The Business Intelligence Promise:
Smarter, more fact-based decision-making as an everyday routine

Presented by Daniel Upton

PerformanceG2, Inc. is a full service performance management consultancy and advisory experts who specialize in helping business leaders drive the value of their data and or analytic environment to ultimately allow them to think smarter, act faster, and grab a competitive edge. Our goal is to guarantee that our clients have the right knowledge and intelligence of their business – customers, products, markets, and people, and we do so by partnering with the best technology companies and hiring the top consultants and trainers.

www.performanceg2.com

Daniel Upton, owner and principal consultant at **DecisionLab LLC**, is a Certified Data Vault Data Modeler. A frequent speaker, blogger, and advocate for agile methods in the business intelligence and analytics space, he has used his unique EDW model-storming method to quickly implement Data Vaults for many organizations nationwide. Daniel is pleased to **partner with PerformanceG2** to deliver the benefits of a data warehouse that is more robust, faster to deliver, easier to adapt, and less costly to maintain than previously considered possible.

www.decisionlab.net

3

Without a Data Warehouse, BI reporting gets bogged down...
* slow query performance
* insufficient drill-down to details, and
* mission critical reports with conflicting data.

A Data Warehouse can resolve those issues, but has been...
* time consuming to build
* slow to adapt to rapid business changes
* costly not only to build, but even to maintain.

Why so?

4

Presented by Daniel Upton

Star Schemas:

Appealing simplicity to less technical users, but to load them, the facts, with their new, hard-coded business rules become dependent on dimensions, with their own set of hard-coded rules, making these models, and their loading code, time-consuming to design and load, and …costly to evolve or change even in small ways once delivered. Meanwhile, source system data, business rules, and reporting requirements continue to evolve faster and faster.

So, with a delivered old school 'Straight to Star Schema' architecture, even without a need for major enhancements …business intelligence maintenance costs increase over time.
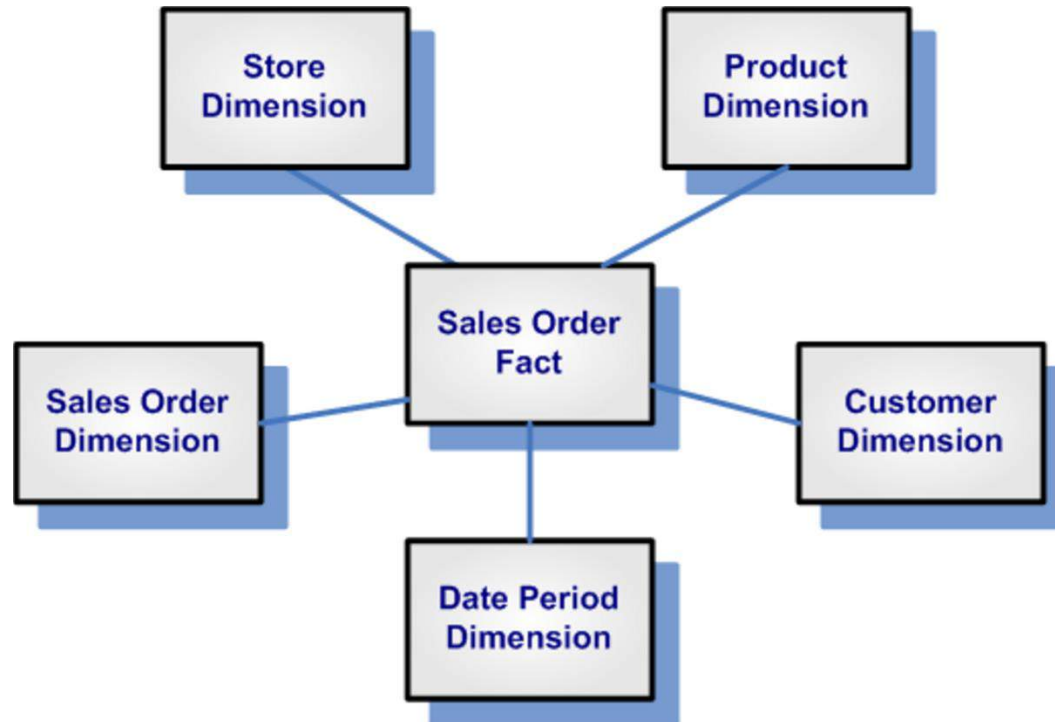
Presented by Daniel Upton

"Build your integrated data warehouse now.
Easily add new operational data, new rules, and new relationships whenever needed"

…would never have been suggested by an honest, experienced practitioner.
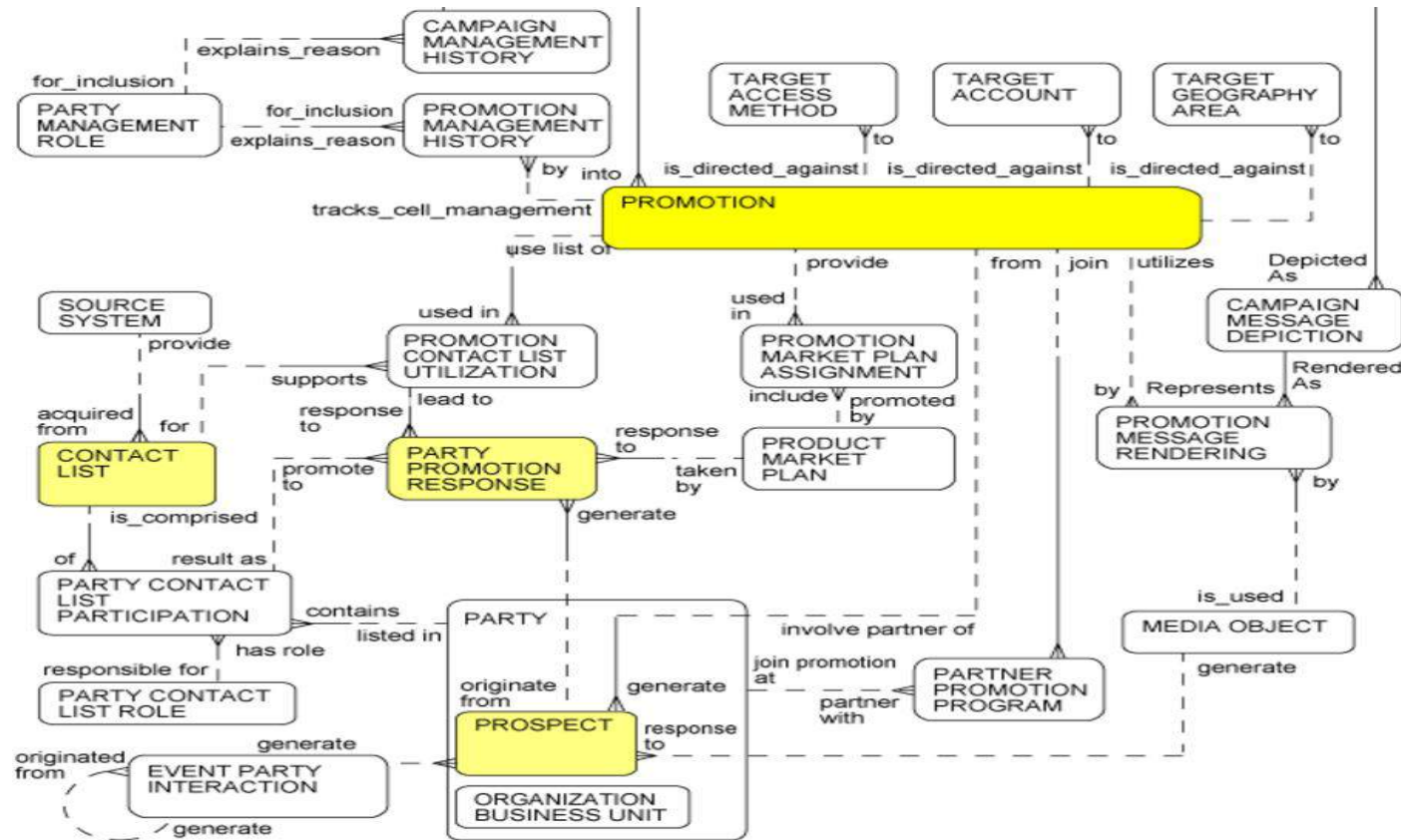
Why do old school data warehouses take so long to deliver?

