Deloitte.

RISK dive

Our approach for support in project’s development process

Project overview and management approach

RiskDive is a R&D project which consists of several parallel independent streams that have to be connected into single IT solution at the end

Key project streams:

- Generic calculation logic
- Machine learning
- Back-end development
- Front-end development

Main risks:

- R&D project implies high level of uncertainty in project requirements and its frequent changes based on results of testing and deliverables from each stream
- It's hard to provide reasonable time/effort estimations due to high fluctuation and frequent changes of requirements
- Mentioned issues lead to risk of mismatch in timeline of integration of deliverables from different streams into single solution

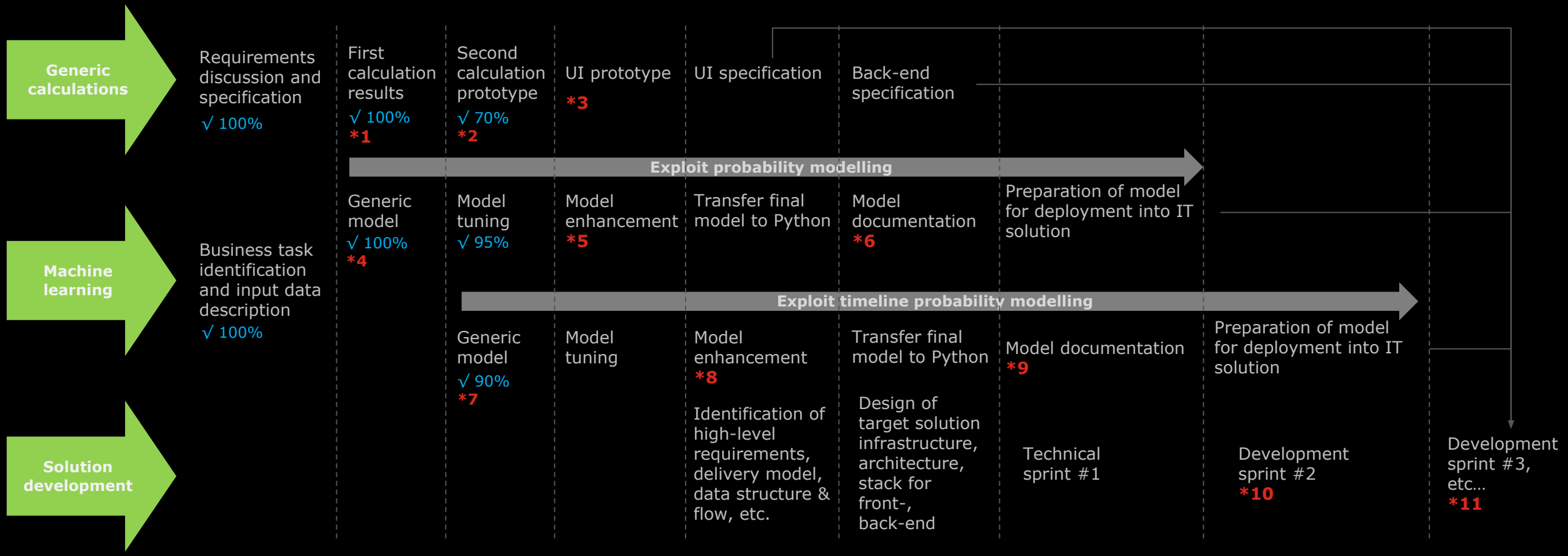
Our approach:

- Start each stream with MVP (minimum value product) and enhance it with additional functionality within small iterations (sprints) of adding pieces of new functionality
- Define / agree requirements for each new sprint at the sprint beginning and test / accept results at the end of each sprint
- Estimations will be provided for each stream and charged on time & material basis

Expected advantages:

- Iterative approach allows to get first results for testing and specification of requirements ASAP.
- Each next sprint will be planned taking into account the results and current status of all other streams.
- DTT Canada team will track and approve efforts and results on regular basis (in-line with sprints duration)
- Short-term sprints reduces risk of wasted resources due to unclear specification or unacceptable results based on testing outcome
- Sprint-by-sprint planning allows to focus only on current critical functionality instead of planning the whole scope, which will be probably changed due to specific of R&D

