Big DATA & ANALYTICS

Le potentiel et les défis du Big Data

Mardi 2 et Mercredi 3 Juillet 2013

CCI PARIS ILE-DE-FRANCE

QUI SUIS-JE ?

CCI PARIS ILE-DE-FRANCE

AMPLEUR, QUELQUES FAITS SAILLANTS

- Mantra vertus magiques, vitesse de propagation, amplitude richterienne… 5 Exaoctets
- 30 milliards de données (messages, photos) ont été ajoutés sur Facebook tous les mois par 600 millions d'utlisateurs, 1 milliard de requêtes Google
- Zynga traite un Petabyte de contenu par jour
- Plus de 2 milliards de vidéos ont été visionnées sur Youtube hier
- Un adolescent envoie 4000 texte message par mois (tous formats confondus)
- 32 milliards de recherche ont été effectués le mois dernier sur Twitter

ENJEUX INITIAUX ET MUTATIONS

- Enfant naturel du cloud : SQL / NoSQL (stockage de masse à plat sans requête, système de requête de B2d) : sysème de fichiers, stockage (HDSF) et calcul distribué massivement parallèle en programmation simple sur machines cluster (OS)

- Convergence d'éléments : HPC, stockage, cloud, provisionning / affectation , pour le dimensionnement de l'infrastructure

- Les 3/5V que relèvent la BI 2.0, informatique décisionnelle réinventée, data mining d'aujourd'hui

- Data en format hybride et du temps réel

- Capteurs coûts faibles, auto-générateur d'informations, recensement mouvements physiques

RUPTURE PARADIGME

- Alternative aux solutions traditionnelles en B2D et BI
- Google, Yahoo
- NoSQL (sans base de données et sans requêtes)
- Pas que du stockage et archivage (au-delà simple considération)
- Hadoop et sa distribution

QUELLES SOLUTIONS CHOISIR

- 3 critères discriminants sur solution de distribution

- Degré open source
- Maturité des solutions
- Compatibilité

CARACTÉRISER

- Alternative aux solutions traditionnelles en B2D et BI
- Convergence de plusieurs éléments
- Google, Yahoo
- NoSQL (sans base de données et sans requêtes)
- Pas que du stockage et archivage

Marketing

CCI PARIS ILE-DE-FRANCE

DISCIPLINE MARKETING CHALLENGÉE

- Marketing calé sur les SI, entrepôts de données : logique de moyenne, instinct, intuition, gamme, silot
- BI 2.0, informatique décisionnelle réinventée, data mining d'aujourd'hui
- Concept granularité / changement de métrique
- Donnée mutante, real time, comportementale, changeante (actif circulant, ubiquité)
- Data date limite de consommation antinomique avec sciences fondamentales marketing
- Contextuel, géolocalisation, autonomie, capacité de prise de décisions, micro-décisions opérationnelles, périmètre local (big data, small data)
- Consommateur générateur, modifie en real time la data

USER EXPÉRIENCE

- Avènement de l'expérience utilisateur (BI, CRM)
- Infrastructures matures pour extrême personnalisation de l'expérience client (revendiquée)
- Interaction omni-canal, omni-segment, verticale
- Bouleverse cycle de prise de décisions

Santé

CCI PARIS ILE-DE-FRANCE

SANTÉ les data forment la chaîne de valeur

- Donner du sens à grande échelle sur choix thépapeutique
- Complexité métier, masse de connaissance, protocoles de soins vêtustes, repose trop sur la chimie
- Dispo data (capteurs, micropuces, quantified self, gélocalisation des applis et supports, habitudes de vie alimentaires)
- Logique de maîtrise des coûts (besoin dynamique vertueuse), levier demodernisation
- Numérisation actes, codification actes, remboursement (DMP), antécédents médiaux de chacun
- Bénéfices : agir sur fraudes, parcours de soins, logistique de soins, coordination

Santé «donner du sens à grande échelle »

- Maladie VS Environnement (champ exploratoire analytique, complexité médecine, analyse multifactorielle)
- Géolocalisation, habitudes de vie, tickets de caisse, valeur nutritionnelles pour comprendre maladies
- 2 patients, une maladie, un même traitement générique
- Rigueur algorithmique VS médecine généraliste ?
- La médecine des 5P : préventive, prédictive, participative (constitution savoir), personnalisée, peu onéreuse
- Patient est générateur de data. Maladie VS environnement

Prévention VS curatif

Avènement d'une médecine personnalisée, prédictive, médecine de Molière

Recentrage valeur patient, environnement multifactoriel

POUR FAIRE QUOI, OBJECTIFS ?