Data dependencies result in task dependencies

Requirements
Data Model
Profile
Stage Data (ETL)
Deliver Data
Integrate Staged Data (ETL)
Load to Star Schema (ETL)
Build Semantic Layer
Build Dashboard
Train Users
Release
Maintain

DecisionLab.Net

7

Presented by Daniel Upton
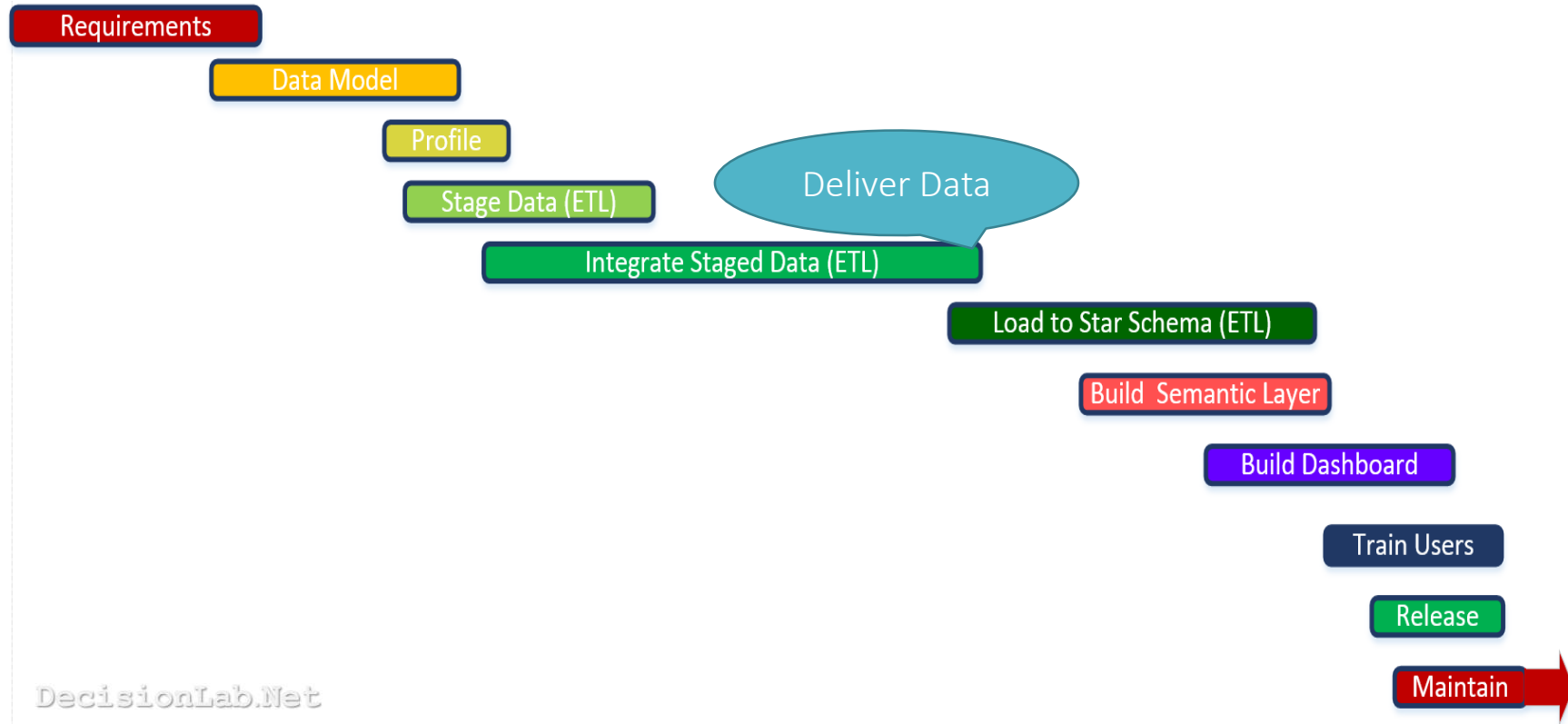
The five most difficult, most sequential tasks

Sprint 1 Ends   Sprint 2 Ends   Sprint 3 Ends

G: Gather Req'mts.

M: Model Data

P: Profile Data

S: Stage Data in ETL

I: Integrate Data in ETL

Deliver Data

DecisionLab.Net

8

Presented by Daniel Upton
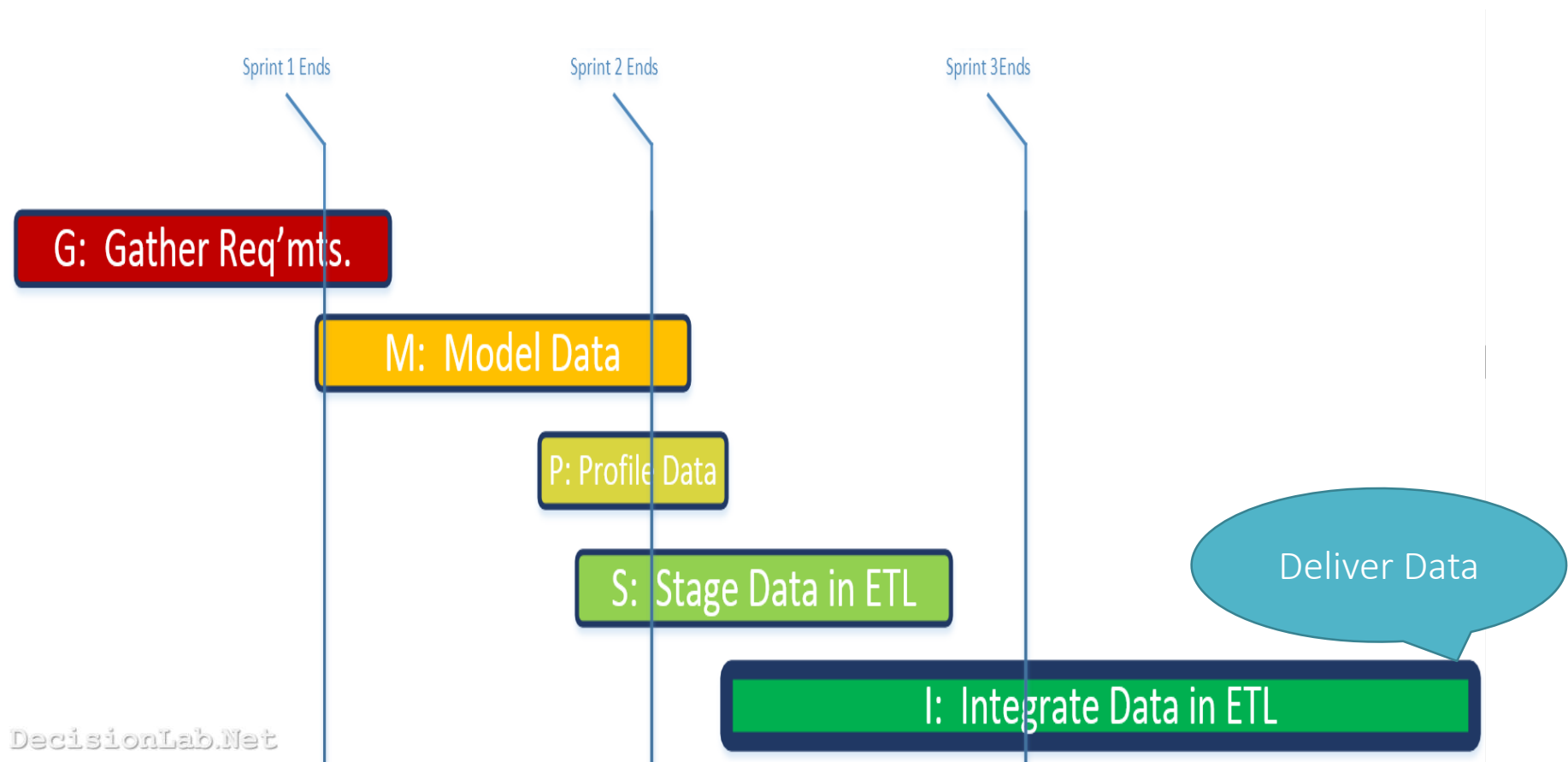
Critical Path Delays
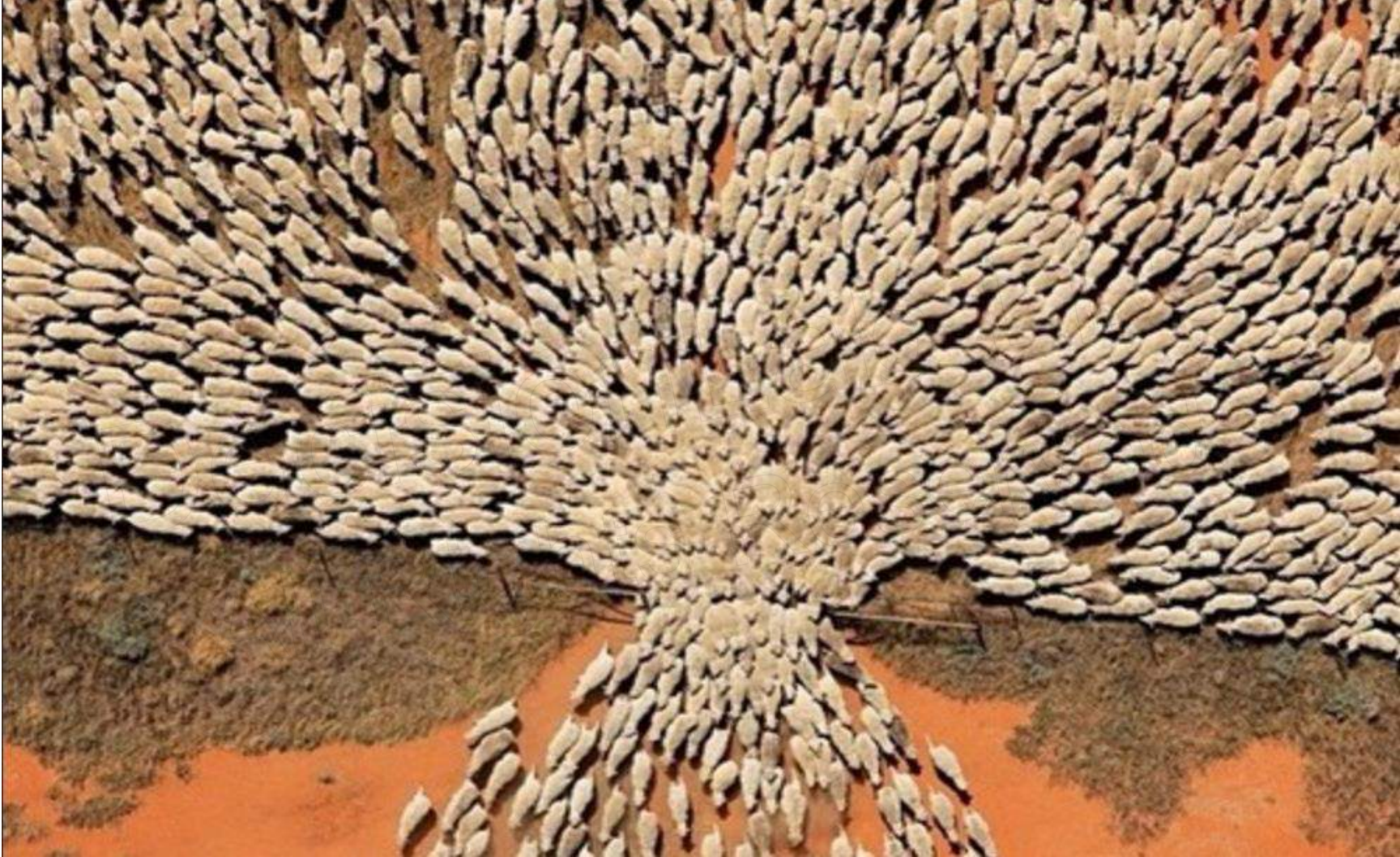
Data Vault Architecture, especially with Modelstorming, shortens tasks and their dependencies into smaller, time-boxed iterative tasks by loosely coupling and quickly warehousing the data before applying business rules for reporting or analytics.

