EXPERIAN
INNOVATION
SUMMIT
2019

DATA ENGINEERING : NOUVEAU MOTEUR DE CROISSANCE

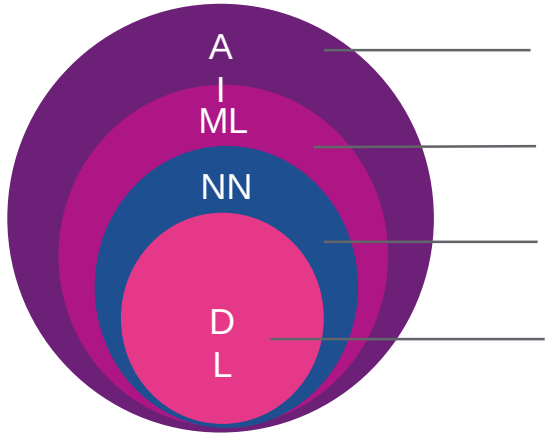
DATATECHNOLOGYANALYTICS

experian

Nouvelles générations de Machine Learning au service de l'analyse réglementaire

Stefan Stoyanov & Gerrit Bekker
Experian

AI evolution
ML, programming, technology



Artificial Intelligence

Machine Learning

Neural Networks

Deep Learning

How to do it	Where you build
Programming Languages python R julia Scala sas	Platforms for Machine Learning Ascend Microsoft ML Amazon Web Services Open Python Open R H2O Tensor Flow

1. Artificial Intelligence is a technique which enables machines to mimic human behavior.
2. Machine Learning is a subset of Artificial Intelligence which uses statistical methods to enable machines to improve with experience
3. Machine learning models goal: best model for the available data and prediction problem
4. Statistical models goal: explain and formalise relationships between variables

experian

ML for regulatory models

Machine learning models are still not accepted by regulators since they are viewed as black boxes and there is lack of strong validation and model governance frameworks

Transparency, **explainability, replicability, monotonicity constrains** and **validation** of machine learning models become increasingly important for their acceptance by regulators

Model Explanation
(PDP, ICE, LIME, LOCO, SHAP)

Traceable

Verification

Reproducibility, e.g. all data stored, model version controlled and re-run capabilities

Validation

Ensure the model performs as expected on out of sample data

Consistency

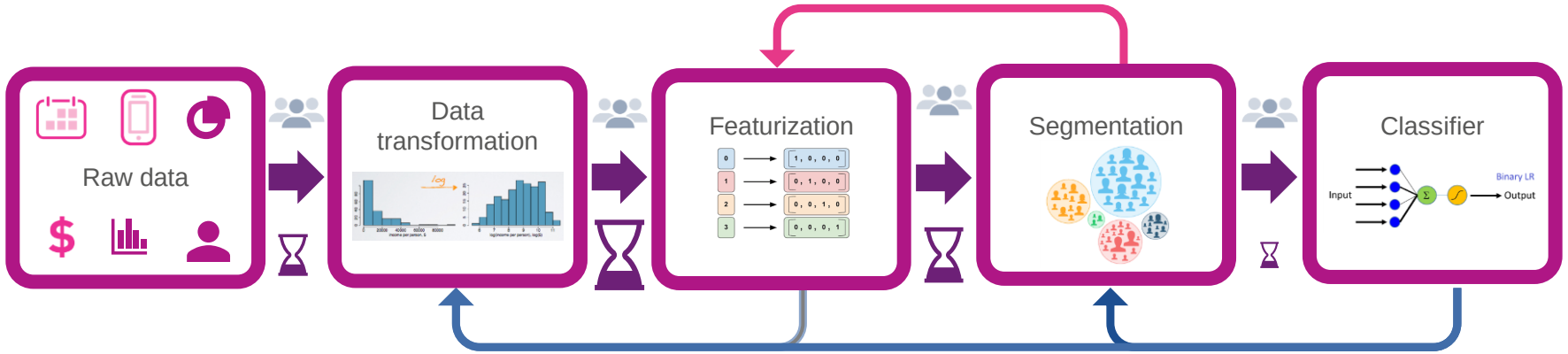
Simulate your decision strategy to see if it delivers *expected* outcomes

+

On-going controls

Reporting, traceability, model performance and updates

Our approach to modelling and deployment



Supplement

Keep existing framework

Looking to improve traditional model performance (speed and/or accuracy)

