Datalab : Enjeux et Bonnes Pratiques

Dataiku, éditeur de Dataiku DSS
plateforme collaborative pour « advanced analytics »

Pays-Bas

France · · Allemagne

· Italie

· États-Unis

Espagne

· Indonésie

Datalabs, Enjeux et Bonnes Pratiques

data iku

Dataiku en quelques dates

Mars 2016
Dataiku DSS 3
Première plateforme intégrée pour le développement et la production d'applications de Data Science

Avril 2015
Dataiku DSS 2
Ajout de fonctionnalités de collaboration et de real-time

Février 2014
Dataiku DSS 1
Premier outil de préparation de données intégrée et de machine learning

Janvier 2016
55 employés
Ouverture d'un bureau dans la Bay Area
Croissance annuelle de 200%

Janvier 2015
20 employés
30 clients
Levée de fonds de 3M€
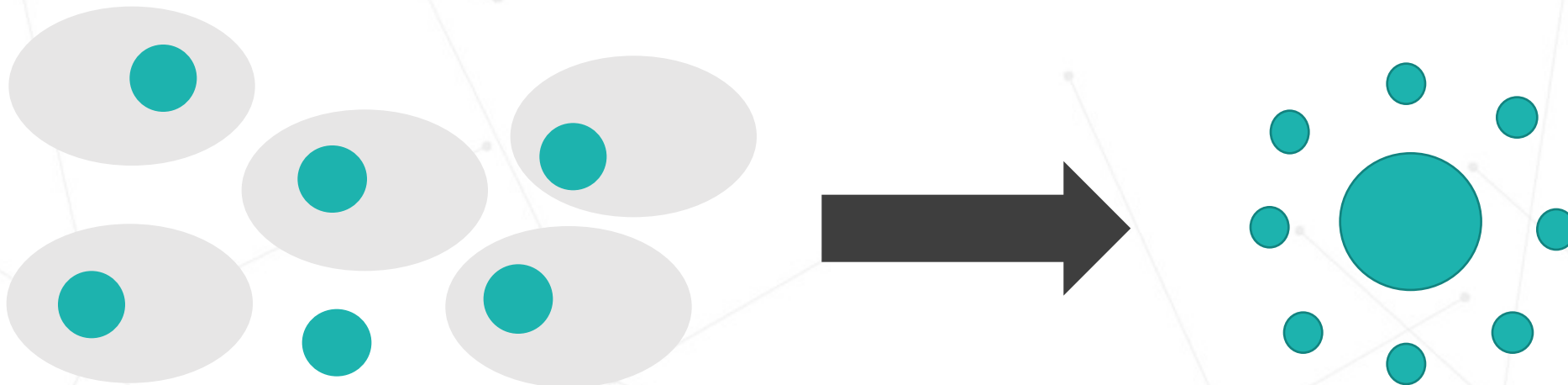
Avril 2015
Ouverture d'un bureau à New York

data iku

Le Datalab : un nouveau must-have dans les entreprises

Datalab : une organisation transversale qui développe des fonctionnalités analytiques pour mettre au point des produits ou services pilotés par la donnée.
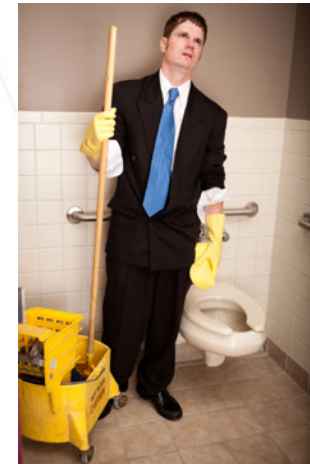
Datalabs, Enjeux et Bonnes Pratiques

Savoir-Faire, Outils et Usages

According to the report, the median salary for a Data Scientist is an impressive $116,000 and there are over 1,700 job openings. For those curious, a "data scientist" typically refers to a mix of skills, part statistician and part computer programmer. For instance, data

Data Scientist vs Data Cleaner?

Quels talents dans mon Datalab? Pour quelles missions et quels modes de management?

dataiku

Savoir-Faire, Outils et Usages

The Python Clan

VS

The R Tribe

VS

The SAS Empire

The Old Elephant Fraternity

VS

The New Elephant Club

Un Datalab polyglotte ou dictateur ?

