

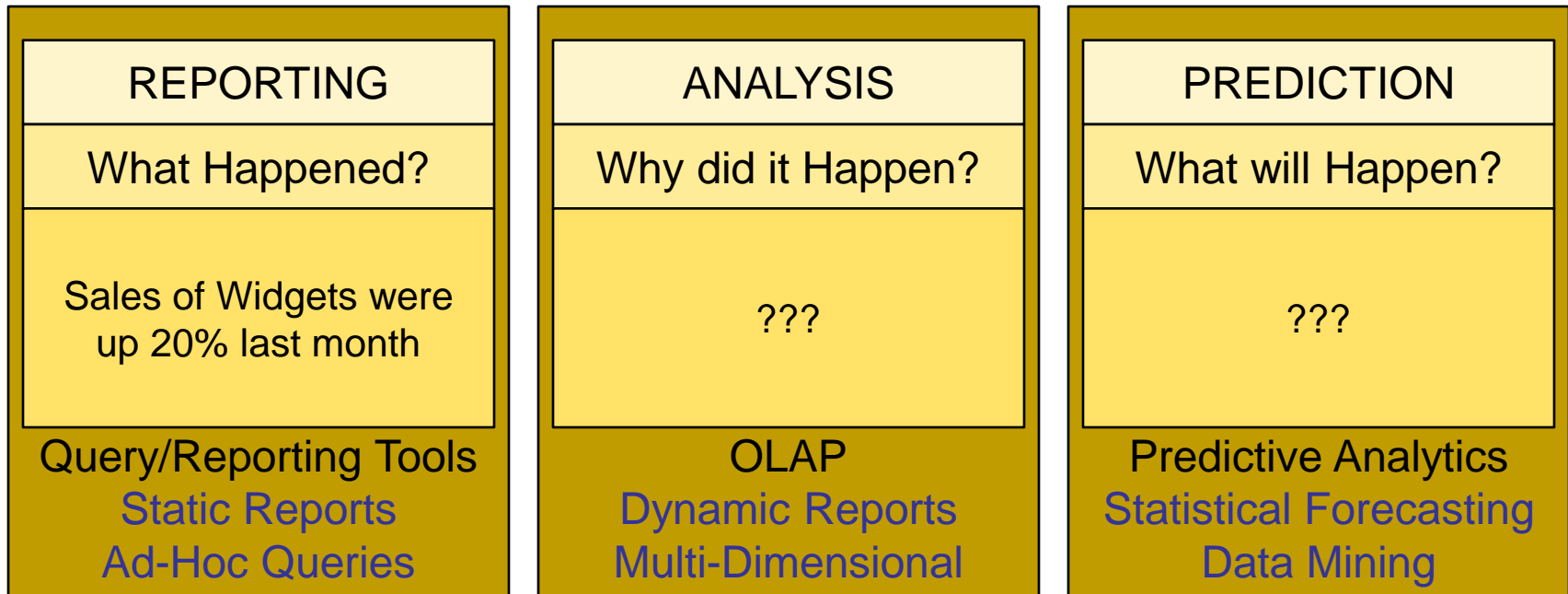
Fast Track to Successful Data Warehousing using RODIN

