Practical Data Governance Use Case

Implementing Proactive Quality Controls to Prevent Reporting Collapses

Alex Salvas, Senior Manager, Risk Management

October 19th, 2011

NATIONAL BANK FINANCIAL GROUP

Agenda

- National Bank of Canada Overview
- Data Quality Issues
- How Did We Fix the Issues

National Bank of Canada Overview

At a Glance

- National Bank Financial Group is an integrated provider of financial services to retail, commercial, corporate and institutional clients. It operates in three business segments — personal and commercial, Wealth Management and Financial Markets—with total assets in excess of $145 billion as at october 31, 2010.
- More than 18,000 employees.
- National Bank of Canada is the leading bank in Quebec and the partner of choice for small and medium-sized enterprises. It is also the sixth largest bank in Canada with branches in almost every province.

NATIONAL BANK FINANCIAL GROUP

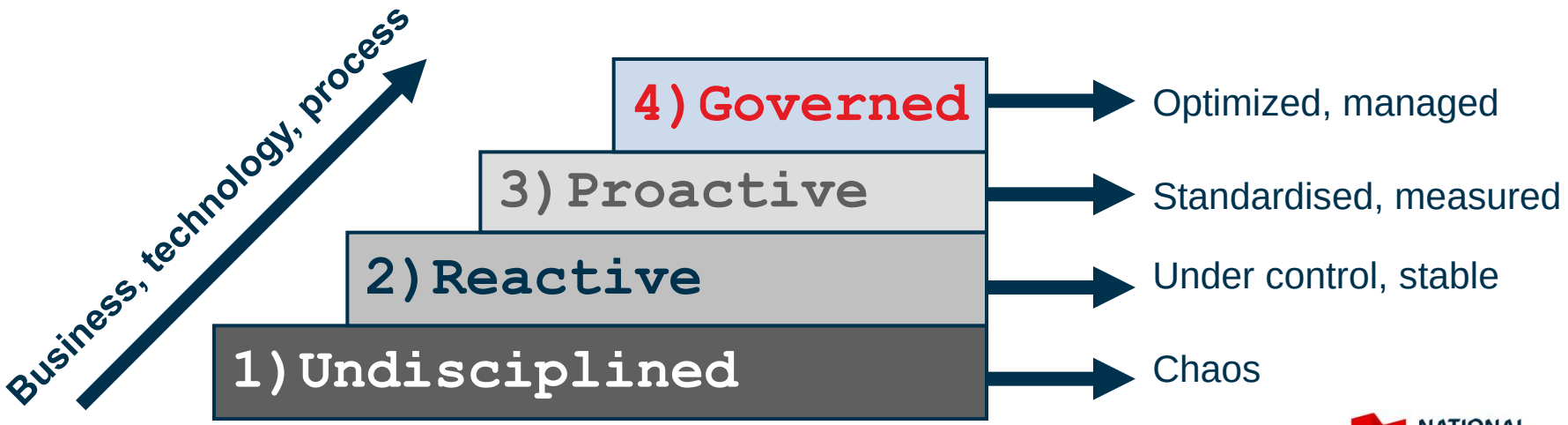
Risk Modeling Team

- The mission of our team is to **develop and implement innovative, quantitative solutions for promoting efficient credit risk management within the National Bank Financial Group**.
- The team consists of 55 professionals and analysts who develop Basel II risk models and business risk strategies, evaluate economic capital, produce business credit risk reports, define risk policies and assess risk information quality (risk data steward).

Data Quality Issues

Risk Data Quality Overview

- At NBC, Data Governance has been in place for three years now. It started at the first maturity level, i.e. undisciplined.
- After a couple of months, some business sectors gained maturity and were starting to be in a reactive mode.
- Then, more and more business units became proactive, which represents the third level of data governance maturity.



NATIONAL BANK FINANCIAL GROUP

Risk Data Quality Issues

- Since regulatory risk reporting, risk parameters modeling and data stewardship are part of the sector's objectives, high information quality is a must.
- Basel II production requires the integration of multiple data sources from all subsidiaries in a very short time.
- In the past years, we've experienced a few problems with some systems or processes :
  - "Year-2000 bug": one of the informational system had a "hard-coded" date table and the new year wasn't in it, resulting in missing values for some of the new transactions;
  - "Credit bureau bug": in a manual process, a 6-month old file was used instead of the current month to request historical credit bureau information, resulting in no information for the newer customers;
  - "Matching problem": due to architecture limitations and improper application of process, there are a few customers and facilities which aren't correctly matched with their transactional and credit information.
- To prevent such problems from happening, or at least to reduce the reaction time, processes are constantly reviewed, controls are strategically implemented and statistical analyses are made in a proactive way…

How Did We Fix the Issues?