Datalabs, Enjeux et Bonnes Pratiques

data iku

Savoir-Faire, Outils et Usages

Churn
Segmentation
Maintenance
Risque
Fraude
Etc.

Quels objectifs (courts et moyens termes) et quels KPIs pour mon Datalab?

Datalabs, Enjeux et Bonnes Pratiques

dataiku

Bonnes pratiques du Datalab en entreprise

Elias Baltassis
Directeur Europe Data Analytics,
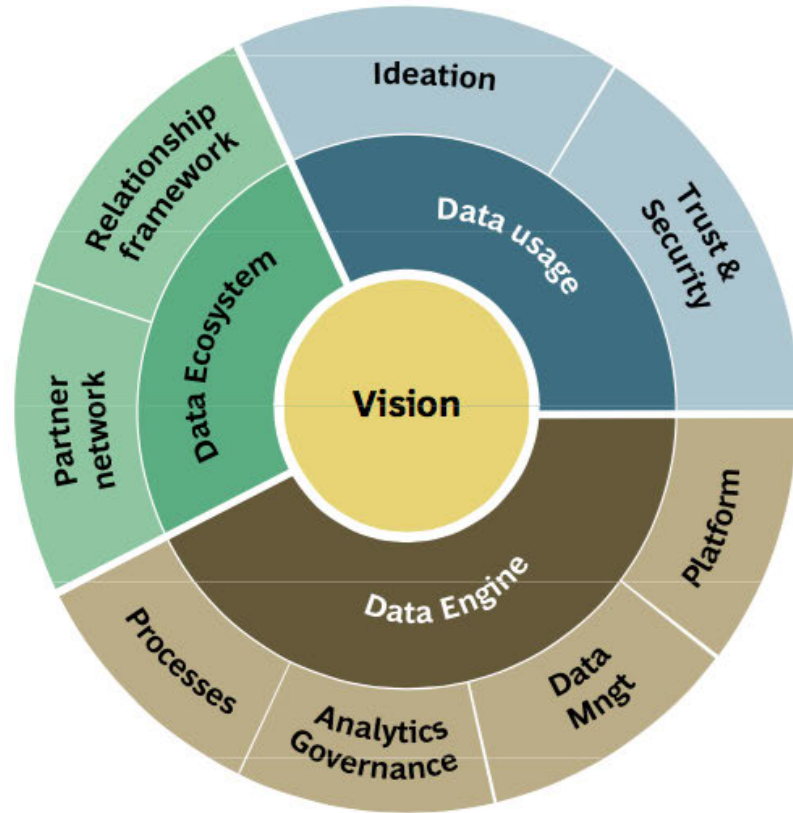*Boston Consulting Group*

Agenda

Background

Observed market practices

A few closing remarks

Appendix

BCG on Data Labs - short - vFF DIFFUSEE.pptx

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data
iku

Many capabilities are necessary for being efficient in Big Data

BCG's Data & Analytics Capabilities Framework

Domains

Vision
What is our vision of the data & analytics role in our **business model**? What is the impact on **value creation**?

Data Usage
How do we generate and manage **new ideas**? How do we **secure** data? Do we use customers **Trust** as a key competitive differentiator?

Data Engine
What are the **key building blocks** (technology and people) of an efficient data engine? What is the **best operating model for us**?

Data Ecosystem
What is the **importance** of a data ecosystem for us? What is the **optimum strategy** for building it? What **role** should we play in it?

