Gdevops

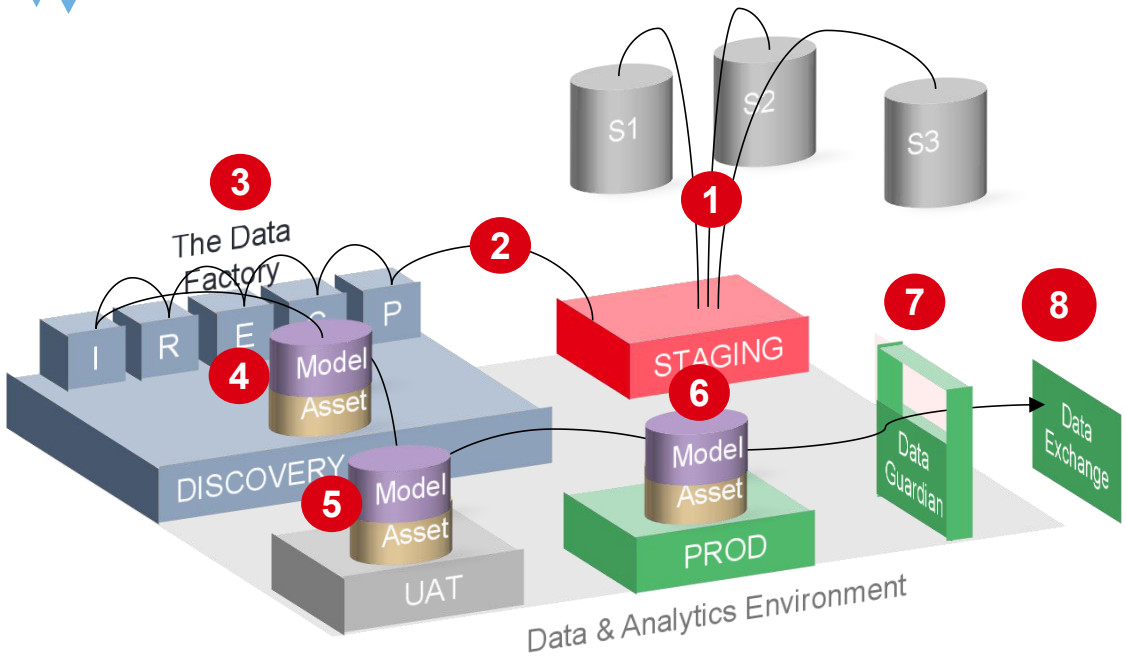
Global DevOps Summit

全球敏捷运维峰会

Big Data Intelligent Processing & Data Visualization

演讲人：吴仕橹

Business Insights & Analytics – How it Works



1) Source systems are ingested into staging (a shared preparation area). Typically through Sqoop (database copy) or CDC (streaming style change updates) or Juniper(in the house platform)

2) System tables are copied into the Discovery environment, where this production data is processed and models/insight are developed post Data Factory

3) The Data Factory takes raw data through a number of steps:
   - i. Profiling : looking at the data to identify its contents and tag it with the correct metadata
   - ii. Cleansing & curating : restructuring the data into the simplest and most efficient form, highlighting errors to revert back to source system owners
   - iii. Enriching : creating new derived fields based on the raw data (e.g. flags) and appending reference data for models to utilise
   - iv. Record linking : using advanced techniques to join up disparate data and masses of separate sources into a single logical model
   - v. Indexing : organising the final data asset into an index, making it quickly searchable

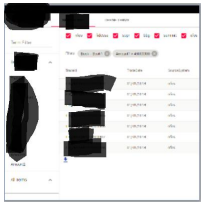
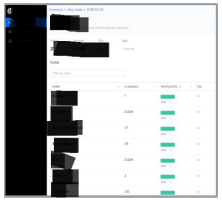
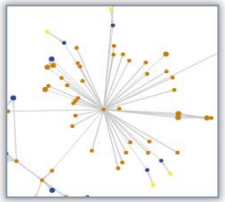
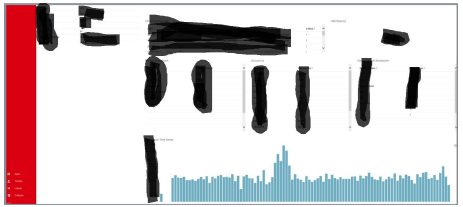
4) Stabilised assets and models are pushed through our UAT environment for testing and data validation from the consuming users

