# PHYTORION

# Academic Data Warehouse

## Agile Analytics in Higher Education

Alliance 2012

Session: 30767

March 20, 1:15 PM

# Presenters

➤ Ashley Silverburg

   Chief Data Architect

➤ Robert Silcher

   Director Business Intelligence

➤ Yiorgos Marathias

   President

PHYTORION

# Background

- Started in Higher Education

- At PeopleSoft 1997-2005

  o Built First Reporting Solutions for PS Higher Ed

- Phytorion established Jan. 2005

  o Pre-Packaged Data Warehouses/BI

  o Hybrid Data Warehouses/BI

  o Fully Custom Data Warehouses/BI

PHYTORION

# Background

We Offer Data Warehouses/BI for:

- ✓ Campus Solutions

- ✓ Human Capital Management

- ✓ Finance

- ✓ Kuali Finance

- ✓ Legacy, Lawson, Spreadsheets, etc.

PHYTORION

# Customers include:

All Maine Public Universities

Loyola Univ. Chicago

Cornell University

Long Beach City College

Cerritos College

Maricopa College District

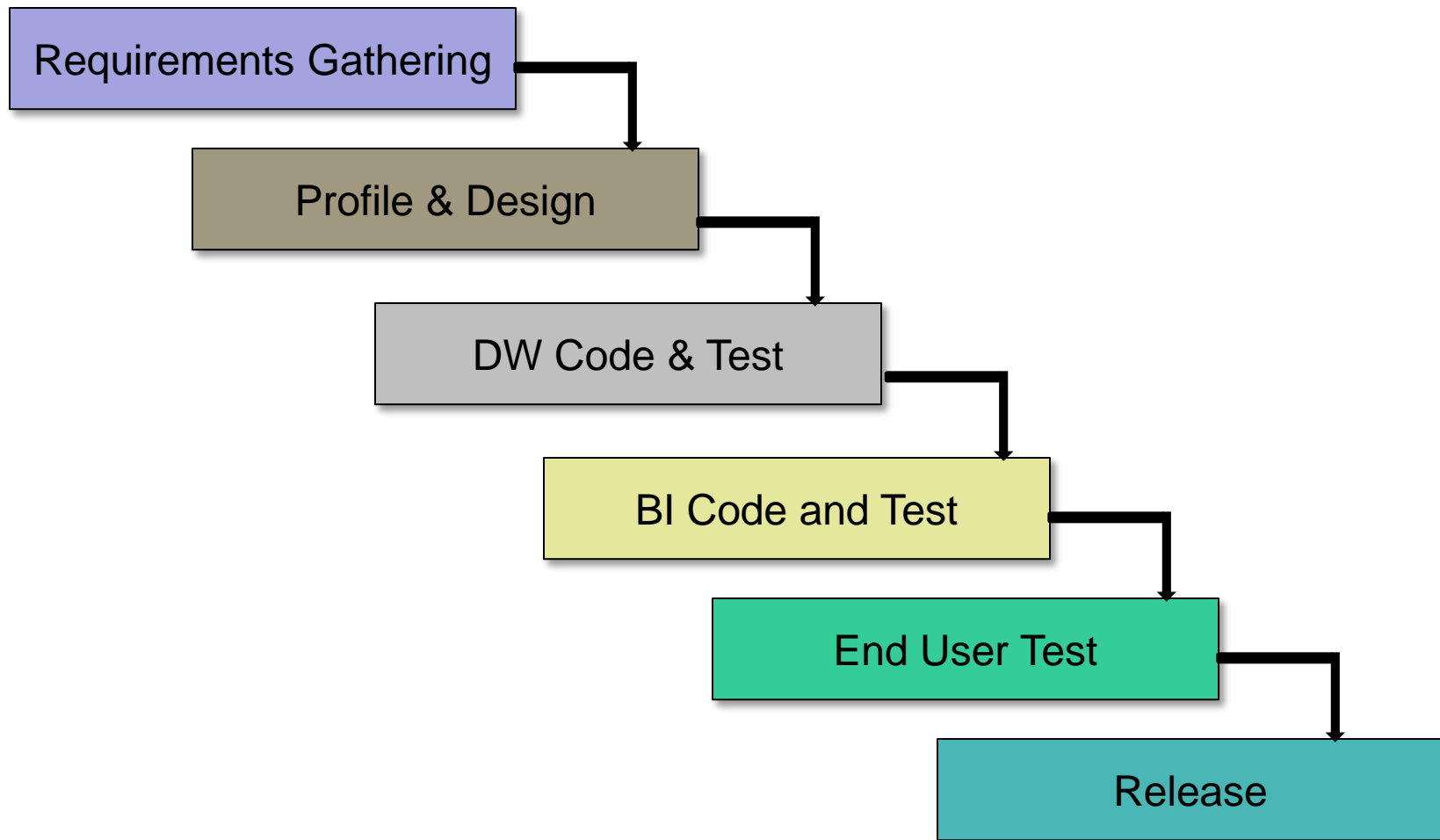Harvard University

University of Cambridge

Los Rios Comm. College

University of Manchester

Waterbury K-12 District

Iowa State University

PHYTORION

# Standard Waterfall Approach

Requirements Gathering

Profile & Design

DW Code & Test

BI Code and Test

End User Test

Release

PHYTORION

# Waterfall issues

➢ Failure rate range: 65% - 86% (DM Review Oct. 2004)

➢ Hard to Scope and Price. For 95% Conf. Int. = Estimate x 4

➢ Communicating needs & Verifying design

➢ Significant time between Requirements and Release

    ➢ Users lose interest, become skeptical

    ➢ Needs may have changed

    ➢ Users forget what they asked for

    ➢ Users truly understand deliverable with BI release

PHYTORION

# Agile Manifesto

We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

Individuals and interactions over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

"Safety" on the right

http://agilemanifesto.org/

PHYTORION

# Principles behind Agile

➢ Early and continuous delivery of working software

➢ Deliver software every 2-8 weeks. Time-box developmt

➢ Welcome changing requirements

➢ Business people and developers work together

➢ Maximize the work not done

➢ The best work emerges from self-organizing teams

➢ Reflect and tune

PHYTORION

# Agile Benefits

➢ 82% Greater Productivity