Few of the companies we serve use them to their full capacity

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data iku

Market context

Most companies have entered the Analytics field and work in testing analytical applications in pilot projects (POCs), or already leveraging data for improving their operations

Hence, most companies employ "data scientists"; sometimes in centralized / single teams, more often spread in various units of the organization

However, very few companies seem to have find an efficient way for organizing their analytics organizations

Our 2016 Big Data Maturity Survey found that the overall maturity of companies (across all Big Data capabilities) stands at 2.7

- **Analytics Governance comes across, at first sight, as one of the most mature capabilities, with an average score of 2.9**
- **However, Analytics Governance has also one of the largest spreads: from 1.5 to 4**

**Many companies work on Analytics,
not all companies have found how to do so effectively**

Source: BCG case experience & research
BCG on Data Labs - short - vFF DIFFUSEE.pptx

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data
iku

Agenda

Background

Observed market practices

A few closing remarks

Appendix

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data
iku

We see four operating models (and severall variances) used in the market today

Maturity →

	"Creative anarchy"	Centralized IT model	"Center of Excellence" (CoE)	Central-platform, local analytics
Operating model elements	• Minimal governance • No consistent technology use: "whatever works" • Limited support; high expertise	• Centralized IT funding and rigid governance • Limited computer science expertise	• Small retained team with specialist expertise • Separate experimental and production environments	• Near-best-in-class central IT platform allowing controlled local access to data • BUs encouraged to innovate • Ability to share innovation
Drivers, key benefits	• Ability to experiment • Businesses control design & operations • Fast, agile set-up	• Leverages IT scale • Standardized governance limits duplication • Tight cost control	• Best use of scarce resource • Gives IT time & benchmarks to build corporate platform	• Minimum duplication • Good control over data quality • Access to wide data
Typical challenges	• Difficult to scale; not easily replicable across BUs • Duplicated platforms & data	• Limited experimental agility • Slow: takes longer to approve projects, set-up experiments	• Capacity issues possible; unmet spikes in demand • Rapidly challenge to support across all use cases asked	• Technologies that few companies master – for now • Requires significant IT effort and investment – **not likely to be a starting point**
Example companies	*Most "new in analytics" companies* Google	*Several EU utilities* Walmart	PHILIPS P&G at&t intel *Several banks*	VISA *Global F I* TESCO *Major US bank*

THE BOSTON CONSULTING GROUP

Draft—for discussion only

Mardi 29 Mars 2016

data iku

We see an increasing use of the "Center of Excellence" (CoE) model in very diverse companies

Selected examples

Global Card Issuer

- Central Binitiative being led by CRO & CMO
- Built shared platform for data access on Hadoop stack
- Several projects implemented using Advanced Analytics approaches; very large pipeline of ongoing tests

Global Technology Co

- Central Big Data initiative being led by CIO
- Looking for opportunities to blend existing data with new
- Built extensive data scientists group
- Deploying projects to drive value in engineering & sales

European Media Group

- Central team of 20 data scientists + ~50 decentralized, reporting to CEO
- Deliver projects across group
- Led by leader with reputation of being 'maverick'
- Dedicated IT resources to build platform capability

Leading UK Bank

- Started in "creative anarchy" mode, but soon moved to a CoE model reporting to COO
- Single, virtual Analytics team, covering Risk, Marketing and Distribution
- Advanced implementation of Hadoop stack

Leading European Bank

- Moved to Hub & Spoke CoE model to rationalize disparate efforts and create synergies
- Commercial focus only, reporting to COO; risk analytics remain separate, reporting to CRO
- Implemented hybrid DWH (Traditional+Hadoop)

US Software Co

- CIO leading Big Data CoE on behalf of Executive Committee
- Focus on linking disparate data sources (large part: public or open data) together to unlock value on Hadoop
- Mix of internal and out-sourced analytics resources

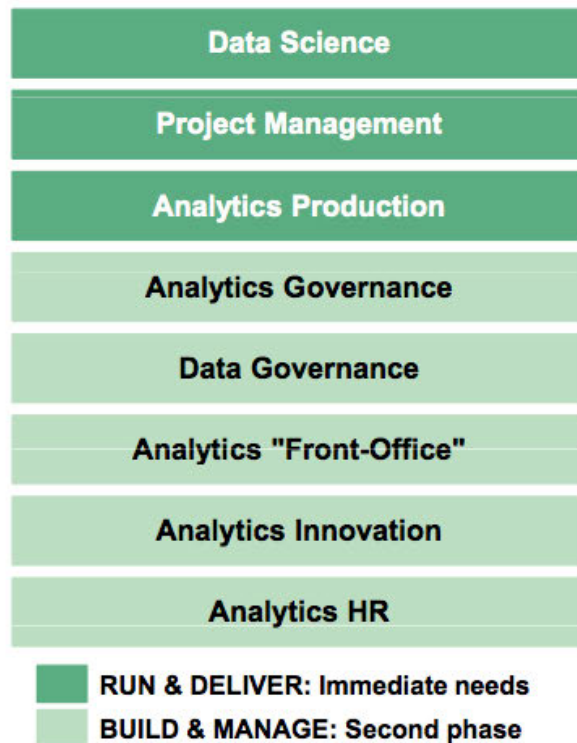**Common goal: operationalize and scale-up often sub-optimally coordinated groups**