Alan Jordan

Coglin Mill, Rochester, Minnesota USA

ajordan@coglinmill.com

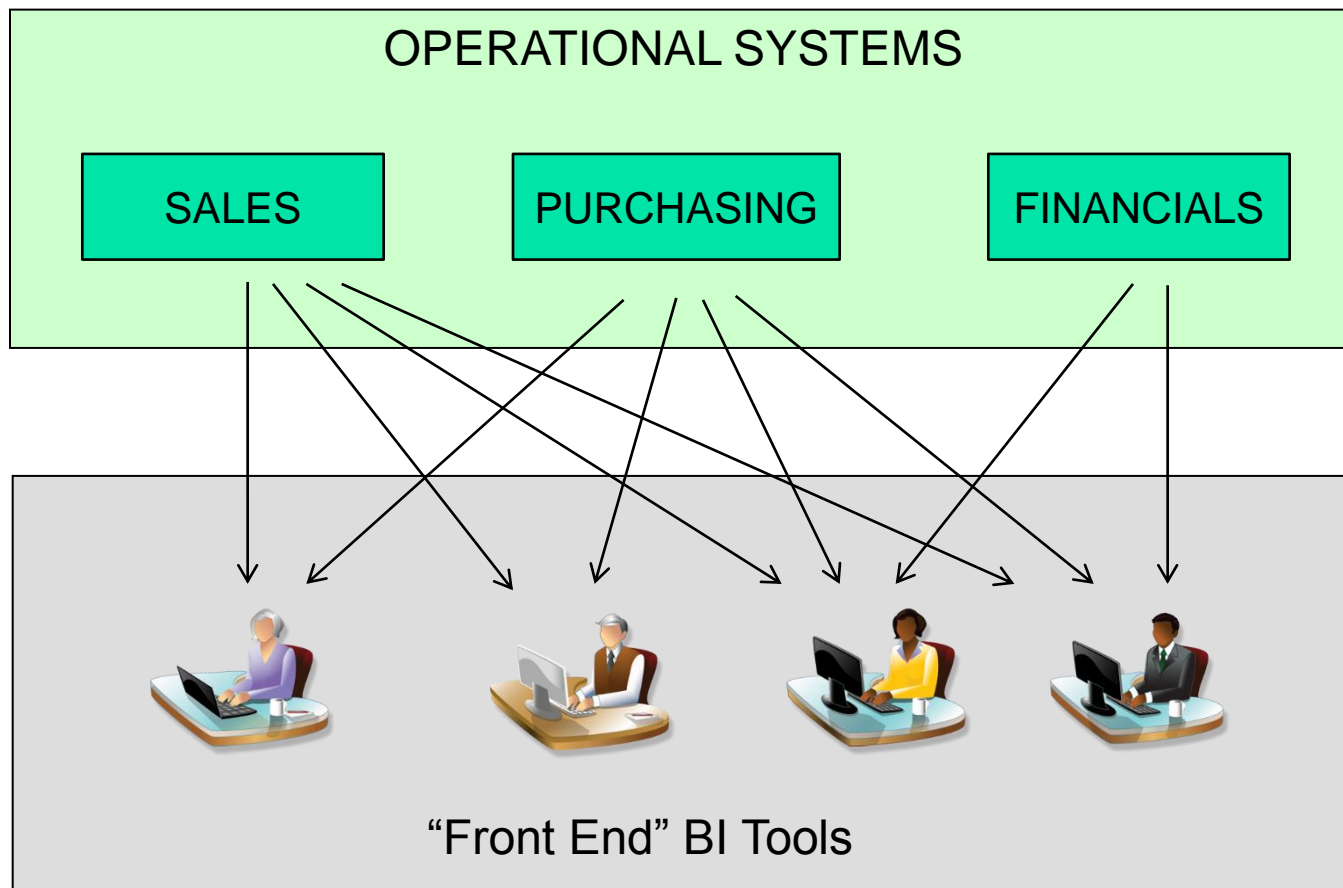
Understanding Business Intelligence



Increasing Business Value

Increasing Complexity

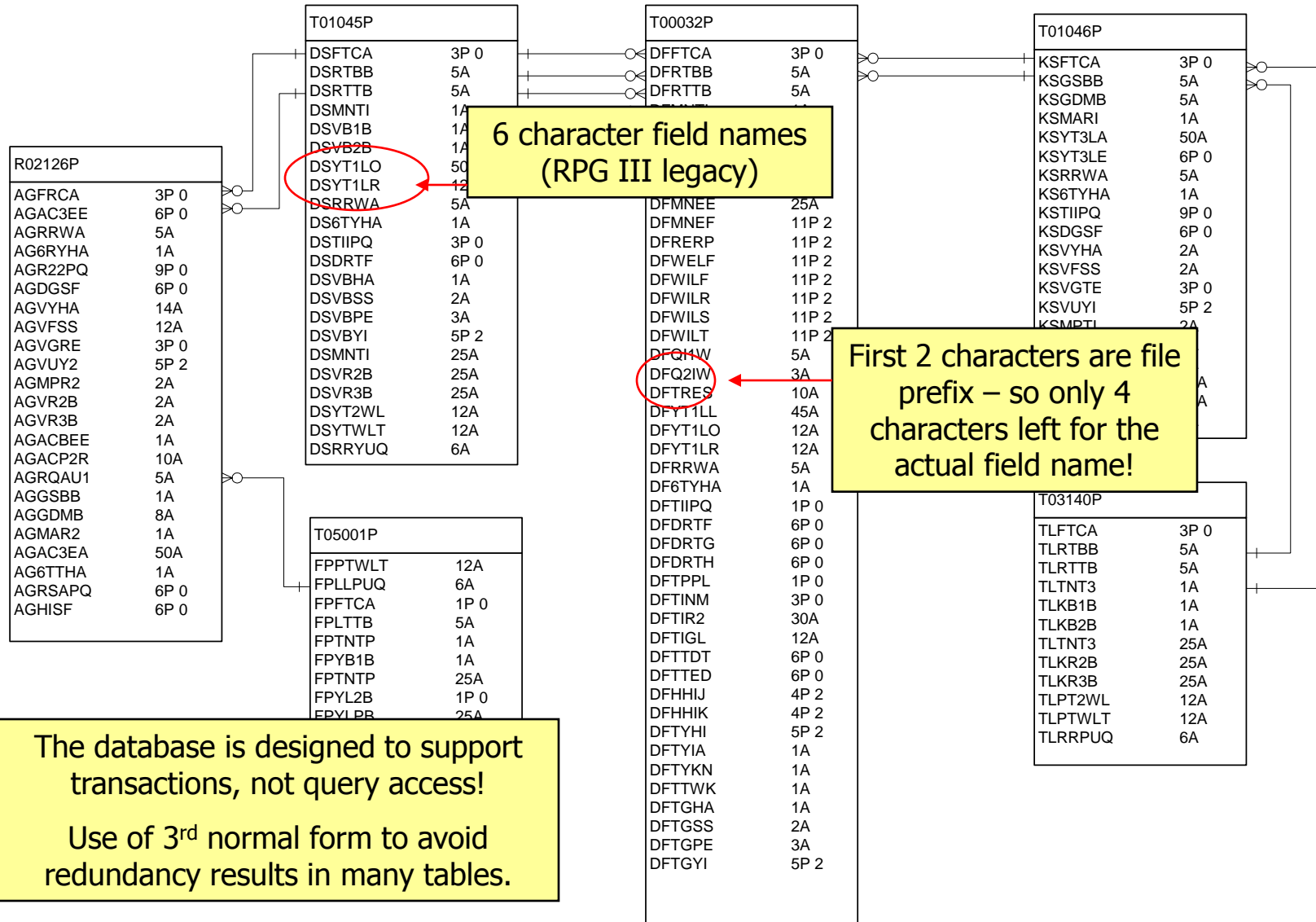
Common BI Implementation



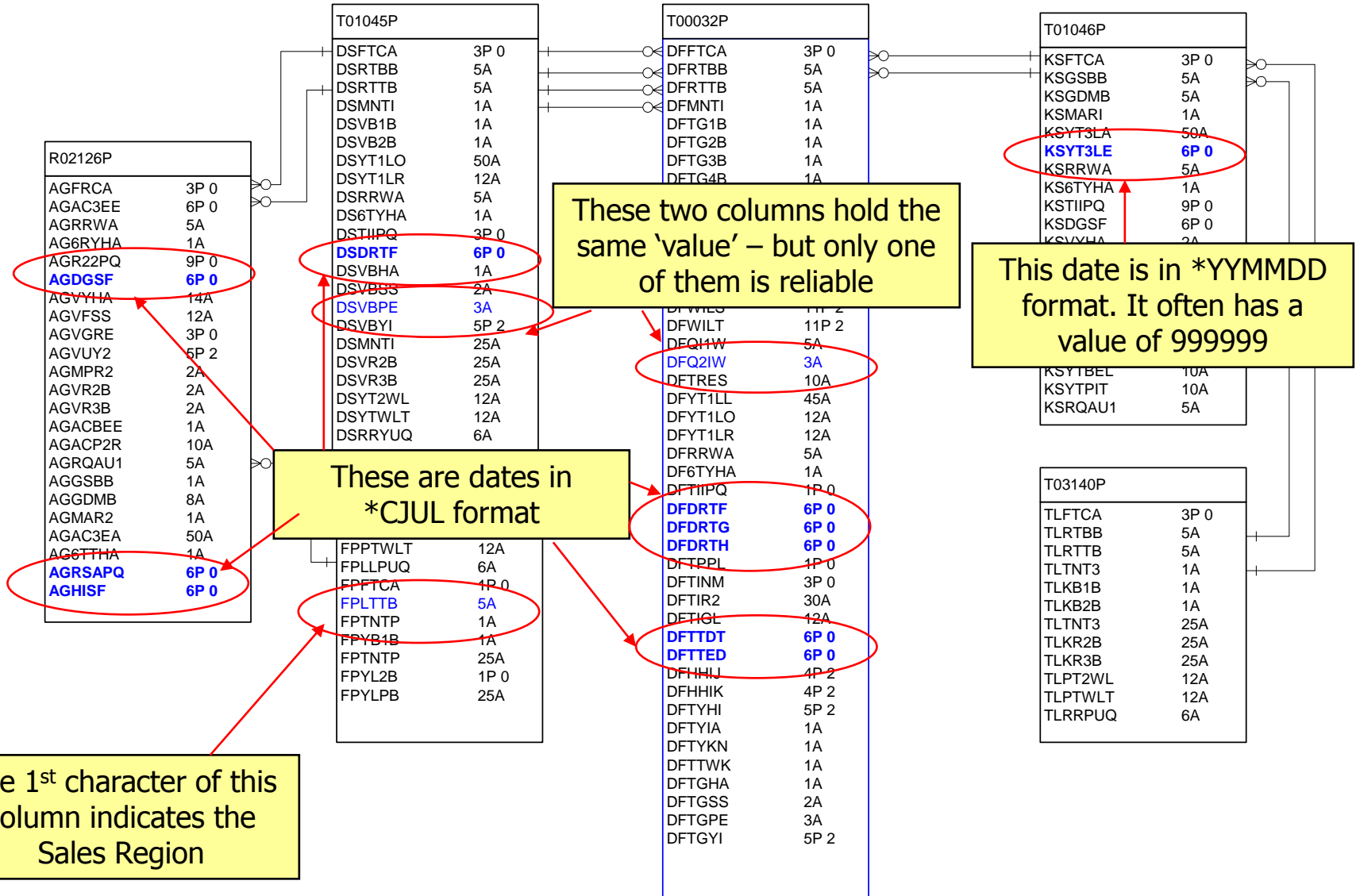
Issues with Operational Data

- The database was not designed for YOU to access
 - It was designed for the software application (ERP, MRP)
 - Column names are not meaningful
- There is probably no 'user manual' for the tables
 - The user manual/documentation is for the application
- There is probably no consistency of design
 - The application grew over many years, with different Developers
 - Redundant fields reused for different purpose
- Joins may require data manipulation
 - Derived values, different data types etc.
- Cryptic or difficult data values
 - 2nd character of column X means ..
 - If column Y = 'S', value Z must be multiplied by -1
- Dates
 - When dates are just numbers, how do you report by quarter or add a month to a date?
 - How do you handle a date of zero, or all 9s or other invalid dates?