Presented by Daniel Upton
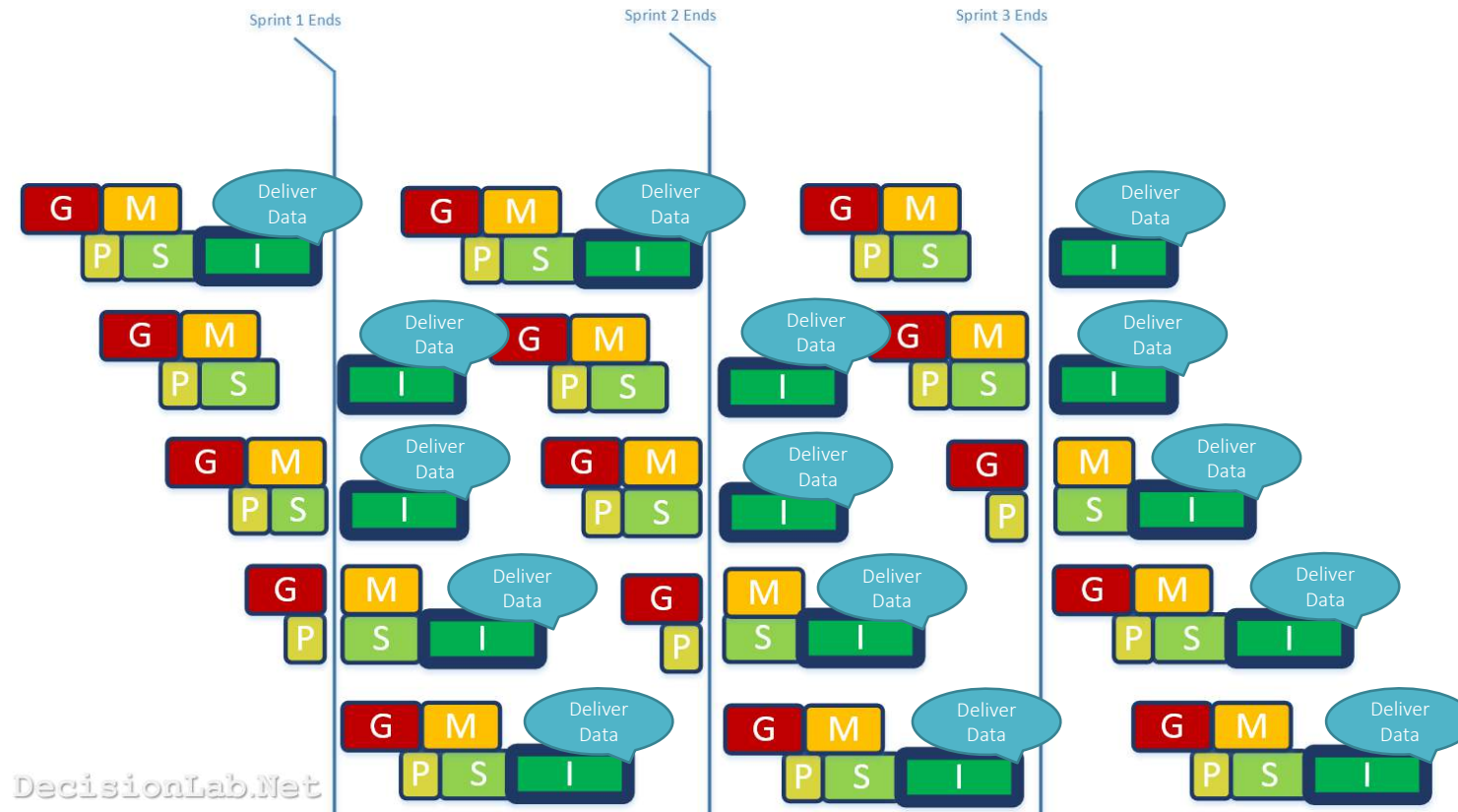
Data Vault is a platform-agnostic architecture.

High Level Data Vault BI-Solution Architecture

Who is using Data Vault? ~ 2,000 worldwide implementations

Simple Business Case: University

Alumni Development specialists must know which donors donate how much and when, in order to ensure donors receive recognition and benefits.