➢ 77% Higher Quality

➢ 78% Higher Customer Satisfaction

➢ 37% Lower Cost

PHYTORION

# How Does it Work?

Team Composition

- ➢ Product Owner – Primary role. Provides requirements and signs off on deliverables

- ➢ Scrum Master – knows the process; keeps team on track.  Not responsible for costs, resources, milestones.

- ➢ Architect

- ➢ Developers – they are engaged from the beginning and assume ownership.  Not just waiting for exact specs

PHYTORION

# How Does it Work?

Sprint: Duration 2-8 weeks

➢ Story driven: "As a <role>, I need the ability to <do something> so that I can <goal statement>"

➢ Story may be too large to complete: Epic or Theme

➢ Something of value that can be delivered

➢ Break it down to smaller parts that you can estimate and deliver

PHYTORION

# How Does it Work?

Examples

➢ As an Admissions Officer, I need to see weekly counts at all Admissions stages so I can track progress towards Admissions goals: EPIC

➢ I need to see weekly counts of Completed Applications for Undergraduate Admissions: USER STORY

➢ DEVELOPER STORIES:

  ➢ I need to build one SCD 1 dimension

  ➢ I need to build one fact

PHYTORION

# How Does it Work?

Estimate – traditional difficulties

➢ We are not good at estimating: 95% Conf Int. requires
    4x effort

➢ We estimate at project start when we know the least

➢ Errors add up over length of project

➢ We don't always account for "distractions"

➢ Estimates become project obligations

PHYTORION

# How Does it Work?

Estimate – Agile approach

➢ Assign story points to each story (not just Hi, Med, Low)

➢ Use numbering scheme:

  ➢ 1,2,4,8,16, (Doubling previous number) or

  ➢ 1,2,3,5,8,13 (Fibonacci)

  ➢ Story 1: 3 points

  ➢ Story 2: 8 points

  ➢ Story 3: 13 points

  ➢ Story 4: 5 points

➢ I can deliver 14 points per sprint: Pick stories 1, 2, 4

PHYTORION

# How Does it Work?

## Sprint breakdown

➢ Goal is to deliver production code

➢ Hold 15 minute stand up meetings each day – progress report

➢ Create enough documentation to start coding.  Update it at end of sprint and not with every change

➢ Test-led development. Think of outliers

➢ To assure quality for design and code, do walk through with another architect/coder

