



TRANSPORTATION
DATA SYMPOSIUM

Integrating Data Sources and Analytics for MAP-21 Initiatives

Shourya Shukla

Florida Transportation Data Symposium 2015



Shourya Shukla, Vice President, Business Solutions Architecture, Rolta



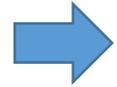
- Over 20 years of Transportation and Solutions Experience
- Lead Rolta Transportation BI Practice
 - Solution Architect and Software Developer
 - Transportation Planner

Rolta Corporate Profile

- 30+ years of experience
- Global company (offices in 9 countries), 4,000 employees & growing
- Annual Revenue of \$580M
- 35% annual growth rate
- Strategic Offering:
 - Industry Solutions in Business Intelligence
 - Geospatial Solutions
 - Systems Integration and Data Services
 - ERP, EAM, EPM, IS and MS
- Focused Industries: Government, Transportation, Utilities, O&G



Outline



1. MAP 21—Moving from activities to outcomes
2. Importance of Data for MAP 21
3. What questions are DOTs trying to answer?
4. Emerging Analytics Landscape
5. Typical DOT stakeholders and their needs
6. How it all fits together



MAP-21 - Moving Ahead for Progress in the 21st Century



[Home](#) / [MAP-21](#)

[Summary](#)

[Q & A](#)

[Fact Sheets](#)

[Cross Reference](#)

[Reports/Publications](#)

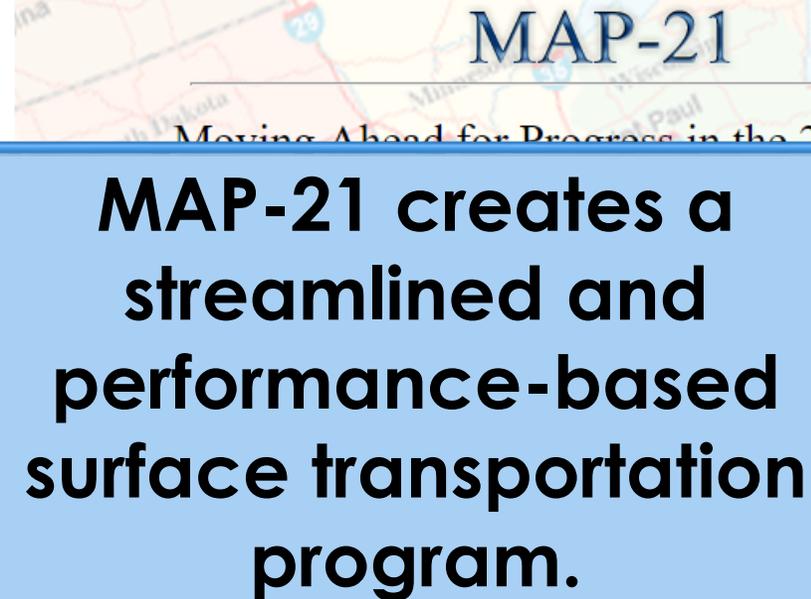
[Presentations](#)

[Legislation](#)

[Funding Tables](#)

[Webinars](#)

[Guidance](#)



MAP-21

Moving Ahead for Progress in the 21st Century

MAP-21 creates a streamlined and performance-based surface transportation program.

Extensions of MAP-21

To allow more time for development and consideration of a long-term reauthorization of surface transportation programs, Congress enacts short term extensions of the expiring law, in this case, MAP-21. Information about MAP-21 extensions may be found on the [Legislation](#) and [Funding Tables](#) pages of this site.

Recently Added

[Tunnel Inspection Final](#)

[CM/CG Contracting](#)

[Tribal Transportation NOFA](#)

[Updated TAP Qs & As](#)

[Private sector surveying and mapping services](#)

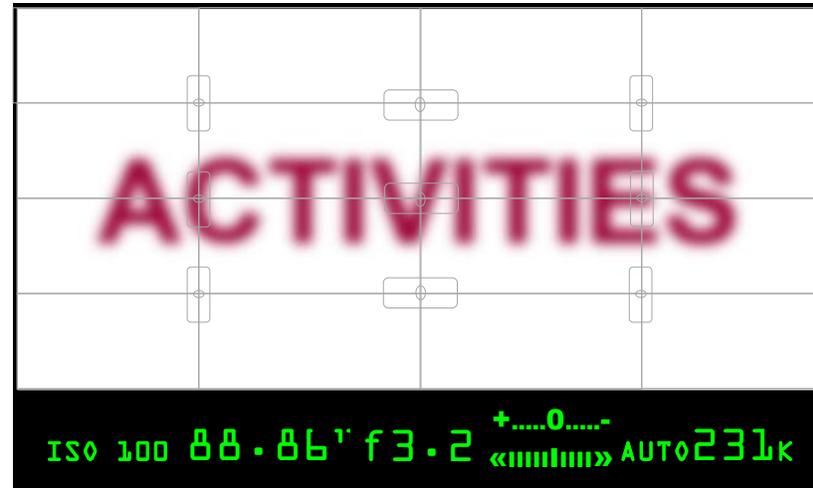
- [4/1 - Asset Management NPRM comment period extended](#)
- [3/13 - Draft Environmental Process Guidance Available for Comment](#)

MAP 21: The Changing Focus

Activities



Outcomes



MAP-21 Questions

- Where is the state with respect to MAP-21 goals?
- Does the state have a full inventory of assets and their condition?
- Which areas of the network need project funding based on MAP-21 goals?
- Did the projects executed have the desired impact?
- Can all of the above be proven with data and consistent analysis?



Moving from Activities to Outcomes

- Lead Indicators vs. Lag Indicators
- Data-driven decision making
- Growth in data volumes
- Technology to store, manage and access data
- Data governance to the fore



Outline

- ➔ 1. MAP 21—Moving from activities to outcomes
- ➔ 2. Importance of Data for MAP 21
- 3. What questions are DOTs trying to answer?
- 4. Emerging Analytics Landscape
- 5. Typical DOT stakeholders and their needs
- 6. How it all fits together