Use ML "behind-the-scenes" to inform traditional approach (segmentation, feature generation, feature selection)

Benchmark

Not yet ready to go to full ML framework

Champion/Challenger approach. Run an ML model alongside a traditional one

Focus on model performance and stability to measure potential benefits of switching

Verify model transparency and traceability

Replace

Replace traditional model with ML (There is infrastructure and stakeholders and regulators buy-in)

Focus on model performance, deployment, on-going validation and regular model updates

experian

Three areas of application of ML models

Alternative data

- Web, transactional, geo and voice data
- Credit granting, EWS, affordability, forbearance and collection models
- Fully transparent
- Easy implementation
- Easy to incorporate into traditional models

Segmentation

- Data-driven
- IFRS9, Basel, credit granting, NPL management
- Highly customisable
- Flexible handling of different types of data formats

Risk models

- Origination, behavioral models
- Accurate
- Fast
- Scalable
- Explainable

Alternative data
Web data - performance

90% of Web Hit Rate

Fresh Data & Wider Insights
9 out of 10 companies present in the web

+20% Gini Traditional Scores

Boosting Traditional Scores
With web data you can augment traditional scores

30+% Gini using only Web Data

Mitigate Credit Risk
Reduce Credit Risk of clients' portfolio of companies

Response <1sec

Real Time Solution
All data processing and score calculation in less than 1 second

experian

Transactional data
Risk and customer insights

- Improve PD Model performance by 28%
- Decrease Credit Losses by 36%
- Improve EWS models performance by 20%
- Improve income estimation accuracy by 50%
- Increase in credit cards spend by 35%

Credit risk data
PD models

Project 1: Sweden
Portfolio: Credit cards

Dataset summary:

- 112k records, 100 variables

Gini index:

- **Scorecard: 0.852**
- XGBoost models:
  - **All variables, unconstrained: 0.883**
  - **All variables, monotonic: 0.873**
  - **Scorecard variables only, unconstrained: 0.856**
  - **Scorecard variables only, monotonic: 0.866**

Speed:

- Scorecard: 7 days
- XGBoost model (all variables): 1-2 days, on laptop

Project 2: Norway
Portfolio: Credit cards

Dataset summary:

- 265k records, 375 variables

Gini index:

- **Scorecard: 0.886**
- XGBoost models:
  - **All variables, unconstrained: 0.891**
  - **All variables, monotonic: 0.911**
  - **Scorecard variables only, unconstrained: 0.886**
  - **Scorecard variables only, monotonic: 0.889**

Speed:

- Scorecard: 10 days
- XGBoost model (all variables): 1-2 days, on laptop

Our way to open the Black Box

Justify

- Transparent, explainable and interpretable models
- Justify modelling choices to the regulator
- Alignment with regulation such as GDPR, Right To Explanation, NPL management, Loan origination and monitoring, IFRS9, Capital Requirements Directive, etc.

Improve

- On-going improvements of the ML algorithms and techniques to infer feature importance and model interpretability
- Apply business and domain knowledge to check and enhance ML prediction. Uncover true drivers and avoid spurious correlation, biases and overfitting

Adopt

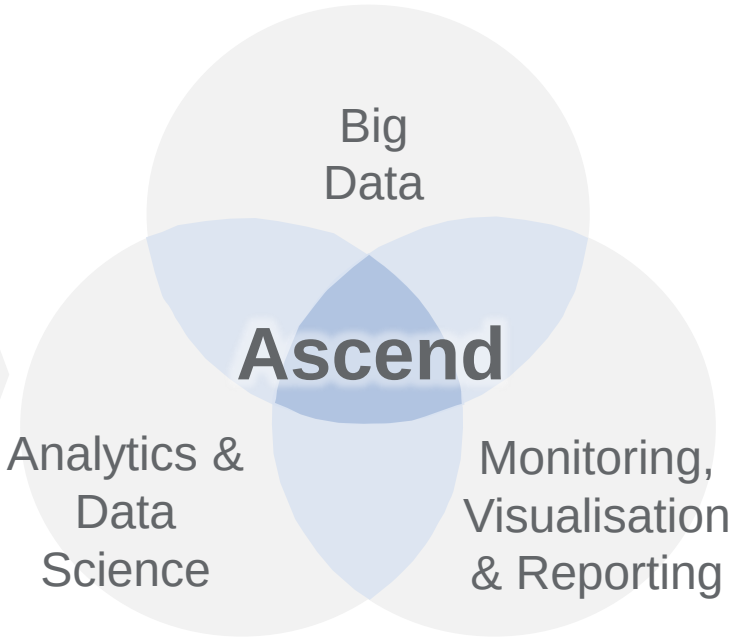
- Validation, MRM and governance frameworks
- Link ML outcomes to business results and actionable insights
- Flexible implementation options

