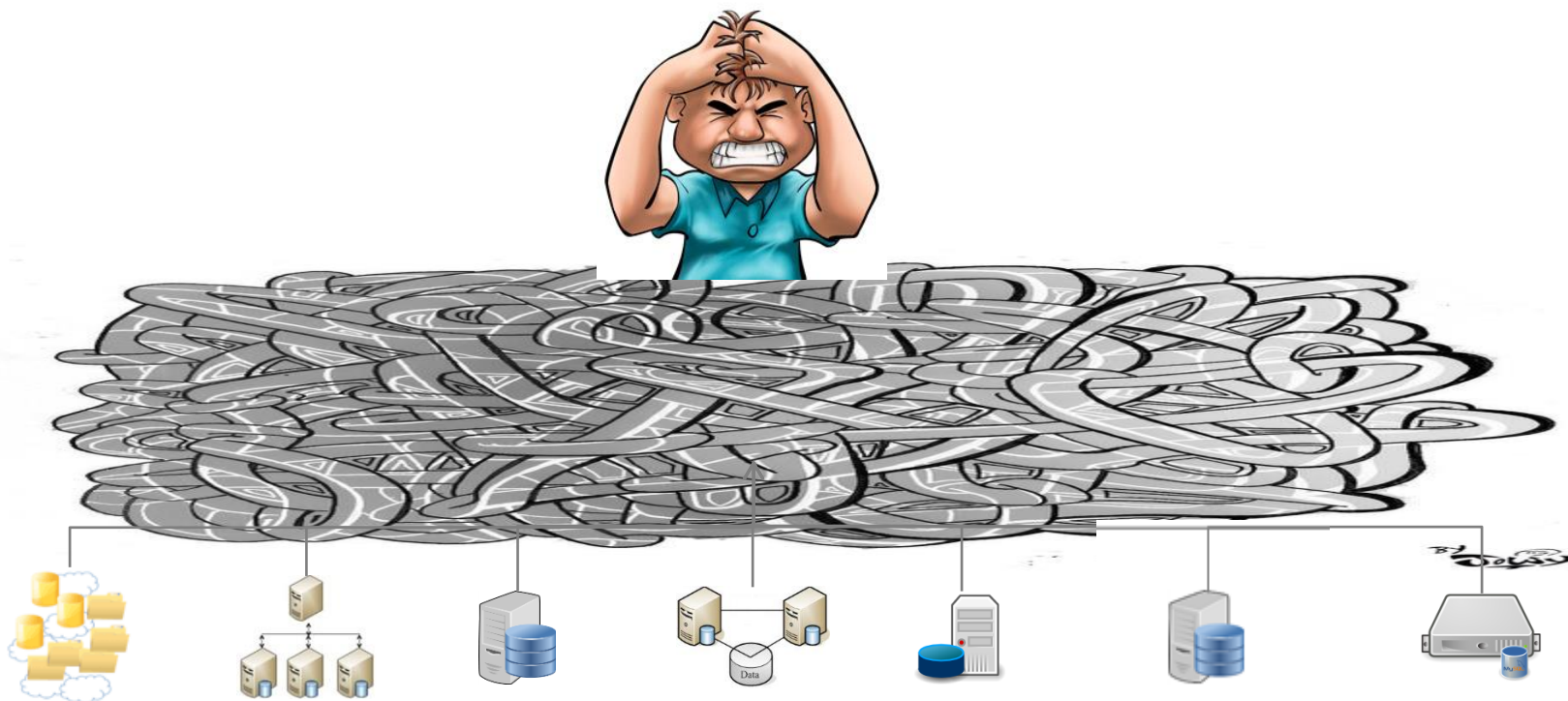# Data Virtualization Overview

Dirk Garner
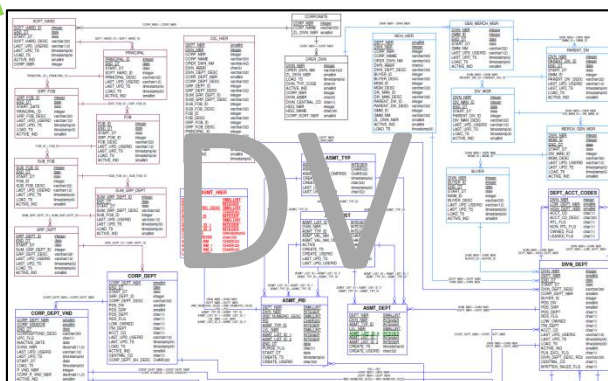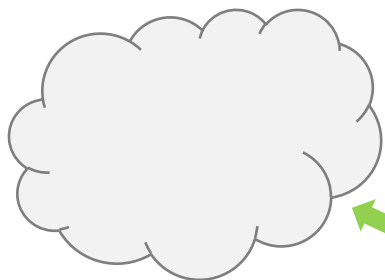
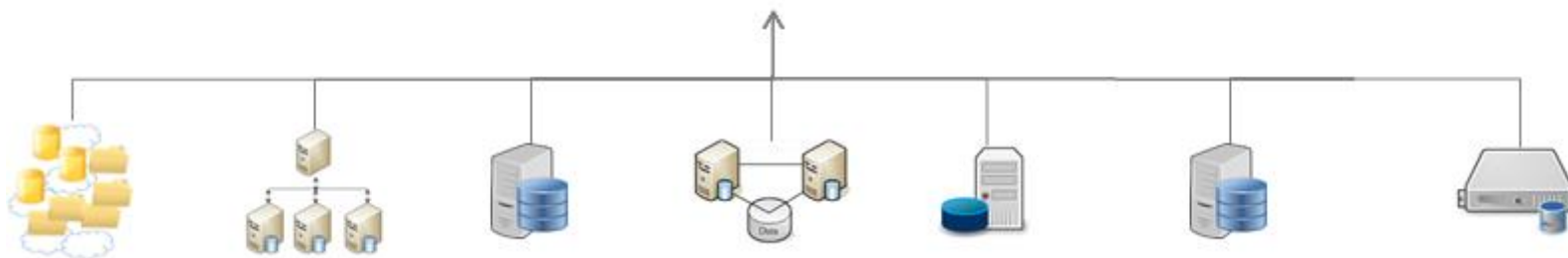# Typical State – 'Dis'-Integration

Partially Integrated Data:

- Data Warehouse/Datamart may or may not have necessary data

- Inaccessible data can derail, postpone or cancel analytical efforts