Example



Example



Data Quality

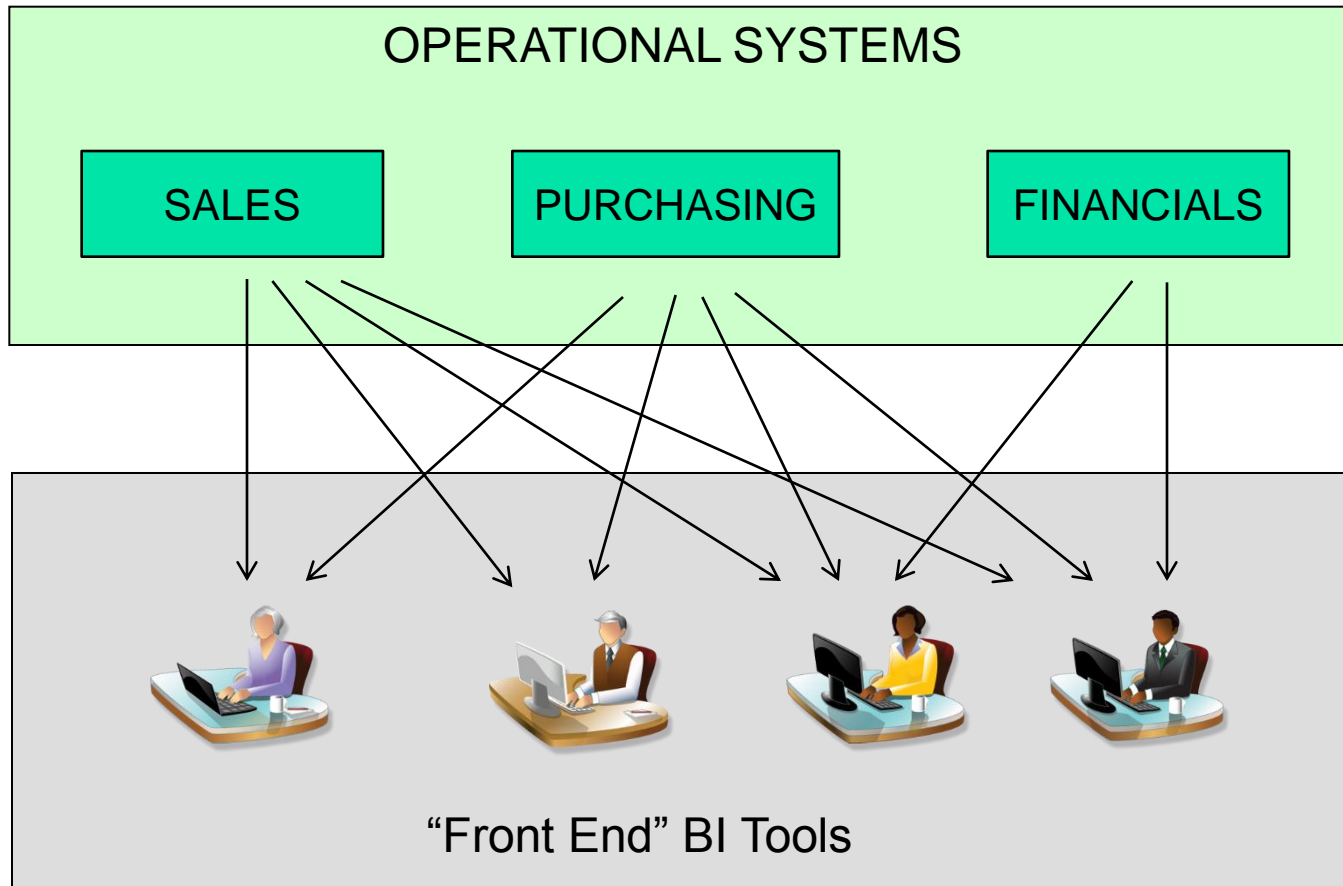
- **Data Quality is one of your most important issues!**
 - Your company will be making strategic decisions based on the data used in reports. You hope it is accurate!
 - Poor data quality is the most common reason for failure of business intelligence initiatives

A 2002 study by TDWI found that
poor data quality costs US businesses

\$600 BILLION per year!

Conservative estimates suggest that
poor data quality costs the typical organization
Up to 10% of potential revenue and
Up to 20% of potential earnings