Fixing Risk Data Quality Issues

- To prevent quality problems, we reviewed a lot of business processes and adapted some to make sure that proper controls were in place.
- Additionally, with the help of SAS, we were able to make good data mining and data analysis. This would help us foresee potential problematic areas.
- As can be seen on the next few slides, we developed some macros to analyze a few million records on a weekly or monthly basis. Monthly production covers all credit information used in Basel production whereas weekly production covers most important fields we're using in the credit risk sector, mainly authorized amount, balance, delinquency and so on.
- The information we wanted to focus on were the following:
  - Basic descriptive statistics (average, std deviation, min, max, missing values, …) and their tendency
  - Stability of the population (KS comparison between the current versus historical variable distribution)
  - Business logic rules (internal rules that data from one system to the other must follow)

Basic descriptive stats

- For the basic descriptive statistics, we simply have a SQL code which gives us the information we want and output it in a SAS file, according to various criteria (source system, product type, scoring model type, …).
- We have automated the comparison of these stats with historical information, so that whenever the statistical value is out of a specific range or trend, a flag is raised and more analysis is required.
- We would then output it to Excel for final reporting purpose.

```
%macro DATA_CONTROL (YEAR_MONTH=);
/*Prepare data for current month*/
/*...*/
/*Get basic stats*/
proc sql;
     create table Stats_&YEAR_MONTH. as
     select distinct
            YEAR_MONTH,
            BNC_MODEL_NAME,
            count(*) as NB,
            mean(CASE WHEN PD>0 THEN PD ELSE . END) as PD_AVG,
            sum(case when PD =0 then 1 else 0 end) as PD_0,
            NMISS (PD) AS PD_MISS,
            sum(case when PD =-1 then 1 else 0 end) as PD_NEG,
            sum(case when PD =100 then 1 else 0 end) as PD_100,
            std(CASE WHEN PD>0 THEN PD ELSE . END) as PD_STD,
                  /*...*/
     from COMMRISK.CI1B_&YEAR_MONTH.
     group by BNC_MODEL_NAME;
quit;
```

We prepare the current month dataset so that it can be compared to historical data

```
/*Add to the historical DB*/
proc sql;
     create table COMMRISK.STATS_TOTAL as
     select *
     from COMMRISK.HISTORICAL_STATS
     union ALL
     select *
     from Stats_&YEAR_MONTH.
     order by YEAR_MONTH, BNC_MODEL_NAME;
quit;

/*Compute historical statistics (less current month)*/
/*...*/
/*Compare to defined thresholds*/
/*...*/
/*Output report to Excel*/
%mend;
```

Compare to historical trends and output "Exceptions" report

NATIONAL BANK FINANCIAL GROUP

Population stability

- Usually, the stability of population analysis is used to evaluate if there's been a shift in the applicants of a specific risk model. In our case, and as was presented by Mark An of CIBC last year[1], we're using it to catch significant distribution shift from the historical population risk profile as well as data discrepancies.
- We've established thresholds for Green-Yellow-Red status. If a variable is in the Yellow or Red area, we make further analysis and dig down deeper to understand what happened.
- This test in conjunction with the basic descriptive statistics allows us to be proactive in our data quality problems detection and saving from a couple of days to a couple of weeks effort to address the issue.
- A sample code is found on next slide.

1: Kolmogorov Smirnov (Max-KS) in Banking Credit Risk Data Quality Control - Mark An, Senior Quantitative Analyst, CIBC Risk Management (Spring 2010)

Population stability (2)

```
/*  DATASET: dataset to analyze KS
    POP_FLAG: variable to determine if the observation is part of the
                          development population (POP_FLAG=1) or part of the current
                          population (POP_FLAG=0)
    VECTOR: list of variables to calculate KS on
*/


%macro KS (DATASET=,POP_FLAG=,VECTOR=);
/*Logistic regression to find the best distribution model*/
proc logistic data=&DATASET.;
    model &POP_FLAG. (event='1')=&VECTOR./ selection=stepwise;
    output out=outp_prob p=probability;
run;


/*KS calculation*/
proc npar1way data=outp_prob noprint;
    class &POP_FLAG.;
    var probability;
    output out=npar_an;
run;


/*Output to global variable "KS" for later use*/
data _null_;
    set npar_an;
    output;
    call symput("KS",round(_D_*100,.001));
run;
/*...*/
%mend;
```

The dataset has been filtered before (product type, portfolio, date, …)

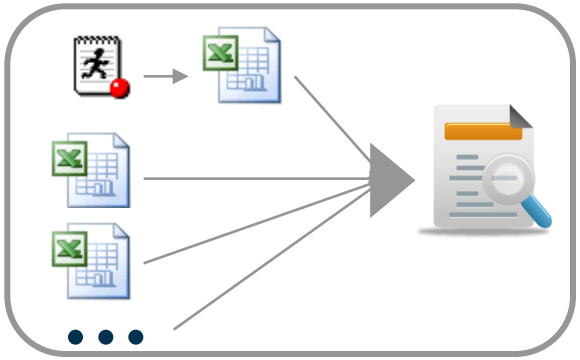
Very simple regression:
More options can be used if wanted