- Encourages siloed work-around tactics, shadow IT

- Leads to ungoverned use of data: inaccuracies, incompleteness, etc.
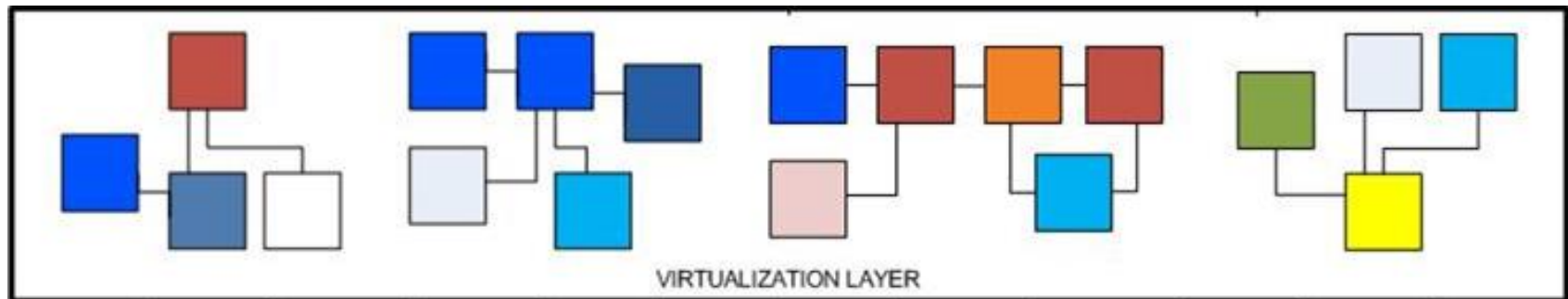
Ideal State: Data Virtualization

- **Immediate access to diverse data sets**
- **Faster time-to-market**
- **Fresher data**
- **Business-friendly semantic view**

- **Supports data governance**
- **Reduces data duplication**
- **Reduced cost**
- **Logical Data Warehouse**