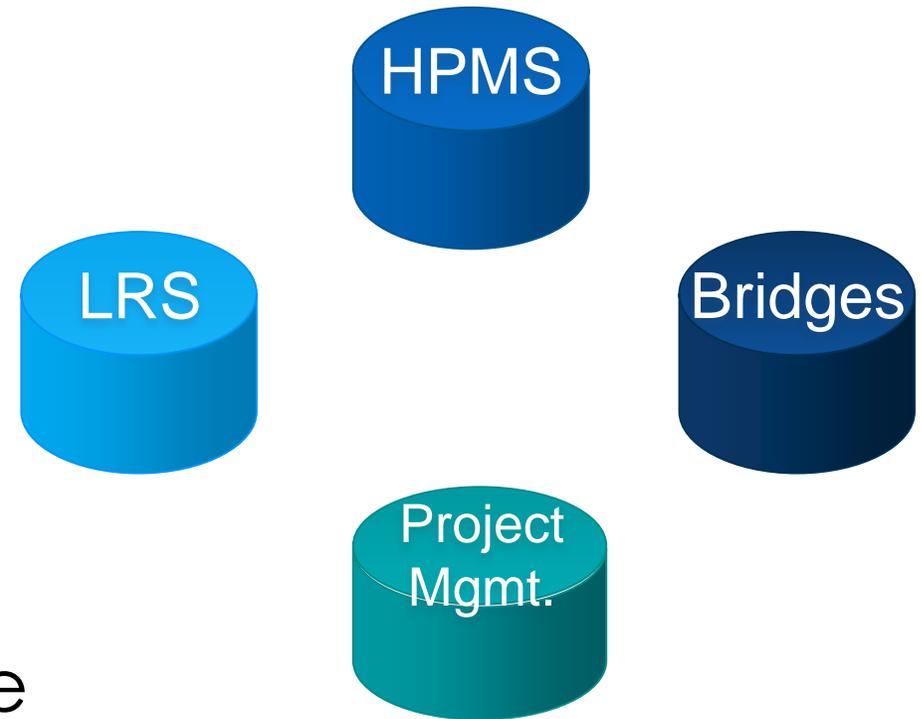
Data is Crucial Because...

- MAP-21 requires proof through data
- Data will dictate funding for DOTs
 - Is the data reliable, consistent and up to date?
 - Is the data understood and accessible to the people in the organization making decisions?



But utilizing current datasets for MAP 21 is hard

- Can data across data systems be related?
- Transport dataset differentiators:
 - Data types – Linear, Spatial
 - Multiple definitions of terms
 - Data at different grains
 - Impact of time on all of the above

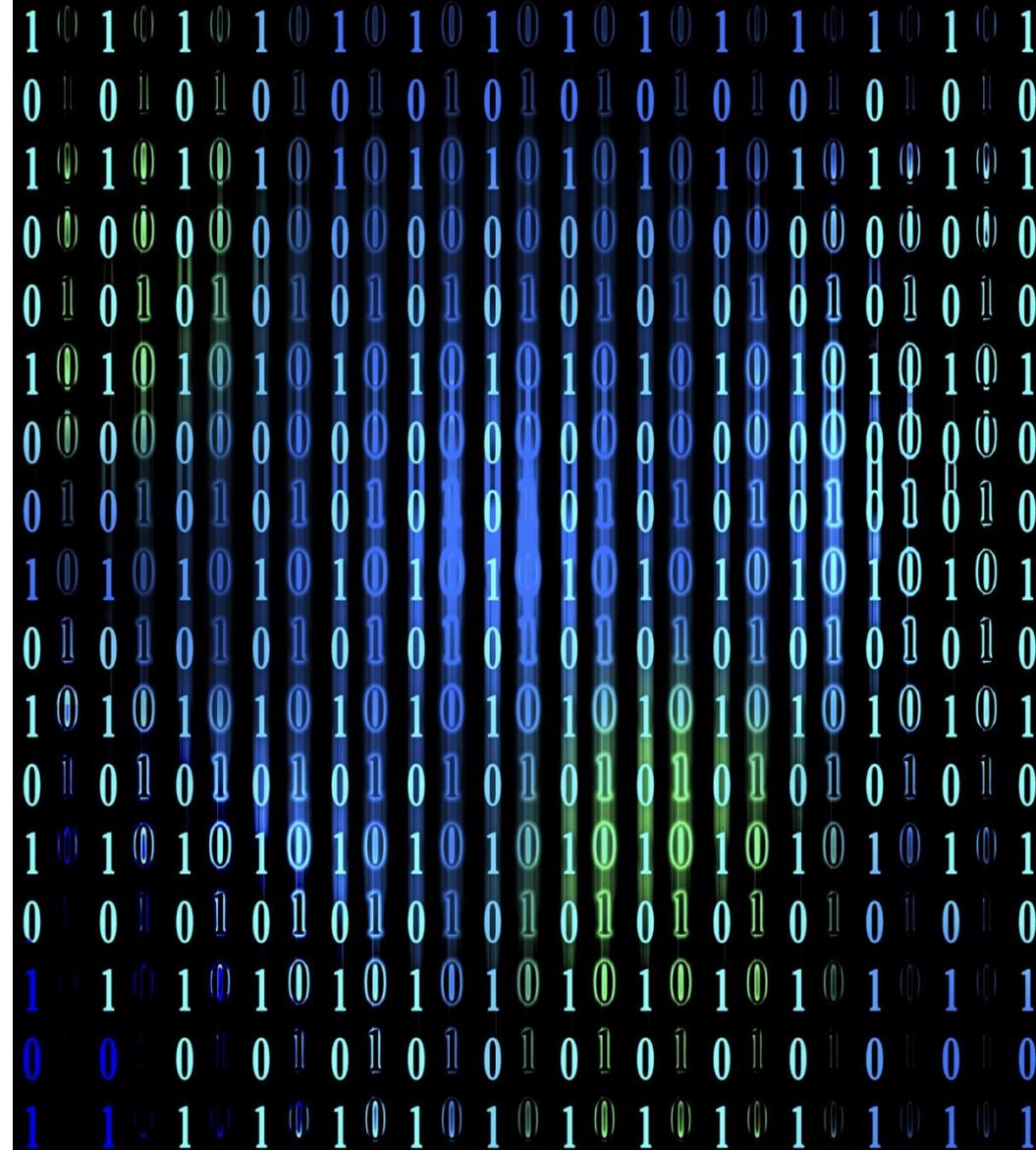


Data needs change based on question being asked...

- Types of questions
 - What happened? What does the trend show? - **Descriptive**
 - What is likely to happen? - **Predictive**
 - What is the best course of action? - **Prescriptive**
 - Can we make sense of real time data? - **Real time**
- How should the answer be presented (visualization)
 - How do you want to present the answer? - **Explanatory vs. Exploratory**

