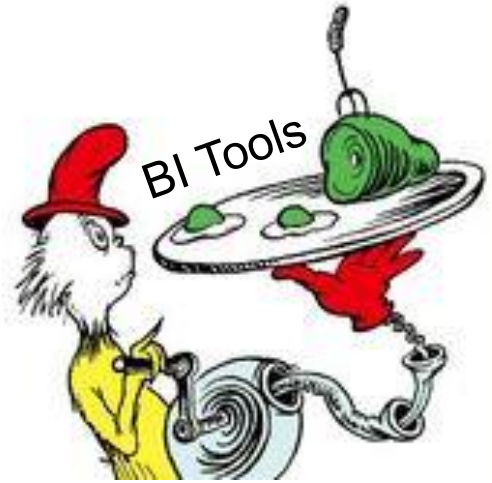
Thinking of Implementing SAS VA8.2 / Viya 3.3:
Real World Lessons Learnt

Steve Clarke

- **What is the world's No 1 BI Tool?**

Would you like a BI Tool?

I do not like them,
Steve -I-am.
I do not like
BI Tools.



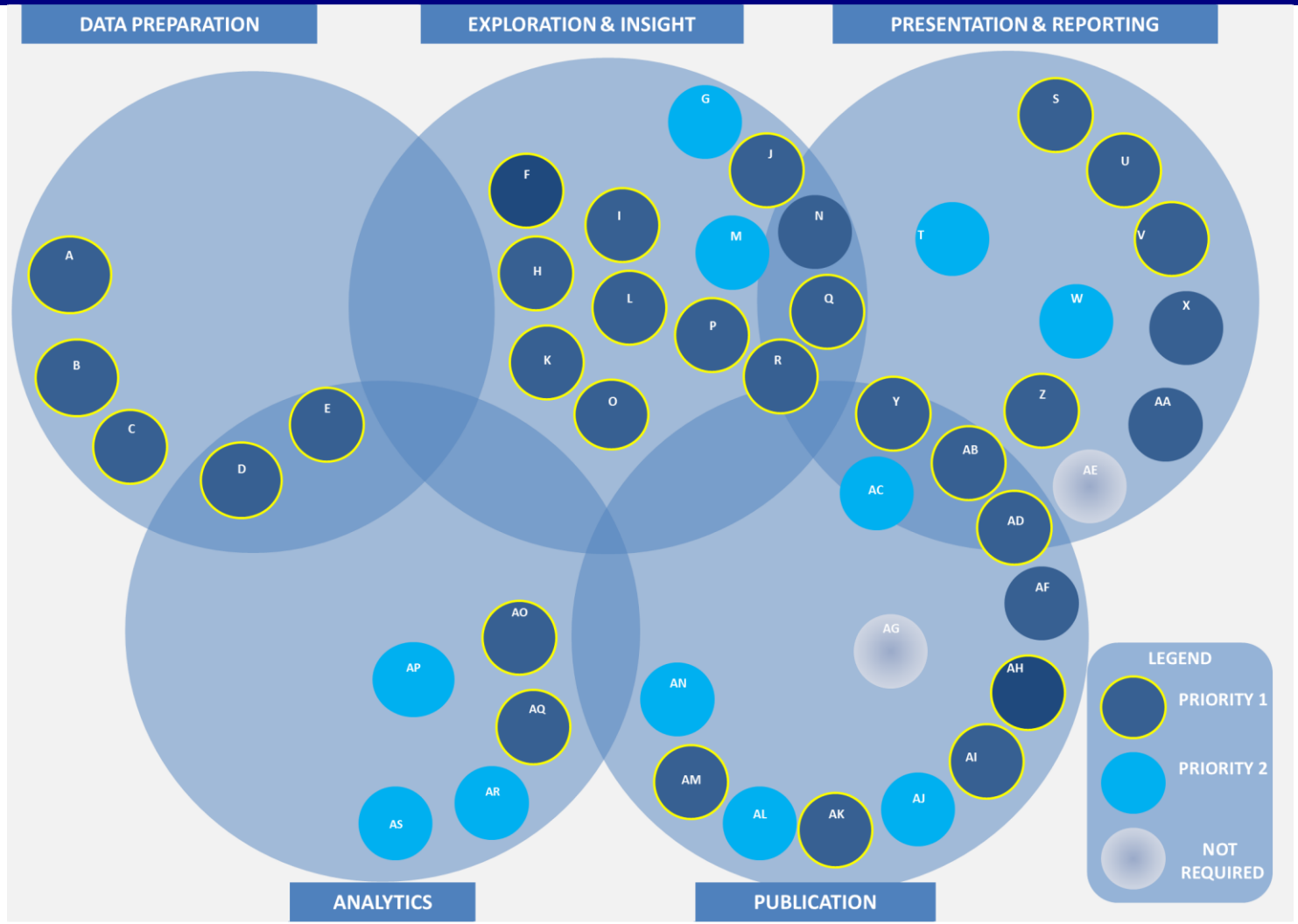
Our Journey

- **BI Selection Process**
- **Solution Overview**
- **Our Approach / Lessons Learned**

BI Selection

What do you want?