Common BI Implementation

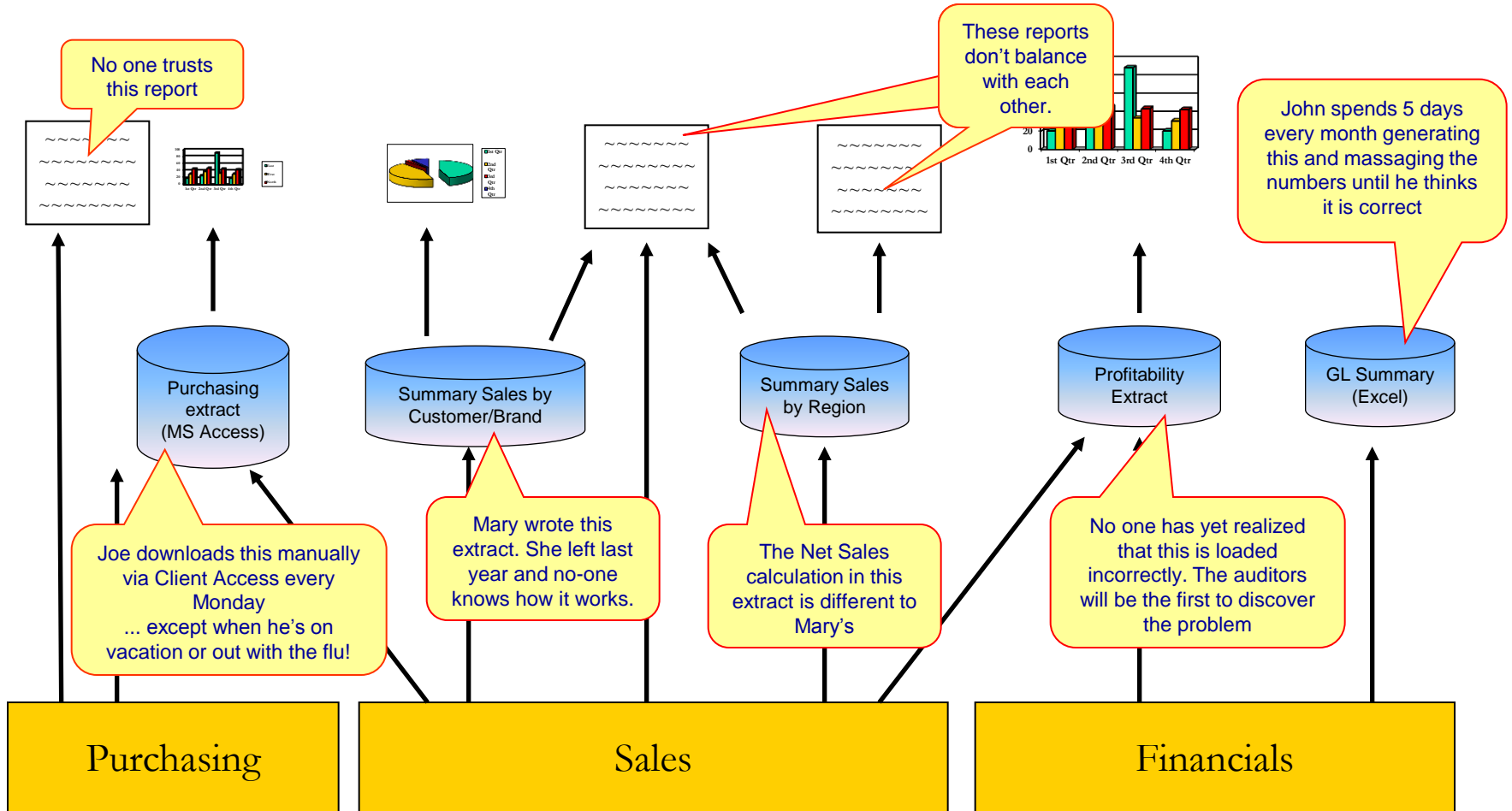


Issues and complexity pushed to the front end tools

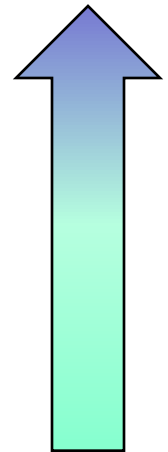
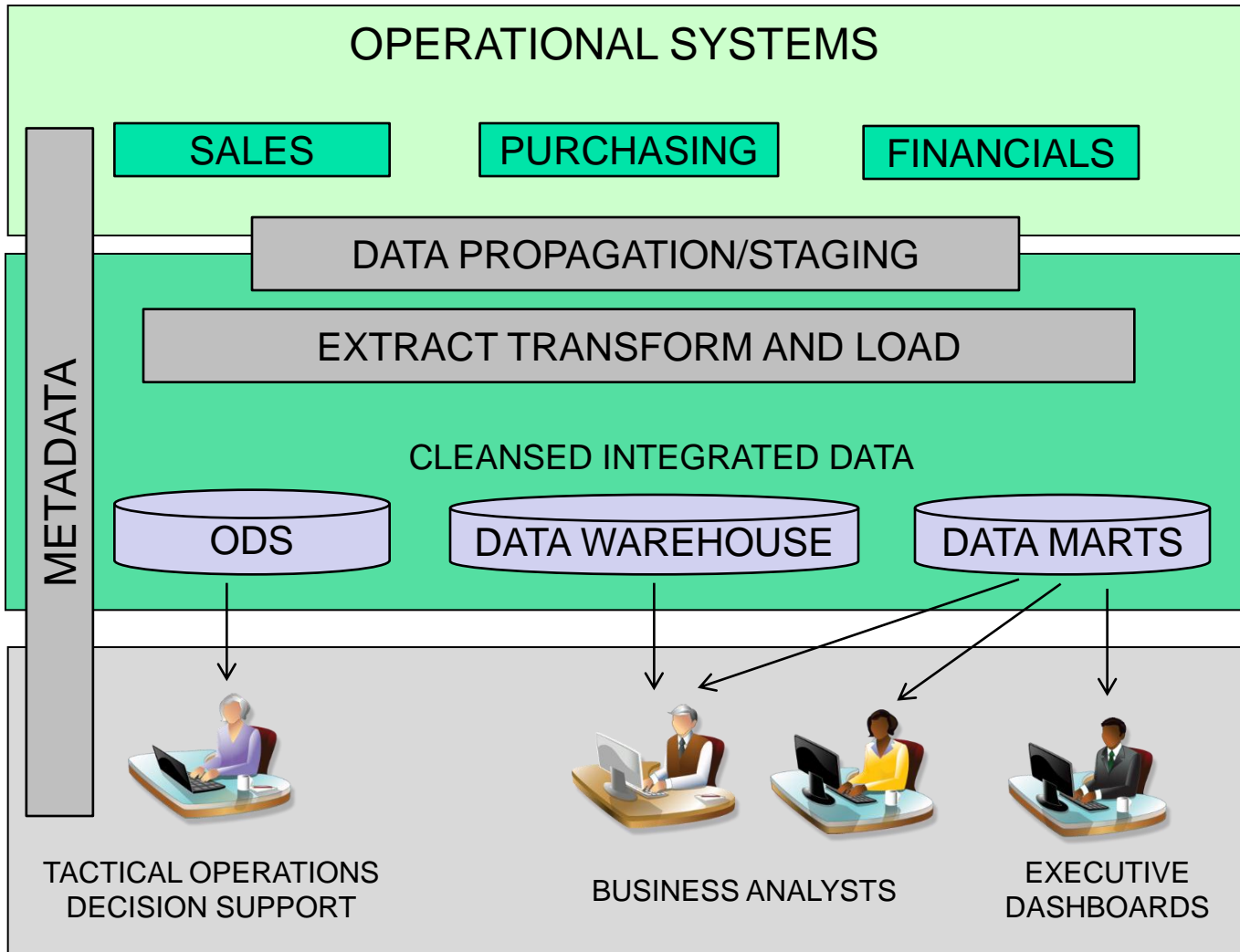
The Usual Result