5) Final models and assets are then landed in our production environment; their insight ready for consumption through agreed patterns (typically APIs or file transfers)

6) The Data Guardian will control all consumption compliance

7) Data Exchange hosts APIs/APPs to source data to consumers

Data & Analytics Execution

	Ingest	Transform	Profile	Link	Analyse	Consume
Pipeline	Automated feed of data, copying the source systems into the GBM Data & Analytics Lake	Data is pre-processed, transformed and optimised by Data Engineers	Data is profiled to tag components for metadata analysis Algorithms used to predict data type and automatically tag	The tagged data is linked and enriched using machine learning, generating unique identifiers for clients	The enriched data is validated against business rules to ensure that it is fit for purpose	The finalised data is passed into a range of MI, analytics and data science applications to generate business value
Example						
	Raw XML Trade Data	Pre-processed Source Data	Metadata Modelling	Record linked Network Graph	Data Validation Results	Time-series Application

Use Cases in Execute

- Case 1
- Case 2
- Case 3
- Case 4
- Case 5
- Case 6
- Case 7
- Case 8

Technologies

Kafka, hadoop, elasticsearch, TensorFlow, Kibana, Grafana, Spark, Spark GraphX, Spark MLlib, Scala, We use more ...

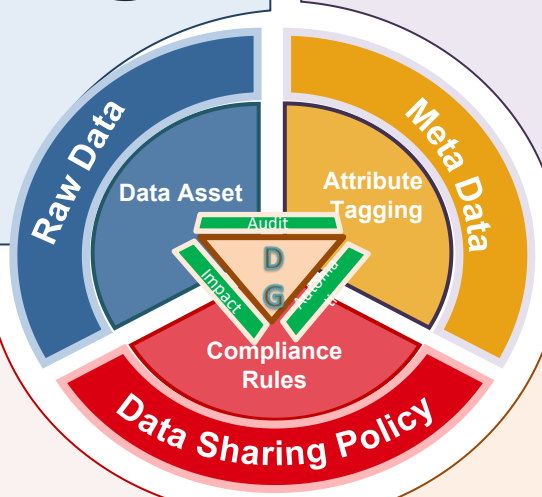
Data Guardian - 1

Source Data

- Data ingested from hundreds of source systems
- Data cleansed via GBM Data Factory
- Data presented in use case assets

Information Asset Registry

- Golden source for physical to logical mappings, mastered in data factory
- Repository for logical attribute hierarchy, containing terms where necessary