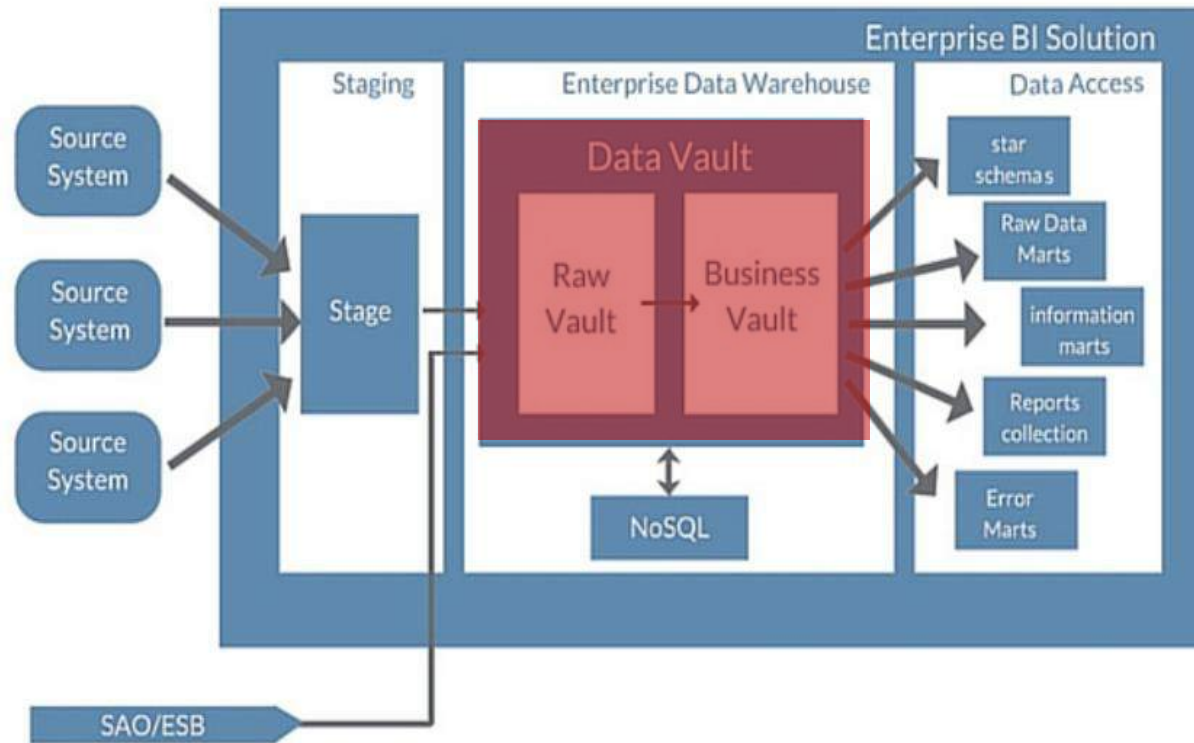Institutional Research must know how many students enrolled weekly to ensure University meets qualifications for student loans.
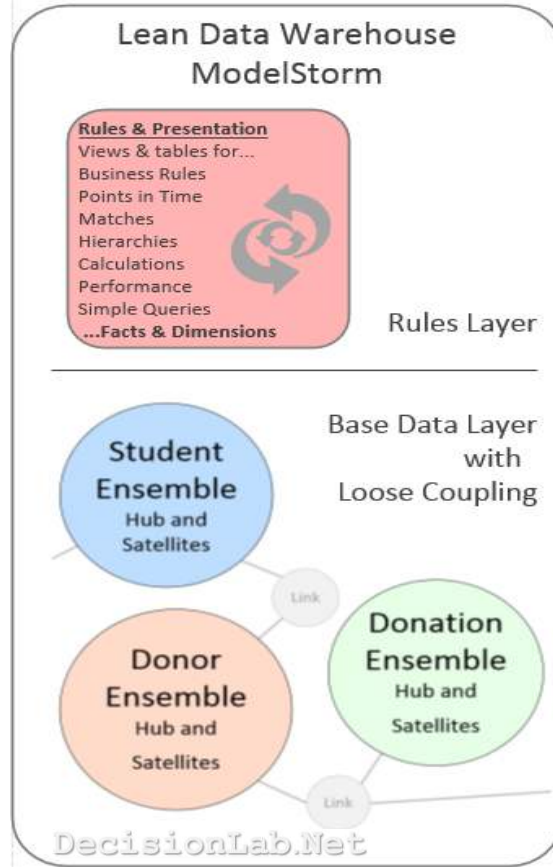
Auditors must know what family relationships exist between donors and students to monitor for conflicts of interest and uphold reputation.

13

PERFORMANCE G2

Data Vault and ModelStorming

User Story:
As a____, I must know____,
so I can do____, and we can achieve____.

Step #1

Lean Data Warehouse ModelStorm

Rules & Presentation
Views & tables for...
Business Rules
Points in Time
Matches
Hierarchies
Calculations
Performance
Simple Queries
...Facts & Dimensions

Rules Layer

Front End: Reports and Visualizations

Data Presentation Layer ModelStorm

Fact ModelStorm
Donor Donates to University

	With Relationship to Students	Makes Donation to University	(With or Without) Attending a Fundraiser	Date of Donation
Donor				

Dimension ModelStorm

Student... has a...	and a...	and a...	and a...
Name	Counselor	Major	Status

Facts & Dimensions Matrix
* Donations
* Student Enrollments

Facts (events of interest)	Who?	What?	How?	When? Relevant Dates
Donor Donates to University	G		x	x G
Student is Enrolled in an Academic Program (Weekly Snapshot)	G x x x			G

#6

#5

#2

Lean Goals:
Releases in ½ Time
Load Data in ½ Time
Remove Dependencies
Delay Rules Decisions
Design is Adaptable
Data is Reloadable
Ingest All Data

Student Ensemble
Hub and Satellites

Base Data Layer with Loose Coupling

Donor Ensemble
Hub and Satellites

Link

Donation Ensemble
Hub and Satellites

Link

#4

#3

Source Data - Fast Profile

Registrar | Fundraising

Student
StudentID
FirstName
LastName
ContactID
CounselorCode

Donor
Donor_Code
Donor_Formal_Name
Rel_To_Primary_Student_ID
Rel_Descrip_To_Primary_Student
Contact_First_Name

Donation
Donation_Code
Donor_Code
Date
Amount

DecisionLab.Net

14

Data Vault and Data ModelStorming:
Fast Profile of Source Data
Data Dependencies Force Sequential Transaction Processing

