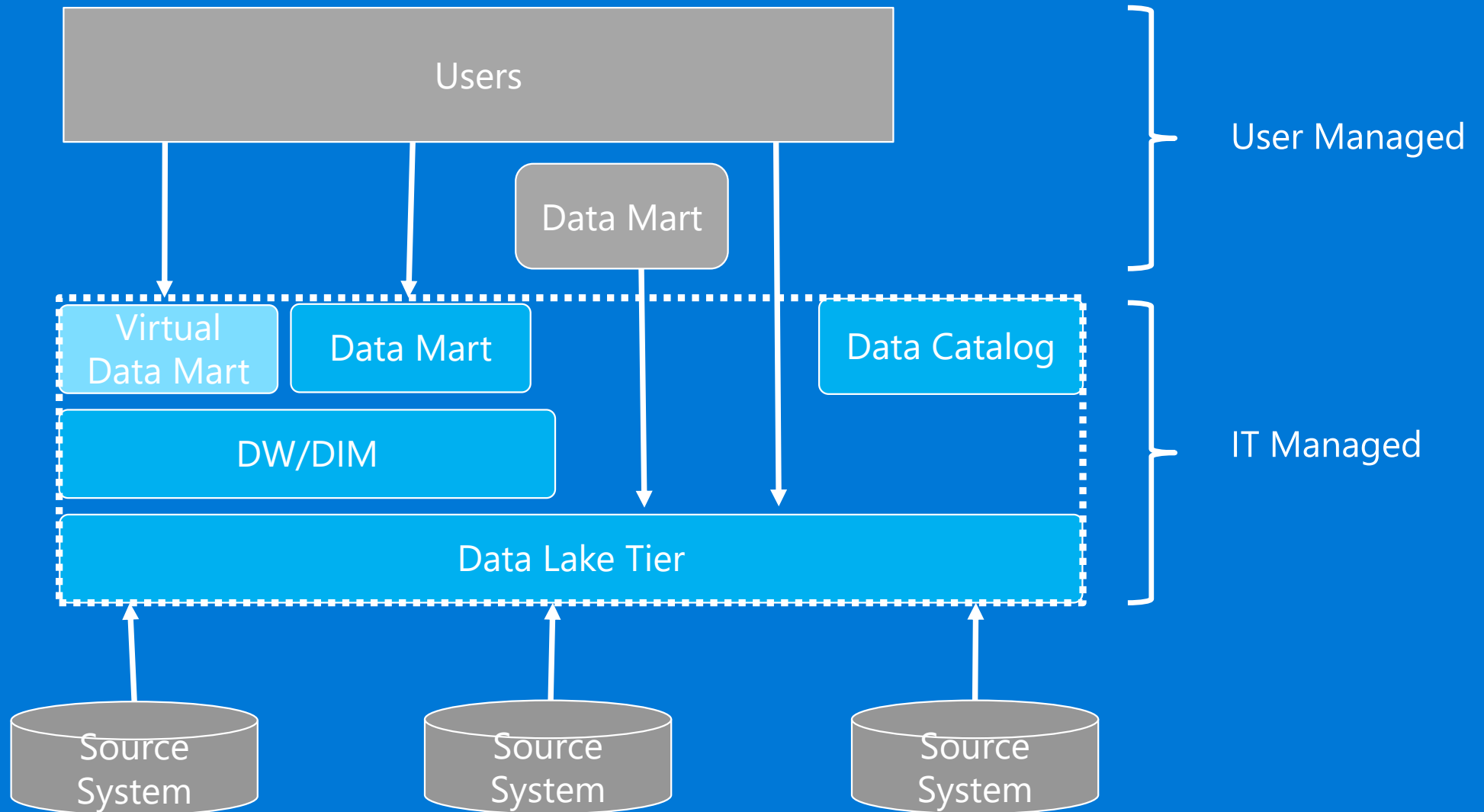
# Modern Enterprise Data Warehouse

- Mix of Relational and non-Relational
- Some Data Virtualization
- Raw Data Access *and* IT-Designed Data Marts
- Supports:
  - Data Scientists
  - BI Professionals
  - End Users



# Modern Enterprise Data Warehouse

## Logical Architecture



# Modern Data Warehouse Components

# Data Lake

- Raw Data: minimal modeling and data cleansing
- Optimized for quickly onboarding new data sources
- Optimized for data extract (vs Query Processing)
- Provisioned access for end-users published in a Data Catalog
- May span multiple physical repositories and technologies