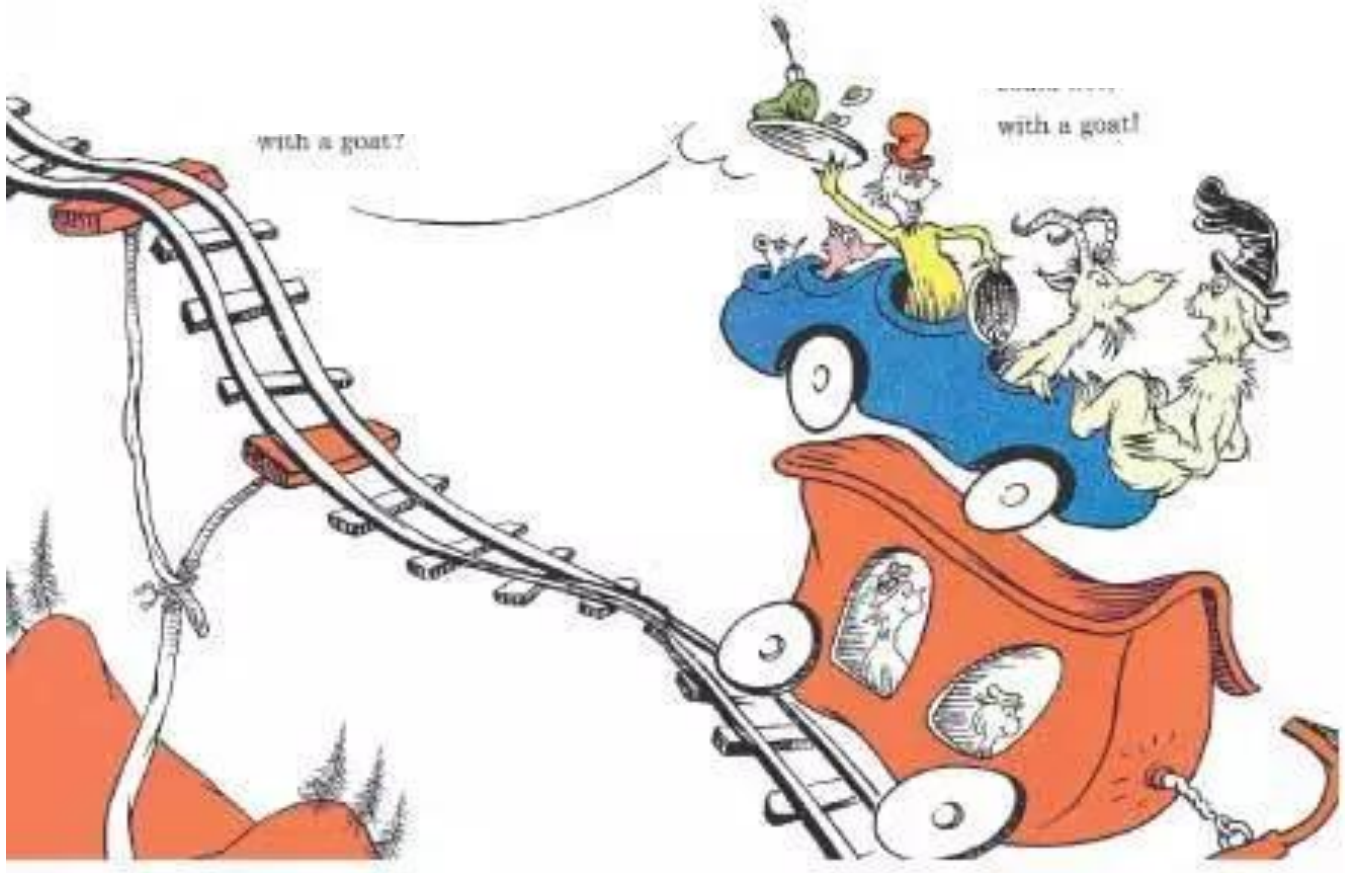
The Infamous Circle Diagrams



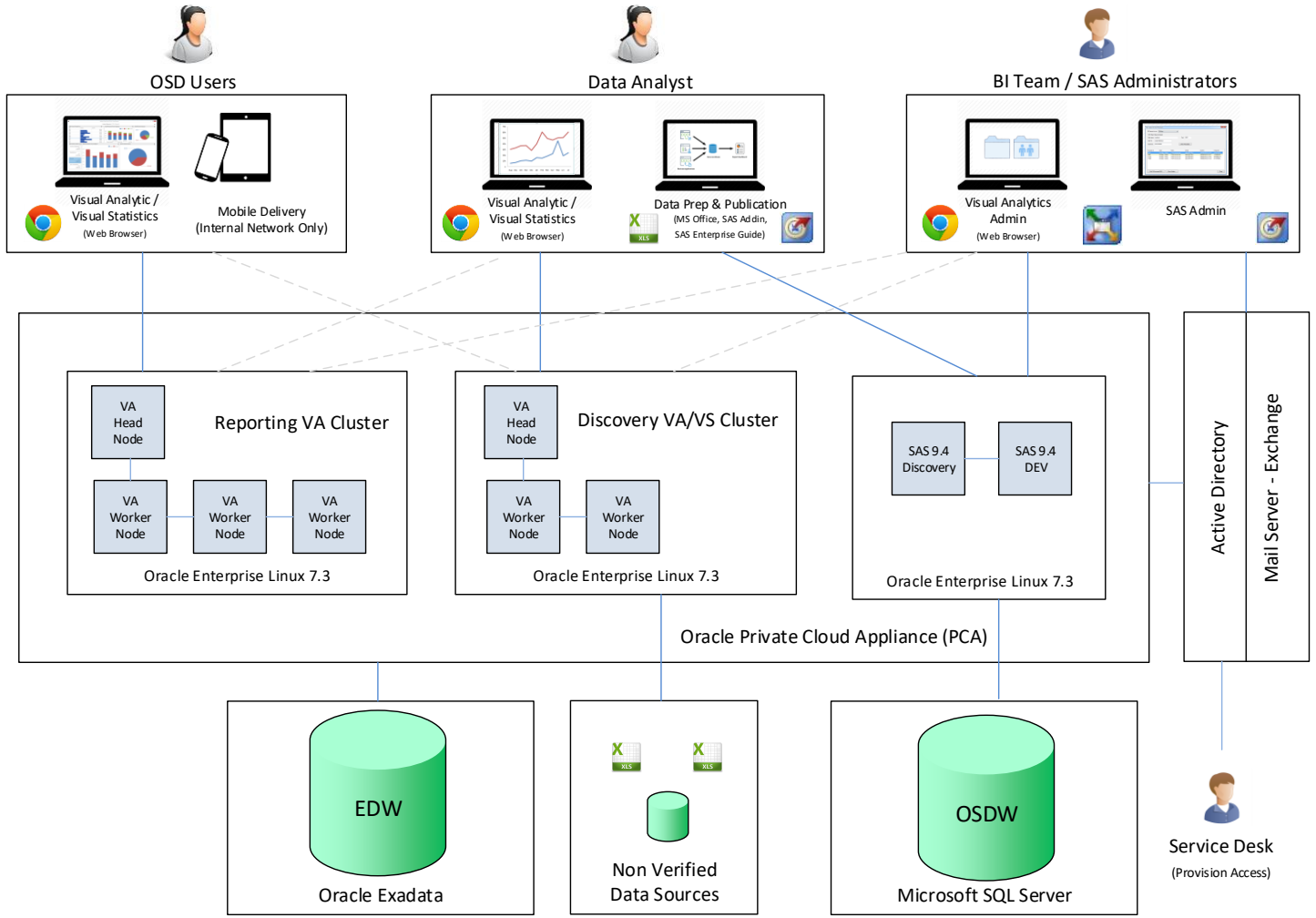
Software Evaluation

- Proof of Concept to test functions
- Testing against two key use cases
- Who is your customer?
- Who is using the tools?

Solution Overview



MOJ SAS Environments



Loading Data into VA