- Démarche d'aide à la décision (individuel), citoyen autonomie
- Participatif, crowdsourcing, collectif, modélisation
- Innovation
- Coûts, rationalisation, fraudes

CNIL

- Clouds souverains, partage de données de santé, captation monopolistique
- Segmentation, logique compartimentée

CCI PARIS ILE-DE-FRANCE

Santé la data a deux niveaux de valeur

- Bénéfice patient, data contextualisée
- Bien commun, participative, base de donnée épidémiologique
- Modélisation et participatif : modèles prédictifs de maladies
- Crowdsourcing (ex.)
- ADR Prism, « quand le e-sentiment fait jurdisprudence » (pharmacovigilance, nettoyage de posts, délivrables, nouveau chainon de surveillance, signaux faibles de 3 variables)

Watson - IBM

- Hypothèse de traitement, capable de les justifier, déf diagnostic
- Capacité à comprendre language courant
- Articles scientifiques, données biologiques, comportementales et contextuelles
- Projet Global Pulse ONU

Predictive Modeling and Data Mining Scientists/Analysts

analyze the campaign data to guide election strategy and develop quantitative, actionable insights that drive our decision-making. Looking for people at both the senior and junior level to join the campaign Analytics Dept through November 2012.

Company: Obama 2012 Presidential Campaign
Location: Chicago, IL
Web: www.barackobama.com

The Obama for America Analytics Department analyzes the campaign's data to guide election strategy and develop quantitative, actionable insights that drive our decision-making. Our team's products help direct work on the ground, online and on the air.

We are looking for Predictive Modeling/Data Mining Scientists and Analysts, at both the senior and junior level, to join our department through November 2012 at our Chicago Headquarters. We are a multi-disciplinary team of statisticians, predictive modelers, data mining experts, mathematicians, software developers, general analysts and organizers - all striving for a single goal: re-electing President Obama.

Using statistical predictive modeling, the Democratic Party's comprehensive political database, and publicly available data, modeling analysts are charged with predicting the behavior of the American electorate. These models will be instrumental in helping the campaign determine which voters to target for turnout and persuasion efforts, where to buy advertising and how to best approach digital media.

Our Modeling Analysts will dive head-first into our massive data to solve some of our most critical online and offline challenges. We will analyze millions of interactions a day, learning from terabytes of historical data, running thousands of experiments, to inform campaign strategy and critical decisions.

Responsibilities include:

- Develop and build statistical/predictive/machine learning models to assist in field, digital media, paid media and fundraising operations
- Assess the performance of previous models and determine when these models should be updated
- Design and execute experiments to test the applicability and validity of these models in the field
- Create metrics to assess performance of various campaign tactics
- Collaborate with the data team to improve existing database and suggest new data sources
- Work with stakeholders to identify other research needs and priorities

Required Experience:

- B.S degree (M.S/PhD for scientist and senior positions) in statistics, machine learning, mathematics, quantitative methods, computer science, or related field
- Experience with political, Nielsen/Arbitron, fundraising or digital media & online advertising data
- Application of advanced statistical, machine learning, and/or data mining techniques (i.e. classification, clustering, association mining, forecasting), to real-world problems with massive data
- Experience with text data, search, natural language processing, social media analytics is a plus - we're also hiring for text mining positions.
- Proven creativity and problem-solving skills

