

Pragmatic Business Intelligence

Steen Dybboe



Pragmatic BI



Agenda

- How did I end here
- Motivation
- Architecture
- Technologies
- Modelling
- The real world



Who am I

- BI Consultant
- Mostly a developer
- Full Stack (On-Premise and Cloud)
- My first professional database : Netware SQL, 1991
 - SQL Server 6.5
- No Machine Learning
- Lots of family, Lots of golf, lots of interests
- No more study 😊



Experience

- TDC
- Vestas
- Private Banking
- Energy
 - Electricity
 - Evida (gas)
- Public services







Gartner on Analytics and business intelligence


Analytics and business intelligence (ABI) platforms

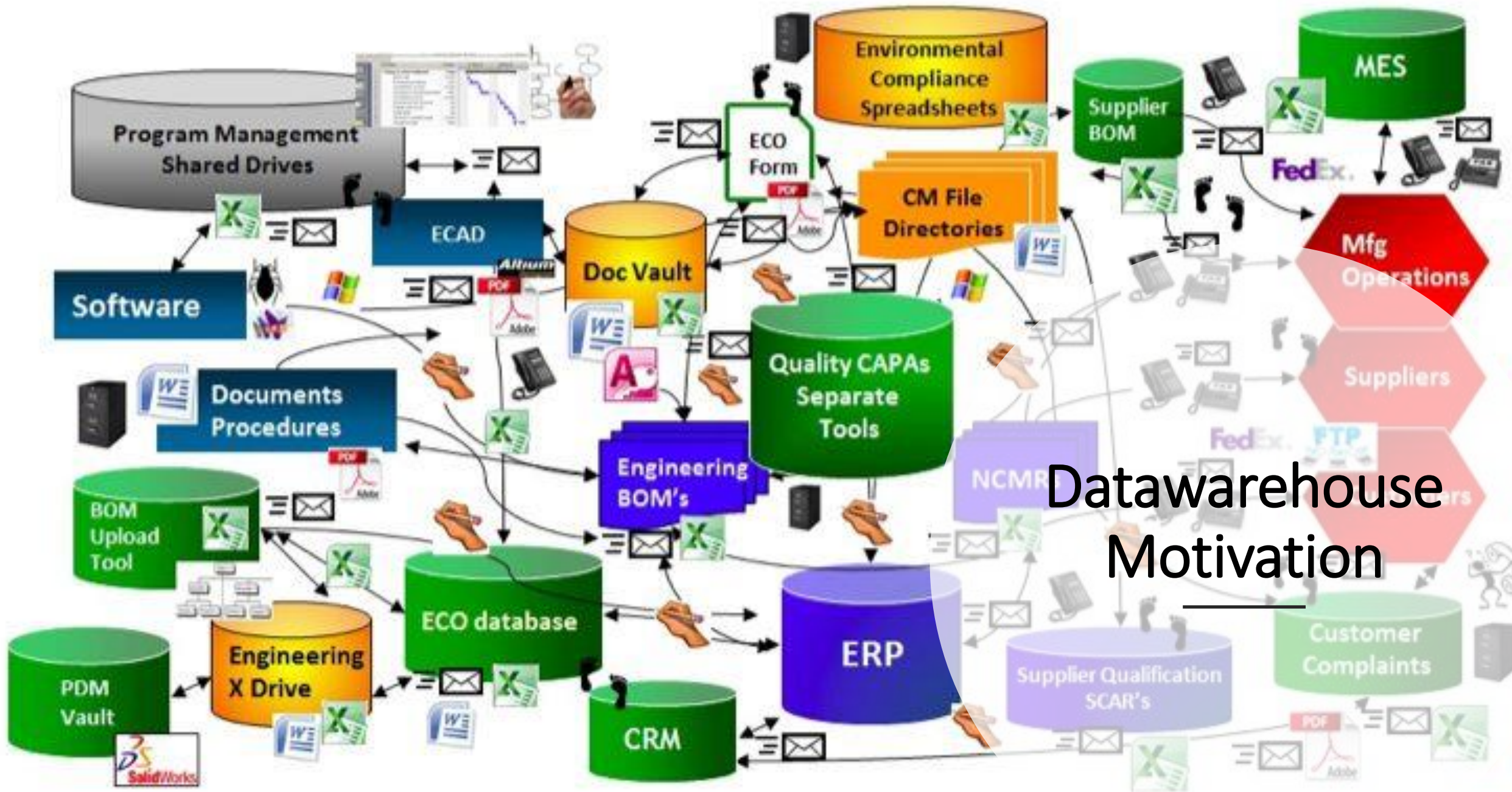
- Enable less technical users, including businesspeople,
- To model, analyze, explore, share and manage data, and
- Collaborate and share findings,
- Enabled by IT and
- Augmented by artificial intelligence (AI).

- ABI platforms may *optionally* include the ability to create, modify or enrich a semantic model including business rules.



Motivation

- Evidence-Based Management
 - Data is resource and must be made available
 - Move away for silo thinking
 - Ability to gain insights
 - Business Operations Visibility
 - Improve efficiency
 - Competativ Advantage
- 