Overall project's streams – our current understanding



Key deliverables and decision points:

*1. Excel file with assets and attacks risk score

*2. Interactive report with details of risk scoring calculation steps

*3. IT-prototype with reactive response to changes & automatic calculation

*4/7. Estimation of possible model accuracy

*5/8. Final model

*6/9. Description of modelling process

*10. MVP with first set of functionality

*11. RiskDive IT-solution

Generic calculation logic with back-, front-end streams

Each step is aimed to get MVP, to test results and adjust / clarify requirements for further steps ASAP.
Sprint duration – 1 week, one task might be done during one sprint ore more than one sprint

1. Requirements discussion and specification

√ Overall requirements (user-story)

2. First calculation results:

- Excel (if applicable)
- Calculation logic with R with saving results to Excel for testing

√ First calculation results

3. Second calculation prototype:

- R+Markdown report

√ Reproducible calculations on big data amounts & interactive charts

4. UI prototype:

- R+shiny dashboard

√ UI prototype & reactive components for online testing

5.1. UI specification:

- Sketchboard + MS Word

√ Tested requirements for IT development stream

5.2. Back-end specification:

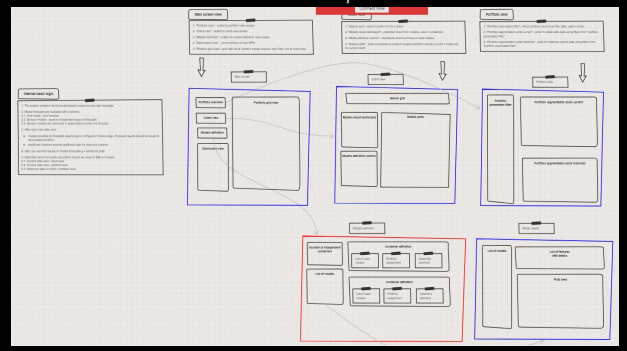
- MS Word

√ Tested requirements for IT development stream

We use **R** as one of the fastest ways to develop comprehensive calculations based on significant amount of data and implement ML or specific calculations (e.g. automatically create network connections map or all possible attack routs)

We use **R+Markdown** for developing of interactive reports (html) which reuse already developed R functionality and provide interactive data visualization and conditional description

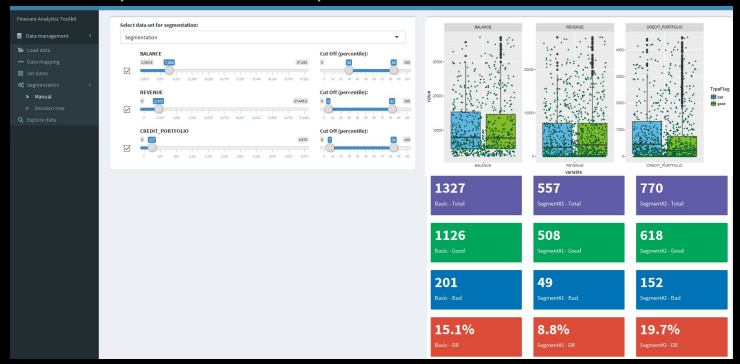
Sketchboard illustrative example:



We use **R+shiny** for prototyping of IT solution with reuse of already developed R+Markdown. R+shiny prototypes provides data reactivity, so we can change system parameters and get calculation output in online mode

We use **Sketchboard** as fast approach for visualization and specification of mockups for graphical interfaces

R+Shiny illustrative example:



Machine learning stream

Each step is aimed to get first ML model's performance estimation ASAP and enrich the model (try to increase accuracy) step-by-step
Sprint duration – 1 week, one task might be done during one sprint ore more than one sprint

1. Business task identification and input data discussion

√ Model target identified

√ Definition of done (target metric) agreed

2. Generic model:

- Data pre-processing & feature engineering
- Basic model development in R

√ First model results

3. Model tuning:

- Model tuning to get close to target metric

√ Tuned model with higher accuracy

3.1. Model enhancement:

- Options for potential accuracy improvement (DTT UA provide options -> DTT CA approves next steps)

√ Improved model

4. Transfer final model to Python

√ Final modelling pipeline developed in Python

4.1. Model documentation:

- MS Word / Jupiter Notebook / Markdown

√ Description of model development process and results

5. Preparation of model for deployment into IT solution:

- Compress developed model object
- Develop processing functions: input data validation, errors catch & processing, prediction function, output functions.