Source: BCG case experience & interviews
BCG on Data Labs - short - vFF DIFFUSEE.pptx

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data iku

6

However, the CoE can be implemented in more than one way – each with its own pros and cons

	Virtual CoE	**Hybrid CoE**	**Centralized CoE**
Org. structure options	*Multiple analytics groups – horizontal coordination*	*Central CoE coordination, mainly decentralized resources*	*Single, central analytics function*
Typically seen	• Standalone BUs with very different analytics requirements (real or perceived)	• Emerging analytics capabilities spread across BUs • Need to evolve through centralized standards and guidance	• Mature organization, with good grasp on data • One BU as dominant user of analytics
Advantages	• Direct Business control over resources and priorities • Content expertise aligned through "virtual council"	• Scalable model • Content expertise aligned through central action • Enterprise capabilities build-up enabled	• Scalable model • Efficient resource (people & platform) management and deployment
Challenges	• Difficult to scale model • Little synergies potentially leading to higher cost & duplicated efforts	• Increased complexity and frequent need for alignment	• Siloed capabilities • Smaller BUs sub-optimally served

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data
iku

Beyond operating models, mature Analytics groups...

... typically fulfill eight analytics functions...

Data Science
Project Management
Analytics Production
Analytics Governance
Data Governance
Analytics "Front-Office"
Analytics Innovation
Analytics HR

- **RUN & DELIVER:** Immediate needs
- **BUILD & MANAGE:** Second phase

... and work along a 7-step analytics process

1. Idea generation & prioritization
2. Data gathering
3. Modeling data and preparation
4. Data modeling & testing
5. Model validation
6. Industrialization
7. Monitoring & maintenance
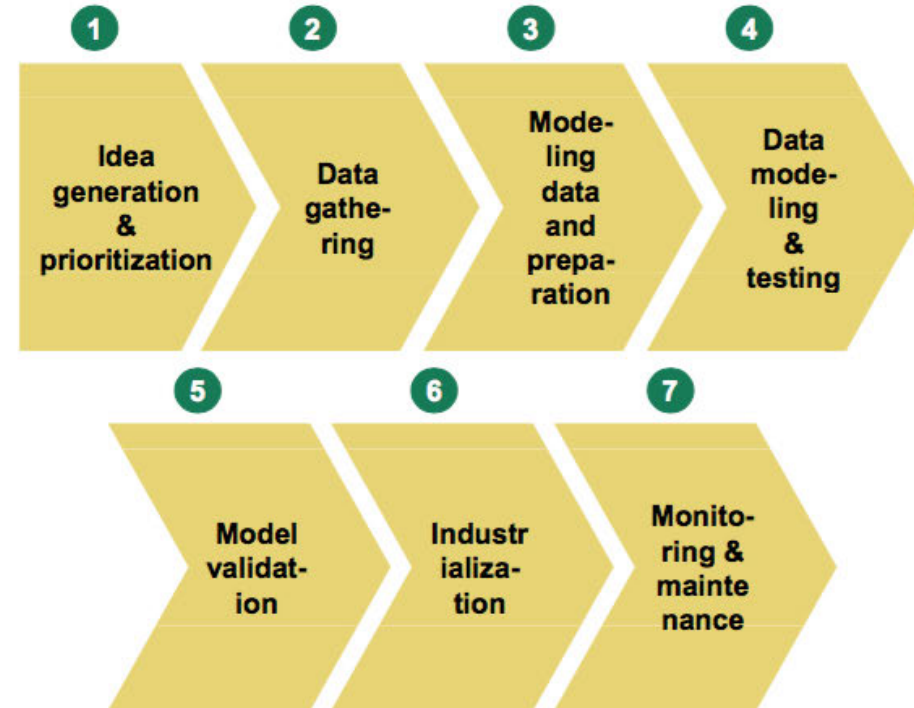
These functions and process steps are independent of the operating model and the size of the team

data iku

As the Analytics organization matures, we see resources moving from one-size-fits-all to specialization