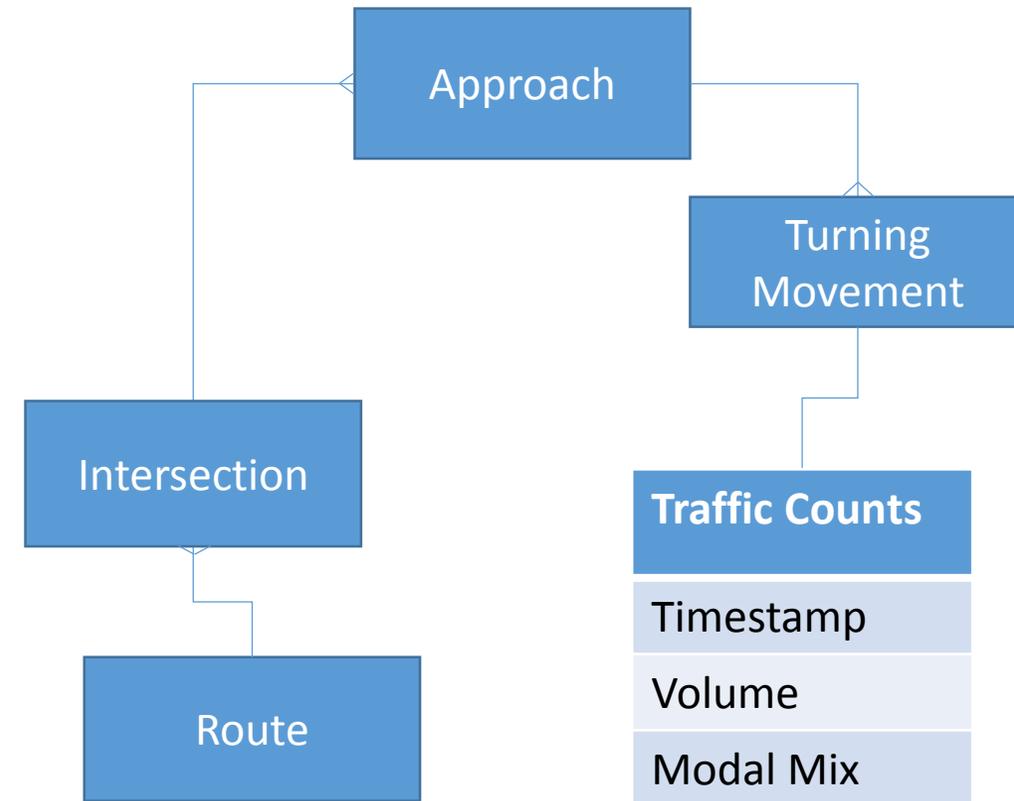
Dimensional Data Warehouse models

- Built for performance and analysis optimization
 - vs. Storage and redundancy in a normalized model
- Built to track history
- Built around dimensional modeling rather than Entity relationship modeling



Analytics Models are Different

Entity Relationship Approach



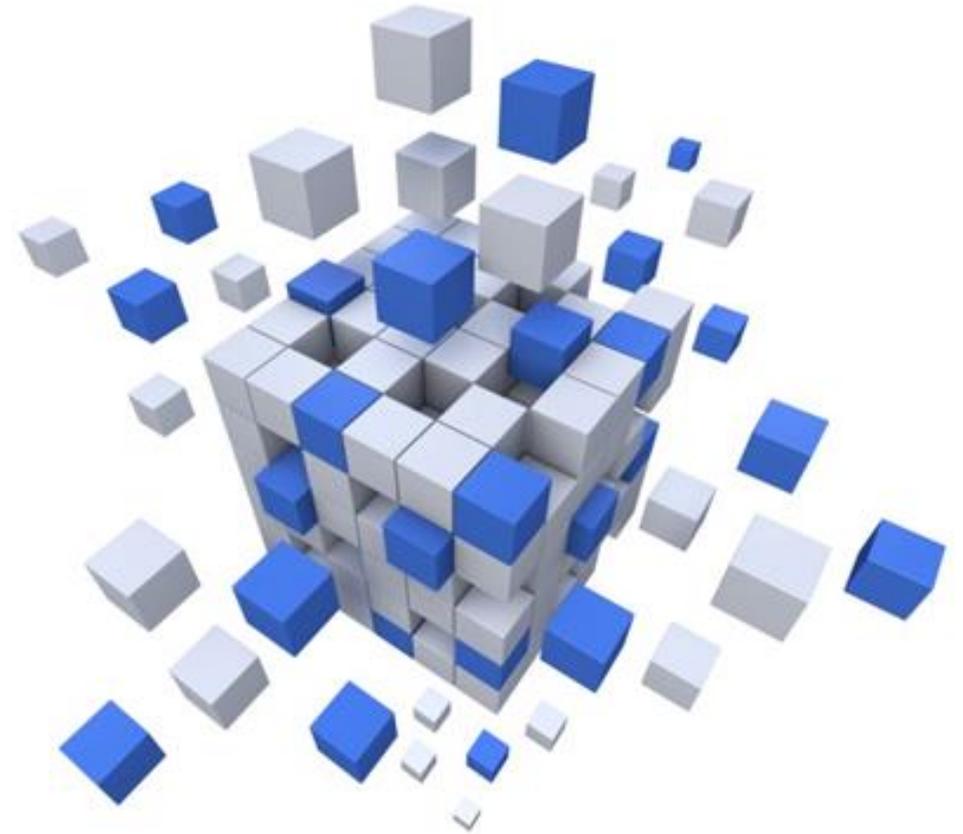
Analytics Models are Different

Entity Relationship Approach

Show me the left turning traffic on Hanson Road along all intersections in traffic moving towards East direction between Hwy 6 and SH 92 during the PM peak on a Wednesday looking at last 3 months data