➢ Define "Done"

PHYTORION

# Design Document

| | | | PROG_EFFDT, EFFSEQ = PROG_EFFSEQ, ACAD_PLAN = 'DEAN SCHLR') |
|---|---|---|---|
| ACADEMIC_LOAD | PS_STDNT_CAR_TERM | ACADEMIC_LOAD | EMPLID, ACAD_CAREER, INSTITUTION = 'UOD01', STRM = TERM |
| DERIVED_ACADEMIC_LOAD | | | {NOTE: FOR CODING, IGNORE COMMENTS IN BRACKETS – THESE ARE FOR CLARIFICATION ONLY}<br>{ADMIN CLASS }IF EXISTS (SELECT EMPLID FROM PS_STDNT_ENRL<br>WHERE PS_STDNT_ENRL.STRM = ACADEMIC_TERM<br>AND PS_STDNT_ENRL.ACAD_CAREER = ACAD_CAREER<br>WHERE STDNT_ENRL.STRM = PS_CLASS_TBL.STRM<br>AND PS_STDNT_ENRL.CLASS_NBR = PS_CLASS_TBL.CLASS_NBR<br>AND PS_CLASS_TBL.SUBJECT = 'ADMN')<br>THEN 'F"<br>ELSE<br>{UNIV_CLASS} IF EXISTS (SELECT EMPLID FROM PS_STDNT_ENRL WHERE<br>PS_STDNT_ENRL.STRM = ACADEMIC_TERM AND PS_STDNT_ENRL.ACAD_CAREER<br>= ACAD_CAREER WHERE STDNT_ENRL.STRM = PS_CLASS_TBL.STRM AND<br>PS_STDNT_ENRL.CLASS_NBR = PS_CLASS_TBL.CLASS_NBR AND<br>PS_CLASS_TBL.SUBJECT = 'UNIV' AND (PS_CLASS_TBL.CATALOG_NBR LIKE ' 55%'<br>OR PS_CLASS_TBL.CATALOG_NBR LIKE ' 35%')) THEN 'F'<br>ELSE<br>{GRAD SUSTAINER} IF GRAD_SUSTAINER IN ('DS','MT','MN') THEN 'F"<br>ELSE<br>{FALL/SPRING ALL OTHERS} IF SUBSTR(STRM,4,1) IN ('3', '8') (i.e. TERM = Spring or<br>Fall ) THEN (all *indented* calcs fall within this 'IF' )<br><br>IF ACAD_CAREER='GRAD' and CUR_REGISTERED_HOURS < 9 AND<br>CUR_REGISTERED_HOURS > 0 THEN 'P'<br>ELSE<br>IF ACAD_CAREER='GRAD' and CUR_REGISTERED_HOURS >= 9 THEN 'F' ELSE<br>IF ACAD_CAREER='UGRD' and CUR_REGISTERED_HOURS < 12 AND<br>CUR_REGISTERED_HOURS > 0 THEN 'P' ELSE<br>IF ACAD_CAREER='UGRD' and CUR_REGISTERED_HOURS >= 12 THEN 'F'<br><br>{WINTER/SUMMER ALL OTHERS} ELSE IF SUBSTR(STRM,4,1) IN ('1','5') AND<br>CUR_REGISTERED_HOURS > 0 THEN 'F' {ANYTHING ELSE}<br>ELSE IF CUR_REGISTERED_HOURS = 0 then 'N' ELSE 'X' |
| DUAL_CAREER_COUNTER | | | '1' IF ACAD_CAREER = PRI_CAREER (SEE TECHNICAL NOTES); 0 OTHERWISE |

