



SAS[®] GLOBAL FORUM 2018

USERS PROGRAM

April 8 - 11 | Denver, CO
Colorado Convention Center

#SASGF

Analytics Centralization

Raj Kannan, Chief Data Scientist, TCS

Analytics Centralization

Agenda

- Model Centralization Criteria
- Model Development vs. Model Centralization
- Definition:
 - Create centralized Enterprise Reusable models
 - Data Engineering to enable analytics
 - Deploy models on IT backbone
 - Monitoring enabled for all statistical output
 - Accessible source code
- Best Practices Team Structure
- Realization Architecture & Workflow
- Other Considerations

Analytics Centralization

Model Centralization Criteria

- Model is repeatedly used by business for decisions
- Consensus between modeling team and business team on sustained model usage
- Model benefits are limited due to scalability issues
- Model has successfully modeled enterprise-specific entities and has potential to be expanded to other business groups
- Diversely talented modeling teams embedded in business

Analytics Centralization

Model Development vs. Model Centralization

- Rigor of a centralized model is not data science work. Data Science is...
 - Problem solving for business objectives
 - Not software engineering
 - Not required to follow best coding practices
 - Can perform data tasks within the modeling platform
 - Not used to adding automated model monitoring constructs

Analytics Centralization

Creating an Enterprise Reusable Model

- 3 P's
 - Parameterized:
 - parameterized and not hard coded.
 - configuration driven
 - Performance Tuned:
 - are optimized to take advantage of underlying data infrastructure
 - data workflows are handled separately from algorithm workflows
 - Parallelized:
 - workflow enabled
 - not a serial script

Analytics Centralization

Data Engineering

- Create and maintain datasets built to enable analytics at enterprise scale
- Closed loop to create derived variables/features at enterprise scale
- Focus on creating entity level variables/features
- Analytics teams should be able to subset model input data instead of devising complicated data workflows
- Separation of “data” and “science”

Analytics Centralization

Deploy on IT backbone

- Automated execution of models on IT infrastructure
- Leverage IT supported scheduling tools
- Configuration file driven

Analytics Centralization

Enable monitoring

- All statistical outputs are programmatically captured, stored and monitored – by user, execution timestamp, execution type
- Customized thresholds defined by user
- Alert model development teams
- Establish retraining criteria
- Proactive model lifecycle management

Analytics Centralization

Accessible Source Code

- Centralized code or model repository
- Resolution of multiple repositories
 - Development version
 - Production version