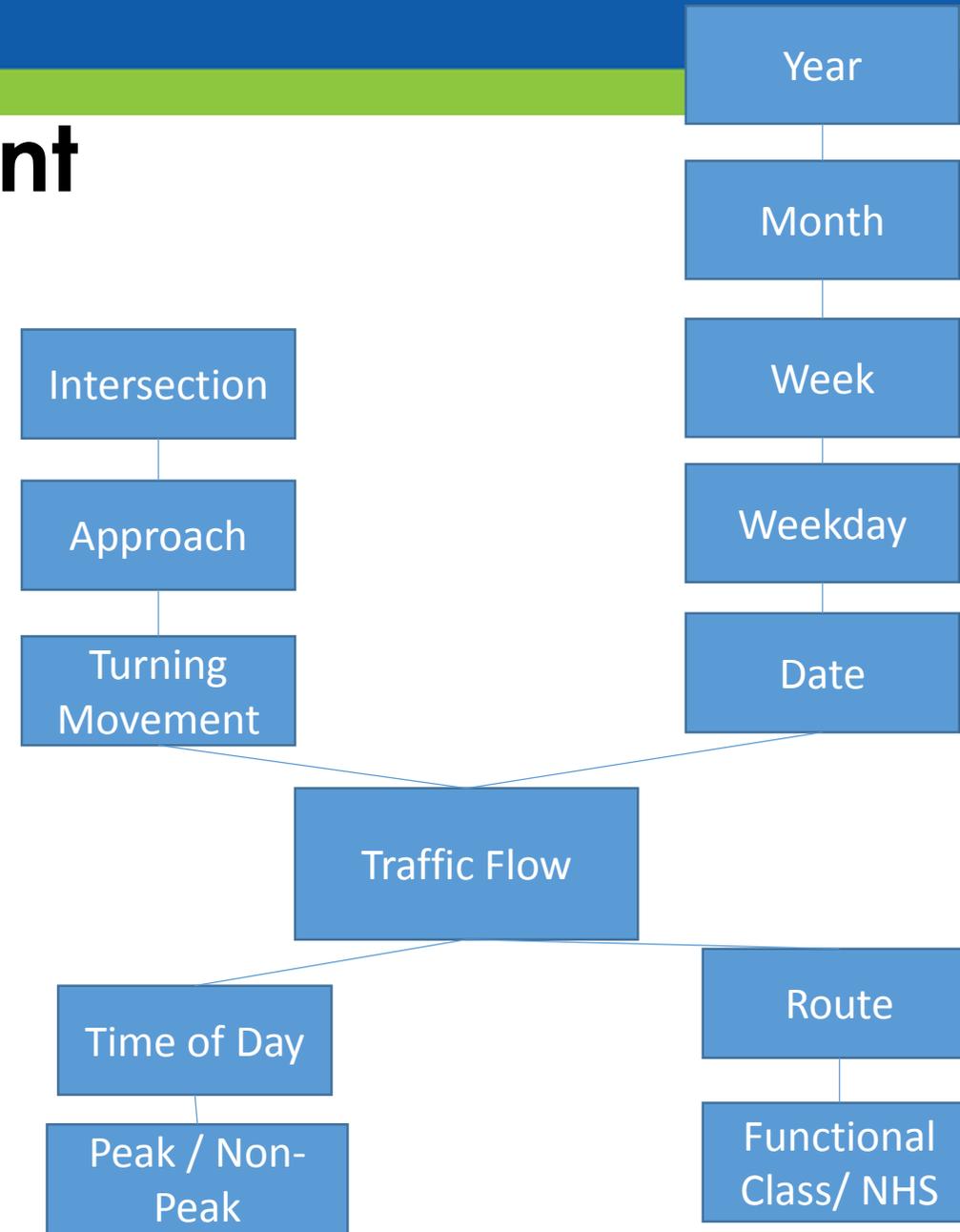
Dimensional Models 101

- Measures
- Facts
- Dimensions
- Conformed Dimensions
- Dimensional Hierarchy



Analytics Models are Different

Dimensional Modeling Approach



Outline

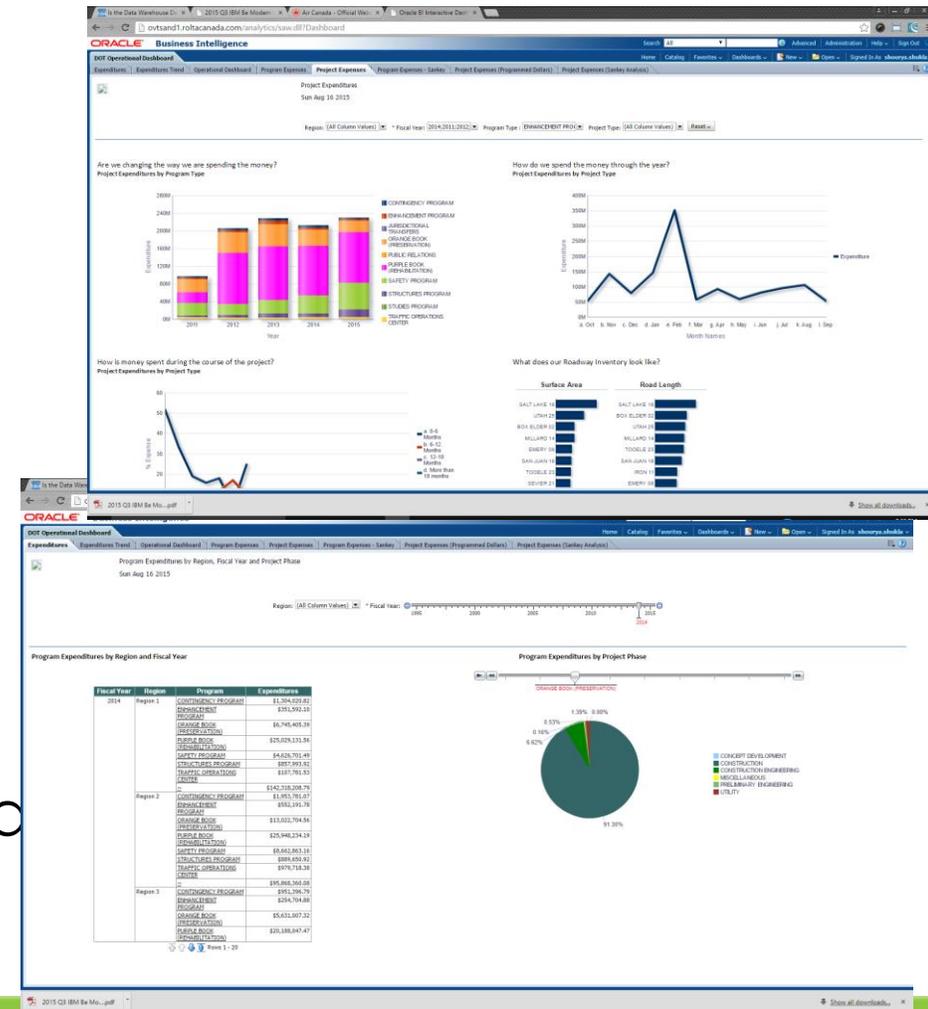
- ➔ 1. MAP 21—Moving from activities to outcomes
- ➔ 2. Importance of Data for MAP 21
- ➔ 3. What questions are DOTs trying to answer?
- 4. Emerging Analytics Landscape
- 5. Typical DOT stakeholders and their needs
- 6. How it all fits together



What questions do DOT executives need to answer?

Case Study: UDOT

- Are we changing the way we spend our money?
- Are we getting better on how we deliver on our outcomes?
- Is there a consensus in the organization on how we should prioritize our spending?
- Did we spend all the dollars we received/ allocated/ obligated?
- If we had one more dollar to spend on our assets, where should we spend it?
- What is the typical cash flow for a project of a certain type and certain value?