Datawarehouse
Motivation

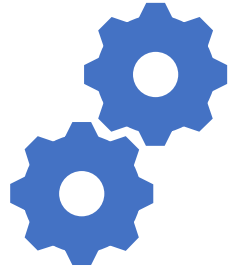




Data Consolidation



Datawarehouse Architecture



Classic Datawarehouse

ETL – Extract Transform and Load

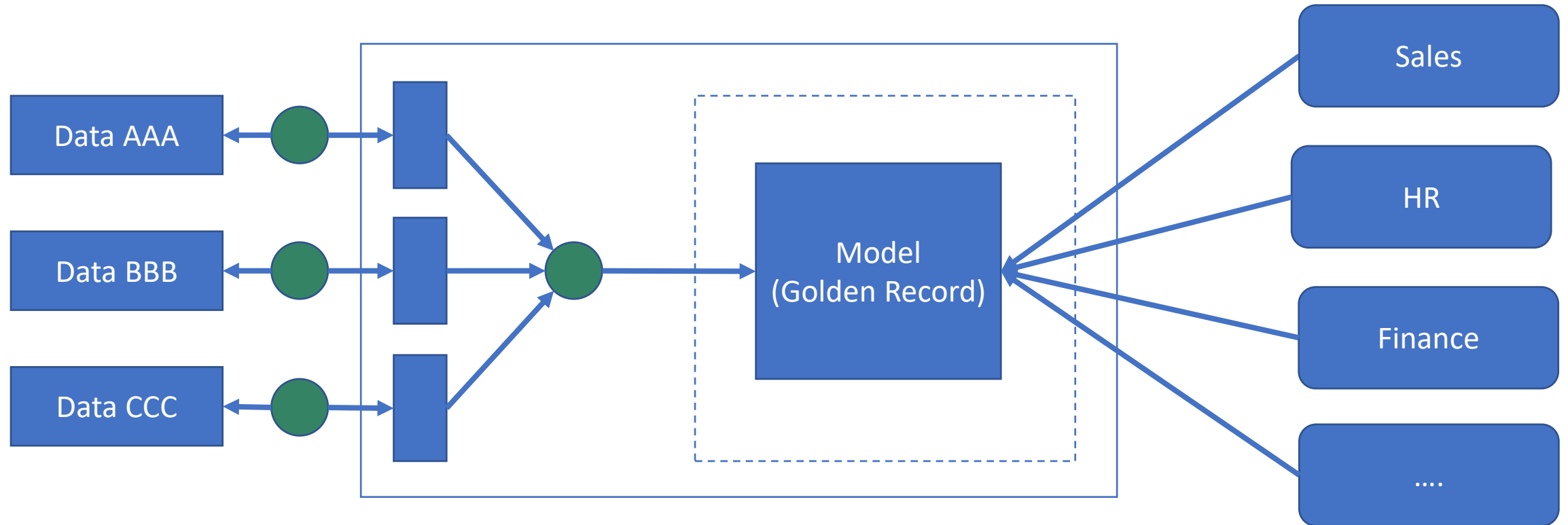


Modern Datawarehouse

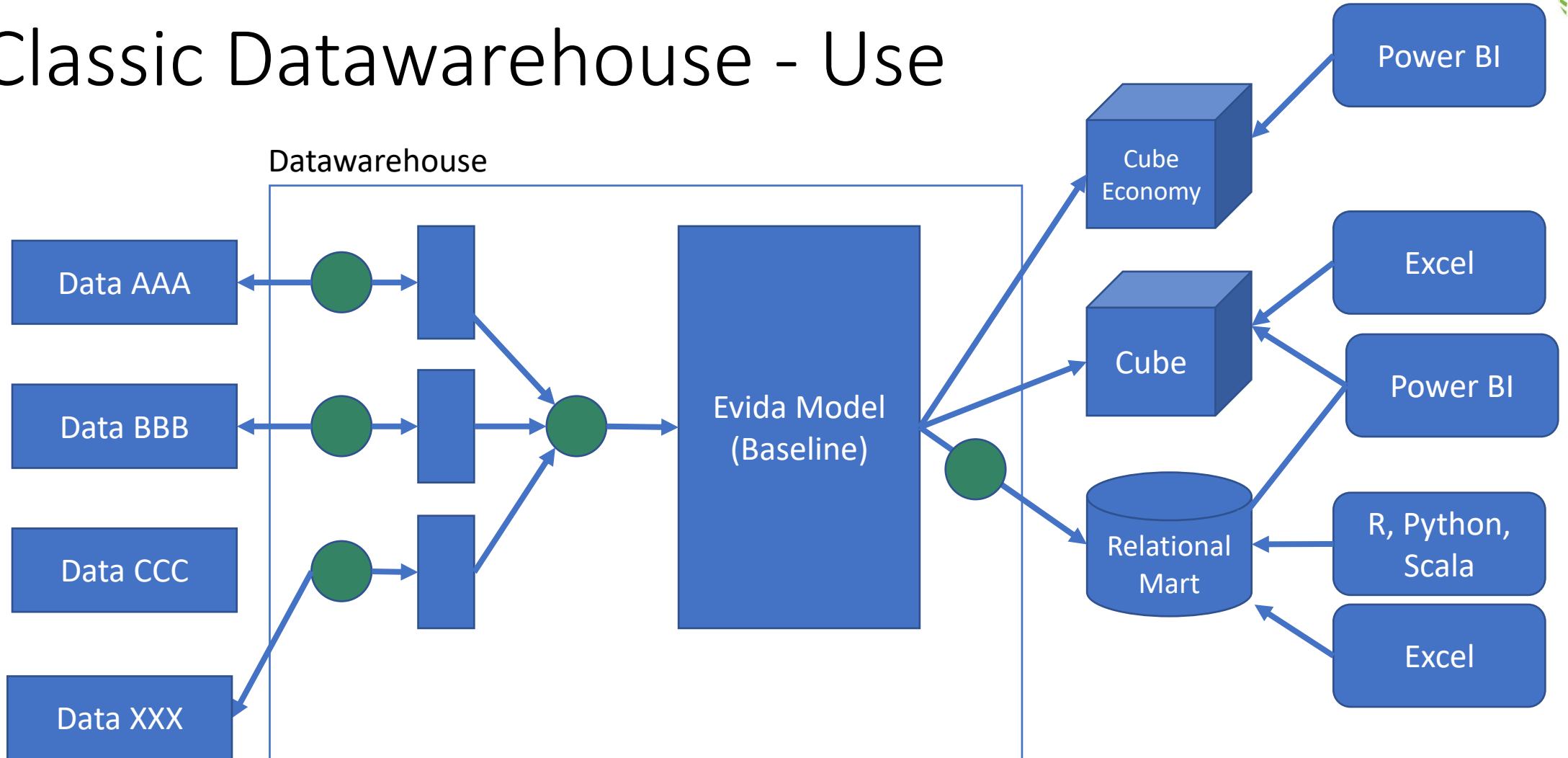
Ingest, Store and Serve



Classic Datawarehouse - Basic

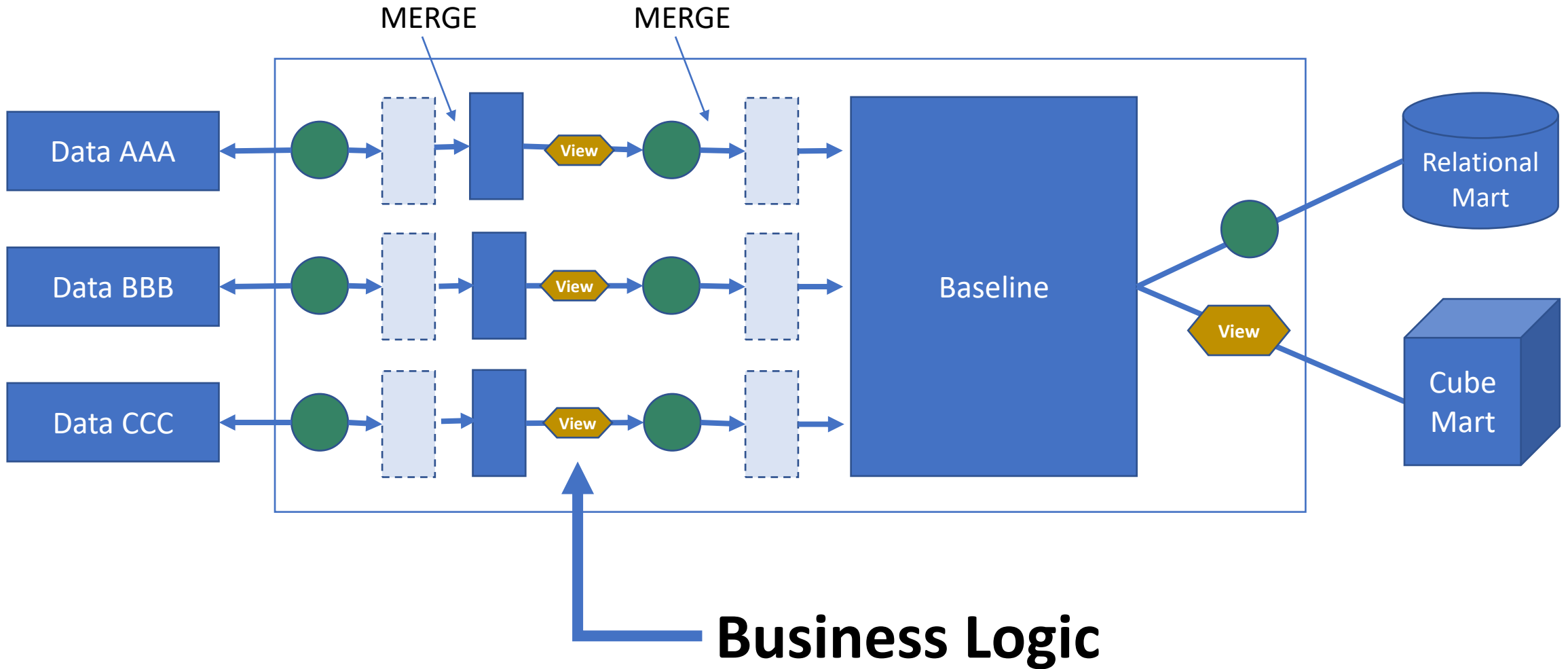


Classic Datawarehouse - Use



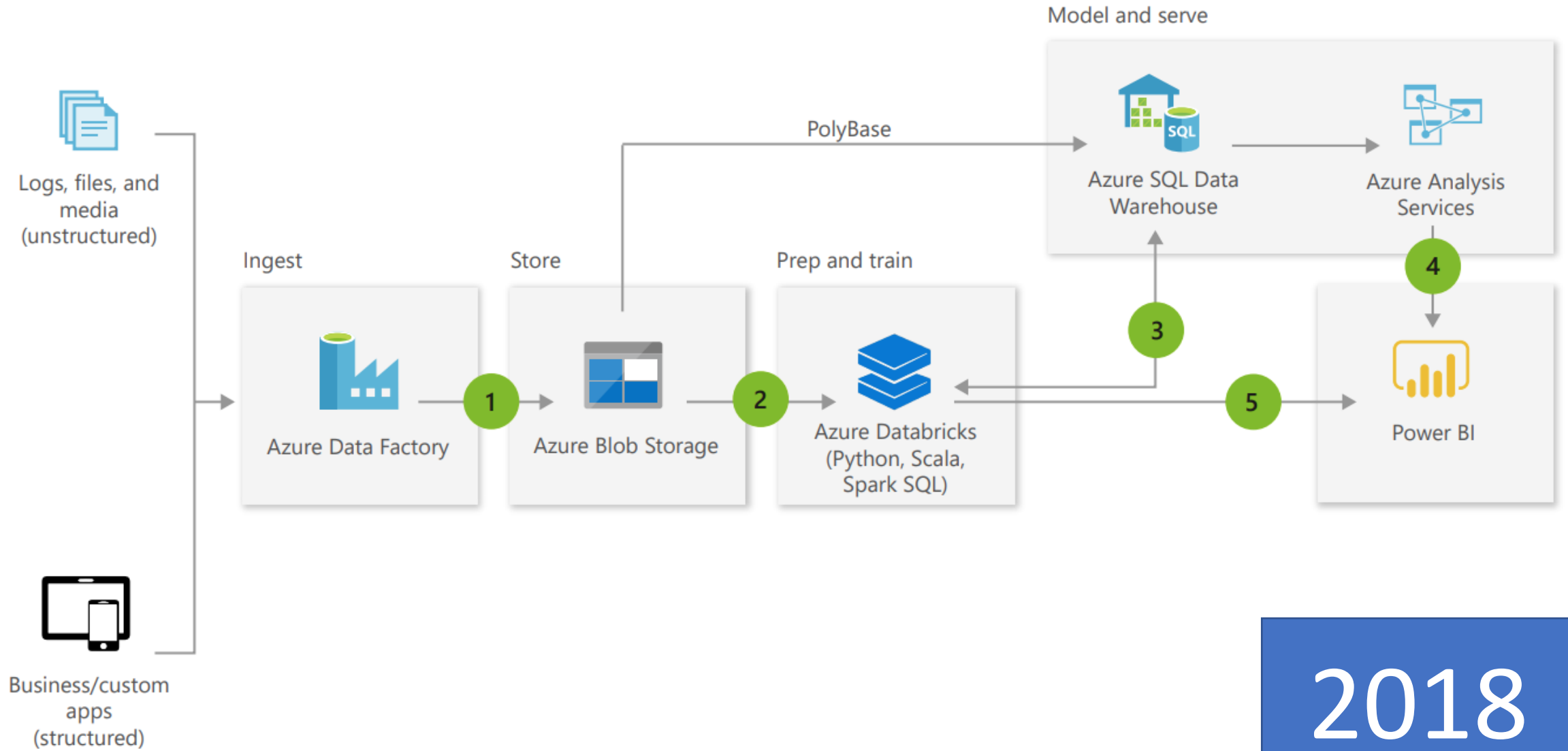


Classic Datawarehouse - Processes





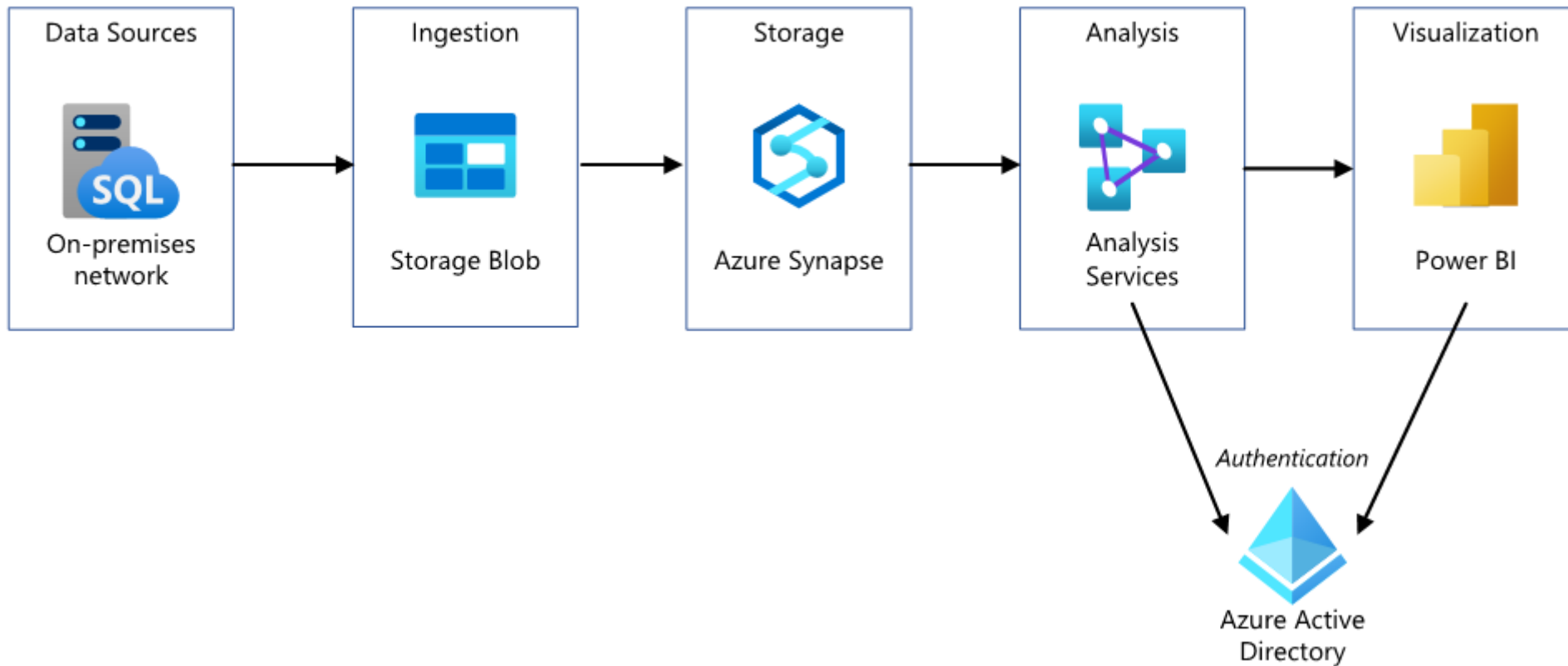
Modern Datawarehouse



2018

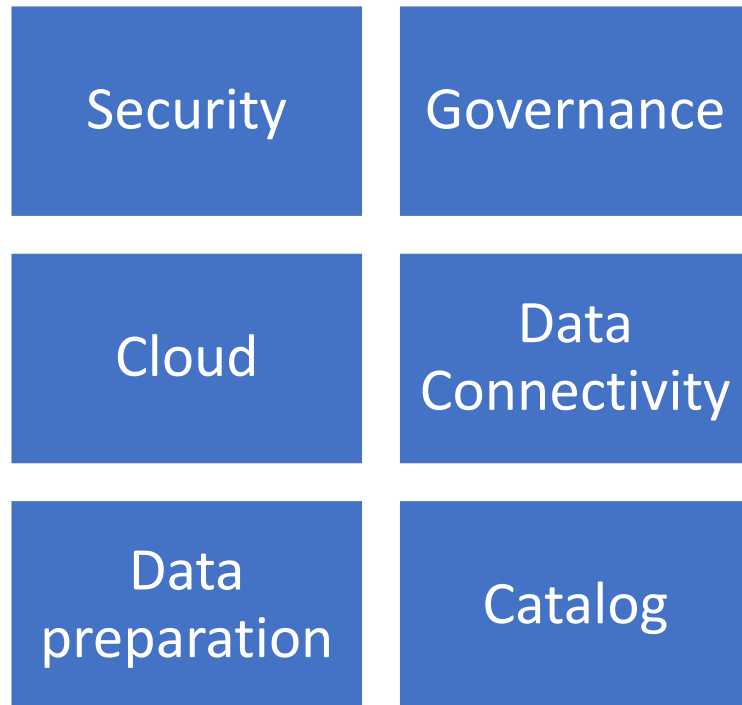