This is where the comparison between historical and current data occurs

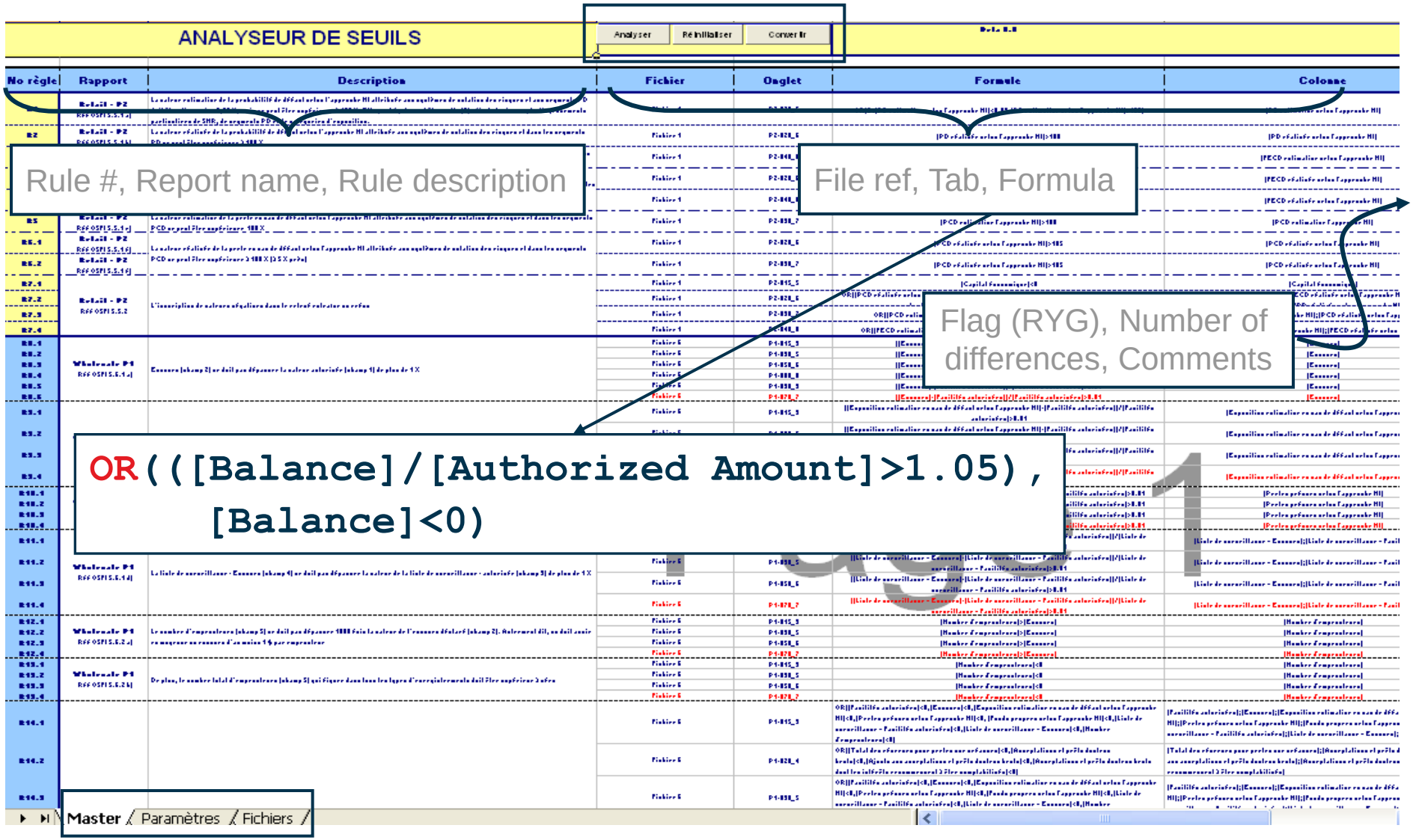
We store the result in a global variable so that we can use it later in another dataset

Business rules

- The preceding tests were statistically-oriented, but sometimes, we have to validate some business logics in the data itself as well as making cross-validation of multiple reports.
- To help us deal with that, we used a mix of SAS and VBA: here comes the “Analyseur de seuils” (Threshold Analysis Tool).
- How does it work?
  1. Produce SAS reports and output it in Excel (and receive other reports such as CSV files);
  2. Adjust business rules and tool parameters;
  3. Execute VBA macro;
  4. Look at the results!



Business rules



NATIONAL BANK FINANCIAL GROUP

Benefits and limitations

- The benefits obtained from this simple solution are numerous, but here are the main ones:
  - Proactivity: we can now assess data quality before regulatory production cycle starts;
  - Adaptability: it's easy to add a new variable or type of control within SAS code;
  - Fast response time: it takes a couple of minutes to get statistics on every fields we want (currently more than a hundred variables tested on a weekly/monthly basis).
- Unfortunately, there are some limitations, but we can live with them (for now!):
  - Not as proactive as we would like: since we don't have access to real time information on all customers, we have to wait before analyzing data;
  - Most business sectors are still reporting using Excel: we have to adapt our SAS code more often then we would like to.

Questions

?

NATIONAL BANK FINANCIAL GROUP

Thank You

- Alex.Salvas@bnc.ca

NATIONAL BANK FINANCIAL GROUP