- These issues are often *solved* in an ad-hoc way
 - IT Creates *extract files* and write RPG programs to load them
 - As each reporting problem occurs, a new extract is written
 - No consistent approach
 - No documentation produced
 - Frustrated users create their own “solutions”
 - Download data to Excel and manipulate it
 - Decide on their own rules

A chaotic reporting environment!



Enterprise Data Warehouse Architecture



Issues and complexity pushed to the "back end"

■ Data Warehouse

- A centralized repository of mostly historical information, built from operational data sources
- Usually contains several different subject areas
- Always in open database tables
- Usually detailed level information, but can include summary data too
- A single version of the truth

■ Data Mart

- Built from the Data Warehouse to support a specific business reporting requirement
- Often summarized, but may be detailed
- If in database tables, often a star schema structure
- May be in a proprietary format - ie MOLAP (cube), such as Essbase
- Updated (or re-built) on a regular basis from the data warehouse

■ Operational Data Store

- A reporting database containing the 'current' view of the operations of the business
- Contains little or no historical data
- Contains incomplete or in-progress entities (ie sales orders not yet fulfilled)
- Completely re-built on a regular (usually daily) basis
- Often creates outputs that feed back into operational systems

■ Metadata

- “Data that describes data”
- Technical metadata – e.g. table and column names, length, data type, decimals
- Business metadata - e.g. validation rules, transformation rules, source/target relationships
- Administrative metadata - e.g. users, authorities, size, usage, performance and data quality statistics, change history

Data Mart Example



PRODUCTS	
PRODUCT_NUMBER	5P 0
PRODUCT_DESCRIPTION	42A
BRAND_CODE	5A
BRAND_DESCRIPTION	20A
ORIGIN_CODE	5A
ORIGIN_DESCRIPTION	20A
FAMILY_CODE	5A
FAMILY_DESCRIPTION	20A
COST	9P 2
BASE_PRICE	9P 2
PRODUCT_WEIGHT	9P 4
PRODUCT_VOLUME	9P 4
LOAD_DATE	DATE
LAST_CHANGE_TIME	TSTP
STATUS_FLAG	1A

INVOICE_LINES	
INVOICE_NUMBER	7P 0
INVOICE_LINE_NUMBER	3P 0
PRODUCT_NUMBER	5P 0
CUSTOMER_NUMBER	10A
SELLING_COMPANY	5A
SUPPLY_WAREHOUSE	5A
QUANTITY_ORDERED	
QUANTITY_SHIPPED	
TOTAL_DISCOUNT	
NET_PRICE	
BASE_PRICE	
UNIT_COST	
EXTENDED_COST	11P 2
EXTENDED_PRICE	11P 2
MARGIN	11P 2
SALES_REP	5A
COMMISSION_VALUE	7P 2
INVOICE_DATE	DATE
SHIP_DATE	DATE
DELIVERY_DATE	DATE
INVOICE_TIME	TIME
MONTH_NUMBER	2P 0
WEEK_NUMBER	2P 0
LOAD_DATE	(DATE)

CUSTOMERS	
CUSTOMER_NUMBER	10A
CUSTOMER_NAME	35A
ADDRESS_LINE_1	35A
ADDRESS_LINE_2	35A
PHONE	15A
PHONE_2	15A
SALES_REP_DEFAULT	5A
CUSTOMER_CATEGORY	5A
CUSTOMER_CLASS	5A
REGION_CODE	5A
LOAD_DATE	DATE
LAST_CHANGE_TIME	TMSTP
STATUS_FLAG	1A