Early Stage

	Data scientist	Data analysts
Data Science	✓	
Analytics Production	✓	✓
Project Management	✓	
Data Governance		✓
Analytics HR	✓	
Analytics Innovation	✓	
Analytics "Front-Office"		✓
Analytics Governance	✓	

Intermediate organization structures

Mature Analytics Organization

	Data Modeller	Data analysts	Project Manager	Analytics Officer	Analytics HR special.
Data Science	✓				
Analytics Production		✓			
Project Management			✓		
Data Governance				✓	
Analytics HR				✓	✓
Analytics Innovation	✓				
Analytics "Front-Office"		✓			
Analytics Governance	✓			✓	

Interacting with CDO

Source: BCG case experience & research
BCG on Data Labs - short - vFF DIFFUSEE.pptx

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data iku

We see **five key profiles** be present in most mature analytics organizations

Data scientists

Senior Modeler	Junior Modeler	Project Manager	Data Analyst	BI Specialists
Build models and perform explanatory analysis • Use specialized languages or tools • E.g. Python, R, Scala, SAS, SPSS-M **Participate to data testing and modeling** **Employ advanced analytics techniques and methods** **Validate analytics models** • Peer validation **Provide R&D support to Junior modelers**	**Prepare data sets for an efficient use in data modeling** **Assist in developing models and, later, in their implementation** **Provide indirect support in setting new data platform**	**Manage end-to-end analytics projects** • Interface with business users, data scientists, analysts and data architects • Ensure fluid workflow and adequate logistics and tools for the data project • Define efficient planning of tasks and monitor it	**Gather raw data from data bases – internal or external** • E.g. use SQL language on DB2 or NoSQL on Mongo DB **Interface with data management resources** • Aware of data management policies and processes	**Gather structured data from data bases or applications** **Use BI tools** (e.g. Spotfire, SAP, Cognos) **to produce BI analysis** • Reporting • Simple pattern analysis • Data visualization

Different views exist on BI vs Data Science, raising questions such as:
- **Frontier between BI and Data Science ?**
- **Common or separate reporting line ?**
- **Common career path ?**

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data iku

Agenda

Background

Observed market practices

A few closing remarks

Appendix

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data
iku

Closing remarks

Lessons learned from 300+ Big Data projects (and 20+ Data Labs)

- **One-size does not fit all for Data Labs – and internal constraints play often an important role**

- **POCs are important; but make sure that your Data Lab does not absorb all available effort / budget in testing**

- **Do not do "*science for the sake of science*"; but go beyond the "*dictatorship of logistic regression*"; it will bring you insight and differentiation**

- **Advanced analytics approaches are important and the way of the future; but don't underestimate the insights offered by descriptive analytics based on granular data**

- **Do not overestimate the quality of your current data: always check and improve**

- **Data is the oil of the 21st century; but Analytics is the refinery – build insights on your processes, customers and markets... before someone else does it for you**

> **Big Data is a journey; not a one-off trip**

Mardi 29 Mars 2016

THE BOSTON CONSULTING GROUP

Draft—for discussion only

data
iku

Appendix

Elias Baltassis
Director, Data & Analytics - Paris

Elias Baltassis is a Director in the Paris office of The Boston Consulting Group. He specializes in Big Data & Advanced Analytics and leads BCG's European team on the topic. He is a core member of the Technology Advantage and Financial Institutions practices.

