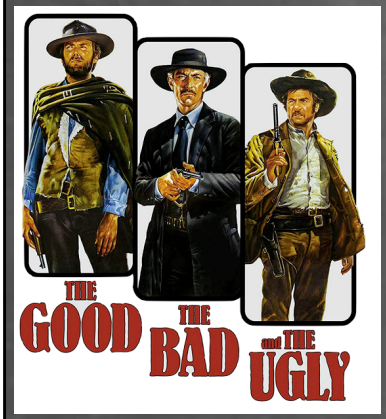
Putting Machine Learning into Production
The Good, the Bad & the Ugly



Brendan Tierney

@brendantierney

www.oralytics.com

ORALYTICS

1

Brendan Tierney

Oracle Groundbreaker Ambassador

ORACLE ACE Director

t. @brendantierney

w. www.oralytics.com

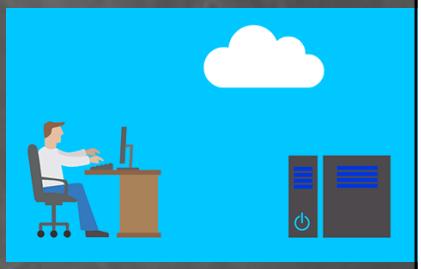
ORACLE MAGAZINE

Japanese edition coming soon

www.oralytics.com

ORALYTICS

2



www.oralytics.com

ORALYTICS

3

Putting ML & AI into Production

www.oralytics.com

ORALYTICS

4

But WHY would you do that ?

When

www.oralytics.com

ORALYTICS

5

Gartner : 85% of Project Fail, remain on shelf

VentureBeat : 87% never make it into Production

Cisco : 74% failure rate

NewVantage : 77% of projects are gathering Dust

Harvard Business Review : 80% of Project Fail

Pactera : 87% of Projects Fail

Failure !

www.oralytics.com

ORALYTICS

6

No clear Objective

No idea How it will be used

Technical Architecture & Integration

Using Traditional Life-cycles

Give me all your Data!

Leading with the Tech/Math

Only in the Lab

Only Works in the Lab

Business Objective vs Oh this is Cool

Not using the Right Tool

Seeking Perfection vs is it Good Enough

Unknow Measures

Not realizing, This (ML/AI) is Not a New Thing

But Why Brendan?

Why do they Fail?

Different Teams not Talking

Re-inventing the Wheel
Ignoring lessons from the past

It's just for the BI or Stats teams

"Think of the Possibilities"
Dream world vs Reality

DS/ML/AI Team Not
working with Other Teams

Only using Cool Tech

Using particular Algorithm vs "No Free Lunch"