Ascend Analytical Sandbox

experian

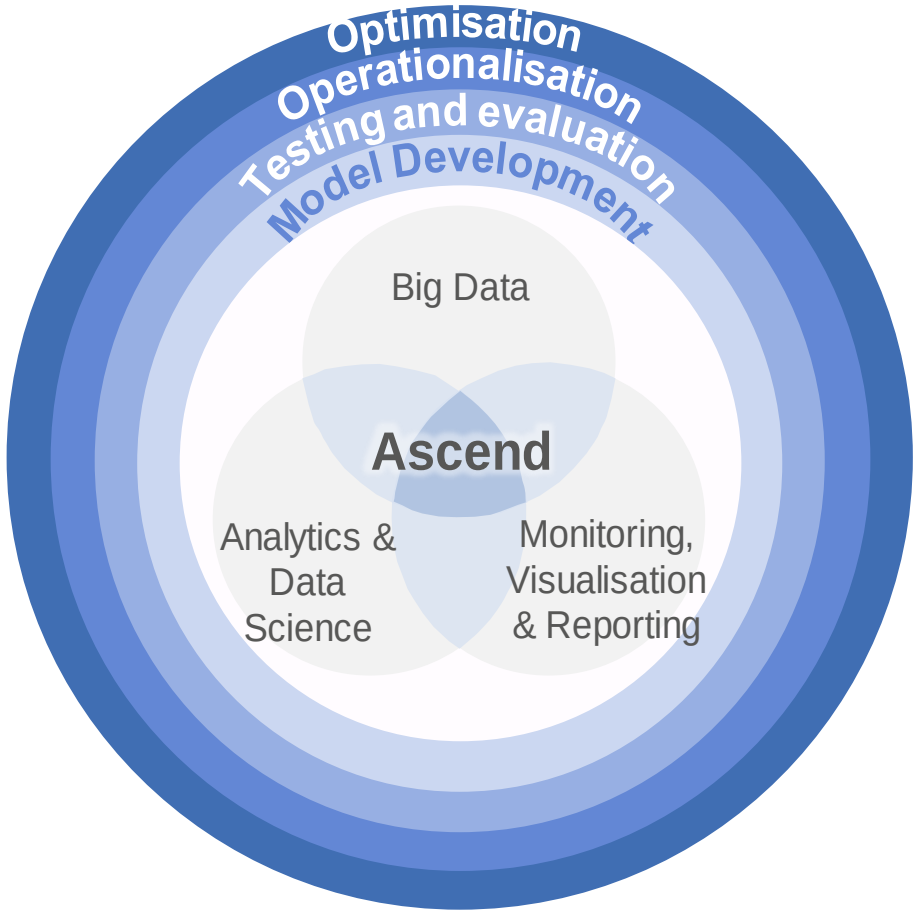
These services are being **further strengthened**

- Business and profitability focussed risk transformation
- Regulatory, economic & data science expertise
- Traditional, non-traditional, and historic data and blended data insights
- Implementation and ongoing evaluation



experian

Introducing Ascend “Analytics on demand”
Three fully integrated secure environments



Big Data

- Manipulate data
- Perform data quality checking
- Automate uploading, aggregation and extraction of data
- Access to Experian data, Bureau data (current and historic), In house data, new non-traditional data