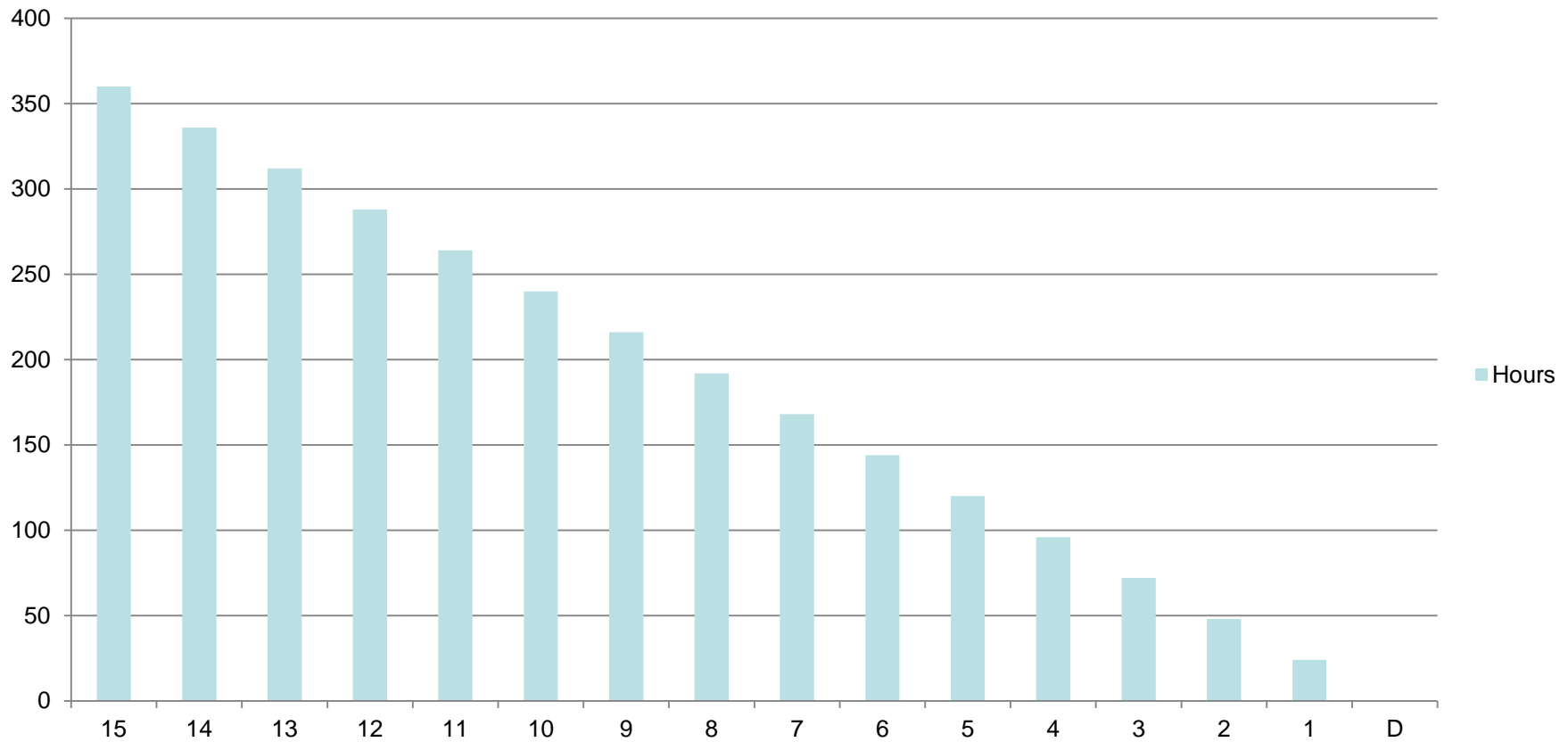| | | | |
|---|---|---|---|
| COM_GPA | PS_STDNT_CAR_TERM | COM_GPA | EMPLID, ACAD_CAREER = ACAD_CAREER, INSTITUTION = 'UOD01', STRM = TERM |
| HIGH_SAT_READING | PS_STDNT_TEST_COMP | SCORE | EMPLID, TEST_COMPONENT=VERB and TEST_ID=HIGHSAT |
| HIGH_SAT_MATH | PS_STDNT_TEST_COMP | SCORE | EMPLID, TEST_COMPONENT=MATH and TEST_ID=HIGHSAT |
| HIGH_SAT_WRITING | PS_STDNT_TEST_COMP | SCORE | EMPLID, TEST_COMPONENT=WRTG and TEST_ID=HIGHSAT |
| HIGH_SAT_TOTAL | PS_STDNT_TEST_COMP | SCORE | EMPLID, TEST_COMPONENT=TOTAL and TEST_ID=HIGHSAT |
| ATHL_CUR_PARTICIPANT1 | PS_ATHL_PART_STAT | CUR_PARTICIPANT | EMPLID, first sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATHL_NCAA_ELIGIBLE1 | PS_ATHL_PART_STAT | NCAA_ELIGIBLE | EMPLID, first sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATLH_SPORT_CODE1 | PS_ATHL_PART_STAT | SPORT | EMPLID, first sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATHL_PARTIC_CODE1 | PS_ATHL_PART_STAT | ATHL_PARTIC_CD | EMPLID, first sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATHL_CUR_PARTICIPANT2 | PS_ATHL_PART_STAT | CUR_PARTICIPANT | EMPLID, second sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATHL_NCAA_ELIGIBLE2 | PS_ATHL_PART_STAT | NCAA_ELIGIBLE | EMPLID, second sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATLH_SPORT_CODE2 | PS_ATHL_PART_STAT | SPORT | EMPLID, second sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| ATHL_PARTIC_CODE2 | PS_ATHL_PART_STAT | ATHL_PARTIC_CD | EMPLID, second sport (sorted alphabetically), max(PS_ATHL_PART_STAT.EFFDT) |