Even more Modern Datawarehouse



Platforms


At some point you must choose a "partner", capable of handling



- Automate insights (ML)
- Data Visualization
- Storytelling
- Reporting



Magic Quadrant for Analytics and Business Intelligence Platforms



Business Intelligence Related Technologies



Database Technologies



Tools and Languages



Presentation and Reporting

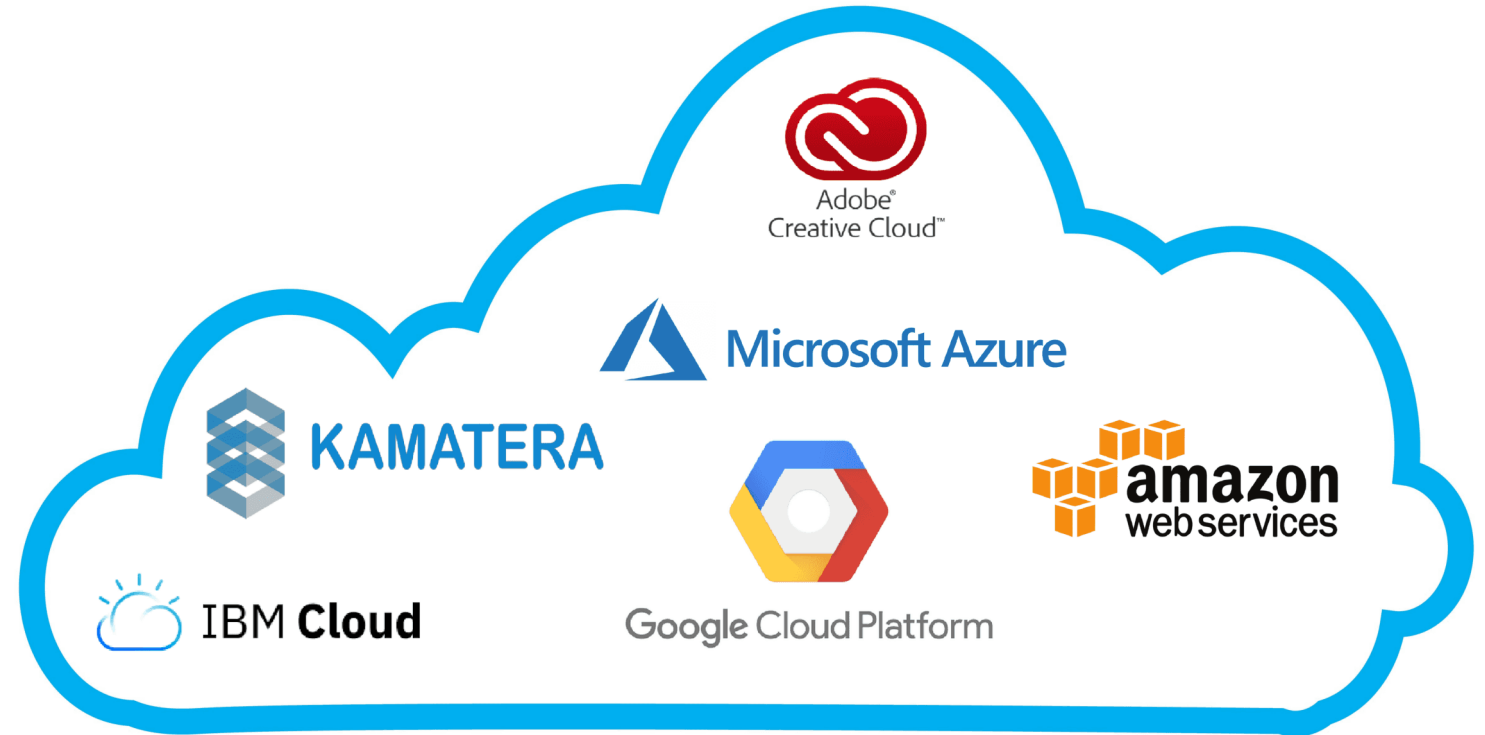


Infrastructure



Infrastructure

- OnPremise
- Cloud
 - IaaS, PaaS
 - SaaS
 - Open Source Solutions
- Hybrid Solutions



Serverless



OLTP

MPP

OLAP / Tabular