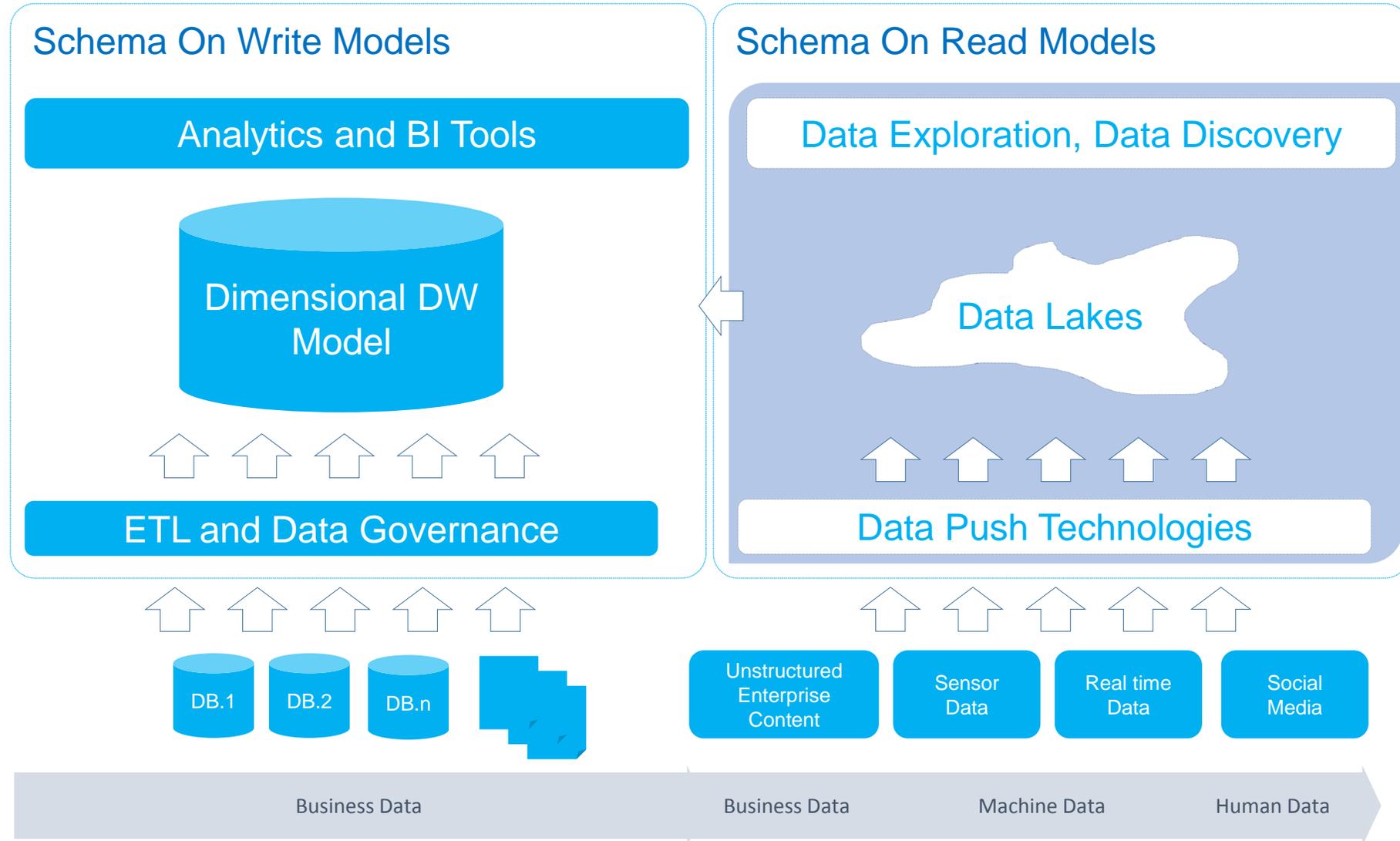


Outline

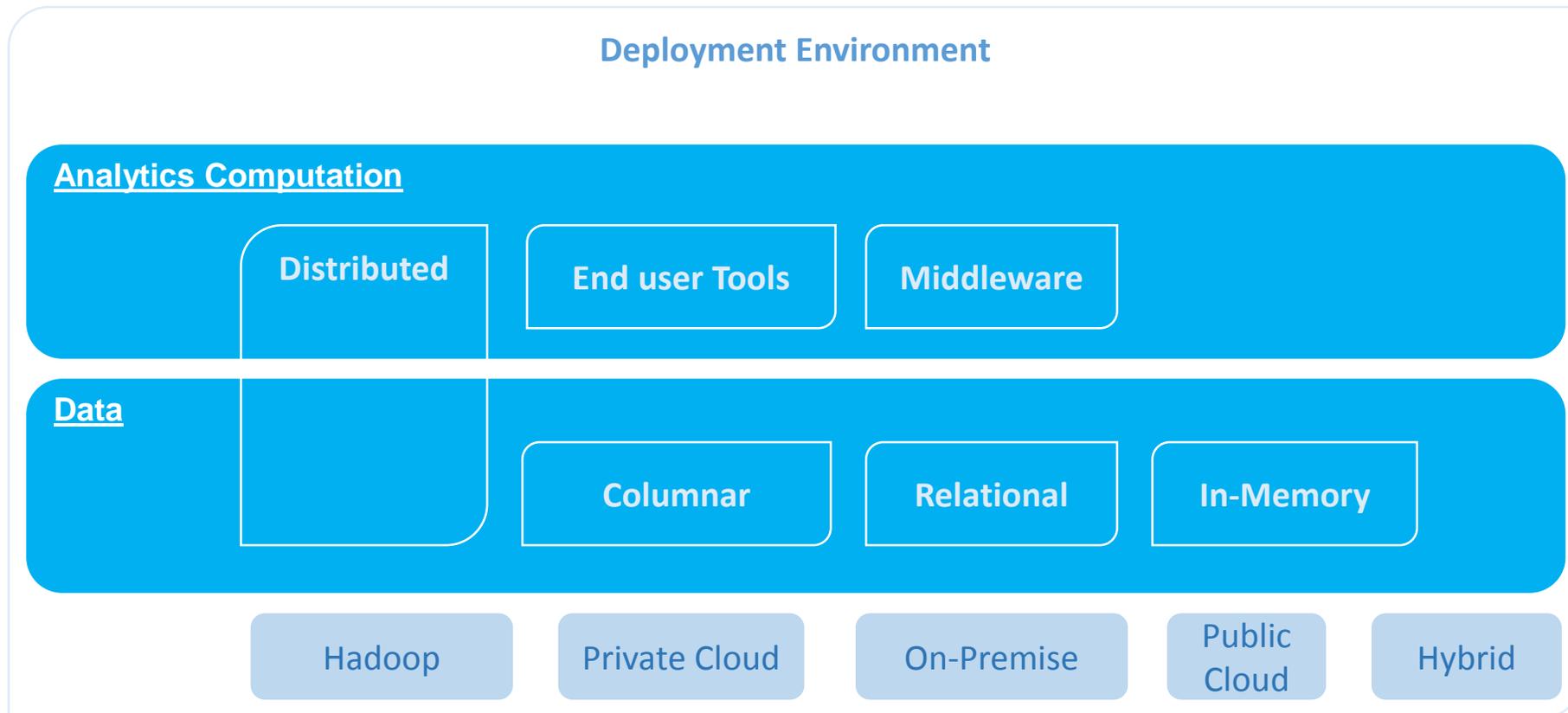
- ➔ 1. MAP 21—Moving from activities to outcomes
- ➔ 2. Importance of Data for MAP 21
- ➔ 3. What questions are DOTs trying to answer?
- ➔ 4. Emerging Analytics Landscape
- ➔ 5. Typical DOT stakeholders and their needs
- ➔ 6. How it all fits together



Changing analytics landscape



Various approaches and models to mix and match



Outline

- ➔ 1. MAP 21—Moving from activities to outcomes
- ➔ 2. Importance of Data for MAP 21
- ➔ 3. What questions are DOTs trying to answer?
- ➔ 4. Emerging Analytics Landscape
- ➔ 5. Typical DOT stakeholders and their needs
- ➔ 6. How it all fits together