√ ML model is ready for integration into IT solution

We use **R** as one of the fasted environments for model development and tuning.

We use **Python** as target environment for serializing and optimizing ML models for production usage

IT solution development

Each stream is aimed to focus only on current most valuable functionality and develop the solution step-by-step with flexibility to cover changes during project
Sprint duration – 2 weeks

Approach of iterative development allows DTT Canada to:

- Get first result (MVP) in the shortest terms
- Keep an eye on progress of development stream
- Get an opportunity for flexible changes in requirements stream-by-stream
- Test delivered functionality with iterative enhancements

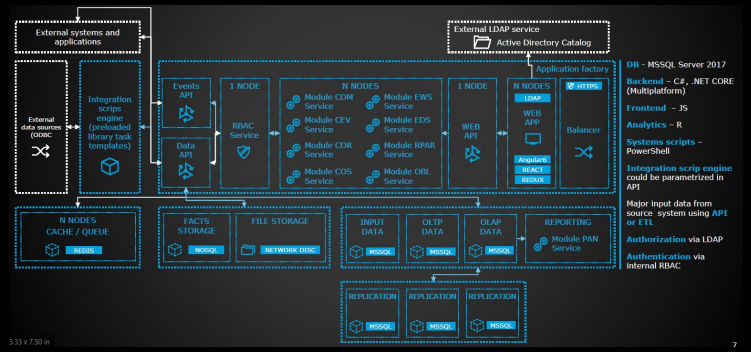
1. Identification of high-level requirements, delivery model, data structure & flow, etc.

√ Overall requirements

2. Design of target solution infrastructure, architecture, stack for front-, back-end

√ High level design

High-level IT architecture illustrative example:



3.1. Sprint planning

- DTT CA takes part in tasks prioritization and estimation for current sprint

√ Estimation and planning for current sprint

3.2. Development

- Development of sprint tasks with daily tracking

3.3. Results presentation:

- DTT UA presents results of last sprint to DTT CA team

√ MVP with enhanced functionality

Example of next sprint planning & estimation

Sprint #1	Start:	10.12.2018	End:	14.12.2018			
Task#	**Stream**	**Task / delivery format**	**Estimation**	**Assignee**	**Grade**	**RPH**	**Cost, $**
1	Generic	Implement calculation logic of iterative exclusion of each CVE with recalculation of overall network risk-score for prioritization of CVE's for remediation plan (R)	3 MDs / 24 hrs	Potseluev, Vasyl (UA - Kyiv)	C	50	1 200,00
2	Generic	Take into account direct connection of a device to Internet (R)	0.2 MDs / 1.5 hrs	Potseluev, Vasyl (UA - Kyiv)	C	50	75,00
3	Generic	Change calculation logic to reflect the case when attack should stop after asset with zero risk-score (R)	0.4 MDs / 3 hrs	Potseluev, Vasyl (UA - Kyiv)	C	50	150,00
4	Generic	Enhance scoring calculation report (R+Markdown / html)	1 MDs / 8 hrs	Skrypin, Mykola (UA - Kyiv)	M	85	680,00
5	ML	Exploit probability model – maximize model fitting to acceptance criteria (Python)	0.5 MDs / 4 hrs	Skrypin, Mykola (UA - Kyiv)	M	85	340,00
6	ML	Prepare options for Exploit probability model enhancement (MS Word)	0.2 MDs / 1.5 hrs	Skrypin, Mykola (UA - Kyiv)	M	85	127,00
7	ML	Exploit timeline model – finalize data processing, feature engineering and generic modeling pipeline. Train baseline model to get first estimation of AUC & precision (email)	4 MDs / 32 hrs	Biliachenko, Yuliia (UA - Kyiv)	C	50	1 600,00
8	IT development	Prepare questionnaire regarding high-level requirements, delivery model, data structure & flow to identify key requirement for the solution (Excel)	0.5 MDs / 4 hrs	Skrypin, Mykola (UA - Kyiv)	M	85	340,00
		Total:	**9.75 MDs / 78 hrs**			**Total:**	4 512, 00