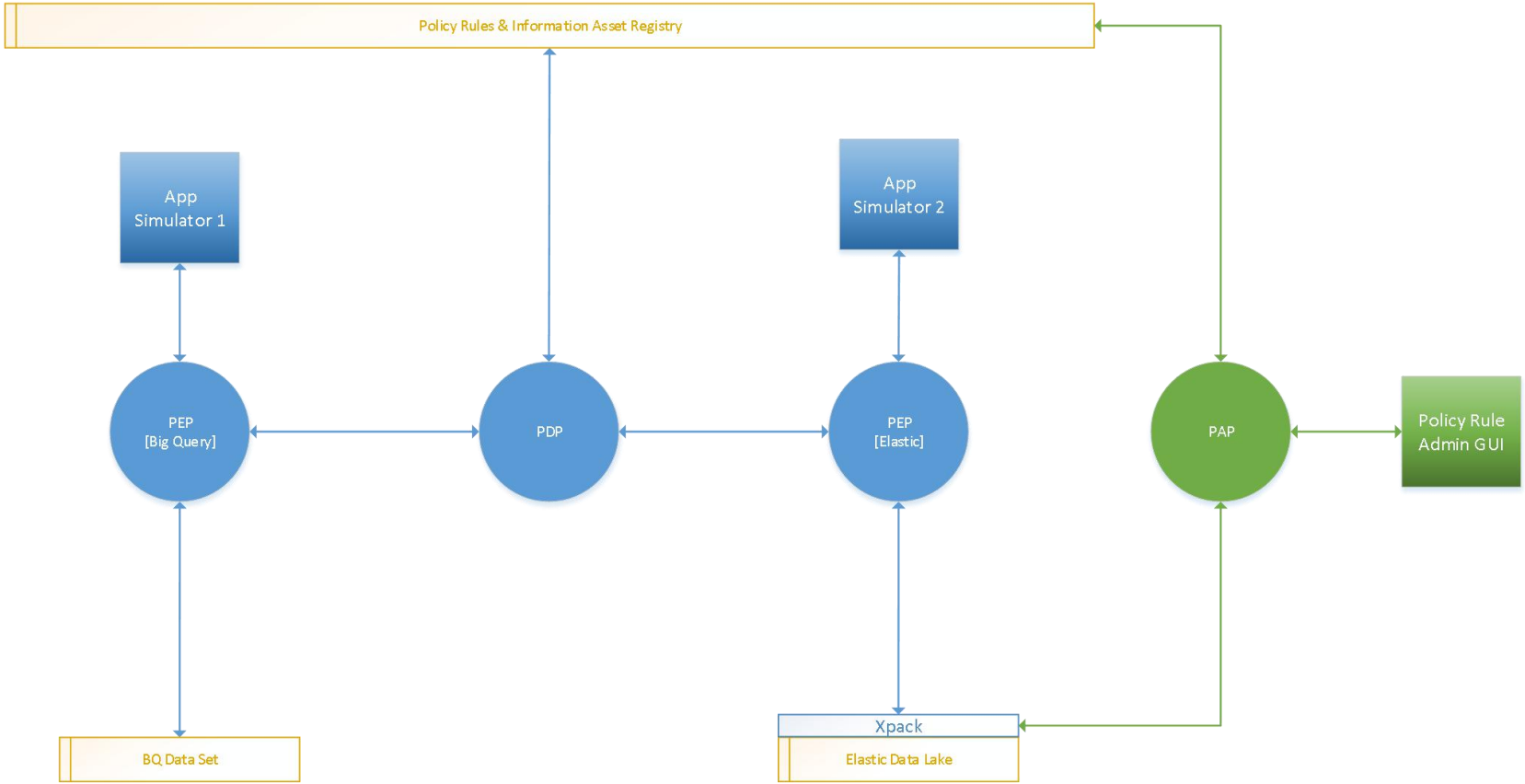
Data Sharing

- Policy's obtained from regional legal and compliance teams
- Policy converted into set of sharing rules
- Rules converted into Standard Rules Template ready for consumption