No/lack of Knowledge Sharing

Needing Big Data

Data Silos

Internal Politics

No Explainability

www.oralytics.com

ORALYTICS

7

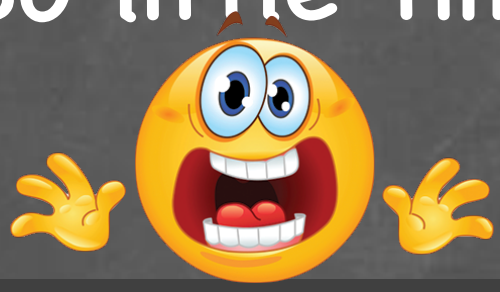
No clear Objective

No idea How it will be used

Technical Architecture & Integration

Using Traditional Life-cycles

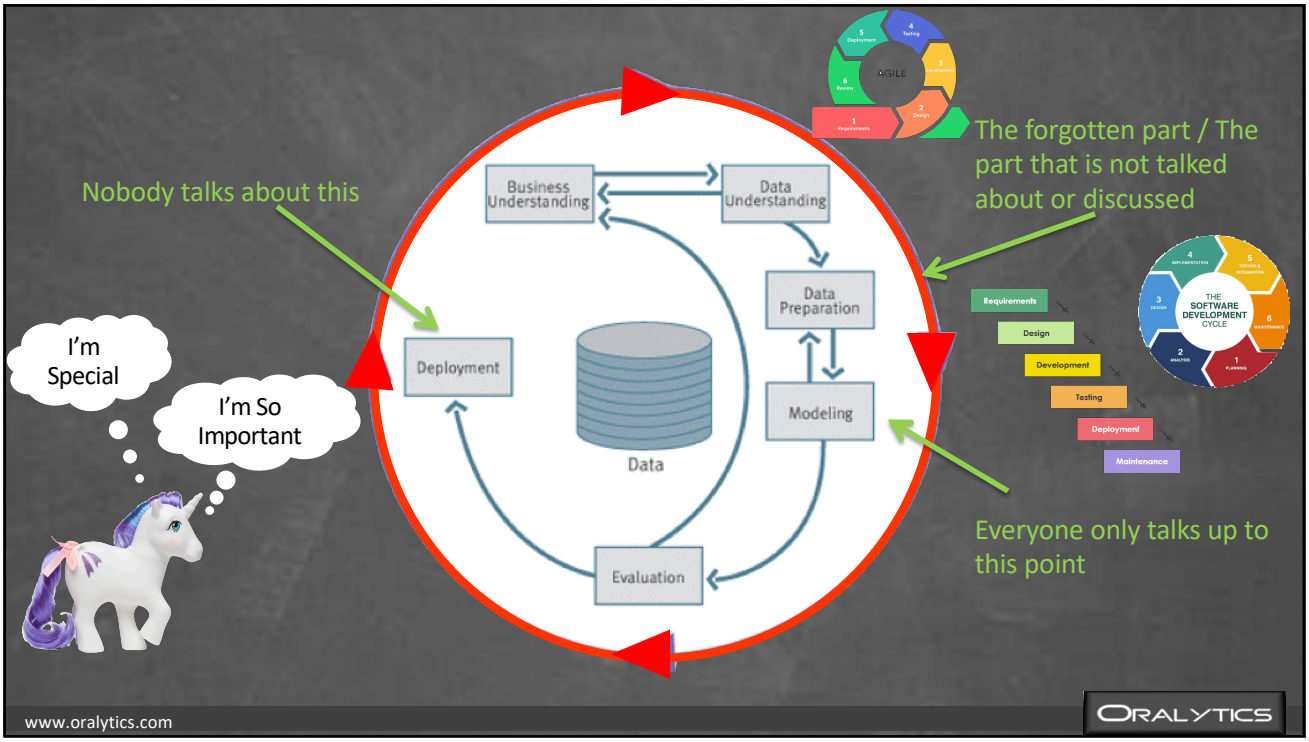
So much to discuss,
So little time



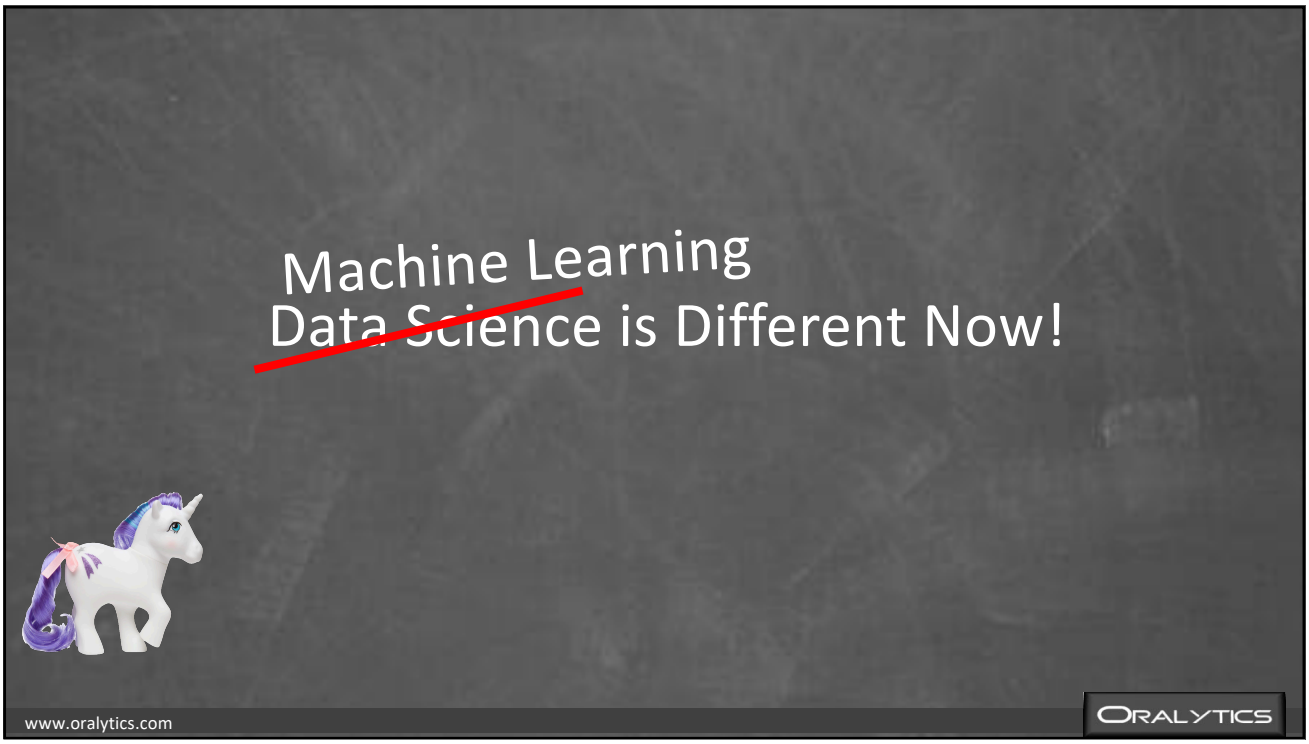