PERFORMANCE G2

DecisionLab.Net

Presented by Daniel Upton
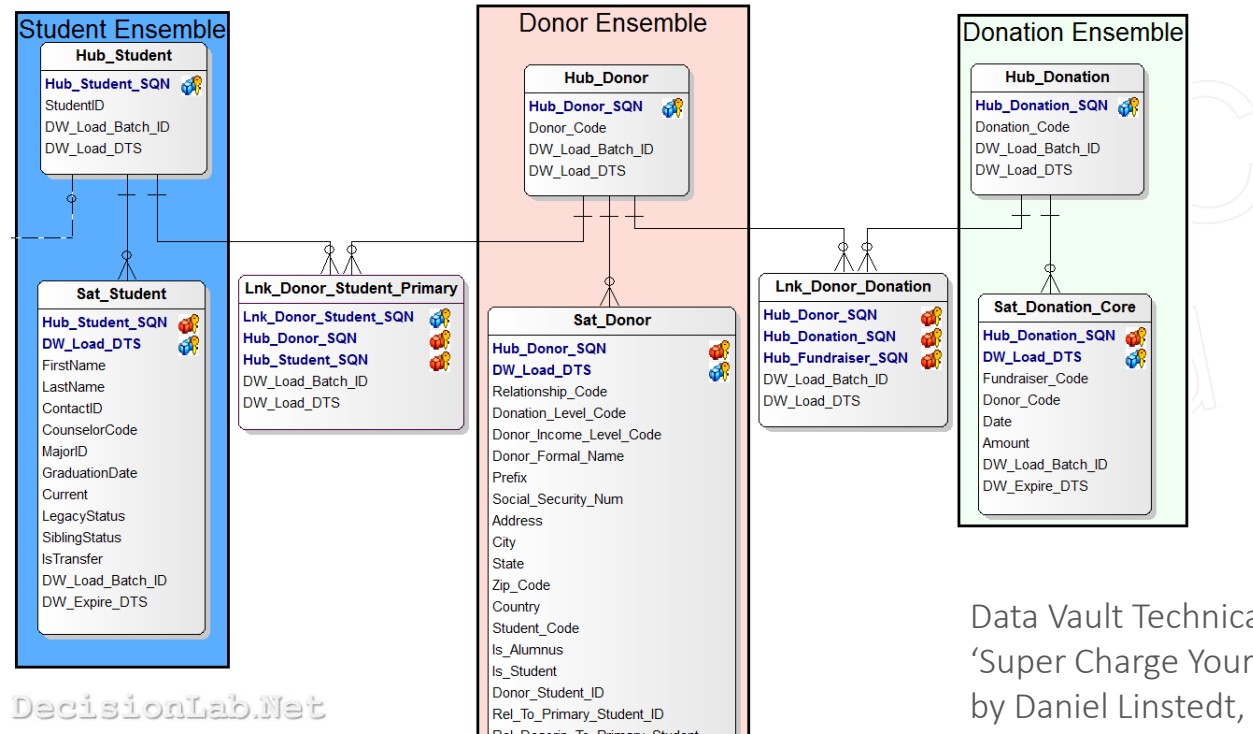
15

Data Vault Modeling, via Modelstorming, is quickly done
High Cohesion and Loose Coupling
Delivered in **1/4** the time of old-school models.
No data dependencies between free-standing ensembles, thus
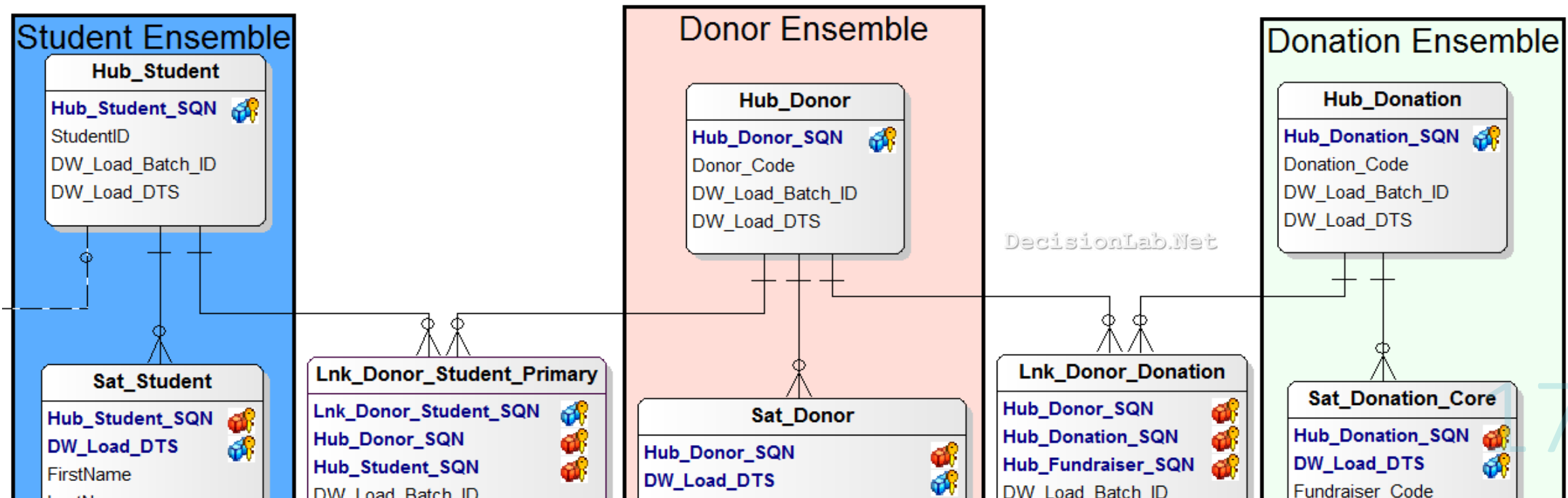load ensembles simultaneously, so loading time is **1/2** of old-school.

Student Ensemble

Hub_Student
Hub_Student_SQN
StudentID
DW_Load_Batch_ID
DW_Load_DTS

Sat_Student
Hub_Student_SQN
DW_Load_DTS
FirstName
LastName
ContactID
CounselorCode
MajorID
GraduationDate
Current
LegacyStatus
SiblingStatus
IsTransfer
DW_Load_Batch_ID
DW_Expire_DTS

Lnk_Donor_Student_Primary
Lnk_Donor_Student_SQN
Hub_Donor_SQN
Hub_Student_SQN
DW_Load_Batch_ID
DW_Load_DTS

Donor Ensemble

Hub_Donor
Hub_Donor_SQN
Donor_Code
DW_Load_Batch_ID
DW_Load_DTS

Sat_Donor
Hub_Donor_SQN
DW_Load_DTS
Relationship_Code
Donation_Level_Code
Donor_Income_Level_Code
Donor_Formal_Name
Prefix
Social_Security_Num
Address
City
State
Zip_Code
Country
Student_Code
Is_Alumnus
Is_Student
Donor_Student_ID
Rel_To_Primary_Student_ID
Rel_Descrip_To_Primary_Student

Lnk_Donor_Donation
Hub_Donor_SQN
Hub_Donation_SQN
Hub_Fundraiser_SQN
DW_Load_Batch_ID
DW_Load_DTS

Donation Ensemble

Hub_Donation
Hub_Donation_SQN
Donation_Code
DW_Load_Batch_ID
DW_Load_DTS

Sat_Donation_Core
Hub_Donation_SQN
DW_Load_DTS
Fundraiser_Code
Donor_Code
Date
Amount
DW_Load_Batch_ID
DW_Expire_DTS

DecisionLab.Net

Data Vault Technical Reference:
'Super Charge Your Data Warehouse'
by Daniel Linstedt, LearnDataVault.com, 2008.

16

Presented by Daniel Upton

Summary of Data Vault Design Fundamentals

Hub: One Business Key (unique list of values); Surrogate Key as Primary Key. A Hub is free-standing, never functionally dependent on another table. Has one or more dependent Satellites.
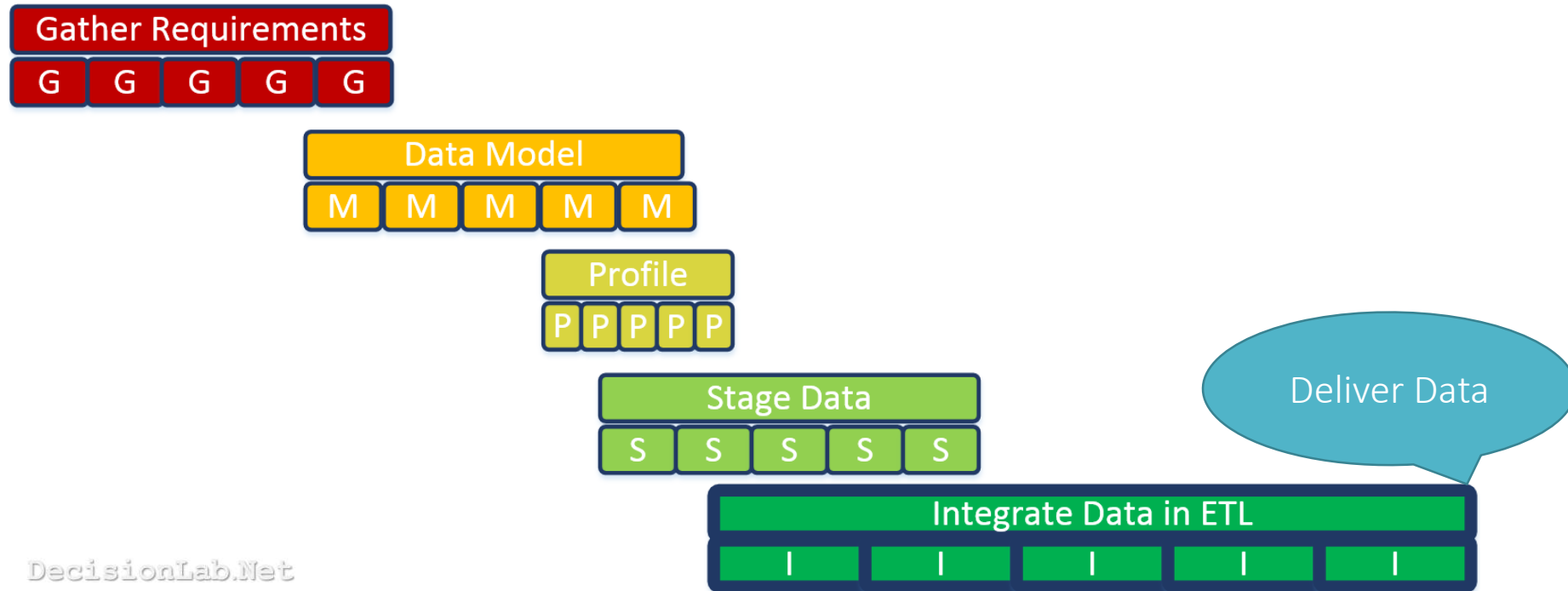
Satellite: Primary key is Hubs Surrogate Key + DateTimeStamp to capture historic versions of all fields from source table or file. Foreign key constraints are removed. Dependent on exactly one Hub.

Link: Historic relationship of records between two or more Ensembles, using new Hub Surrogate Keys. Used instead of source system's foreign keys, thus preserving correct cardinality relationships among records between tables, while removing cascading data dependencies between Ensembles, allowing faster data loads and allowing for future enhancement without re-work in existing tables.


Student Ensemble

Hub_Student
- Hub_Student_SQN 🔑
- StudentID
- DW_Load_Batch_ID
- DW_Load_DTS

Sat_Student
- Hub_Student_SQN 🔑
- DW_Load_DTS 🔑
- FirstName

Lnk_Donor_Student_Primary
- Lnk_Donor_Student_SQN 🔑
- Hub_Donor_SQN 🔑
- Hub_Student_SQN 🔑
- DW_Load_Batch_ID

Donor Ensemble

Hub_Donor
- Hub_Donor_SQN 🔑
- Donor_Code
- DW_Load_Batch_ID
- DW_Load_DTS

Sat_Donor
- Hub_Donor_SQN 🔑
- DW_Load_DTS 🔑

Lnk_Donor_Donation
- Hub_Donor_SQN 🔑
- Hub_Donation_SQN 🔑
- Hub_Fundraiser_SQN 🔑
- DW_Load_Batch_ID

Donation Ensemble

Hub_Donation
- Hub_Donation_SQN 🔑
- Donation_Code
- DW_Load_Batch_ID
- DW_Load_DTS

Sat_Donation_Core
- Hub_Donation_SQN 🔑
- DW_Load_DTS 🔑
- Fundraiser_Code

DecisionLab.Net

Modelstormed Data Vault Architectural Process reduces task dependencies to a fraction, by quickly warehousing and loosely coupling the data prior to applying business rules for analytics or reporting.