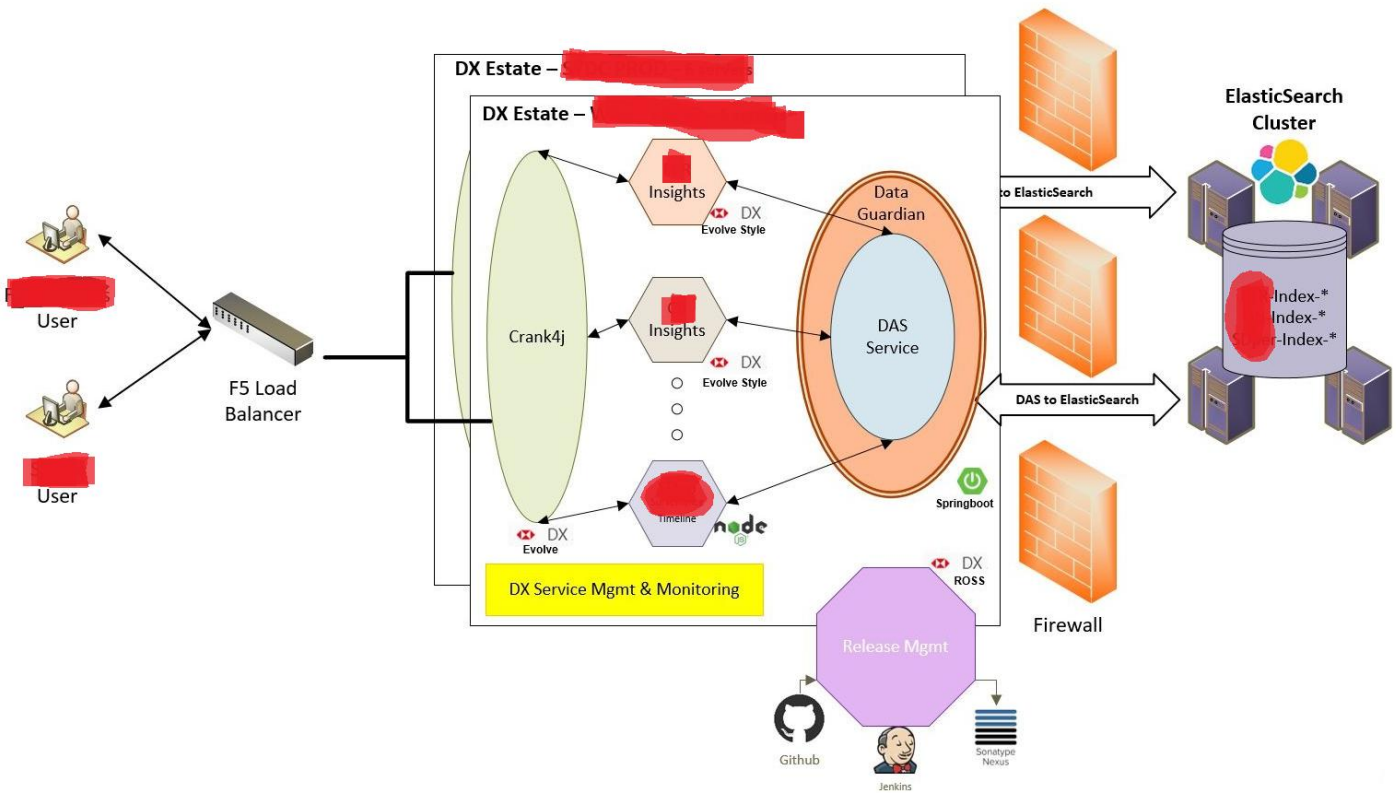
Data Guardian

- Policy Administration tool linked up with meta data store, allows policy rules to be entered in logical terms
- Each “data access request type” is assed by Policy Engine in order to produce a Policy Decision Point summarizing the resultant compliant dataset
- Automatic adaption of queries and in process filters in order to produce compliant data view