www.oralytics.com

ORALYTICS

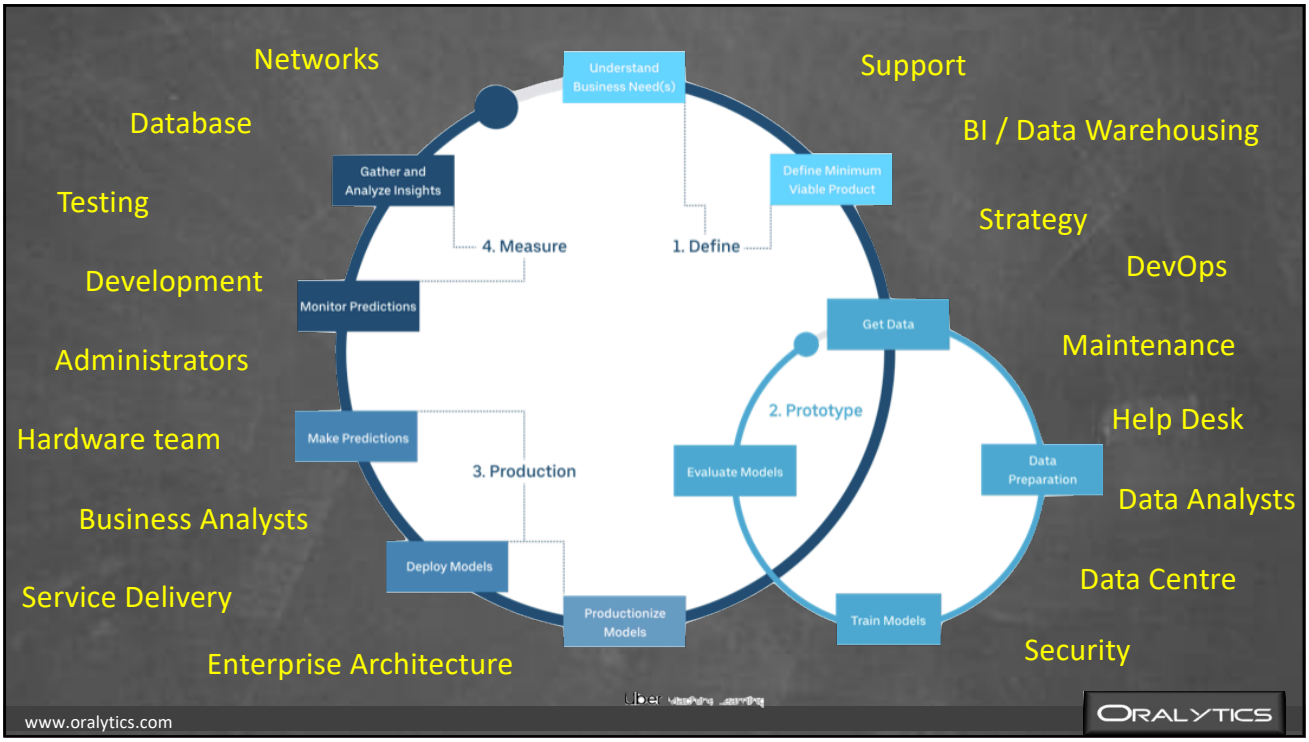
8


Nobody talks about this
The forgotten part / The part that is not talked about or discussed
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data
I'm Special
I'm So Important
Everyone only talks up to this point
AGILE
Requirements
Design
Development
Testing
Deployment
Maintenance
THE SOFTWARE DEVELOPMENT CYCLE
www.oralytics.com
ORALYTICS