Presented by Daniel Upton
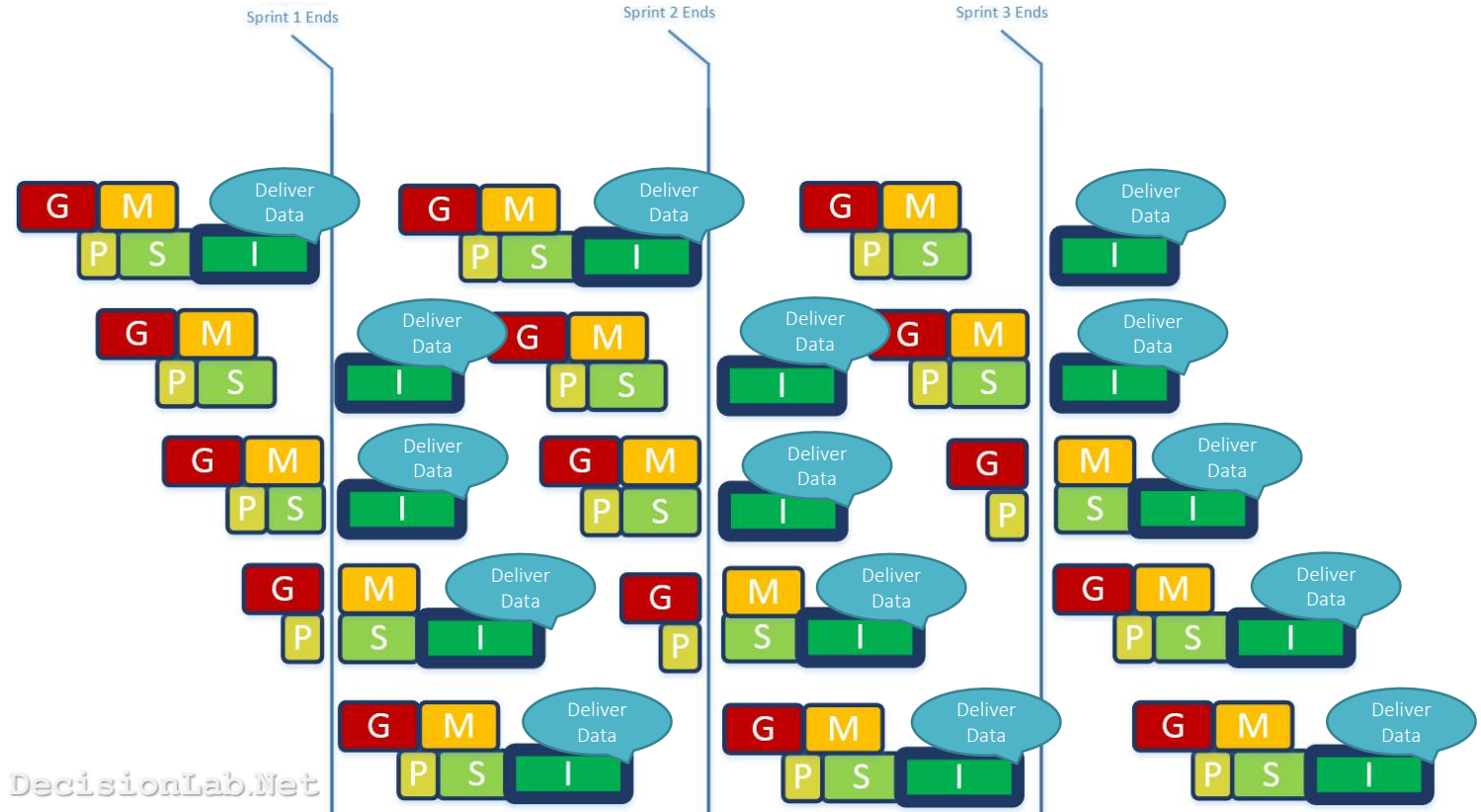
In 11ᵗʰ Hour,
Two User Stories Change

Institutional Research must know how many students enrolled weekly, with what <u>relationship to a Counselor</u>, to ensure University meets qualifications for student loans.

Auditors must know what family relationships exist between donors and students <u>and Counselors</u> to monitor for conflicts of interest and uphold reputation.

_____

In a straight-to-star schema architecture, such requirement changes require <u>time-consuming, costly re-engineering</u> within the data warehouse. Here, <u>a fact table</u>, first built without reference to Counselor, <u>now needs re-engineering</u>. DecisionLab.Net

20

Presented by Daniel Upton

Data Vault Model: Prior to 11ᵗʰ Hour Change

Presented by Daniel Upton

21
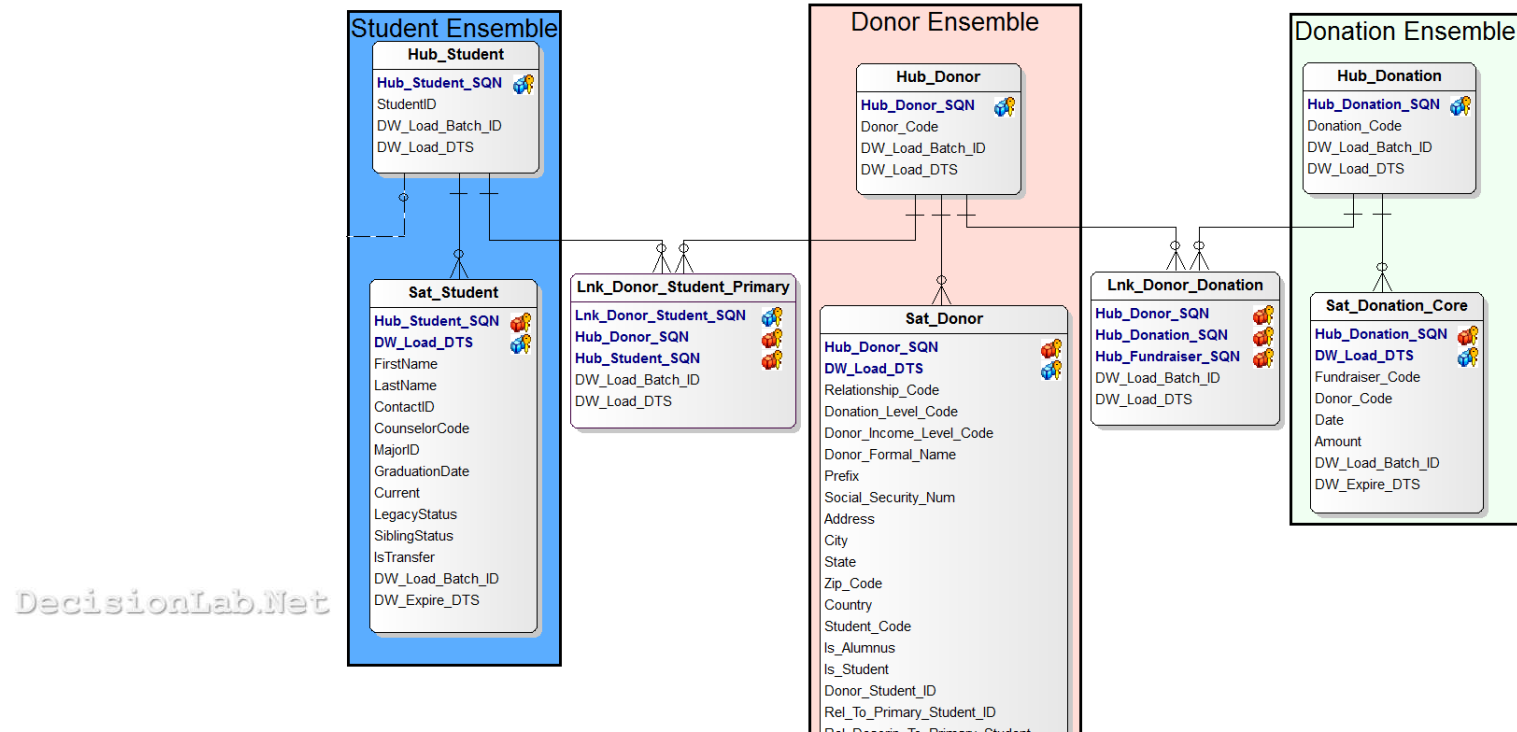
# Data Vault Model: After 11th Hour Change
Lean ModelStorm Additions Complete
Zero changes to any existing objects, thus fast and no interference with other coding efforts. Delivered this time in ¼ the time of old-school methods.