- **Provided Three Mechanisms**
  1. Scheduled batch load
  2. Enterprise Guide macro
  3. Introduced a concept of a “Drop Box”

     (Clearly needed to state to Security and Architects we weren’t implementing the Dropbox file share ☺)
- **Made a choice to disable the Import feature**

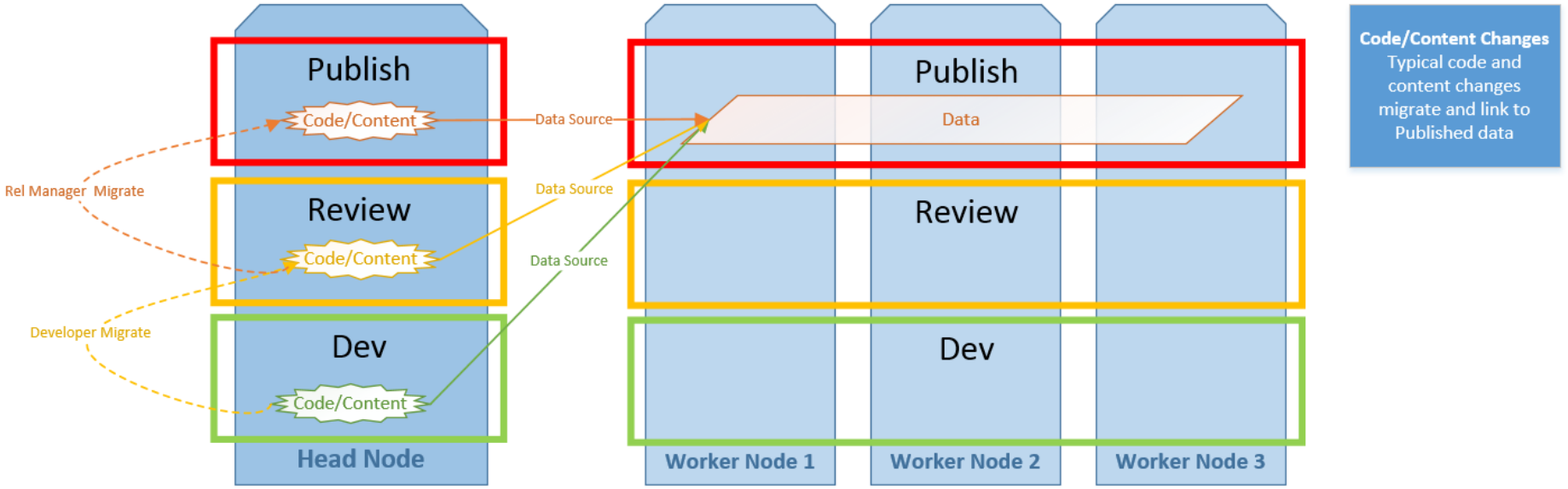
Dropbox Setup Steps

1. **Create shares on our VA server**
2. **Created VA Libraries to the shares**
3. **Place data in Dropbox file locations**

   Note: Can import SAS Datasets, XLS, CSV
4. **Users can load and unload data as required**
5. **Supports lazy loading**

   (Requires write access to CAS library!)