Prior to joining BCG, Elias was a founding member and Managing Director of Opera Solutions, a leading pure-player in Big Data. Prior to that, he was a Partner with Bain & Co

Selected recent experience
- Behavioral client segmentation and pricing strategy for a large European insurer
- New generation e-CRM for a leading French bank and a leading EU electronics manufacturer
- Data & Analytics strategy for a global top-2 payment card issuer and a global retailer
- Client Life-Time Value Management for a large UK financial institution
- Fraud detection and management for a leading European insurer
- Analytics governance for two leading G-SIB financial institutions
- New product design strategy for specific behavioral clusters for a global CPG company
- Analytics governance and Data Lab build-up for several banks, 2 retailers and one telco
- Cross- and Up-sell marketing campaign design for a top-3 French universal bank
- Data management for three G-SIB financial institutions

Elias graduated *cum laude* in Economics and Econometrics from the University of Fribourg (Switzerland) and has earned a *magna cum laude* Master in Applied Mathematics and Computer Sciences from IAUF / the Swiss Federal Institute of Technology in Lausanne (EPFL). He also holds an MBA from INSEAD

Contact
baltassis.elias@bcg.com

Sector expertise
- Retail Financial Services
- Payment & Transactions
- Digital pure players
- Media & Telecom

data iku

Disclaimer

The services and materials provided by The Boston Consulting Group (BCG) are subject to BCG's Standard Terms (a copy of which is available upon request) or such other agreement as may have been previously executed by BCG. BCG does not provide legal, accounting, or tax advice. The Client is responsible for obtaining independent advice concerning these matters. This advice may affect the guidance given by BCG. Further, BCG has made no undertaking to update these materials after the date hereof, notwithstanding that such information may become outdated or inaccurate.

The materials contained in this presentation are designed for the sole use by the board of directors or senior management of the Client and solely for the limited purposes described in the presentation. The materials shall not be copied or given to any person or entity other than the Client ("Third Party") without the prior written consent of BCG. These materials serve only as the focus for discussion; they are incomplete without the accompanying oral commentary and may not be relied on as a stand-alone document. Further, Third Parties may not, and it is unreasonable for any Third Party to, rely on these materials for any purpose whatsoever. To the fullest extent permitted by law (and except to the extent otherwise agreed in a signed writing by BCG), BCG shall have no liability whatsoever to any Third Party, and any Third Party hereby waives any rights and claims it may have at any time against BCG with regard to the services, this presentation, or other materials, including the accuracy or completeness thereof. Receipt and review of this document shall be deemed agreement with and consideration for the foregoing.

BCG does not provide fairness opinions or valuations of market transactions, and these materials should not be relied on or construed as such. Further, the financial evaluations, projected market and financial information, and conclusions contained in these materials are based upon standard valuation methodologies, are not definitive forecasts, and are not guaranteed by BCG. BCG has used public and/or confidential data and assumptions provided to BCG by the Client. BCG has not independently verified the data and assumptions used in these analyses. Changes in the underlying data or operating assumptions will clearly impact the analyses and conclusions.
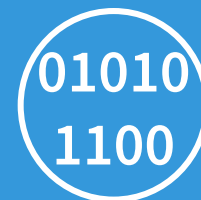
data iku