Big Data

Database
Technologies



Tools and Languages

- ETL Tools
 - SQL Server Integration Services
 - Informatica
 - Pentaho
- Languages
 - SQL
 - R
 - Python
 - C#



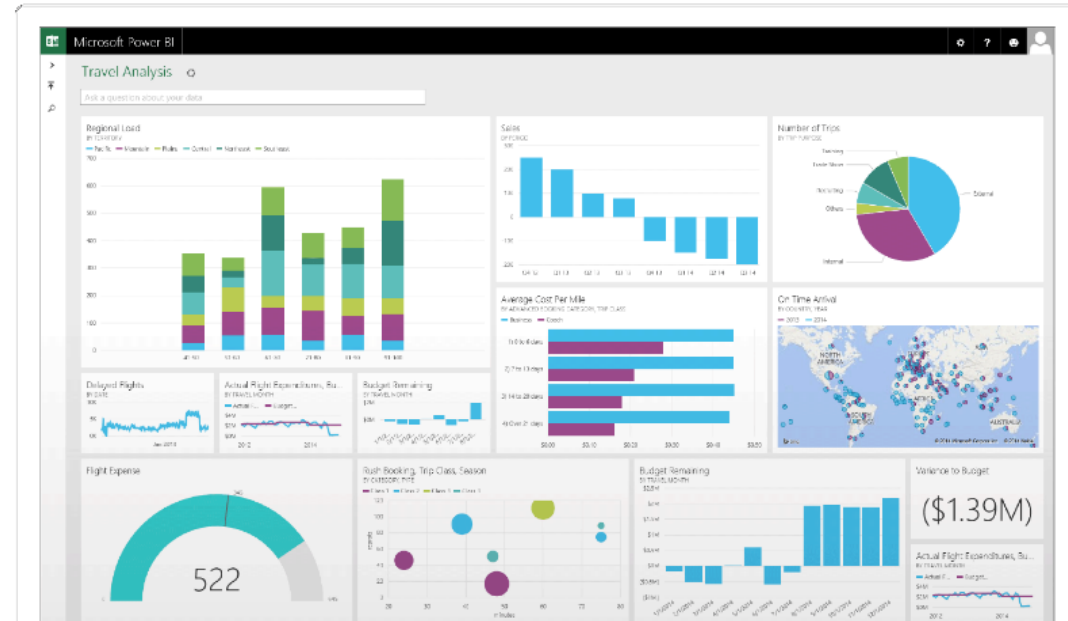
SQL - Transact SQL

- SQL = Projection of data
- SQL is by nature declarative
- “Everything” is tabular and relational
- Transact SQL Microsoft implementation of procedural elements

```
SELECT fname, lname  
FROM dbo.members  
WHERE city = 'Aarhus'
```

Presentation and Reporting – Power BI

- Power BI is a brand
- Microsoft Self Service BI Tool
- Desktop, Cloud, Mobile Apps