Meaningful table and column names

Complex calculations already done

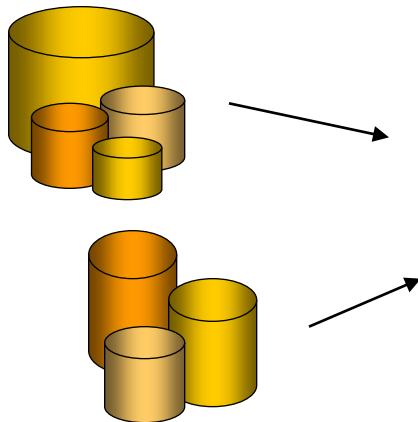
Dates are true date columns

De-normalized design reduced to only 3 tables

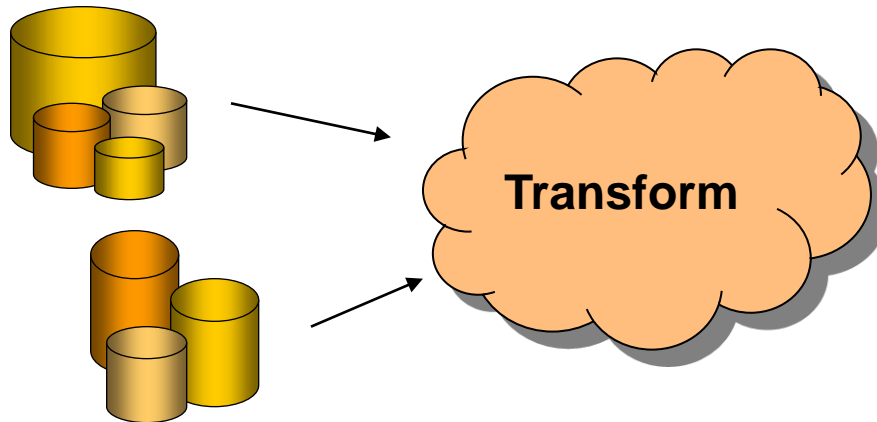
Only includes the columns we care about

■ Extract data from sources

- Database tables (IBM i)
- Remote Databases (e.g. DB2, MS SQL Server, Oracle)
- Text/delimited files
- Change Data Capture from journal images



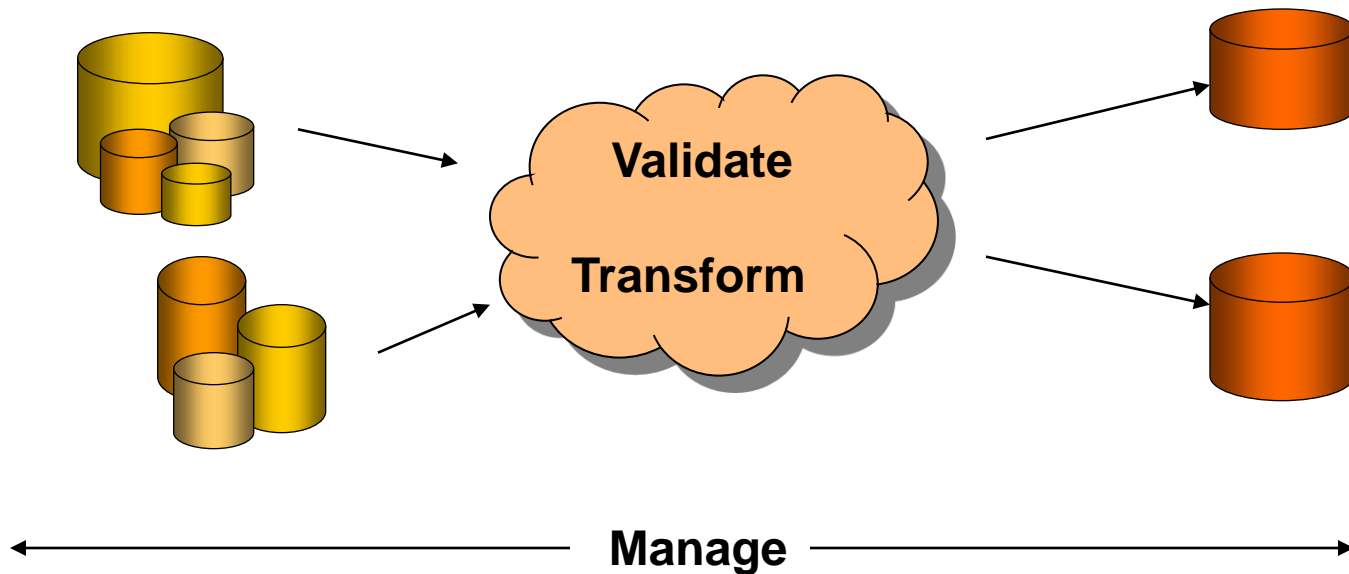
- **Transform** the extracted data
 - arithmetic calculations
 - string operations
 - lookup/replace
 - date/time conversions and calculations
 - data quality rules
 - integrate disparate data



- **Load** the transformed data
 - Into one or more target tables
 - detail or summary level
 - insert or update



- But... There are two VITAL additional requirements
 - Validate – define business rules
 - Manage – data errors
 - the overall environment



What is RODIN?

- **RODIN is a Data Management tool to define and maintain Data Warehouses and Data Marts**
 - Builds the Data Warehouse and Data Mart tables
 - Manages changes
 - Provides security and operational environment

- **RODIN is an ETL tool**
 - Allows ETL processes to be built without any complex programming
 - Extensive business rule capabilities to manage data quality
 - Extensive error management and auditing functionality