Promotion – Code / Content



Lessons Learned / Key Points

- 64bit Browser – Chrome performs best
- VM and infrastructure tuning
- Good security design
- Proof of Concepts leads to better / faster outcomes
- Address areas of risk early

Happy Customer

- **Contact Us**

  Steve.Clarke@justice.govt.nz

Additional Slides

Loading Data – Establish CAS Sessions

```
/***********************************/
/* Set run parameters              */
/***********************************/
%alloc_va_env(va_report_env=Development,va_report_subject_area=OSD);
%let input_name = COURT;
```

- CAS – Cloud Analytics Services
- CAS works with sessions, you need to establish a connection to CAS with a set user id.
- We have a custom macro that establish a session and based on parameters sets the target CAS Library
- Scheduling is external using our Oracle Scheduler straight after Facts and Dimensions
- Multiple VA tables can be loaded in parallel from our flattened

Load Code – Stage Data using CASUTIL

```
proc casutil;

   load casdata="&input_name._VW" incaslib="&va_dim_lib" casout="&input_name._stg" outcaslib="&va_ext_report_env&va_ext_report_subject_area"
        replace copies=0
        label="Criminal Final Disposal Workload List"
        varlist=(
             (name="SNAPSHOT_DATE"                label="Snapshot Date" format="DATE9."),
             (name="SNAPSHOT_MTH_OFFSET_NO"       label="Snapshot Mth Offset No"),
             (name="PERIOD_DATE"                  label="Period Date" format="DATE9."),
             (name="PERIOD_MTH_OFFSET_NO"         label="Period Mth Offset No"),
             (name="SOURCE_NAME"                  label="Source Name"),
             (name="JURISDICTION_CODE"            label="Jurisdiction Code"),
             (name="CASE_AGE_DAYS_ACT_PRIORACTV"  label="Case Age Days Actual Prior Active"),
             (name="CASE_AGE_DAYS_ACT_ADMSTG"     label="Case Age Days Actual Admin Stage"),
             (name="CASE_NUM_OCC_ADMSTG_EVTS"     label="Case Num Occ Admin Stage Evts")
        );

quit;
```

- CASUTIL server to server copying
- View to flatten data – better striped over Workers
- Stage table – minimise outage to in-memory dataset
- Copies=0, default COPIES=1 for redundancy
- Recommended reading https://communities.sas.com/t5/tkb/articleprintpage/tkb-id/library/article-id/2092

Load Data – Stage Data using a datastep

```
/*******************************************/
/* Load data to staging table              */
/*******************************************/
data mycas.&input_name._stg (duplicate=yes);

  set edw_dim.&input_name._vw (keep=COURT COURT_TYPE_CODE COURT_LOCATION COURT_CLUSTER COURT_REGION
                                    COURT_LATITUDE COURT_LONGITUDE REPORTING_SORT_ORDER_NO);

       label
            COURT                = "Court"
            COURT_TYPE_CODE      = "Court Type Code"
            COURT_LOCATION       = "Court Location"
            COURT_CLUSTER        = "Court Cluster"
            COURT_REGION         = "Court Region"
            COURT_LATITUDE       = "Court Latitude"
            COURT_LONGITUDE      = "Court Longitude"
            REPORTING_SORT_ORDER_NO = "Reporting Sort Order No"
       ;
run;
```

- **Datastep is a Client / Server operation invokes a Workspace Server**
- **Duplicate/Repeat option copies a table in whole to each worker**
- **Minimise inter worker node communication for joins** https://blogs.sas.com/content/sgf/2017/08/14/cas-data-modeling-for-performance/

Load Data – Promote into CAS

```
/*********************************************/
/* Replace reporting table with latest data   */
/*********************************************/
proc casutil incaslib="&va_ext_report_env&va_ext_report_subject_area";
   droptable casdata="&input_name" quiet;

   promote casdata="&input_name._stg" casout="&input_name" outcaslib="&va_ext_report_env&va_ext_report_subject_area";

   contents casdata="&input_name";
quit;

cas firstsession clear;
```

- Drop the In-Memory table
- Ensure you promote the content (or else it is only available to your session)
- If you load data directly add the promote key word
- Ensure you close your session so you don't waste resources