PERFORMANCE G2

DecisionLab.Net

Presented by Daniel Upton

22
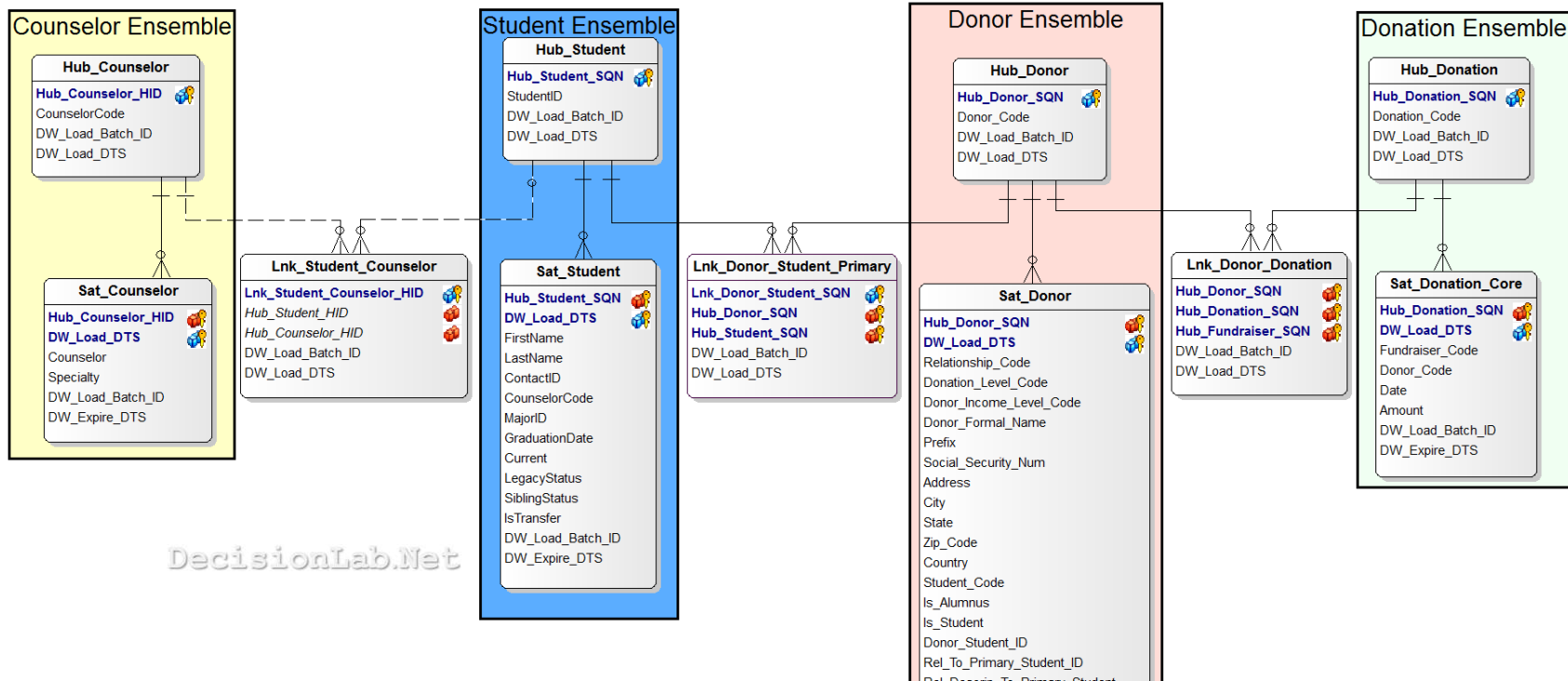
Data Vault Architecture
Robust Integration with Big Data from
Social Media, Machine Sensors, etc.

High Level Data Vault BI-Solution Architecture
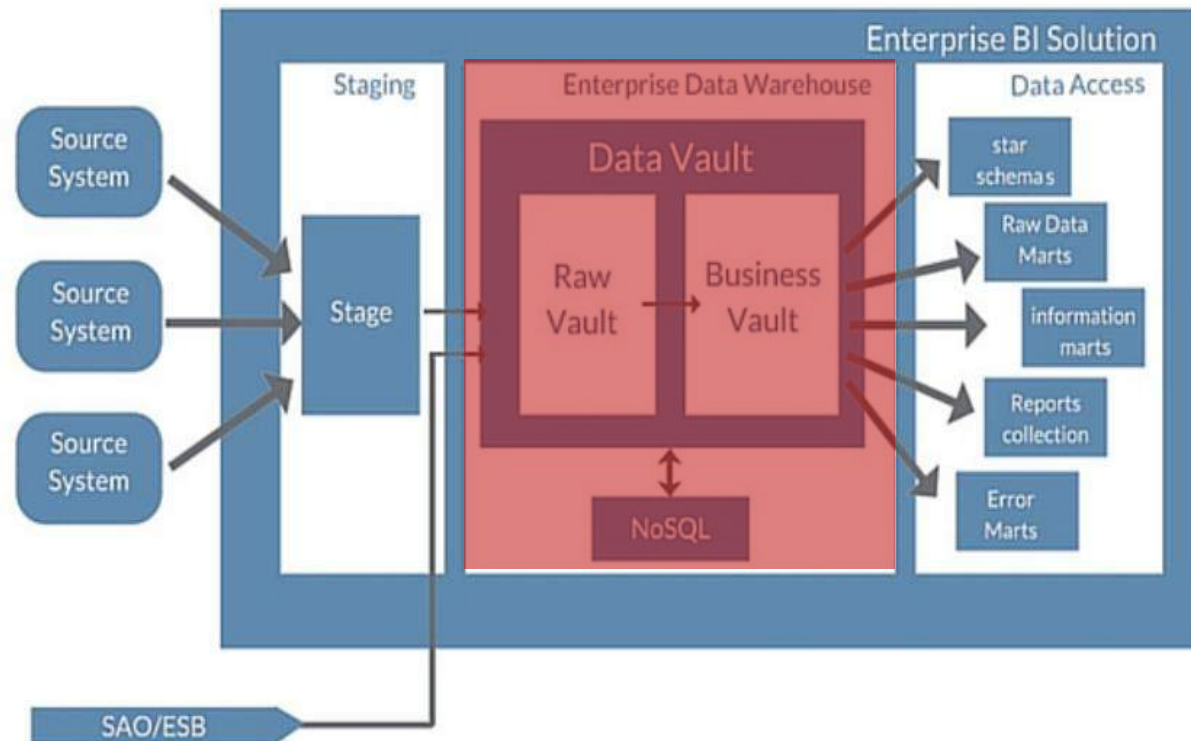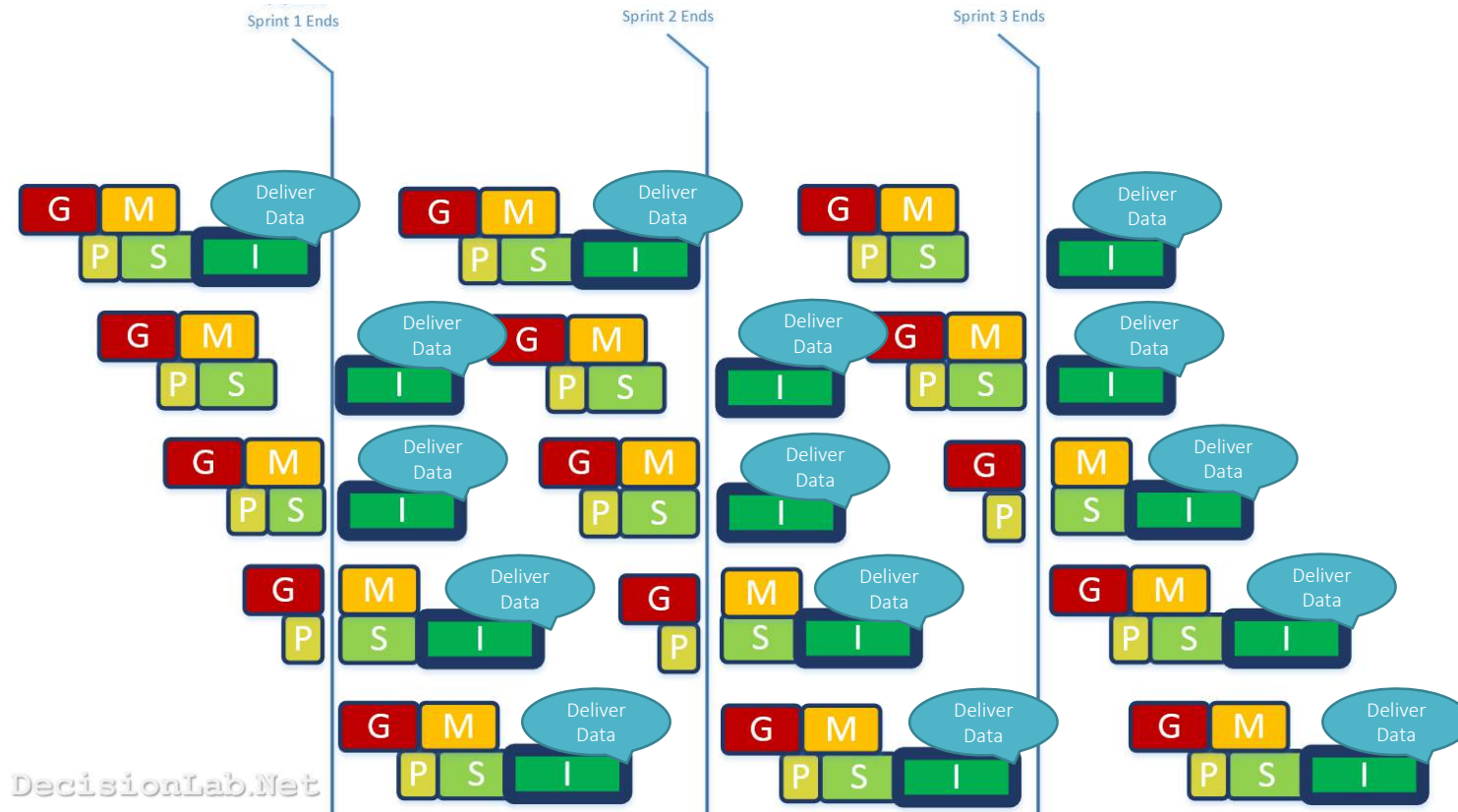
"Build your integrated data warehouse now.
Easily add new operational data, new rules, and new relationships
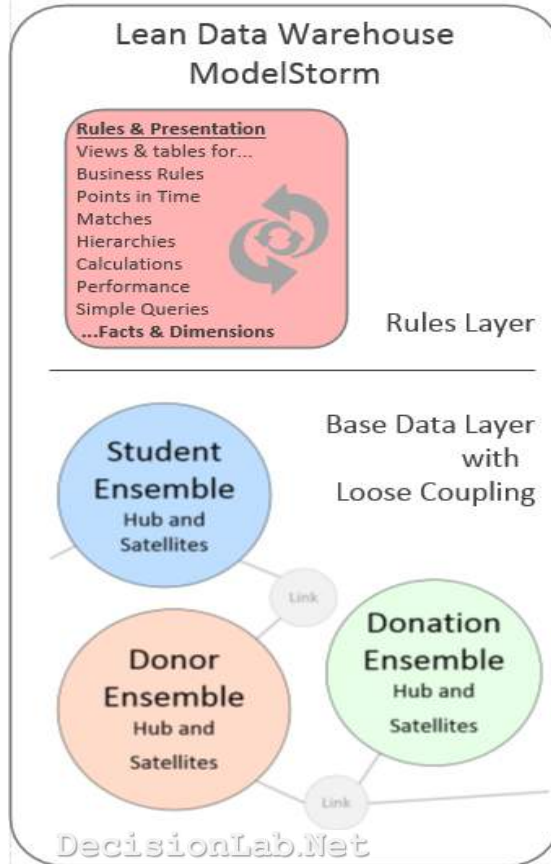# whenever needed" *…is now a routine expectation.*