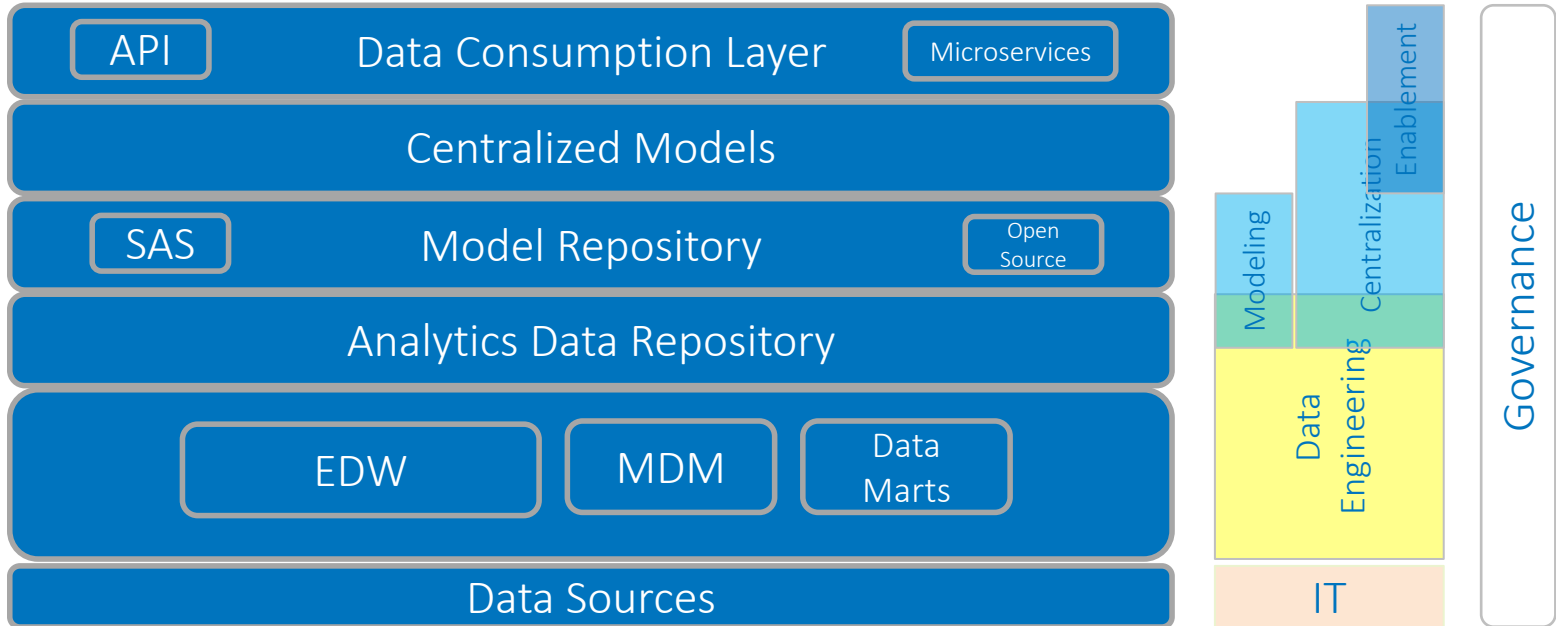
Analytics Centralization

Best Practices Team Structure

- **Model Centralization Team**
 - To convert dev caliber models to prod caliber models by infusing software engineering
- **Data Engineering Team**
 - with special skills in data engineering and knowledge of IT platforms
- **Enablement Team**
 - Integrate analytics output with operational systems
 - Create APIs, Microservices encapsulating model outputs
 - Create reports on model outputs on IT Consumption Layer

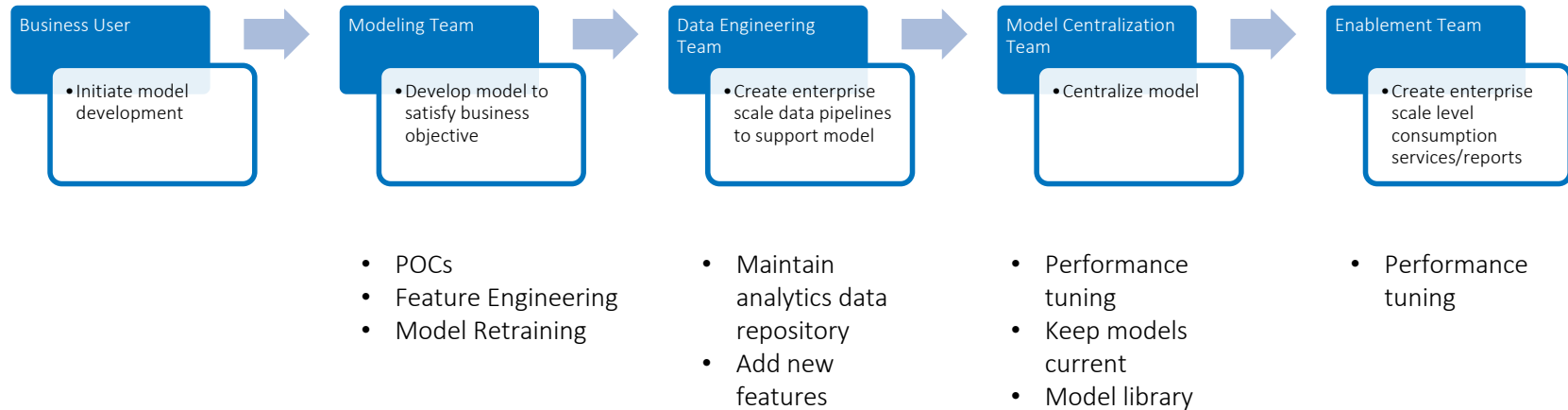
Analytics Centralization

Realization Architecture Discussion



Analytics Centralization

Realization Workflow



Analytics Centralization

Other Considerations

- Evolving data management strategy – EDW, Data Fabric
- Evolving data infrastructure strategy – DataCenter, Cloud
- Silo'ed business facing analytics teams
- No rigor in maintaining models – speed vs results
- Centralized organization responsible for analytics enablement
- Collaboration with IT teams:
 - Integrating with Agile IT teams
 - Integrating with IT processes, toolsets

Presenter

Raj Kannan, Chief Data Scientist, TCS

Raj Kannan works for Tata Consultancy Services as Chief Data Scientist for their Retail Business Unit.

In his current capacity, he works with customers in the US to elevate the data science capabilities. His past experience includes working as data scientist for Dell Technologies, Freescale Semiconductors and General Motors.

Raj graduated from the University of Toledo in Manufacturing Management (Master's) with specialization in Operations Research.

Raj is SAS Certified Predictive Modeler in addition to getting certified in Base SAS and Advanced SAS Programming

Your feedback counts!

Don't forget to complete the session survey
in your conference mobile app.

1. Go to the Agenda icon in the conference app.
2. Find this session title and select it.
3. On the sessions page, scroll down to Surveys and select the name of the survey.
4. Complete the survey and click Finish.

#SASGF

SAS[®]
GLOBAL
FORUM
2018

April 8 - 11 | Denver, CO
Colorado Convention Center