Analytics & Data Science

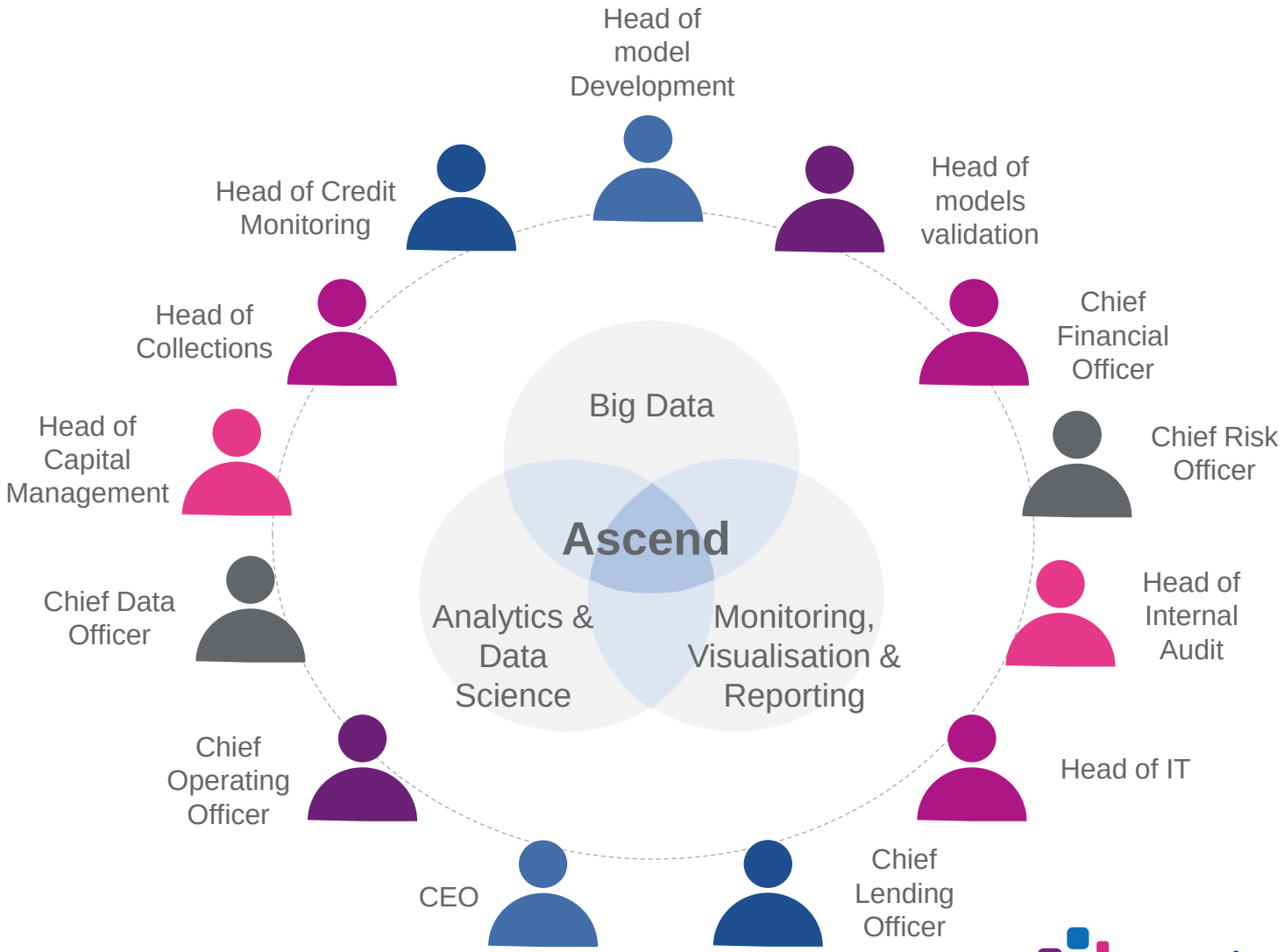
- Develop and manage multiple analytics projects.
- Develop machine learning models using Python, R or Scala
- Perform data engineering tasks connected directly with HUE data strage environment