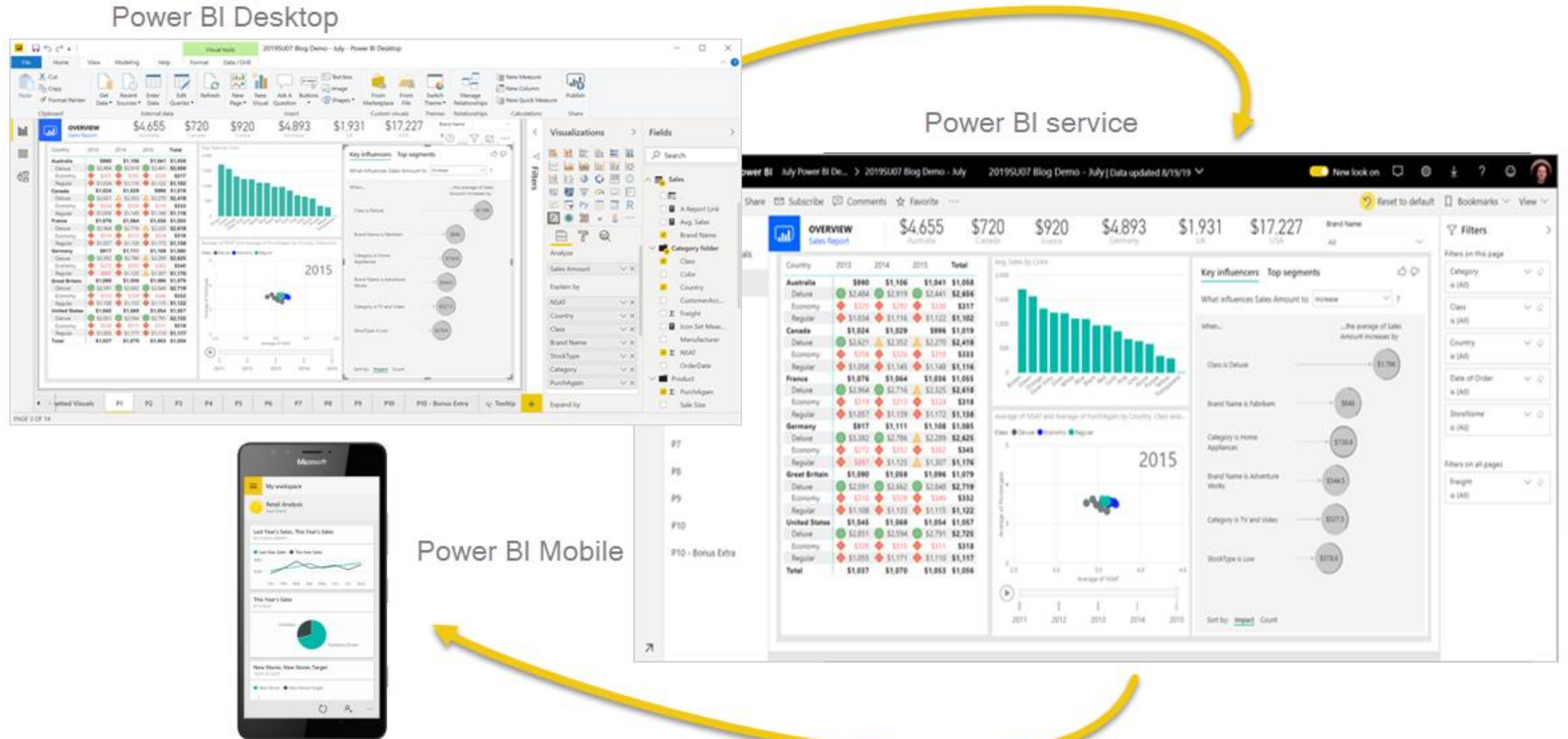


Power BI Desktop

- Free standalone Windows Desktop application
- Combines Power Query, Power Pivot Data Model and Power View into a seamless experience
- Access data from hundreds of supported on-premises and cloud-based sources
- Provide advanced analytics with the familiarity of Office
- Publish to Power BI Service



Three Parts of Power BI





Power BI – Data Behind



Data Mash Up

The screenshot displays the Microsoft Power BI interface for a workspace named "Northwind Trader". The top navigation bar includes "Home" and "Table tools" tabs, with options for "Get Data", "Transform Data", "Enter Data", "New Query", and "New measure".

The left-hand navigation pane shows a list of items: Home, Favorites, Recent, Create, Datasets, Goals, Apps, Shared with me, Deployment pipelines, Learn, Workspaces, and Northwind Trader (selected).

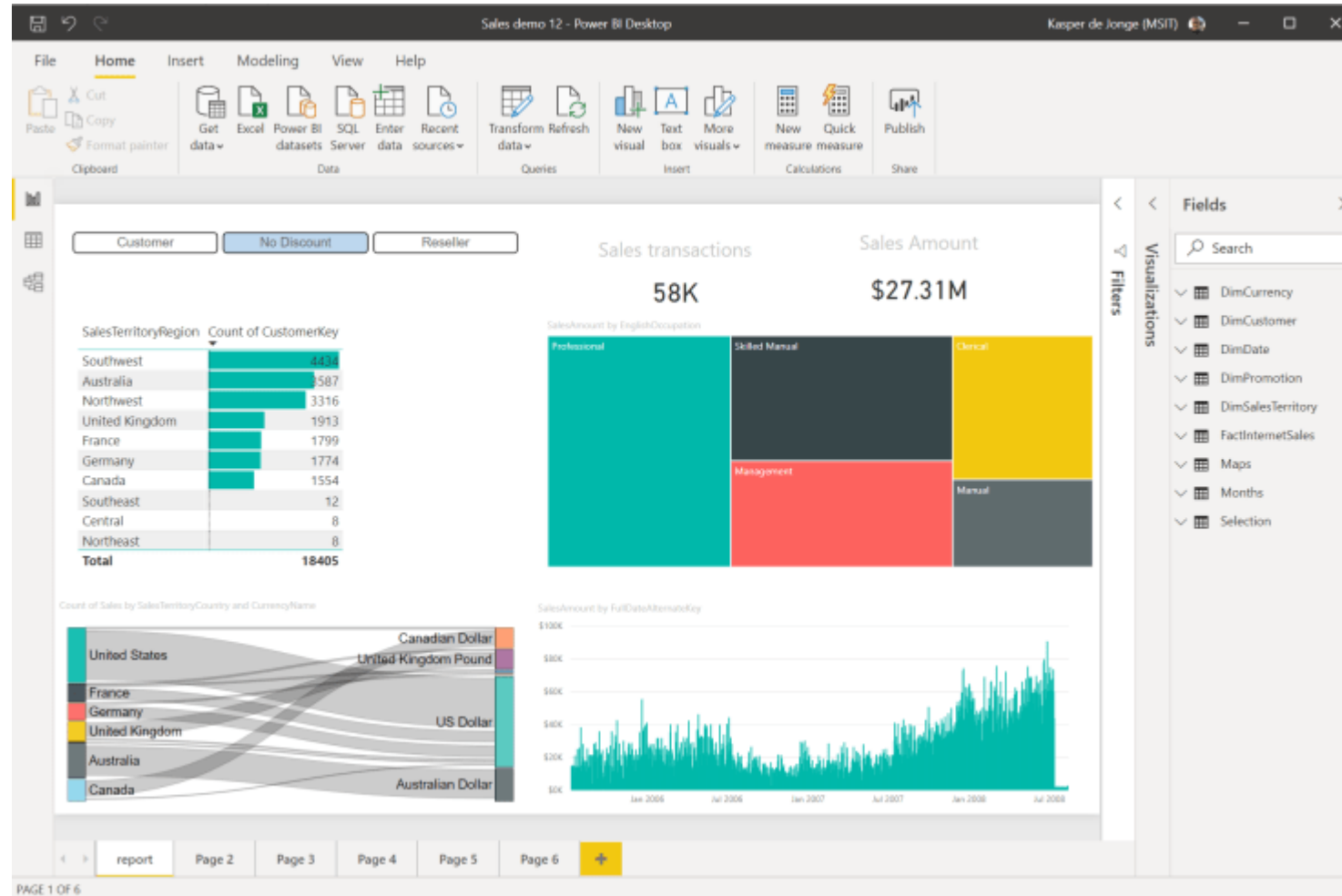
The central area shows a data model diagram with the following tables and their relationships:

- Categories**: CategoryID (PK), CategoryName (UK), Description.
- Products**: CategoryID (FK), ProductID (PK), ProductName (UK), QuantityPerUnit, UnitPrice, UnitsInStock, UnitsOnOrder, LastModifiedDate, SupplierID (FK).
- Order_Details**: OrderID (FK), ProductID (FK), Quantity, UnitPrice, Discount, ShipDate.
- Orders**: OrderID (PK), CustomerID (FK), EmployeeID (FK), OrderDate, RequiredDate, ShippedDate, ShipVia, ShipCountry, ShipRegion.
- Suppliers**: Address, City, CompanyName, ContactName, ContactTitle, Country, Name, Phone, HomePage, HomeFax, HomePageURL.
- Regions**: RegionID (PK), EmployeeID (FK), RegionName.
- Shippers**: ShipperID (PK), Name, Phone, ShipCountry, ShipRegion.
- Territories**: TerritoryID (PK), RegionID (FK), Country, ShipRegion.

The right-hand pane shows the "Properties" window for the selected "Orders" table, with the "General" tab active. It displays fields for Name (Orders), Description, Row label (Select a row label), and Is hidden (No).