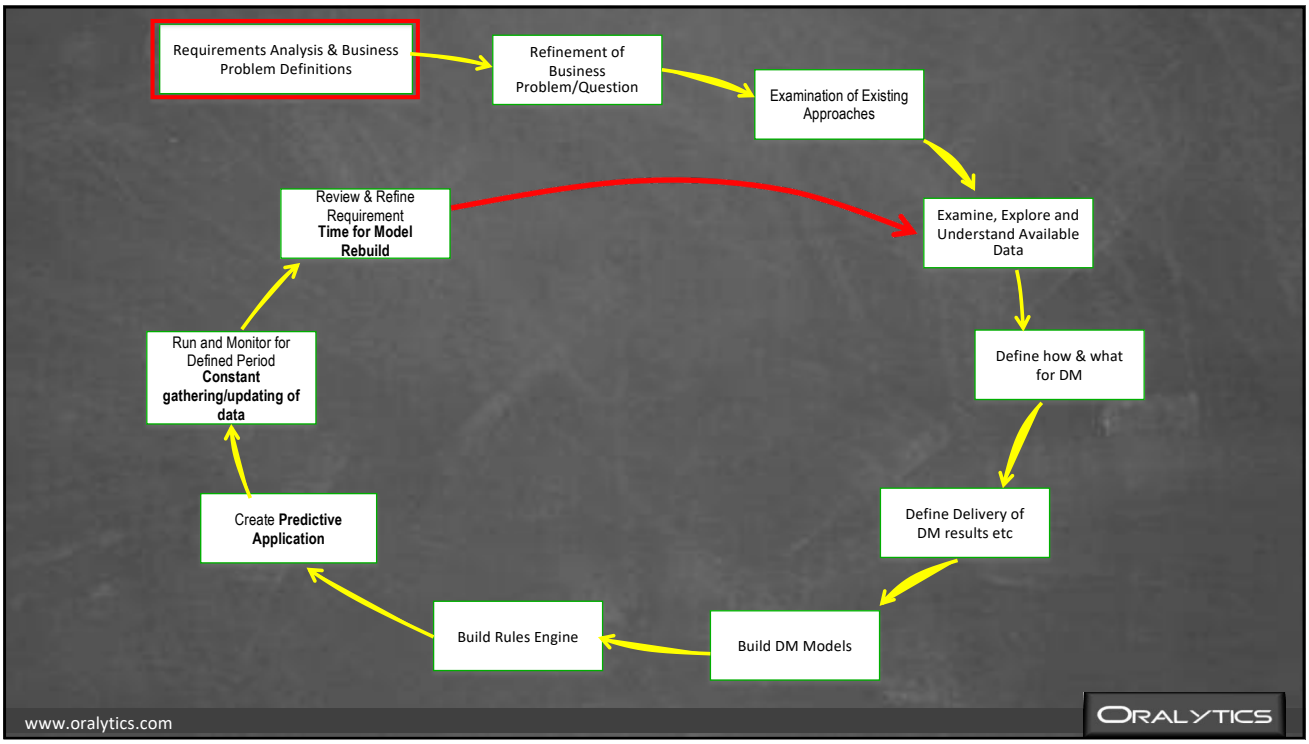
9


Machine Learning
Data Science is Different Now!
www.oralytics.com
ORALYTICS

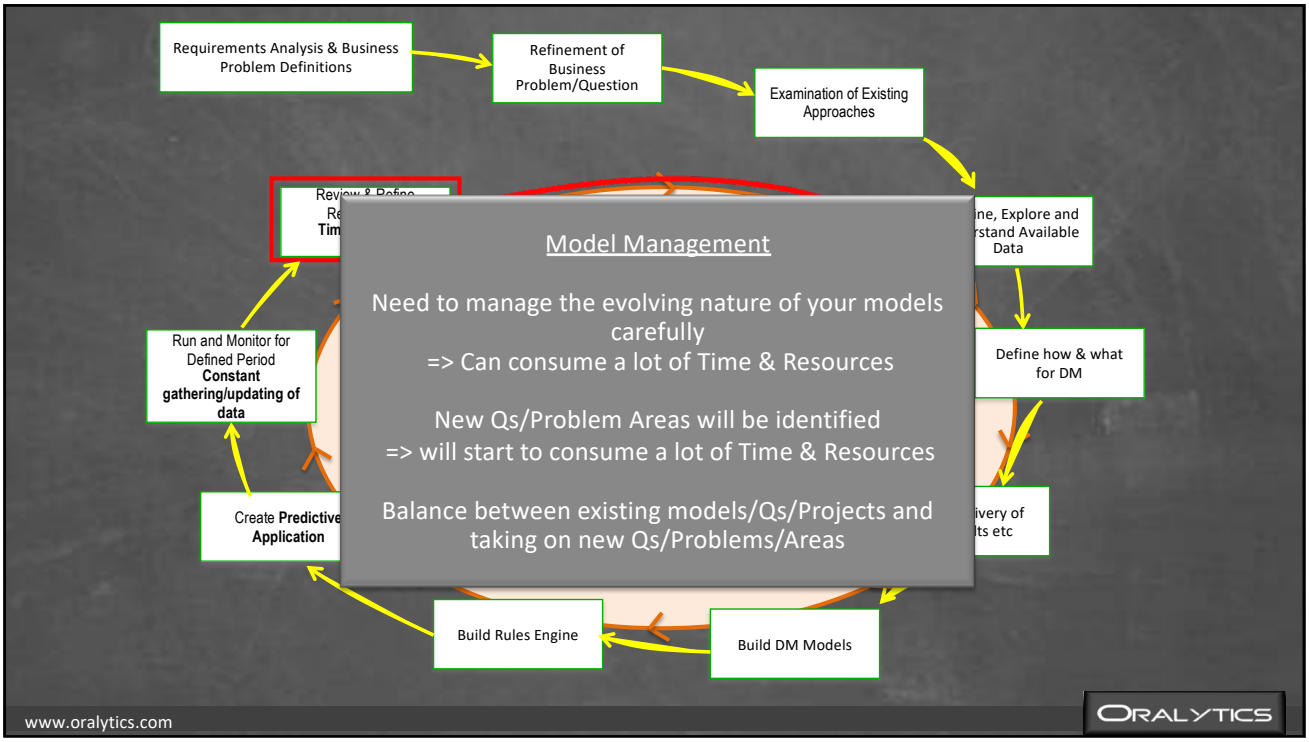
10


Networks
Database
Testing
Development
Administrators
Hardware team
Business Analysts
Service Delivery
Enterprise Architecture
Support
BI / Data Warehousing
Strategy
DevOps
Maintenance
Help Desk
Data Analysts
Data Centre
Security
Understand Business Need(s)
Define Minimum Viable Product
Get Data
Data Preparation
Train Models
Evaluate Models
Productionize Models
Deploy Models
Make Predictions
Monitor Predictions
Gather and Analyze Insights
1. Define
2. Prototype
3. Production
4. Measure
www.oralytics.com
ORALYTICS

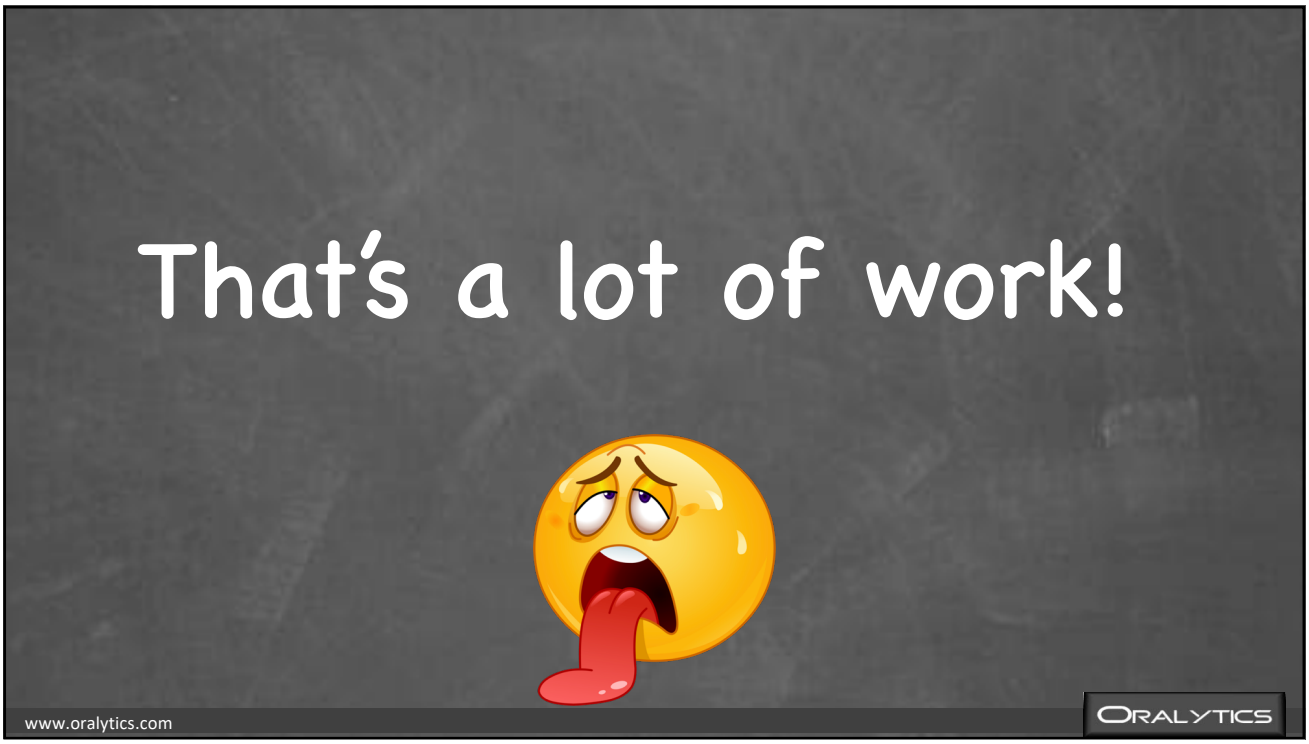
11


Requirements Analysis & Business Problem Definitions
Refinement of Business Problem/Question
Examination of Existing Approaches
Examine, Explore and Understand Available Data
Define how & what for DM
Define Delivery of DM results etc
Build DM Models
Build Rules Engine
Create Predictive Application
Run and Monitor for Defined Period Constant gathering/updating of data
Review & Refine Requirement Time for Model Rebuild
www.oralytics.com
ORALYTICS