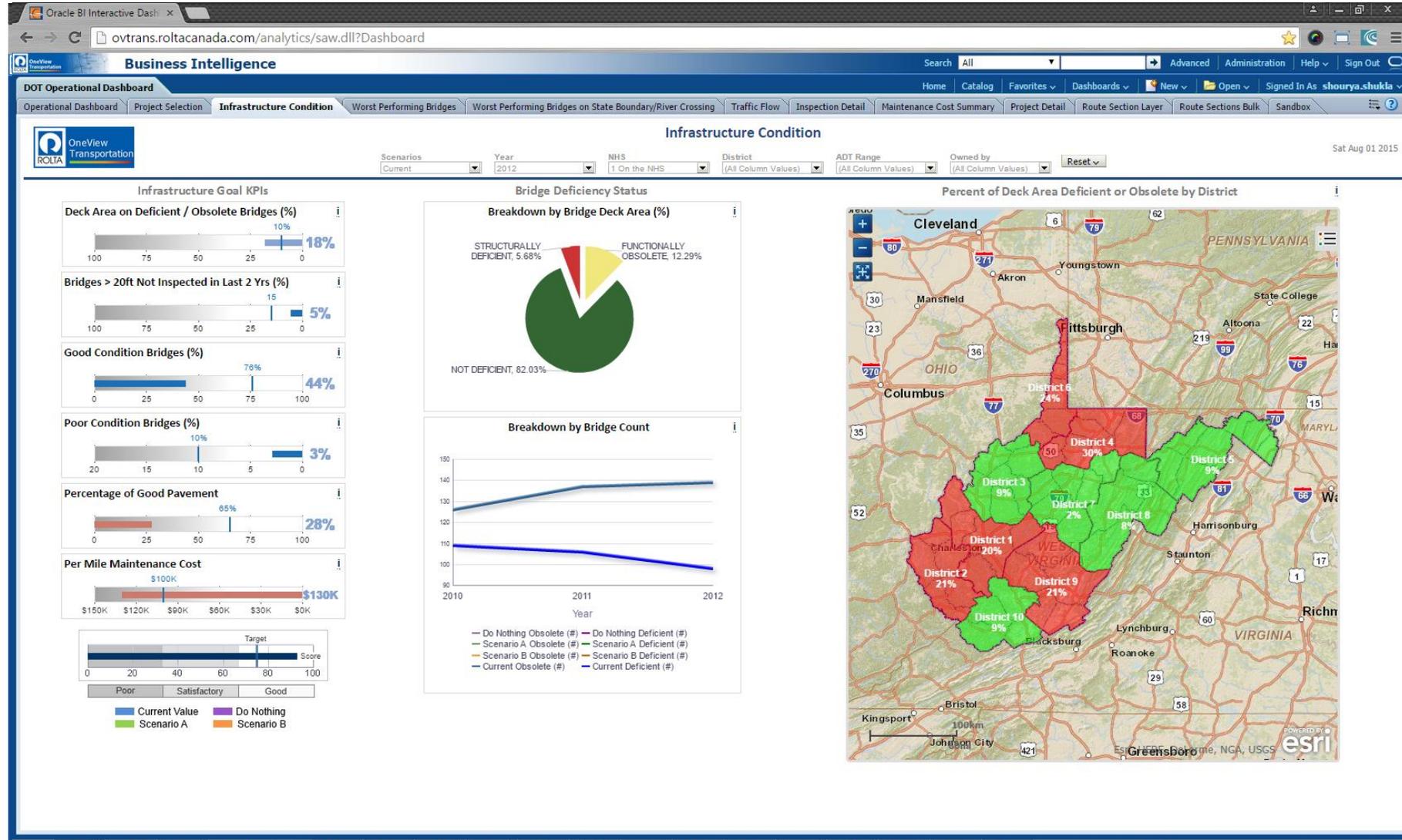
Different user segments have different needs

<u>Executive Users</u>	Schema On Write	<ul style="list-style-type: none"> • Consistent reports • Dashboards • Ability to drill down • Ability to apply filters • Ability to look at historical data in a consistent way • Ability to understand alternatives without worrying about implementation details
<u>Self Service Users</u>	Schema On Write	<ul style="list-style-type: none"> • Intuitive Data model irrespective of source • Ability to build own analysis using domain terms • Ability to share analysis with other users
<u>Business Analysts</u>	Schema On Read	<ul style="list-style-type: none"> • Ability to mine data and define schema at time of use • Ability to combine DW and non-DW data sources and ask non-standard questions • Ability to execute predictive and prescriptive models
<u>Data Scientists, IT Developers</u>	Schema On Read	<ul style="list-style-type: none"> • Ability to combine unrelated data in different ways • Deep understanding of data algorithms and sifting and sorting data • Define data structure at the time of use