Power BI - Report





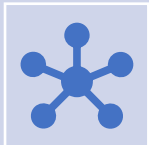
Modelling



Data modelling in general



3NF, Kimball, Inmon, Data Vault



Relational, Dimensional

Relational, Dimensional

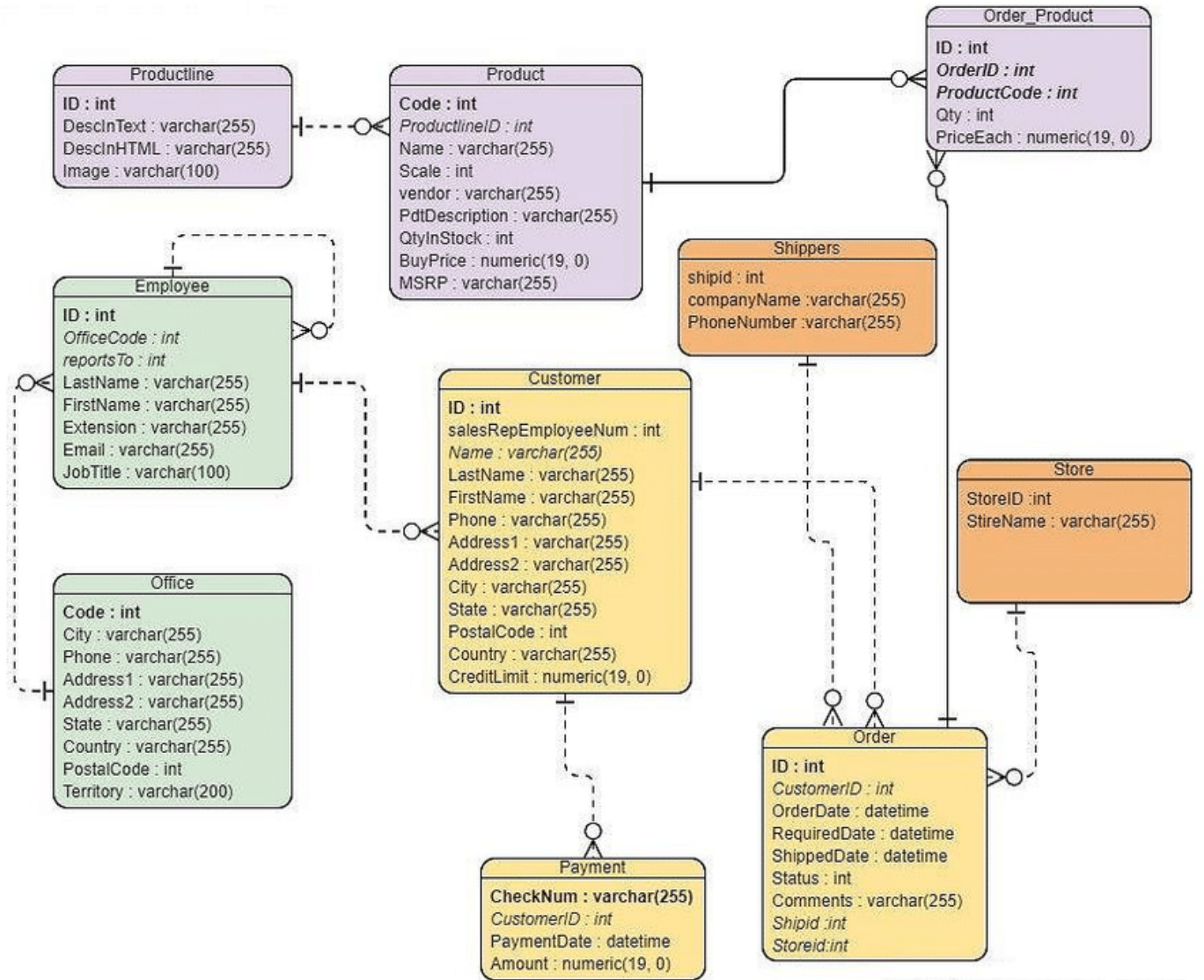
Relational

- Data is normalized (xNF)
- Tables and relationships (E/R)
- Volatile data changes
- SQL is used to manipulate data
- Data designed for business systems
- Data designed for concurrent handling of many small transactions by many users
- User is usually constrained by an application that understands the data design

Dimensional

- Data is more de-normalized
- Several Fact tables related to Dimensions
- Non volatile data
- MDX/DAX is used to manipulate data
- Data designed for analysis systems
- Suited for analyzing large amounts of data
- Designed for do-it-yourself data analysis
- Users are typically analysts, company strategists, or executives