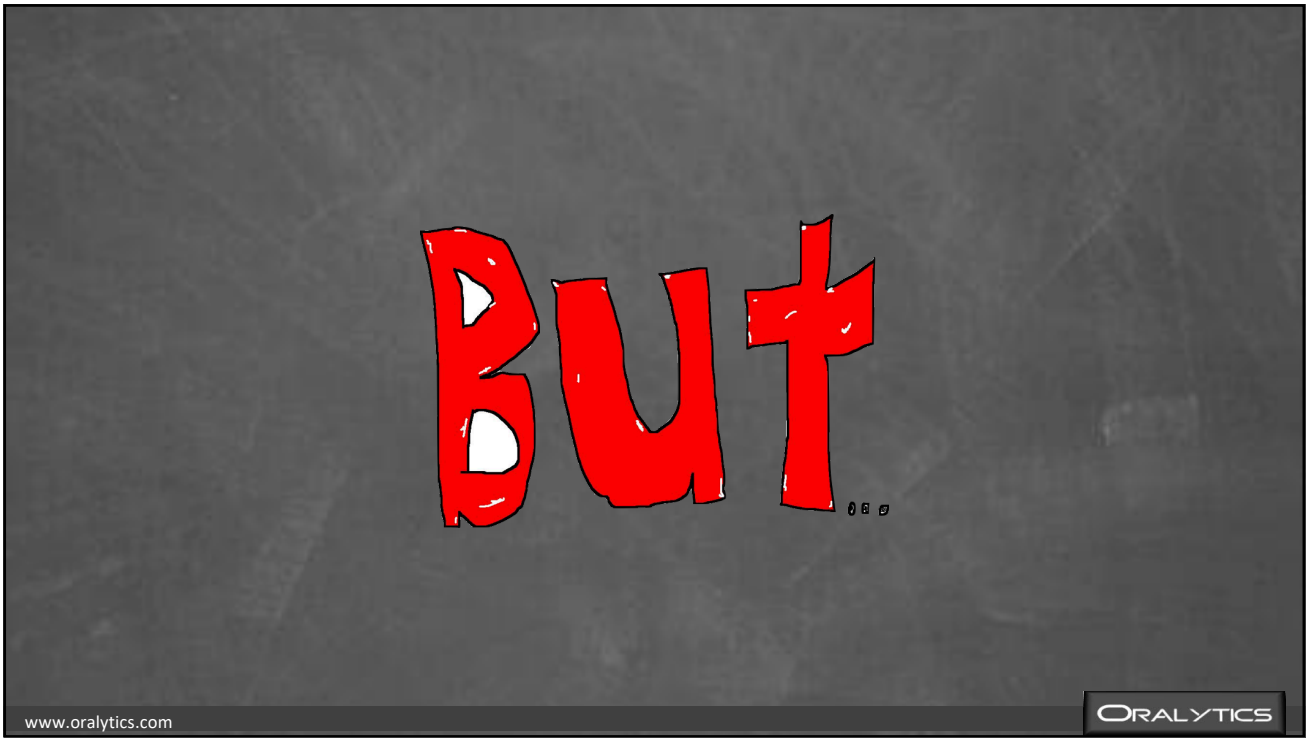
12


Requirements Analysis & Business Problem Definitions
Refinement of Business Problem/Question
Examination of Existing Approaches
ne, Explore and rstand Available Data
Define how & what for DM
ivery of ts etc
Build DM Models
Build Rules Engine
Create Predictive Application
Run and Monitor for Defined Period Constant gathering/updating of data
Model Management
Need to manage the evolving nature of your models carefully
=> Can consume a lot of Time & Resources
New Qs/Problem Areas will be identified
=> will start to consume a lot of Time & Resources
Balance between existing models/Qs/Projects and taking on new Qs/Problems/Areas
www.oralytics.com
ORALYTICS