Executive Dashboards

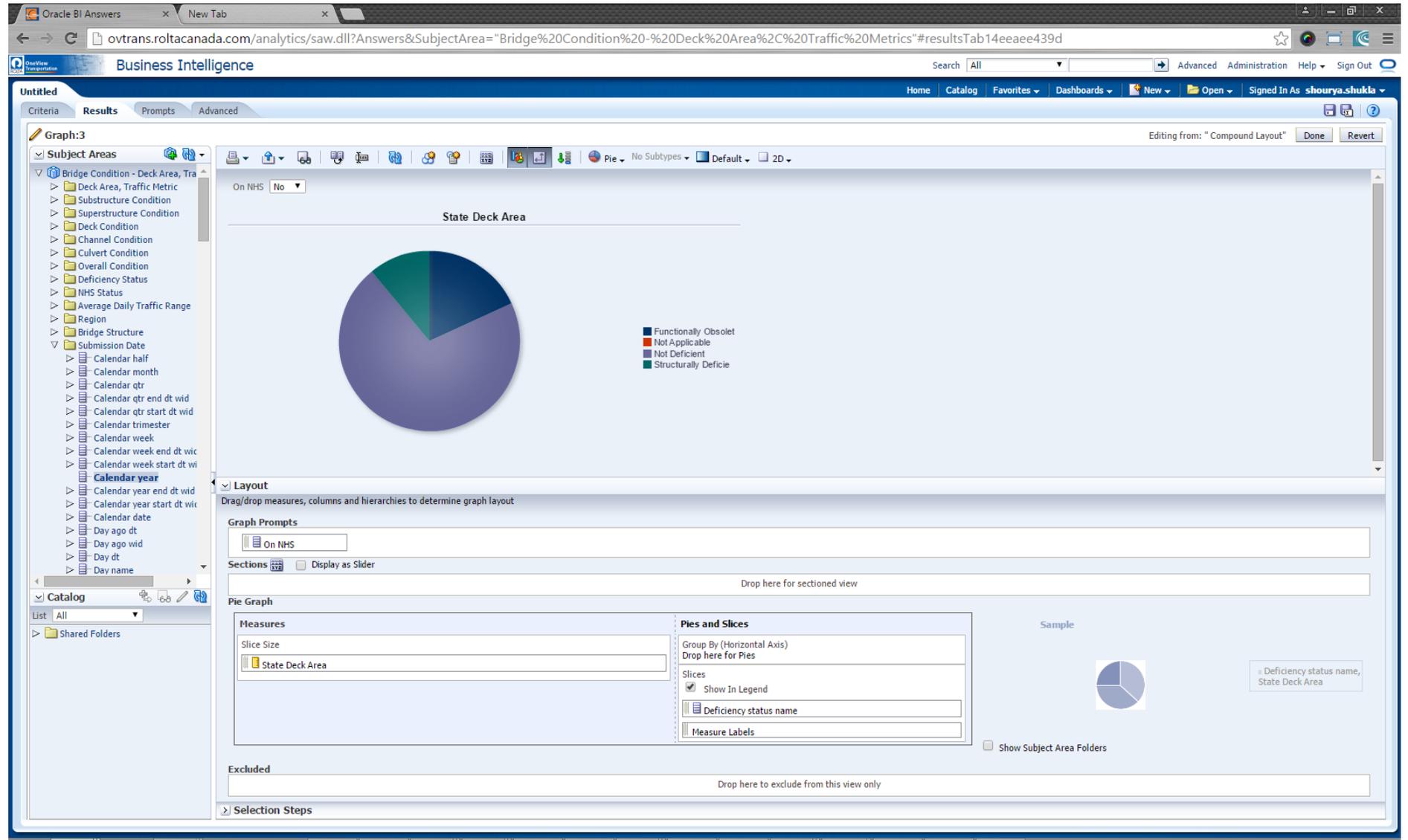
Everything that matters on one screen

Ability to drill down and across



Self Service Analysis

Building dynamic analysis using a semantic model



The screenshot displays the Oracle BI Answers interface. The browser address bar shows the URL: `ovtrans.roltacanada.com/analytics/saw.dll?Answers&SubjectArea="Bridge%20Condition%20-%20Deck%20Area%2C%20Traffic%20Metrics"#resultsTab14eeeee439d`. The interface is titled "Business Intelligence" and shows a user signed in as "shourya.shukla".

The main view is a pie chart titled "State Deck Area" with a legend indicating four categories: Functionally Obsolete (dark blue), Not Applicable (orange), Not Deficient (purple), and Structurally Deficient (green). The chart is set to "On NHS" with a value of "No".

The configuration panels include:

- Layout:** Drag/drop measures, columns and hierarchies to determine graph layout.
- Graph Prompts:** Includes a prompt for "On NHS".
- Sections:** Includes a "Display as Slider" checkbox.
- Pie Graph:**
 - Measures:** Includes "Slice Size" and "State Deck Area".
 - Pies and Slices:** Includes "Group By (Horizontal Axis)", "Drop here for Pies", "Slices", "Show in Legend" (checked), "Deficiency status name", and "Measure Labels".
- Excluded:** Drop here to exclude from this view only.
- Selection Steps:** (Empty panel)

A "Sample" view of the pie chart is shown on the right, with a label "Deficiency status name, State Deck Area".

Data Discovery

Look for patterns in structured and unstructured data

UT14 | NBI Utah | Oracle E

vmovtodi.rolta.com:8101/eid/web/nbi-utah/new-page?p_p_id=endecachartportlet_INSTANCE_2hmH&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view

ORACLE Endeca Information Discovery | NBI Utah

UT14 + Add Component

Search Box

Selected Refinements

Available Refinements

General

- ADT_029
- APPR_KIND_044A
- APPR_RAIL_036C
- APPR_RAIL_END_036D
- APPR_ROAD_EVAL_072
- APPR_SPANS_046
- APPR_TYPE_044B
- APPR_WIDTH_MT_032
- BASE_HWY_NETWORK_012
- BRIDGE_IMP_COST_094
- CHANNEL_COND_061
- COUNTY_CODE_003
- CULVERT_COND_062
- DATE_OF_INSPECT_090
- DECK_COND_058
- DECK_GEOMETRY_EVAL_068
- DECK_PROTECTION_108C
- DECK_STRUCTURE_TYPE_107
- DECK_WIDTH_MT_052
- DEGREES_SKEW_034
- DESIGN_LOAD_031
- DETOUR_KILOS_019
- DIRECTION_005E
- FACILITY_CARRIED_007
- FEATURES_DESC_006A
- FEDERAL_LANDS_105
- FED_AGENCY
- FRACTURE_092A
- FRACTURE_LAST_DATE_093A

Tag Cloud

'AMERICAN FORK CREEK' 'ASHLEY CREEK' 'BEAR RIVER' 'BEAVER CREEK' 'BEAVER RIVER' 'BIG COTTONWOOD CREEK' 'CHALK CREEK' 'CLEAR CREEK' 'COAL CREEK' 'CORINNE CANAL' 'COTTONWOOD CREEK'

'COUNTY ROAD' 'DEER CROSSING' 'DRY WASH' 'DUCHESNE RIVER' 'EAST JORDAN CANAL' 'FREMONT RIVER' 'HIGHLINE CANAL' 'HOBBLE CREEK' 'HUNTINGTON CREEK' 'I-15 (SR-15)' 'I-15 (SR-15)'

'I-15 (SR-15) NBL & SBL' 'I-80 (SR-80) EBL & WBL' 'JORDAN & SALT LAKE CANAL' 'JORDAN RIVER' 'LITTLE COTTONWOOD CREEK'

'LOGAN RIVER' 'MALAD RIVER' 'MAMMOTH CREEK' 'MILL CREEK' 'NO DATA ENTERED' 'OGDEN RIVER' 'PLEASANT CREEK' 'PRICE RIVER' 'PROVO RIVER' 'SALINA CREEK' 'SAN PITCH RIVER'

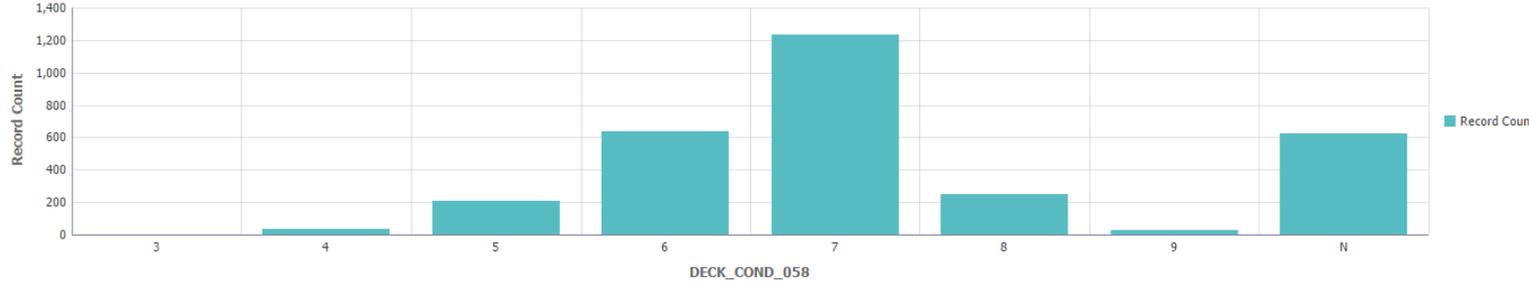
'SANTA CLARA RIVER' 'SEVIER RIVER' 'SEVIER VALLEY CANAL' 'SPANISH FORK RIVER' 'STRAWBERRY RIVER' 'UNION PACIFIC RAILROAD' 'UNNAMED WASH'