| GRAD_SUSTAINER | | | IF ACAD_CAREER <> 'GRAD' THEN 'N'<br>ELSE<br>IF EXISTS (SELECT FROM PS_STDNT_ENRL for this EMPLID, STRM, ACAD_CAREER WHERE INSTITUTION = 'UOD01' AND PS_STDNT_ENRL_STATUS = 'E' AND PS_STDNT_ENRL.CLASS_NBR = PS_CLASS_TBL.CLASS_NBR AND PS_CLASS_TBL.SUBJECT = 'UNIV' AND PS_CLASS_TBL.CATALOG_NBR = ' 999' AND PS_CLASS_TBL.CLASS_SECTION = '011'<br>(IGNORE PS_CLASS_TBL KEYS: CRSE_ID,CRSE_OFFER_NBR, STRM,SESSION_CODE,CLASS_SECTION)<br>) THEN 'DS'<br>ELSE<br>IF EXISTS (SELECT FROM PS_STDNT_ENRL for this EMPLID, STRM, ACAD_CAREER WHERE INSTITUTION = 'UOD01' AND PS_STDNT_ENRL_STATUS = 'E' AND PS_STDNT_ENRL.CLASS_NBR = PS_CLASS_TBL.CLASS_NBR AND PS_CLASS_TBL.SUBJECT = 'UNIV' AND PS_CLASS_TBL.CATALOG_NBR = ' 899' AND PS_CLASS_TBL.CLASS_SECTION = '011'<br>(IGNORE PS_CLASS_TBL KEYS: CRSE_ID,CRSE_OFFER_NBR, STRM,SESSION_CODE)<br>) THEN 'MT'<br>ELSE<br>IF EXISTS (SELECT FROM PS_STDNT_ENRL for this EMPLID, STRM, ACAD_CAREER WHERE INSTITUTION = 'UOD01' AND PS_STDNT_ENRL_STATUS = 'E' AND PS_STDNT_ENRL.CLASS_NBR = PS_CLASS_TBL.CLASS_NBR AND PS_CLASS_TBL.SUBJECT = 'UNIV' AND PS_CLASS_TBL.CATALOG_NBR = ' 895' AND PS_CLASS_TBL.CLASS_SECTION = '011'<br>(IGNORE PS_CLASS_TBL KEYS: CRSE_ID,CRSE_OFFER_NBR, STRM,SESSION_CODE)<br>) THEN 'MN'<br>ELSE 'N' |
| COUNT_FIELD | | | 1 |