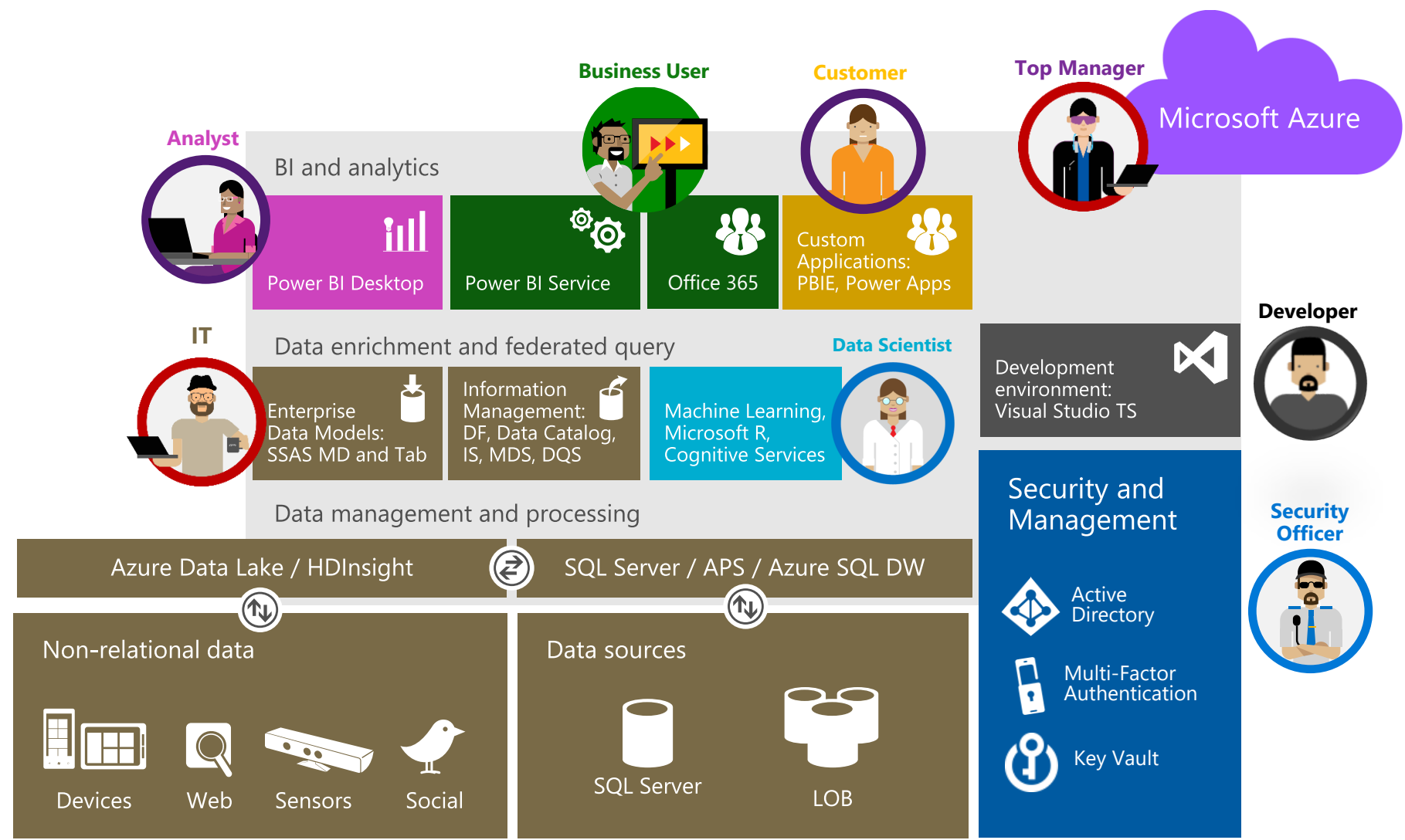
# Data Marts / Models

- Fully cleaned and modeled data
- Narrow scope
- May be built by IT or by end users
- May span multiple physical repositories and technologies

# Data Catalog

- Many Repositories, but one Data Catalog
- Publish Data Sets and Data Marts
- Discoverable by Analysts and End Users
- Technical and Business Metadata

# Reference Architecture



Questions