Monitoring and reporting

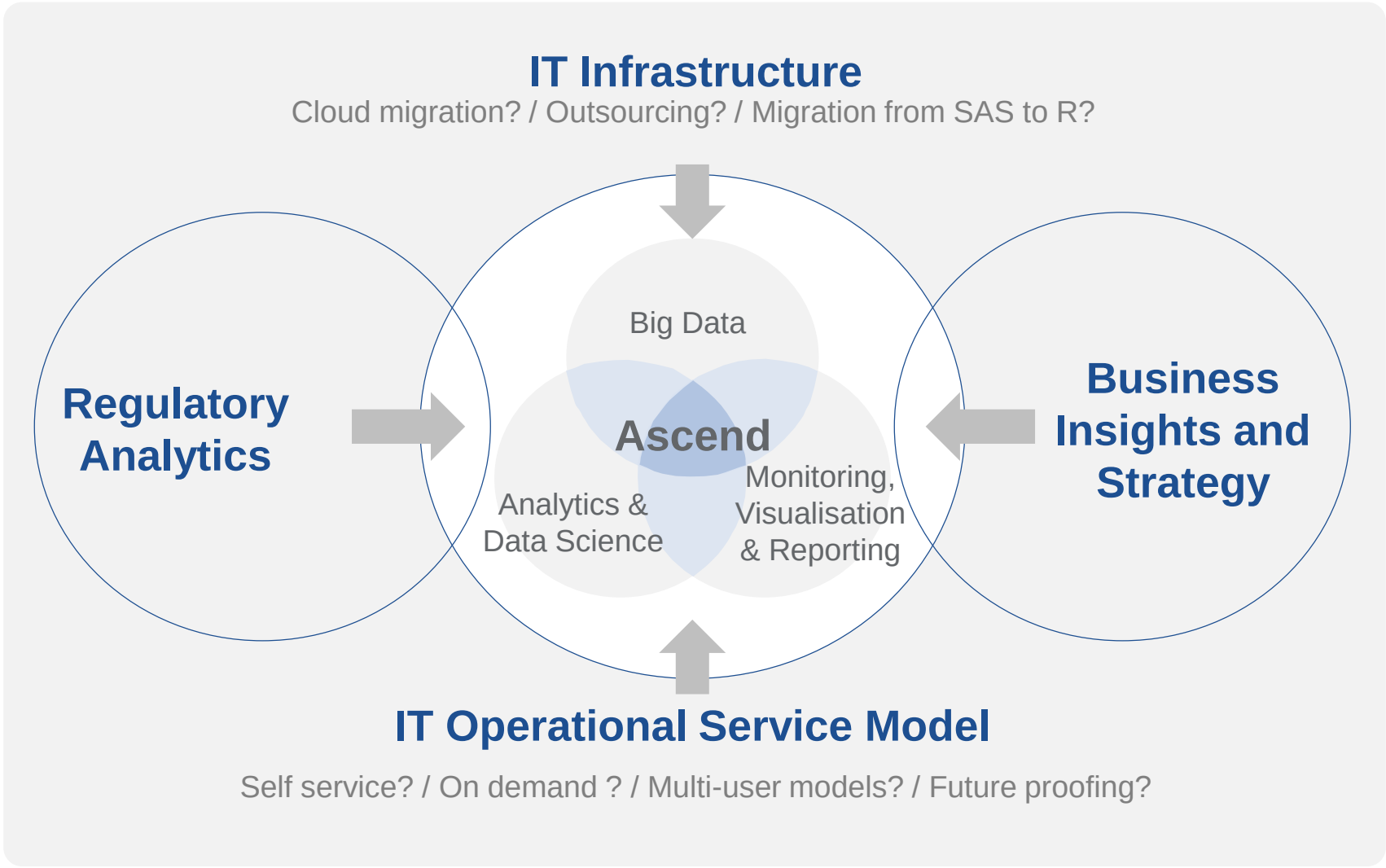
- Create, share and export of reports directly from source data
- Prepare powerful, interactive dashboards, charts and other visualisations

Helps BUILD understanding and REMOVE silos

- Self service empowers all stakeholders across the enterprise
- Multi-user access to source data
- Aligned behaviours
- Promotes transparency and understanding



Helps OPTIMISE capabilities and value



Business and profitability focussed risk transformation

- Regulatory insights and models support day to day business decisions
- More accurate alignment of capital with corporate strategy
- Improved risk management and more profitable growth

experian

The additional benefits created by Ascend

Cutting IT Costs

- Outsourced alternative to costly in house build
- No CAPEX requirements

Accelerating Delivery

- Estimating 50% saving in time to deliver models
- Cutting time for manual reports

Cutting costs with Migration from SAS

- Remove costs of SAS licenses
- Access Ascend R / Python Toolbox

Improving integration of Functions and Department:

- Single platform connecting
  - Model development
  - Model Validation
  - Risk Management

Cut costs of Machine Learning Migration

- Self service access to Machine Learning tools
- Avoid hiring new data scientists

Our Ascend roadmap covers each of these areas

- Internal development programme
  - **7 identified modules**
  - **Bespoke Templates**
  - **Supporting functionality**
- Ongoing evolution
  - **Additional capabilities will be identified and added over time**

EXPERIAN
INNOVATION
SUMMIT
2019

DATA ENGINEERING : NOUVEAU MOTEUR DE CROISSANCE

DATATECHNOLOGYANALYTICS

experian