Normalized E/R model

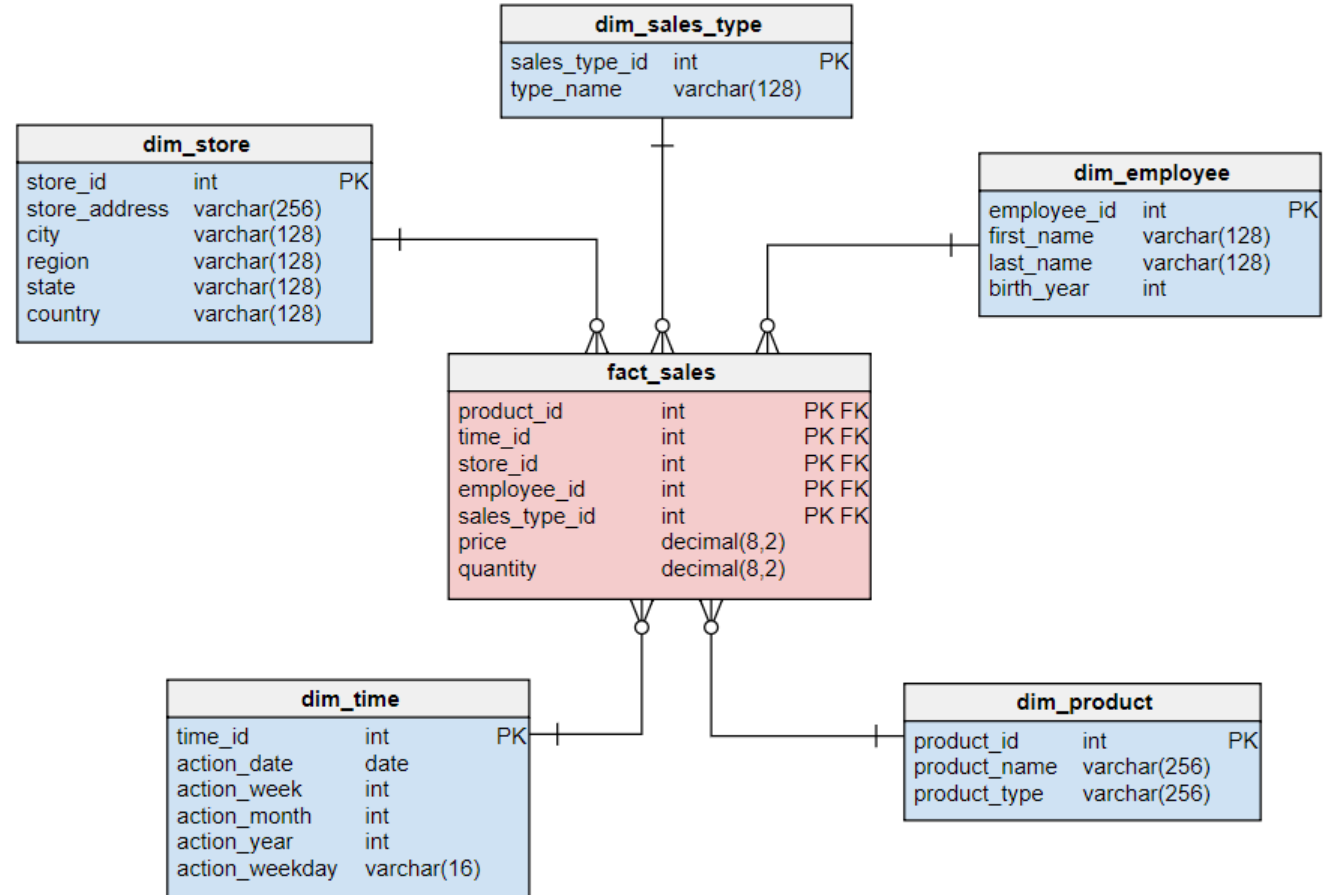


Dimensional Model

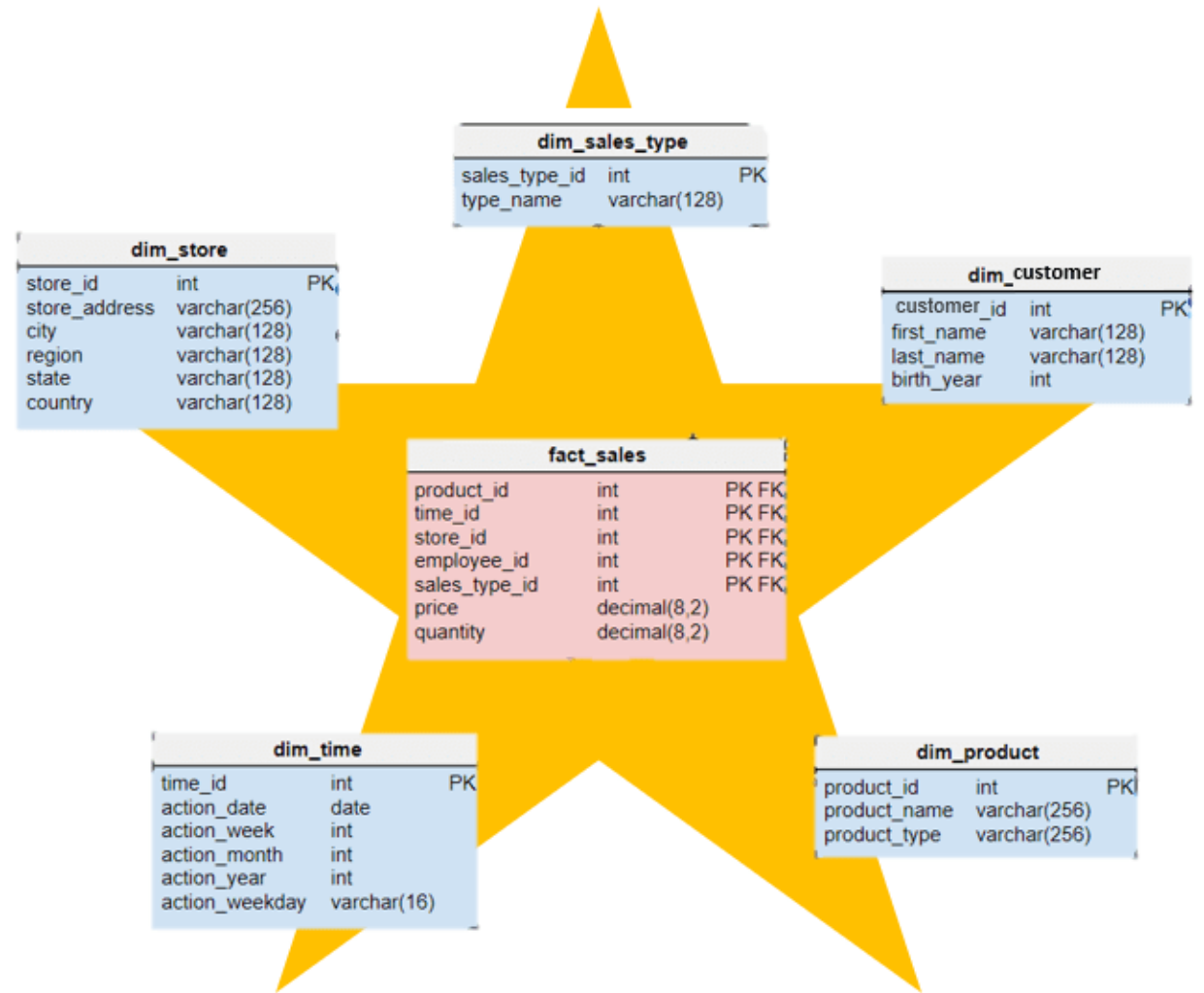
- The dimensional data model (star schema) is a denormalized alternative to a entity/relations E/Rmodel
- A dimensional model is user friendly and provides good performance on summarized queries
- A dimensional model typically consists of fewer facts and many dimensions
- Central to reporting solutions (Business Intelligence)



Dimensional Model



Dimensional Model “Star Schema”





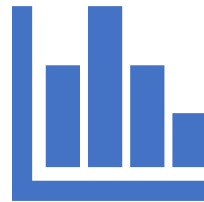
Fact tables

A fact is usually something that is not known in advance.

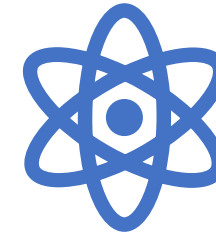
A Fact is “an observation” in the market place



The fact tables contains our measures, which is everything that can be measured and weighted



Facts are everything that can be measured and weighted



Fact tables are often defined by their *grain*. The grain of a fact table represents the most atomic level by which the facts may be defined

Dimension Tables

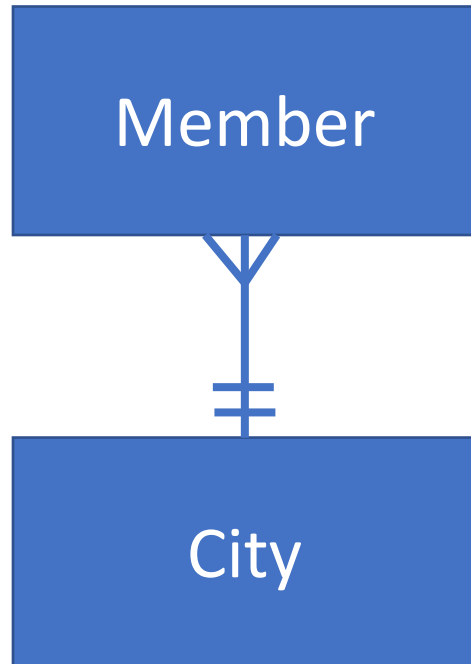
- Contains data that categorize the measures
- A dimension consists of many attributes
- A dimension can contain one or more hierarchies
- Dimension tables are the entry points to datawarehouses

Lets revisit SQL





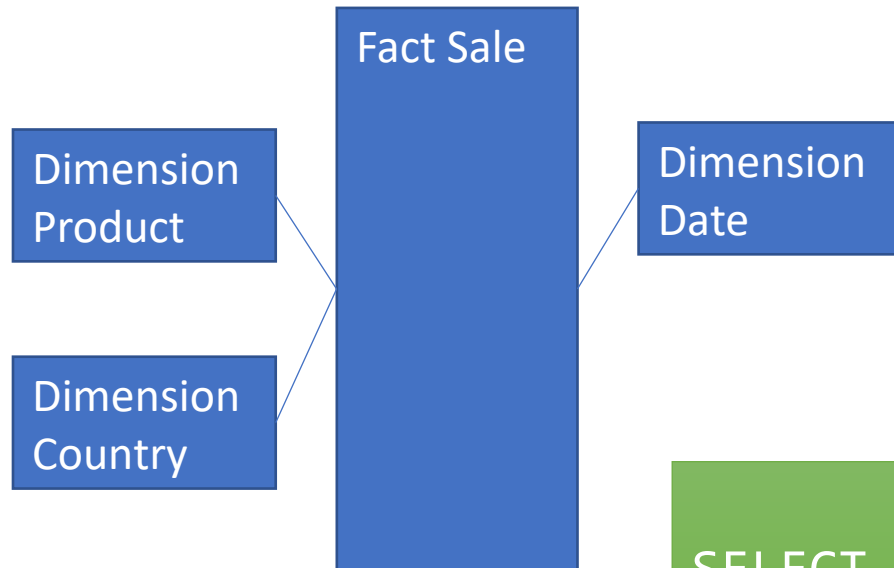
A little bit advanced queries



```
SELECT m.fname, m.lname  
FROM dbo.member m  
INNER JOIN dbo.city c ON m.ID_City = c.ID_City  
WHERE c.city = 'Aarhus'
```



An even more advanced query – Dimensional

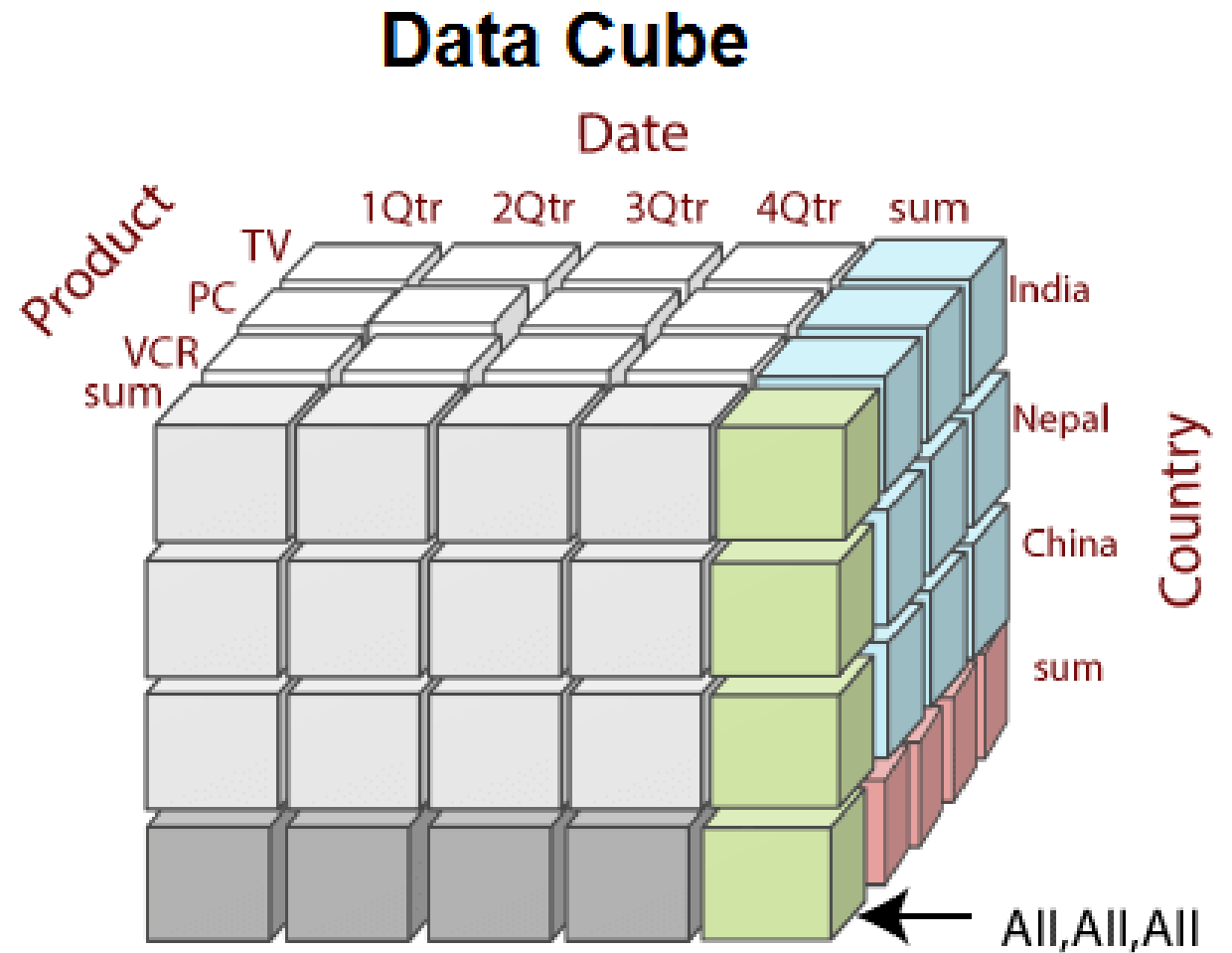


```
SELECT p.ProductName, d.DateName, s.Count
FROM fact.Sale s
INNER JOIN dim.Product p ON s.ID_Product = p.ID_Product
INNER JOIN dbo.Date d ON s.ID_Date = d.ID_Date
INNER JOIN dbo.Country c ON s.ID_Country = c.ID_Country
WHERE p.Product = 'Piano'
```



