SyncOasis
Bring Your Information into Sync

Creating an Analytic Data Platform
CAS Spring Meeting 2019

Platform Technology Principles

1. Structured Environment
   - Every person for themselves (i.e. self service) is sub-optimal
2. Microservices and APIs
3. Workloads
   - Loading data
   - Transforming data
   - Reporting data
   - Analyzing data

Workflow 5 Steps

| Get the data | Load the data | Transform the data | Analyze the data | Present & Visualize Steps 2-4 |

- *Do ELT versus ETL (Extract, Load, Transform)*
- *Use Many Different Tools versus Monolithic Tool*

Step 1 - Get the Data

Data Movement Software

- Scheduled movement of data independent of sources

- What types of sources?

- Open Source versus Commercial Tools

- On-Premises versus Cloud Tools
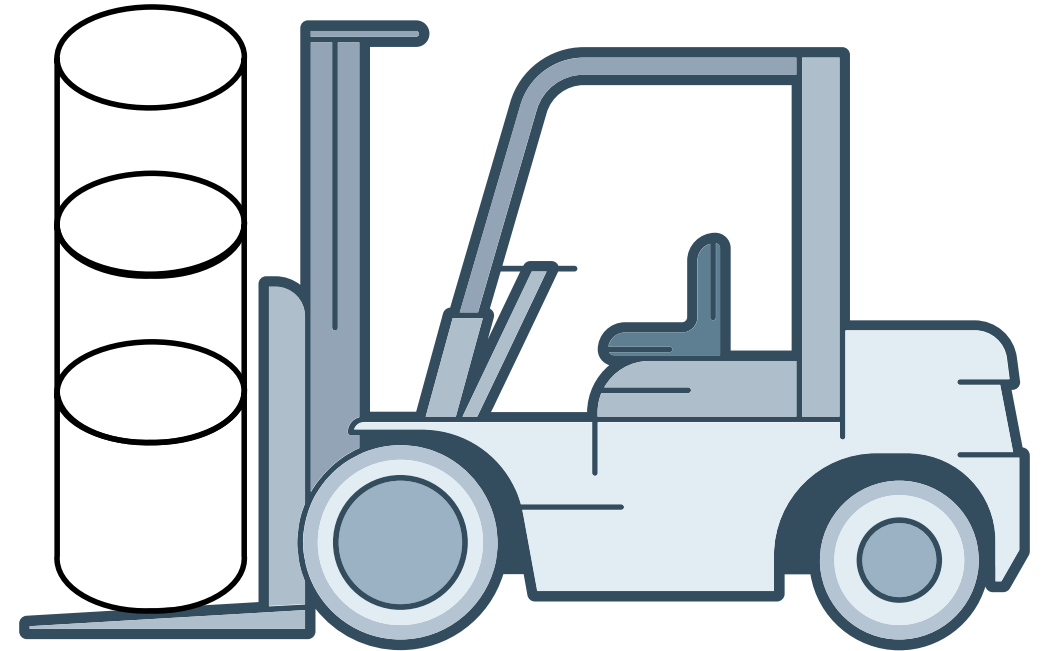
Step 2 - Load the Data

Where?

- Object Storage

- Hadoop

- Relational Databases

- Document / Graph DBs

- "One Size Fits All": An Idea Whose Time Has Come and Gone (Stonebraker and Cetintemel)

Step 3 - Transform the Data

- Integrate / Standardize as if from the same source

- Apply Data Quality

- Merge / Match data (Customer 360)

- Flexible to add new sources and data elements

- Ability to run both point-in-time and period-in-time queries

- Ability to get the same data when running the same query now or in the future

- Business logic (i.e. mappings) managed through table driven logic

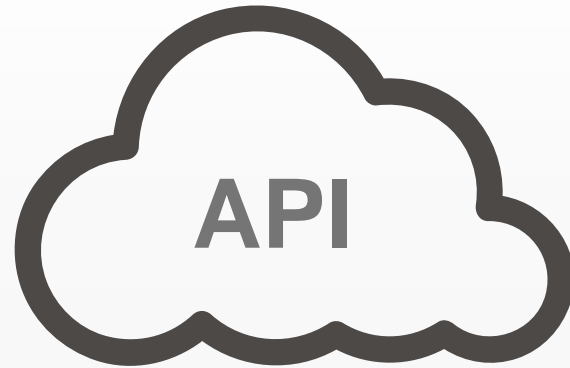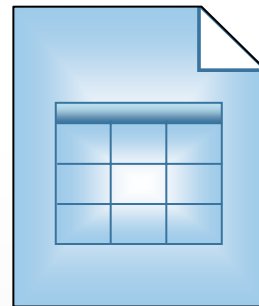- Organize data by physical entities

Step 4 - Analyze the Data

▪ In database analytics? Table extracts for analysis?

▪ Embed analytical data with operational data?

▪ Share analytical results with other users?

Step 5 - Present / Visualize the Data

Conclusions

- Need structure and discipline

- ELT versus ETL

- Need many tools, use best tool for task and need

- Cloud provides more options than On-Premises

- Likely need multiple data stores

- Core data store should meet your key requirements, may not be where end-users access the data