# How Does it Work?

## Sprint breakdown

➢ Demo to Product Owner

➢ What is not complete, becomes Tech Debt to start next sprint

➢ Retrospective

- ➢ What worked

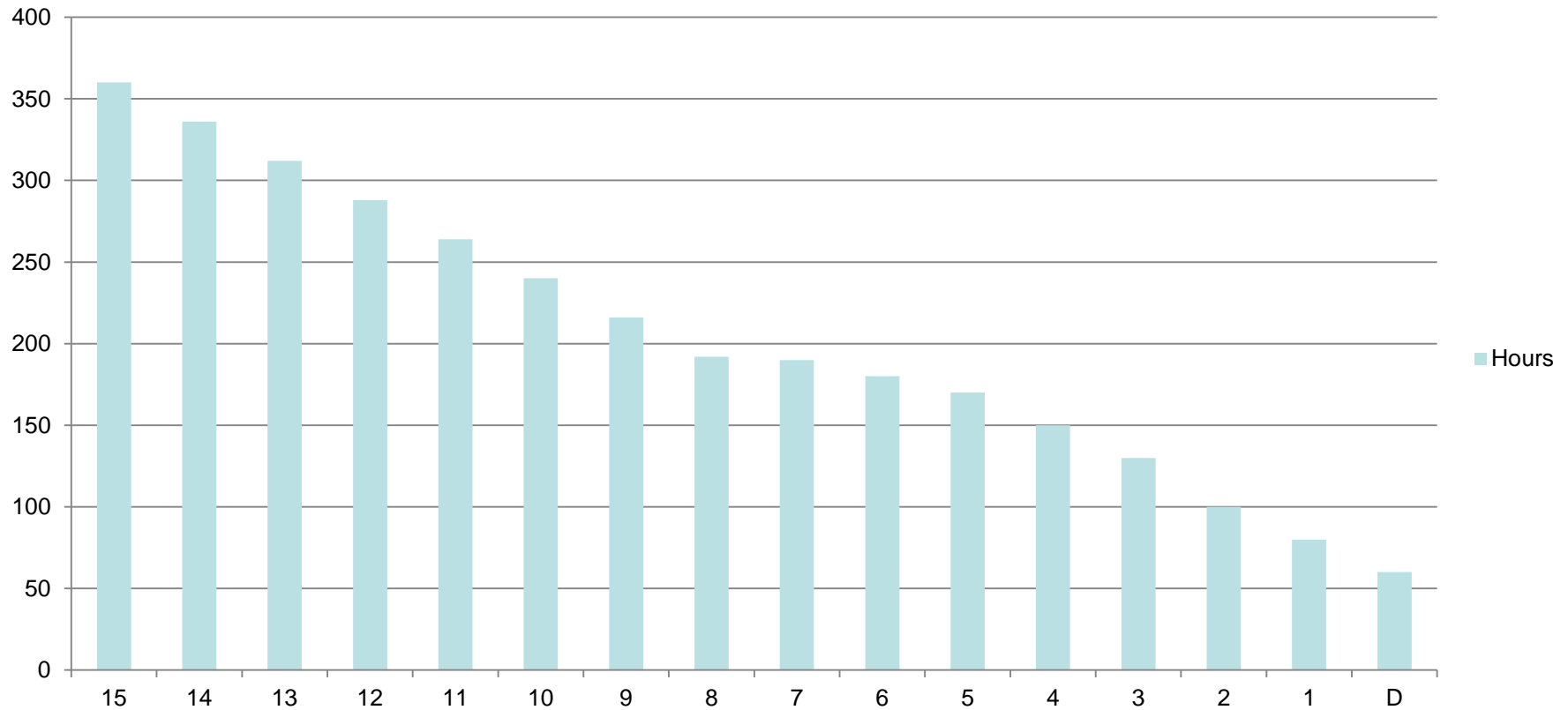- ➢ What didn't

- ➢ How good were our estimates?