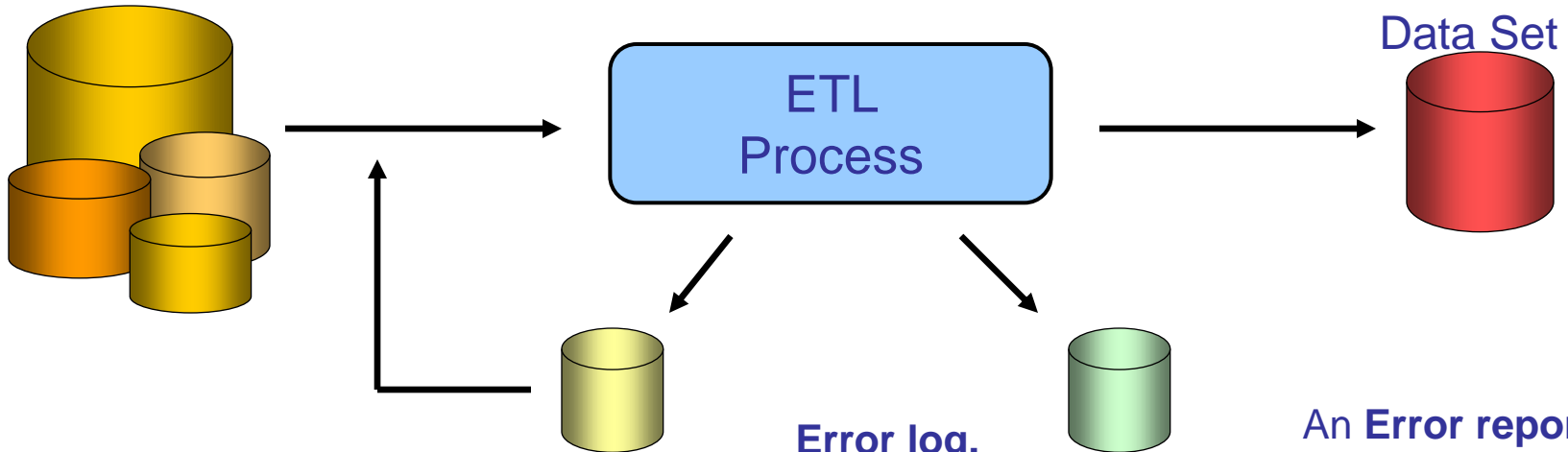
- **RODIN is a Metadata tool**
 - Many types of metadata automatically captured and enhanced
 - Unlimited free format descriptive text allowed at any level
 - Provides accurate, complete documentation

RODIN Demonstration

- We only have around 20 minutes – but that's plenty of time to create and load a data warehouse table:
 - Define and build a CUSTOMERS table in the data warehouse
 - Identify and generate metadata for the source tables
 - Define the ETL definition
 - Include selection criteria
 - Include data quality rules
 - Include transformations
 - Source to target mappings
 - Run ETL and review result
 - Data in target table
 - Review error report
 - Review audit report
 - Revise to handle errors

RODIN Error Management

Source Tables



Error suspense table.

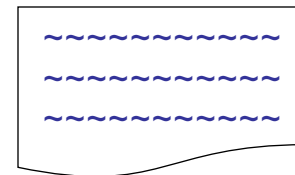
Rejected data is written to this table. After correction, this data can be re-processed by the extract in **error recovery mode**.

Error log.

Reason for error is written out to this table.

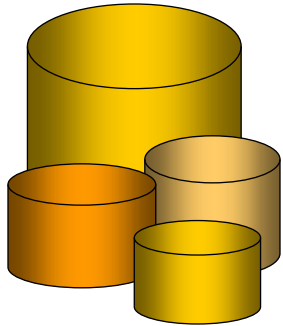
An Error report

is produced at end of processing if any errors occurred.

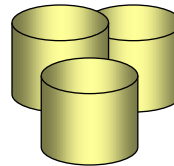
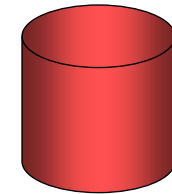


RODIN Auditing

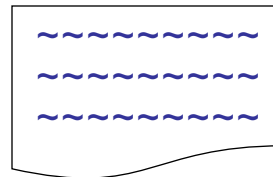
Source Tables



Data Set



Audit Tables



Audit report produced at end of extract.

Audit information includes

- Number of source rows read, omitted and rejected.
- Hash totals for up to 3 numeric source columns.
- Number of data sets updated and inserts/updates to each.
- Elapsed time for all jobs.
- etc.....

Audit information retained in meta-data indefinitely

Interesting quote

I just received this email in response to an article I recently had published via MC Press:

Alan,

I just read the article you wrote about business intelligence (BI). I wanted to let you know that was the most honest article I've ever read on BI.

I'm an independent consultant/contractor whose skills are typical for an IBM i professional; SQL, RPG, CL, DDS, etc. Many times I have found myself in the situation your next to last paragraph describes:

“Those organizations that forego an architected data warehouse and ETL almost always end up building independent data marts, with quick-and-dirty load processes. This usually results in multiple versions of the truth, duplicated effort, no metadata, and no error management, which in turn leads to a maintenance nightmare and eventual failure.”

Some of my clients in the past have been sold on the idea that BI is nothing more than a good reporting tool. I'll be using your article in the future to help warn future clients from venturing down the wrong path. I don't think a more clear, concise, and honest article has been written.

Thank you,

Article link: <http://www.mcpressonline.com/database/business-intelligence/unraveling-the-mysteries-of-data-warehousing-and-business-intelligence-on-ibm-i.html>

A Few RODIN Success Stories



RODIN

Data Engineering

EXTRACT ⇨ TRANSFORM ⇨ LOAD ⇨ MANAGE

For more information:

call **1-866-RODIN-DW**

visit www.thinkrodin.com