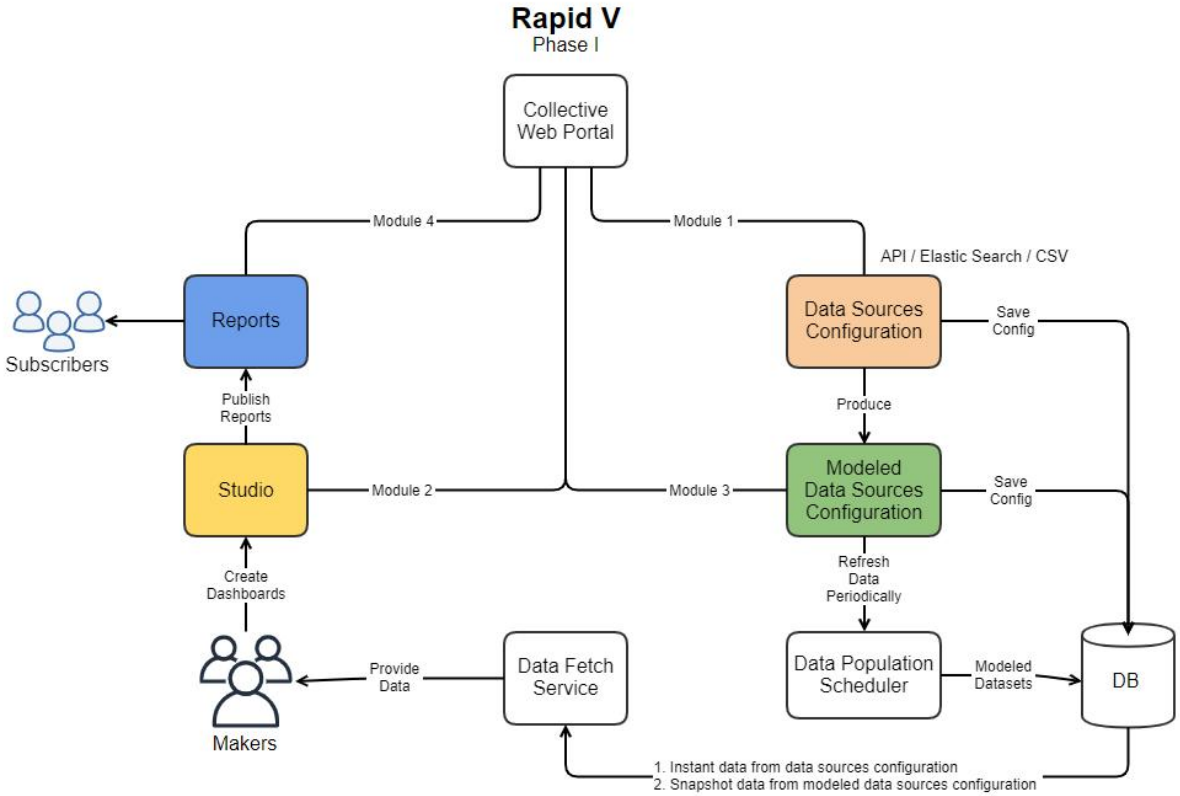
Data Guardian - 2



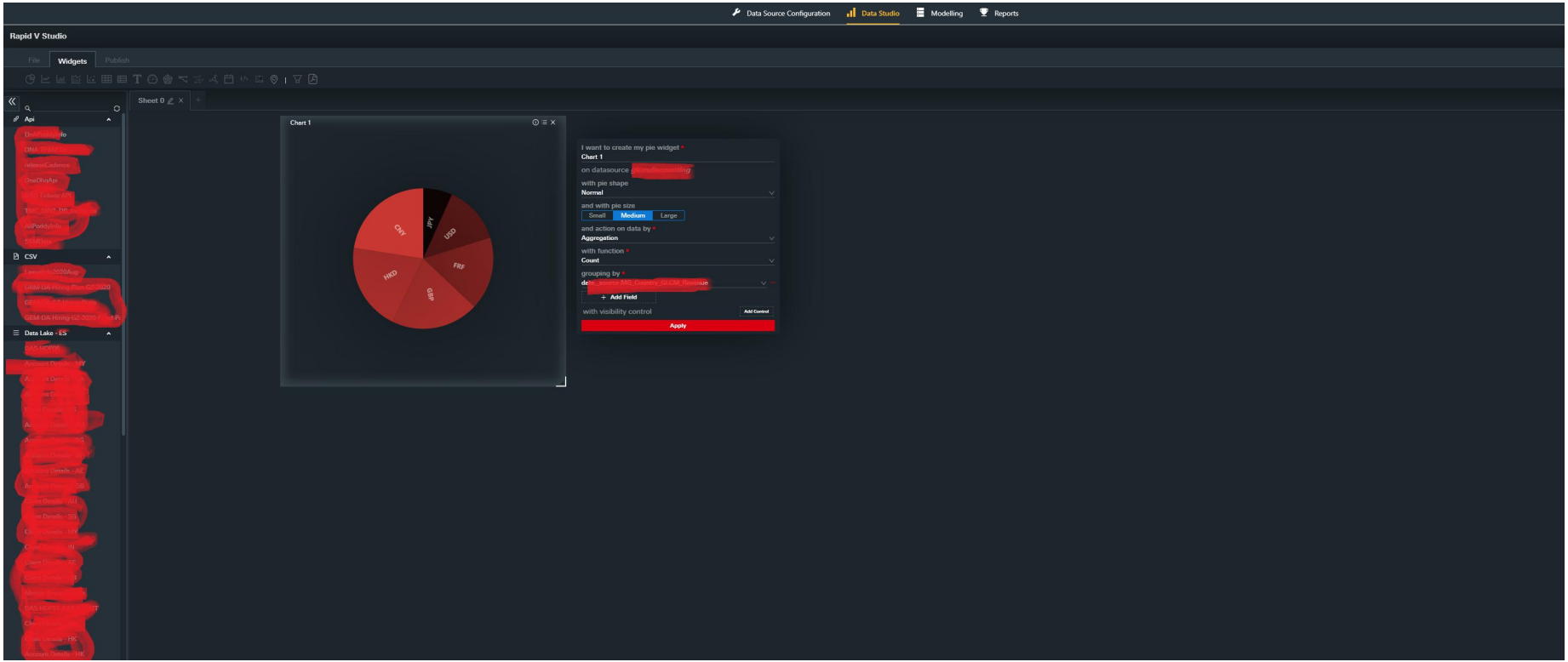
Data Exchange



Rapid-V Design



Rapid-V Demo


Data Source Configuration
Data Studio
Modelling
Reports
Rapid V Studio
File
Widgets
Publish
Sheet 0
Api
CSV
Data Lake - ES
Chart 1
JPY
USD
FRE
GBP
HKD
CNY
I want to create my pie widget *
Chart 1
on datasource
with pie shape
Normal
and with pie size
Small
Medium
Large
and action on data by *
Aggregation
with function *
Count
grouping by *
+ Add Field
with visibility control
Add Control
Apply

Rapid-V Sample

Join Us

Gdevops

全球敏捷运维峰会

THANK YOU !