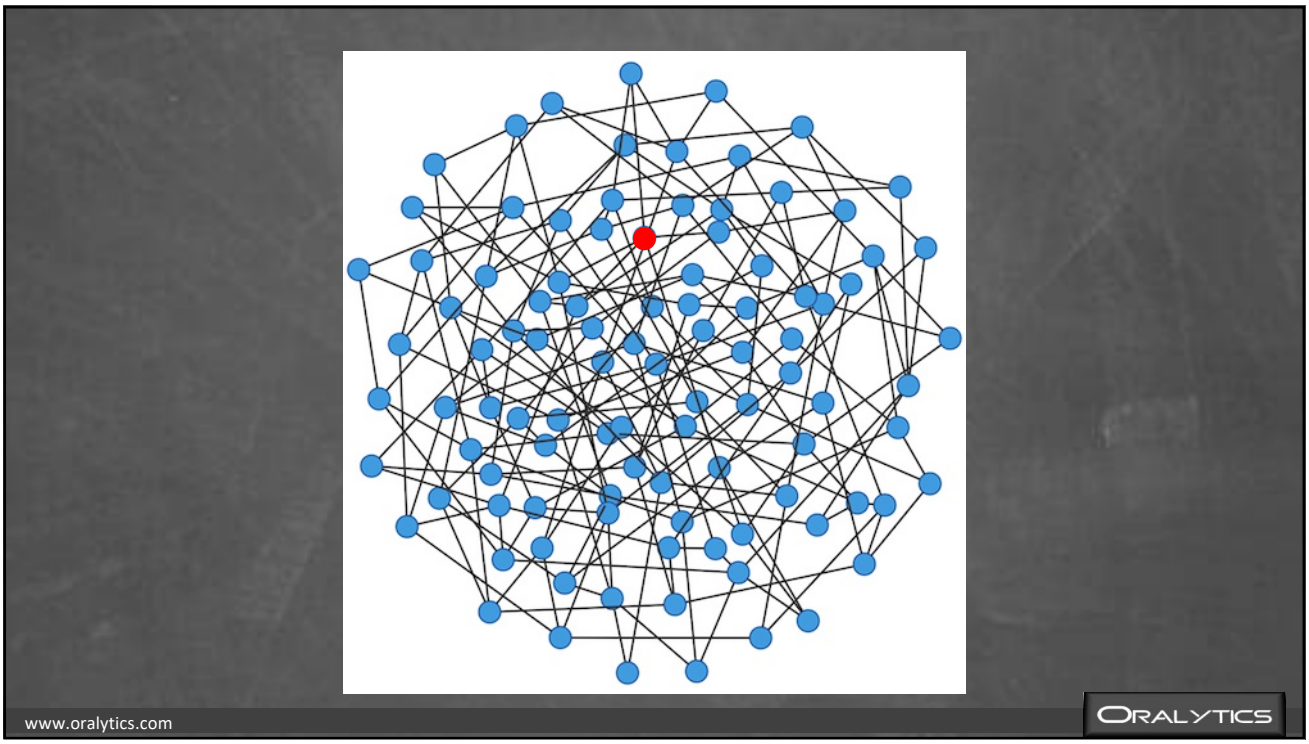
13


That's a lot of work!
www.oralytics.com
ORALYTICS

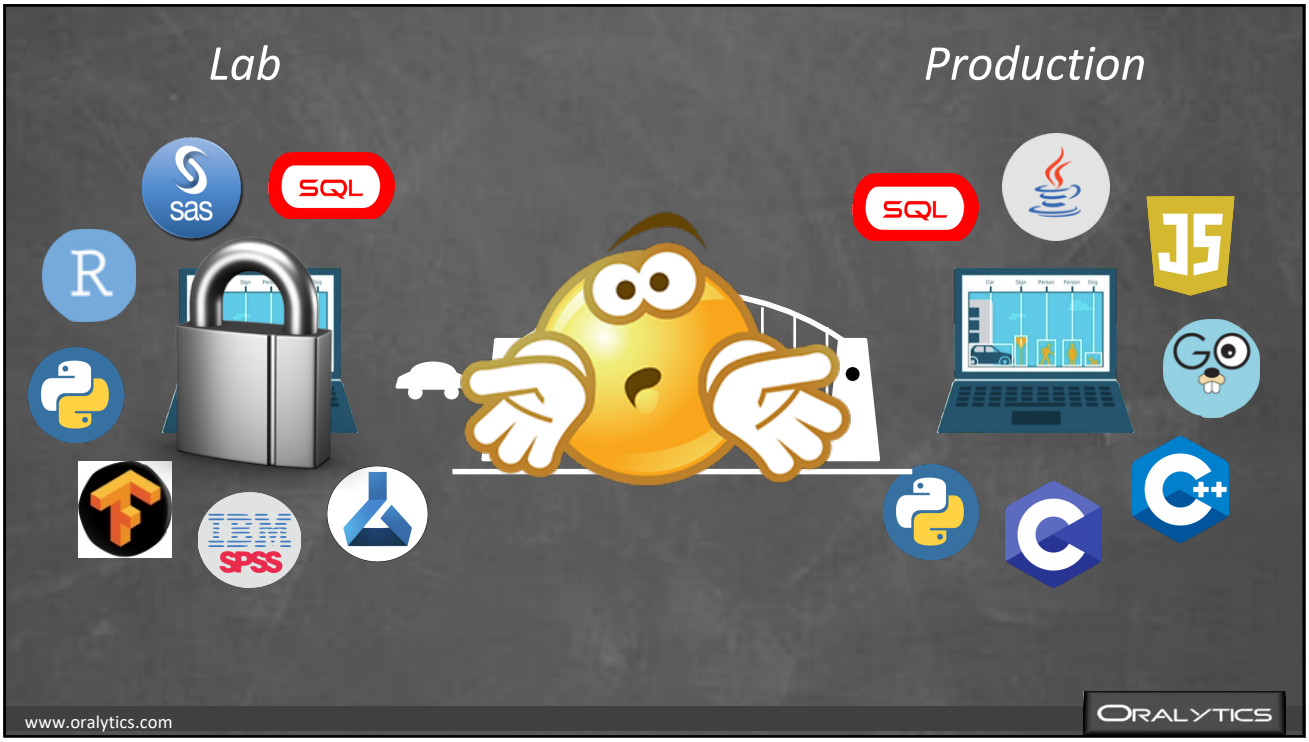
14


But...
www.oralytics.com
ORALYTICS

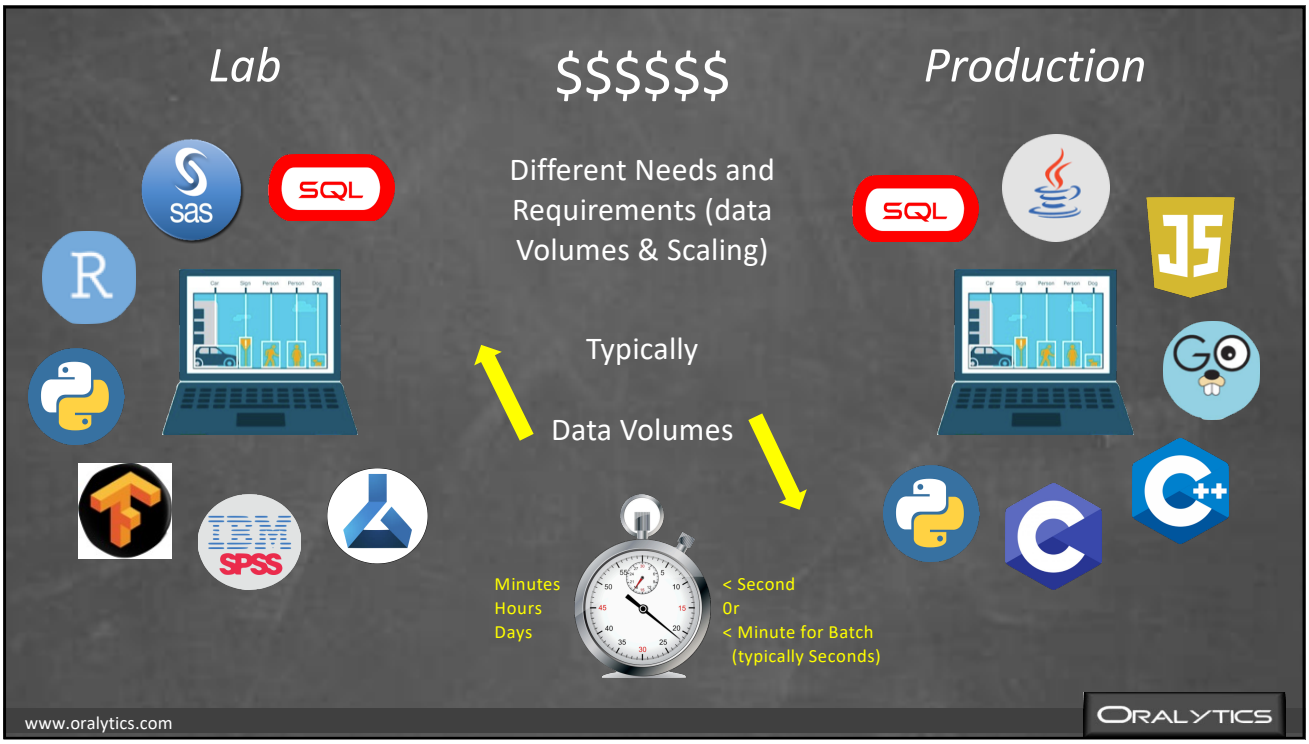
15


www.oralytics.com
ORALYTICS

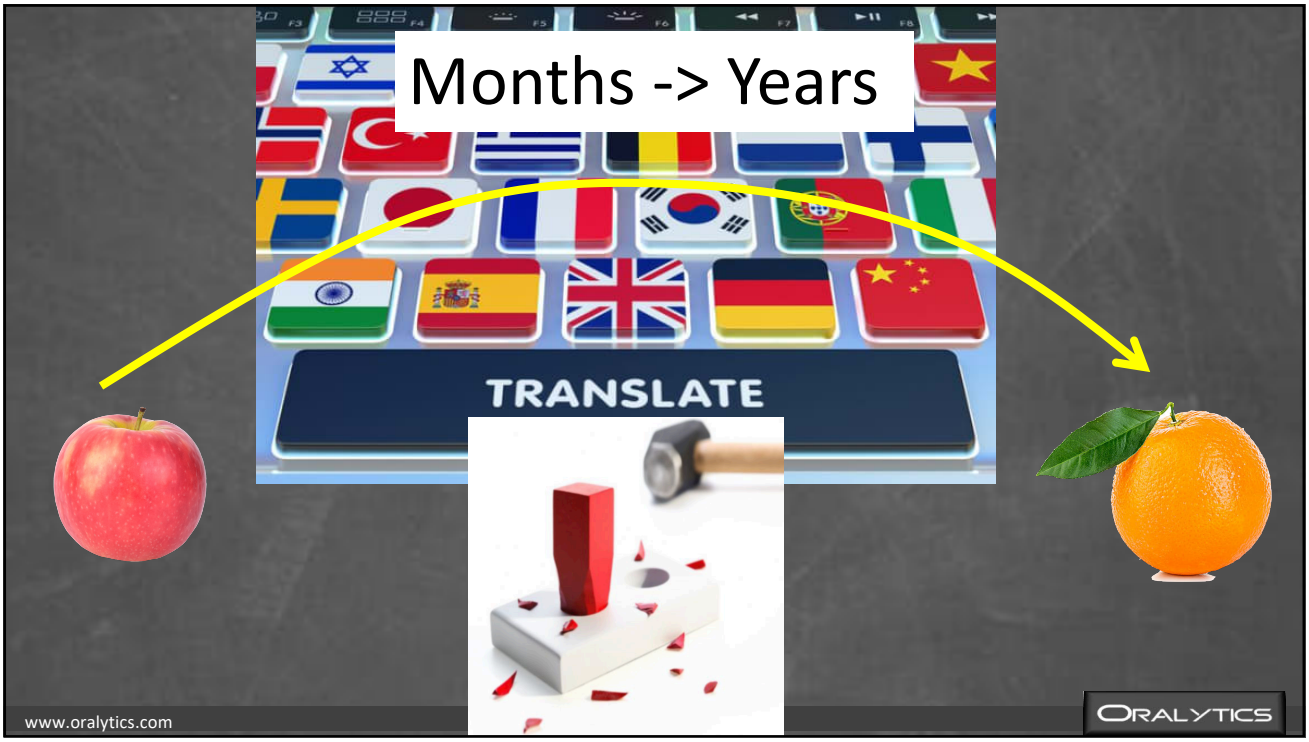
16


Lab
Production
sas
SQL
R
IBM
SPSS
SQL
JS
GO
C++
C
www.oralytics.com
ORALYTICS

17


Lab
$$$$$$
Production
sas
SQL
R
IBM
SPSS
Different Needs and Requirements (data Volumes & Scaling)
Typically
Data Volumes
Minutes
Hours
Days
< Second
Or
< Minute for Batch
(typically Seconds)
SQL
JS
GO
C++