PHYTORION

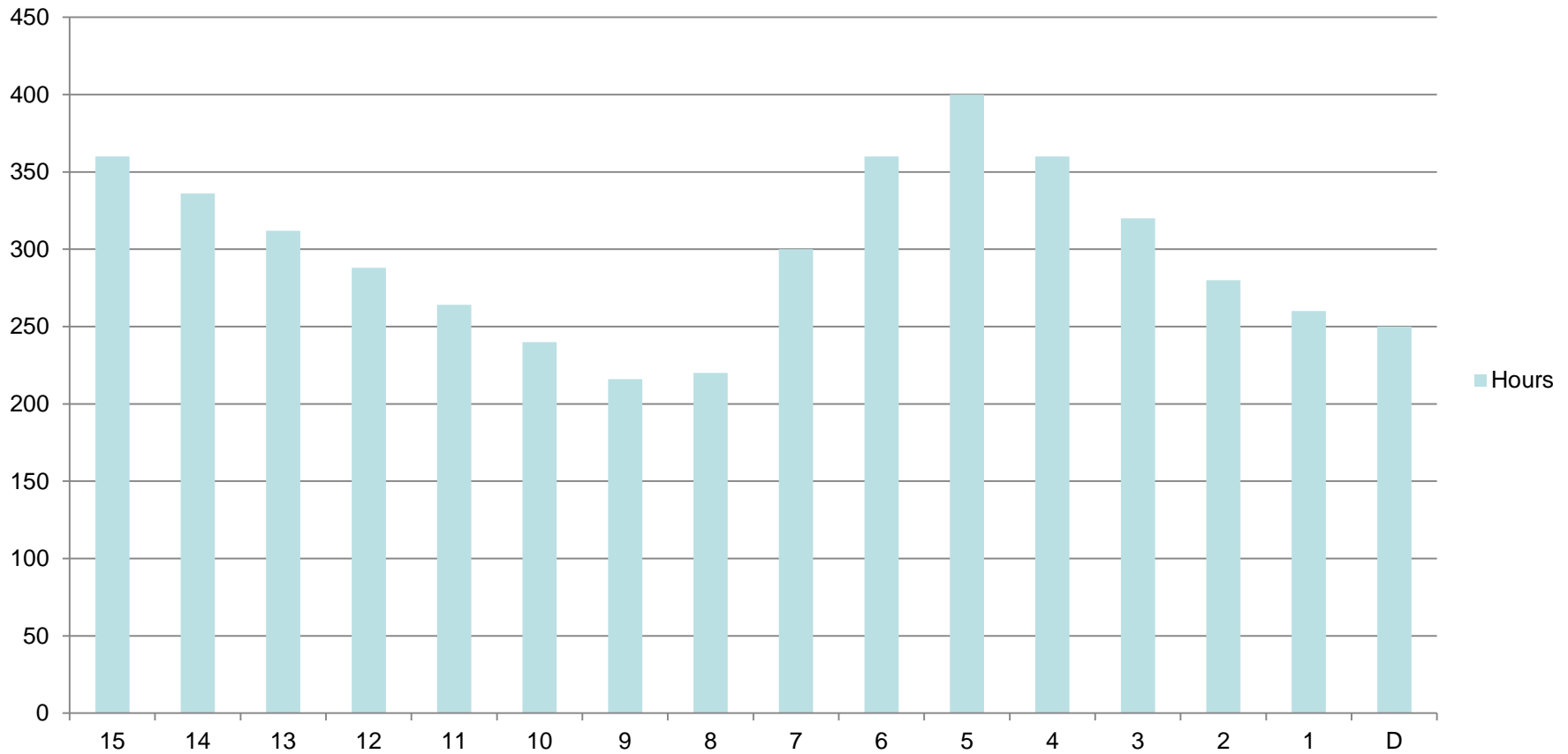# How Does it Work?

**Sprint**

# How Does it Work?



Sprint — Burndown chart showing Hours remaining over sprint days (15 down to D).

# How Does it Work?

# Additional Sprints

➢ Sprint 0: Project start, infrastructure

➢ Architecture: Make sure you see the big picture

➢ Spike Sprint: Team focuses on difficult problem

➢ Reclamation: Fix shortcuts

➢ Go Live

PHYTORION

# Potential Difficulties

➢ Schools focus on Full Data Warehouse, not pieces

➢ Funding is justified for full project only

➢ People prefer big, detailed plans in order to provide some certainty

➢ Schools have many Product Owners with different or competing definitions – ex: Deferrals

➢ It's not always easy to get people's full attention

PHYTORION

# Advantages

➢ No Shelf life for requirements

➢ People solve problems together: Dimension not needed

➢ Estimation improvement: What if you were building 20 metrics and you discovered they were much more complex

➢ Users understand what they are getting and *find* ways to be involved even in testing!

➢ Early successes that evangelize solution

PHYTORION

# References

You can learn more…

➢ Ralph Hughes: Agile Data Warehousing

➢ Ken Collier: Agile Analytics

➢ TDWI Conference classes

PHYTORION

# Q&A

Questions & Answers

PHYTORION

# Next Session & To Contact Us

Wednesday @10:30 AM,  Hermitage D

Session #30768:

Campus Solutions, HCM, Finance & K-12

Data Warehouse/BI Demo


Yiorgos Marathias

President, Phytorion
773-255-3861
yiorgos.marathias@phytorion.com

PHYTORION