'UTAH & SALT LAKE CANAL' 'VIRGIN RIVER' 'WEBER RIVER' 'WEST CANAL' 'WILLOW CREEK'

Explore FEATURES_DESC_008A by Number of records

Chart

Record Count by DECK_COND_058



DECK_COND_058	Record Count
3	0
4	~50
5	~200
6	~650
7	~1250
8	~250
9	~50
N	~650

Sort: DECK_COND_058

Page 1 of 1

1-8 of 8 | 50 per page

Results Table

R Statistical Programming

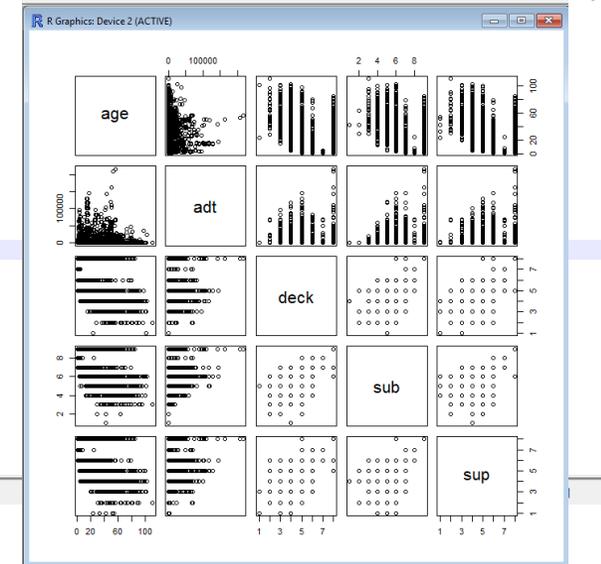
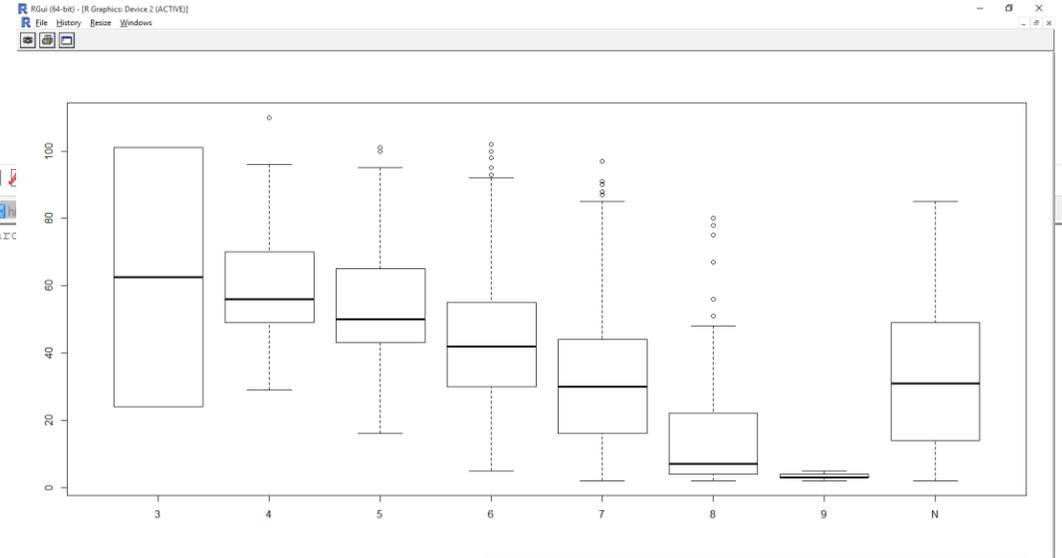
Powerful language for data scientists

```

1 Bridges <- read.csv("C:\\Work\\Transportation\\UtahDOT\\BIDashboard\\RFI\\FromRolta\\scripts\\bridges.r - Notepad+
2 dim(Bridges)
3 names(Bridges)
4 age<-2015-Bridges$YEAR_BUILT_027
5 hist(age)
6 adt<-Bridges$ADT_029
7 deck<-Bridges$DECK_COND_058
8 sup<-Bridges$SUPERSTRUCTURE_COND_059
9 sub<-Bridges$SUBSTRUCTURE_COND_060
10
11
12 deck<-as.factor(Bridges$DECK_COND_058)
13 sup<-as.factor(Bridges$SUPERSTRUCTURE_COND_059)
14 sub<-as.factor(Bridges$SUBSTRUCTURE_COND_060)
15
16 plot(deck, age)
17
18 plot(sup, age)
19
20
21 pairs( ~ age + adt + deck + sub + sup, Bridges)
22 summary(Bridges)
23
24 plot(sub, age)
25 plot(deck, adt)
26 x=seq(1,100)
27 y=seq(1,225000)
28 f = outer(x, y, function(x,y) cos(y)/1+x^2)
29 contour(x,y,f)
30 image(x,y,f)
31 persp(x,y,f, theta=30)
32
33 lm.fit1 = lm(deck~age,Bridges)
34 summary(lm.fit1)
35 coef(lm.fit1)
36 confint(lm.fit1)
37 lm.fit2=lm(deck~adt, Bridges)
38 summary(lm.fit2)
39 coef(lm.fit2)
40 confint(lm.fit2)
41

```

R programming language length : 925 lines : 47



Outline

- ➔ 1. MAP 21—Moving from activities to outcomes
- ➔ 2. Importance of Data for MAP 21
- ➔ 3. What questions are DOTs trying to answer?
- ➔ 4. Emerging Analytics Landscape
- ➔ 5. Typical DOT stakeholders and their needs
- ➔ 6. How it all fits together



Gartner Recommendations for Analytics

Business analytics leaders should:

- Determine the BI and analytics use cases needed by your organization.
- Consider the skills level of the users when choosing a product or vendor.
- Provide a balanced BI and analytics portfolio. This may require expanding BI and analytics delivery and roles from a centralized, IT-driven approach to a more decentralized, agile process.
- Assess customer satisfaction measures in platform/ technology selection
 - Support quality, product quality, upgrade difficulty, sales experience, ease of use, user enablement programs and achievement of business benefits
 - Supplements an evaluation of functionality, integration and cost-of-ownership requirements.

Putting it all together for MAP 21

- Start with the questions that executives need to answer?
- What datasets are available in the organization to answer these questions?
- Are there better alternative datasets available in the market place that can answer the questions reliably?
- How will this data be acquired and where will it reside?
- How will you ensure data consistency and quality?
- Who needs to access this data and what are their skillsets?
- What resources exist to acquire the technology?
- What resources exist to sustain the technology?
- How will you let your data and analysis drive your business?



Shourya Shukla

shourya.shukla@rolta.com

J. Michael Smallcombe, P.E.

michael.smallcombe@rolta.com