Datalab : Enjeux et Bonnes Pratiques

Pour Conclure

data
iku

Dataiku, pour toutes vos applications sur la donnée
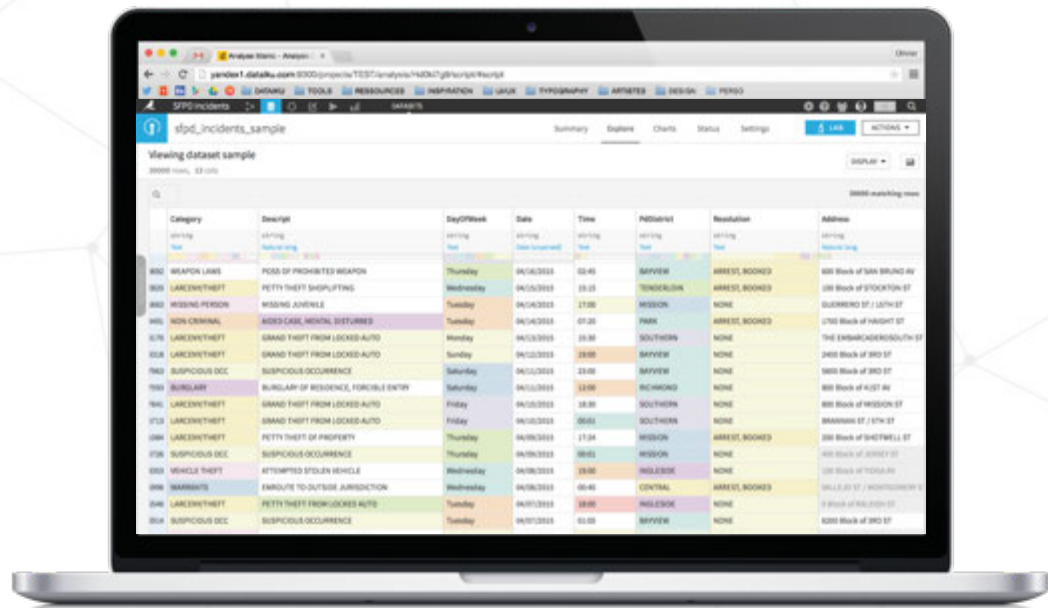
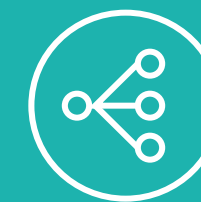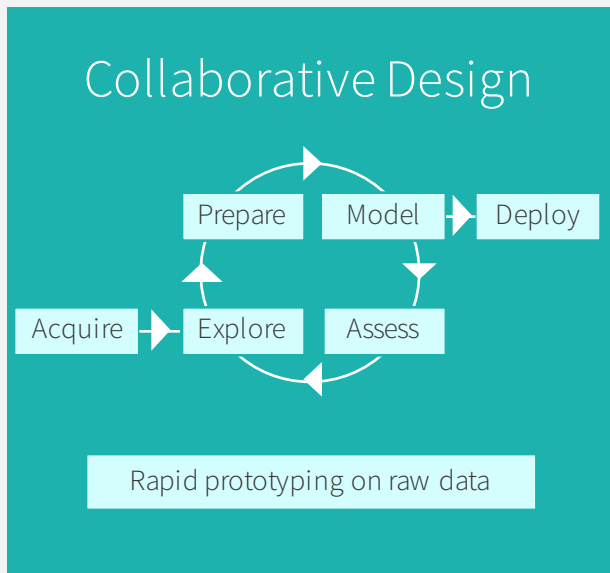|			
DESIGN	**01010 1100** **PRÉPARER** Chargez et préparez vos données	**ANALYSER** Visualisez et partagez vos découvertes	**MODÉLISER** Construisez vos modèles
PRODUCTION	**MONITORER** Suivez votre production	**AUTOMATISER** Ré-exécutez tous les jours…	**EXECUTER** … ou en Temps Réel

Datalabs, Enjeux et Bonnes Pratiques

dataiku

Data Science Studio

One product for a complete, design-to-production workflow

Design

Collaborative Design

Prepare → Model → Deploy

Acquire → Explore → Assess

Rapid prototyping on raw data

Production

Automated Applications

Advanced Scheduler

Monitoring & Dashboarding

Model Lifecycle Management

API Builder

Runtime

Real-Time Predictions

Dedicated scoring engine

REST API

Self-contained environment

Scalable for high availability

dataiku

Nos clients (60+)

Web

- ✓ Analyse des parcours web et segmentation comportementale
- ✓ Anticipation du 'churn' d'abonnés
- ✓ Prévision des ventes
- ✓ Pricing dynamique

Industrie & Infrastructure

- ✓ Maintenance prédictive et diminution de l'impact des pannes matérielles
- ✓ Optimisation logistique
- ✓ Smart Cities

Banque & Assurance

- ✓ Détection de fraude
- ✓ Anticipation des risques (défaut de paiement…)
- ✓ Détection des moments de vie

Prochaines dates

- « Comment engager la révolution du data-driven Marketing? » Prochain Petit-Déjeuner des Datalabs le 26 Mai avec Voyages Privés et Cap Gémini Consulting

- Sortie de Dataiku DSS 3.0 aujourd'hui!