OLAP Database

- Often named Cubes
- Based on dimensional model
- MDX Language (older)
- DAX Language (modern)
- Used for
 - Measures
 - Queries
- Columnstore Technology





DAX Queries

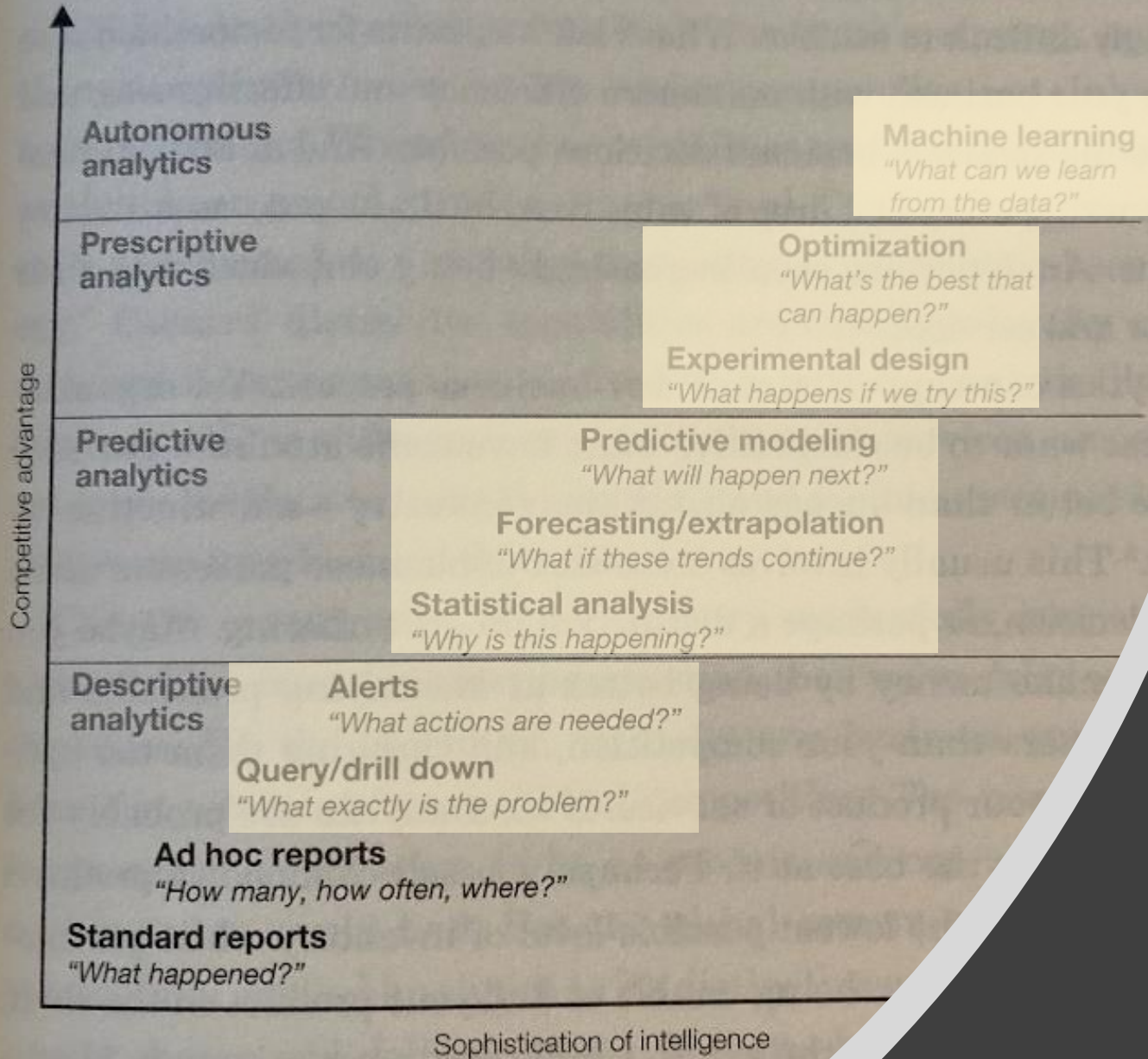
```
EVALUATE
  FILTER(
    SUMMARIZE(FactResellerSales
      , DimDate[CalendarYear]
      , "Total Reseller Sales"
      , sum(FactResellerSales[SalesAmount])
    )
    , DimDate[CalendarYear]
      IN (2011, 2012)
  )
```

DimDate[CalendarYear]	[Total Reseller Sales]
2011	16288441.7675
2012	27921670.5182

```
evaluate(
  summarize(
    FactInternetSales
    , FactInternetSales[OrderDateKey]
    , FactInternetSales[CustomerKey]
    , DimCustomer[FirstName], DimCustomer[LastName], DimCustomer[EmailAddress]
    --, DimCustomer[Full Name]
    , "Total Tax Amount", SUM(FactInternetSales[TaxAmt])
    , "Total Sales Amount", SUM(FactInternetSales[SalesAmount])
    , "Total Amount", SUM(FactInternetSales[SalesAmount])+SUM(FactInternetSales[TaxAmt])
  )
  order by
    FactInternetSales[OrderDateKey]
    , FactInternetSales[CustomerKey]
)
```

FactInternetS...	FactInter...	DimCusto...	DimCusto...	DimCustomer[...	[Total Tax A...	[Total Sale...	[Total Amount]
20010701	11003	Christy	Zhu	christy12@a...	271.9992	3399.99	3671.9892
20010701	14501	Ruben	Prasad	ruben10@ad...	55.9279	699.0982	755.0261
20010701	21768	Cole	Watson	cole1@adve...	286.2616	3578.27	3864.5316
20010701	25863	Sydney	Wright	sydney61@a...	271.9992	3399.99	3671.9892
20010701	28389	Rachael	Martinez	rachael16@...	271.9992	3399.99	3671.9892
20010702	11005	Julio	Ruiz	julio1@adve...	269.9992	3374.99	3644.9892
20010702	11011	Curtis	Lu	curtis9@adv...	271.9992	3399.99	3671.9892
20010702	16624	Albert	Alvarez	albert7@adv...	286.2616	3578.27	3864.5316

Potential competitive advantage increases with more sophisticated analytics



The Real world
As it should be



Risks

- Swamp of information
- Plethora of solutions (small solution)
 - Multiple versions of truth (again....)
 - No sharing of data
 - Silos



The real world

- It's messy out there
- Open Source has been limited
 - Though Cloud is opening up the world



Questions





Pragmatic BI

Pragmatic BI Aps

Åhusene 4,9,3

8000 Aarhus C

Telefon +45 42 606 202

E-mail steen@prabi.dk

Internet www.pra_bi.dk

CVR nr. 36974133