- **RDBMS**
- **Cloud**
- **Columnar**

- **Hadoop**
- **Teradata**
- **Mainframe**
- **ESB**

- **Flat file**
- **Web service**
- **Web scraping**

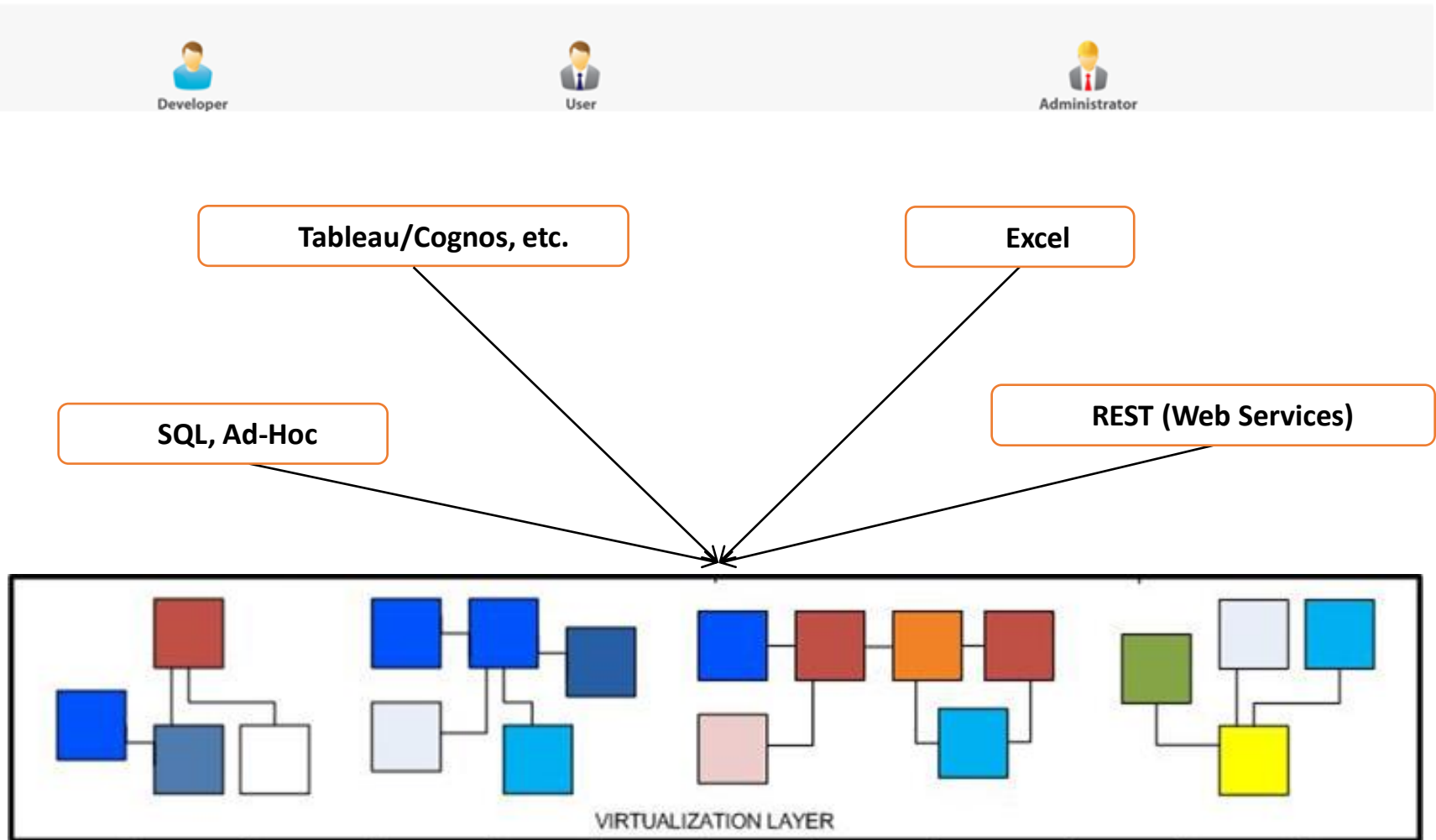# DV: Usage & Data Sourcing Strategies

VIRTUALIZATION LAYER

Exploration, Prototyping, & Discovery:
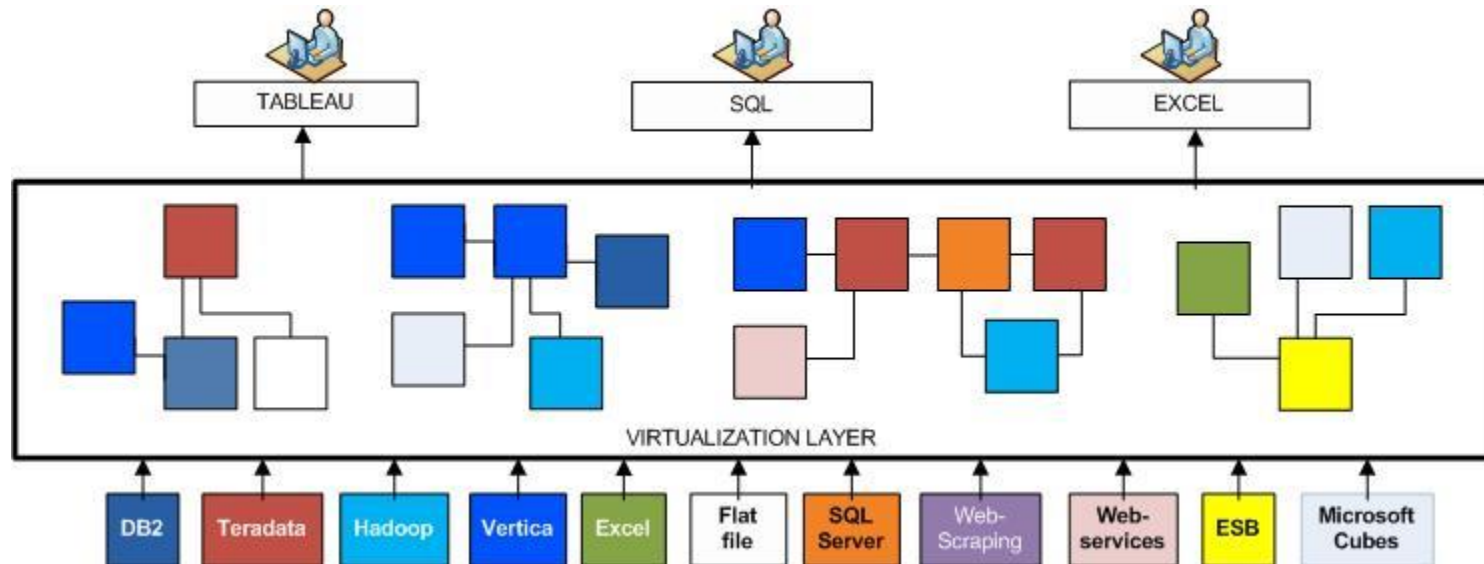Real time from source system

Data Source and Productionalization:
Data Warehouse, Data Mart, Data Lake,

Quick Integration and Query Performance:
Cached from source system

DV: Consumption Strategies

Consumer Perspective: Simplicity

- <u>Acceleration</u> immediate access to integrated (on-the-fly) data across physically and logically diverse data sources
- <u>Analyst Self Service</u> flexible user defined data models
- <u>Hide Complexity</u> from application developers and business users
- <u>Decouple</u> consuming applications and data sources
- <u>Make It Easy</u> to create, extend and use data services
- <u>Single Point</u> of access across diverse databases

Comparison

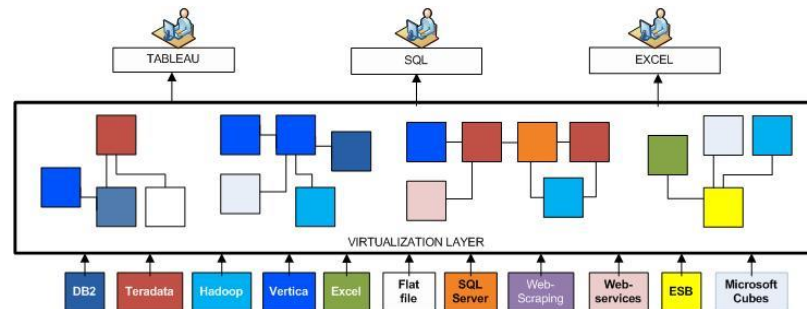| Without Data Virtualization	With Data Virtualization
•A typical analytical prototype may undergo 2-3 months or more of data sourcing, permission requesting and ETLing to get to the point at which analysis could begin.	•Typical analysis can begin after data integration efforts as brief as 2-3 days.
•A data Integration point is required such as Excel, Access, Hadoop, etc. and is bound to performance limitations of those platforms.	•Single integration point is the DV platform leveraging performance aspects of each source system and the DV.
•Managed through Interface appropriate to the integration point.	•Options to manage/view your data: the simple and easy to navigate DV GUI, Tableau, Excel, or your favorite query or visualization tool.
•Data is as fresh or stale as the last ETL	•Data is as fresh as your last query
•New/additional data elements require re-work to ETLs or new ETLs.	•New/additional data elements require a few mouse clicks and moments to complete.

Questions? Comments?

Dirk Garner
dirkgarner@garnersoftware.com
http://www.linkedin.com/in/dirkgarner