24
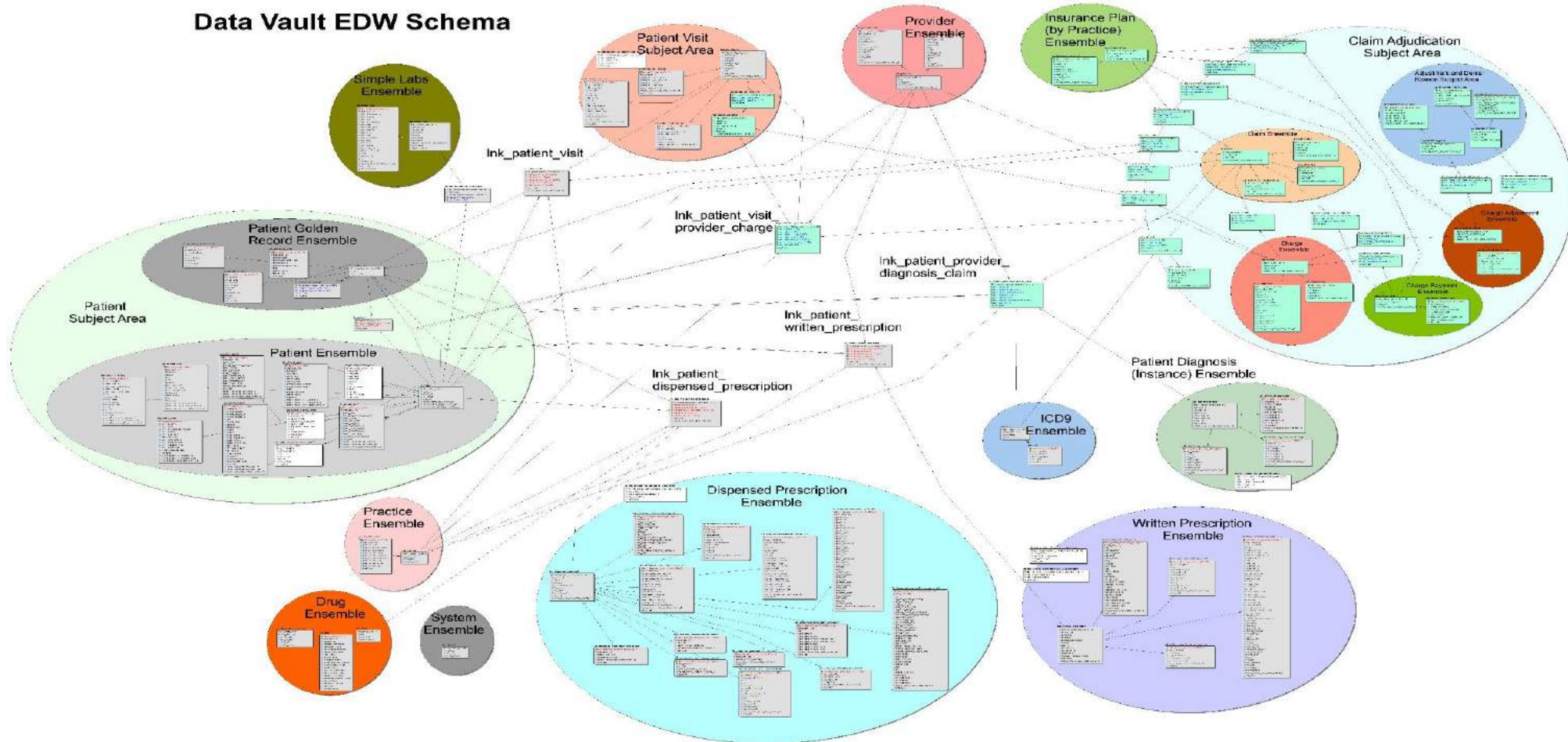
Data Vault Architecture, especially with Lean Modelstorming, reduces task dependencies to a fraction, warehousing and loosely integrating the data prior to applying your business rules for analytics or reporting.

Data Vault Modelstorming

User Story:
As a___, I must know___,
so I can do___, and we can achieve___.

Step #1

Lean Data Warehouse ModelStorm

Rules & Presentation
Views & tables for...
Business Rules
Points in Time
Matches
Hierarchies
Calculations
Performance
Simple Queries
...Facts & Dimensions

Rules Layer

Front End: Reports and Visualizations

#6

#5

Lean Goals:
Releases in ½ Time
Load Data in ½ Time
Remove Dependencies
Delay Rules Decisions
Design is Adaptable
Data is Reloadable
Ingest All Data

#2

#4

#3

Data Presentation Layer ModelStorm

Fact ModelStorm
Donor Donates to University

	With Relationship to Students	Makes Donation to University	(With or Without) Attending a Fundraiser	Date of Donation
Donor				

Dimension ModelStorm
Student... has a... Name — and a... Counselor — and a... Major — and a... Status

Facts & Dimensions Matrix
* Donations
* Student Enrollments

Facts (events of interest)	Who?	What?	How?	When? Relevant Dates
Donor Donates to University	G		x	x G
Student is Enrolled in an Academic Program (Weekly Snapshot)	G x x x			G

Base Data Layer with Loose Coupling

Student Ensemble Hub and Satellites

Link

Donor Ensemble Hub and Satellites

Link

Donation Ensemble Hub and Satellites

DecisionLab.Net

Source Data - Fast Profile

Registrar | Fundraising

Student
StudentID
FirstName
LastName
ContactID
CounselorCode

Donor
Donor_Code
Donor_Formal_Name
Rel_To_Primary_Student_ID
Rel_Descrip_To_Primary_Student
Contact_First_Name

Donation
Donation_Code
Donor_Code
Date
Amount

26

Presented by Daniel Upton

The Business Intelligence Promise:
Smarter, more fact-based decision-making as an everyday routine

27

Presented by Daniel Upton

For next steps…
speak with Joe Aanas

joe@PerformanceG2.com | Cell: 608-469-2190

www.PerformanceG2.com

Thank You