C
www.oralytics.com
ORALYTICS

18

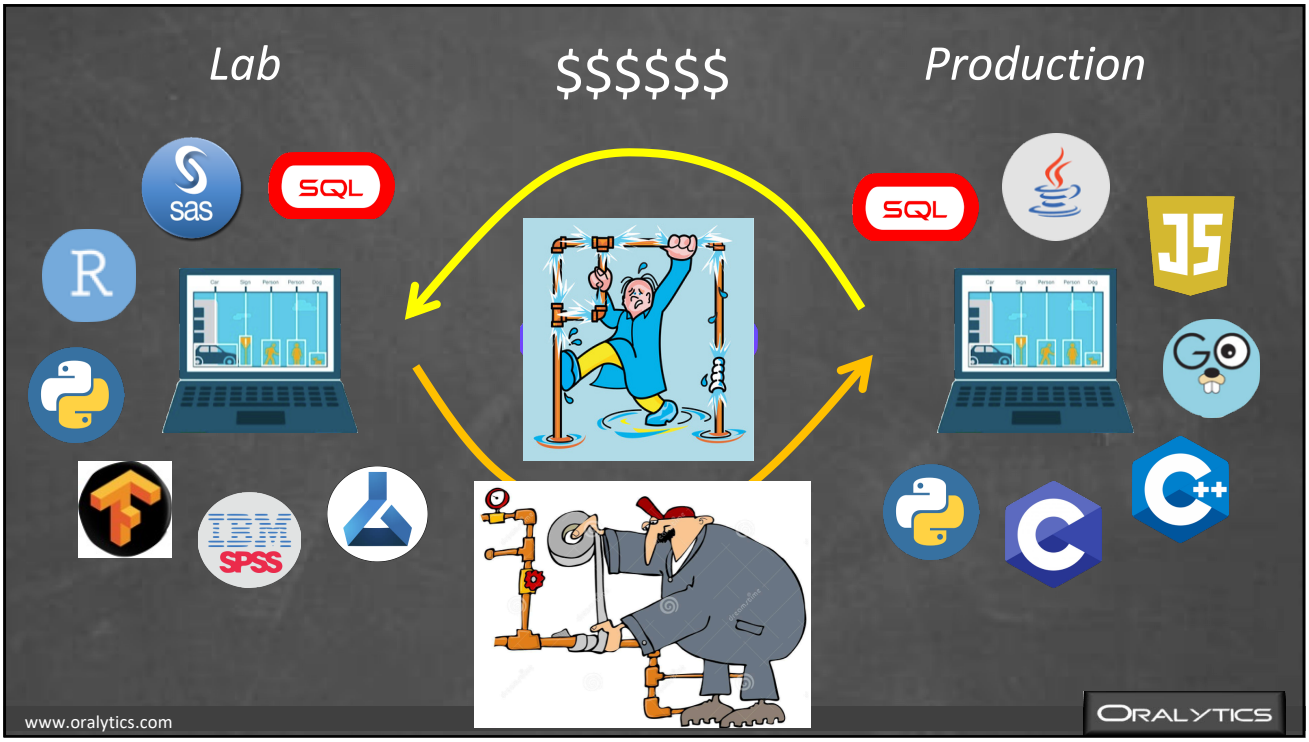


19

ML Model Interchange Languages

- PMML (Predictive Model Markup Language) is an XML-based language that enables the definition and sharing of predictive models between applications.
- The PMML specifications establish a vendor-independent means of defining these models, so that problems with proprietary applications and compatibility issues can be circumvented.
- Train once, deploy anywhere
- Supports
  - Pipeline
  - Transformations
  - Missing data treatment
  - Algorithms
- Hyperparameter settings

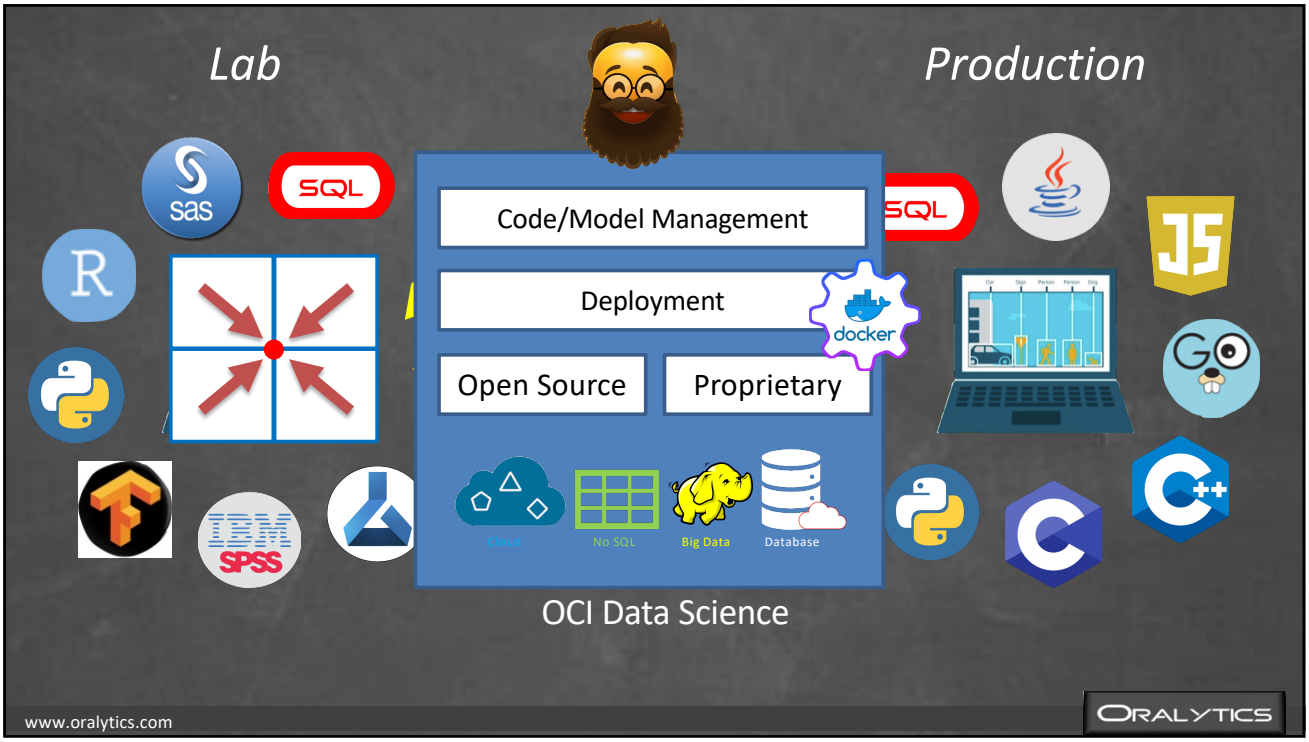
20


Lab
$$$$$$
Production
SQL
SQL
www.oralytics.com
ORALYTICS

21

Changed Everything!
www.oralytics.com
ORALYTICS

22


Lab
Production
Code/Model Management
Deployment
Open Source
Proprietary
Cloud
No SQL
Big Data
Database
OCI Data Science
www.oralytics.com
ORALYTICS

23

Changed Everything!
www.oralytics.com
ORALYTICS

24

- **Know your problem, Know why it is needed, Know how it will be solved, Know who will use it, Know how/where it will be deployed**
- Designed with Deployment/Production Deployment from the Start
- Know your data, what type, volume, etc. (not everything is BIG or Images or DL). But do you have a mixture
- Multi-Language + Multi-Data Source + Visualisation = Integrated Tool
- Easy Model Deployment & Easy Updating (with new models)
- Is it good enough! Deploy quickly, iterate, learn, improve, …
- Quick predictions
- Explainable Predictions
- Reduces or Remove the Manual, Repetitive & Boring tasks – **Automate !**

- No Explainability or limited
- Batch Processing / Only at certain times / Only when Fred is in the Office
- Mixed Complex Architecture ($$$$)
- Poor (leaking) API design ($$$$)
- No or little cross platform or data support ($$$$)
- Slow
- High Maintenance – Time & $$$$
- Focusing on wrong measures

- Not knowing How or Where it is to be deployed ($$
- Random Exploring ($$$$)
- Trying to solve a problem that isn't a problem ($$$$)
- Using Locked Software – No re-use ($$$$)
- Needing someone to Press a Button ($$$$)
- Having to Re-Write Everything ($$$$)
- Look at Me! Look at Me! Look at Me! (Not about the Problem) ($$$$)
- It's all about the Accuracy ($$$$)

www.oralytics.com ORALYTICS

25

Faidcases !!

www.oralytics.com ORALYTICS

26

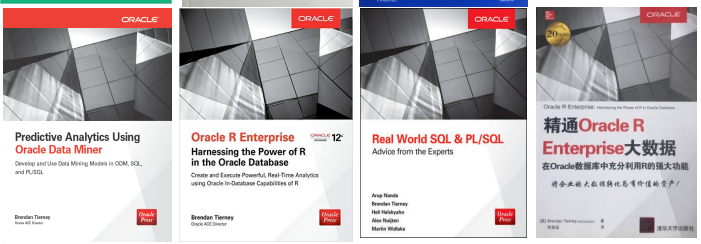
ORACLE MAGAZINE

Brendan Tierney

Oracle Groundbreaker Ambassador

ORACLE ACE Director

t. @brendantierney

w. www.oralytics.com

Japanese edition coming soon