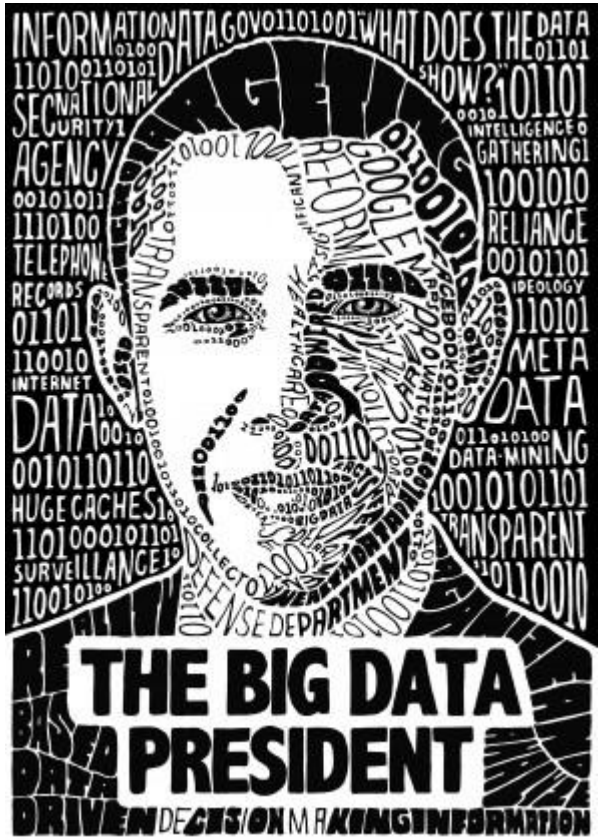
Required Software:
Applicants must have demonstrated, extensive experience (professional or academic) with any major statistical or data mining package (R, STATA, SPSS, SAS, Enterprise Miner, Matlab, KNIME, Weka). Other desired software skills would include:

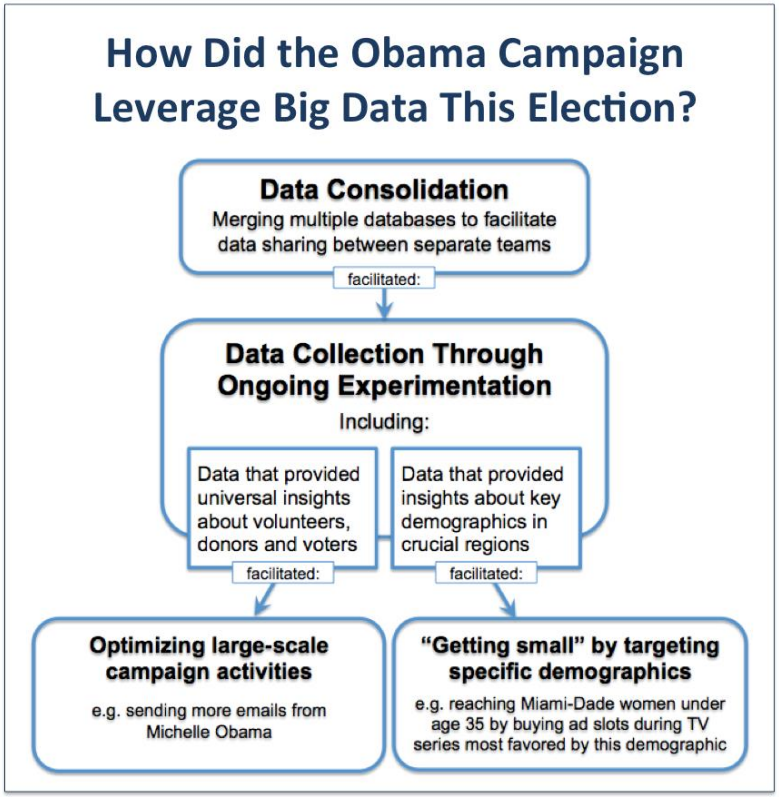
- Any SQL-based query language (MySQL, PostgreSQL, etc.)
- Programming skills desirable but not required for all positions (C#, C++, Java, Python, Ruby, Perl)
- Strong MS Excel skills also desired

Contact:
Please send resumes to **analyticsjobs@barackobama.com** and mention kdnuggets.

Obama for America is committed to diversity among its staff, and recognizes that its continued success requires the highest commitment to obtaining and retaining a diverse staff that provides the best quality services to supporters and constituents. Obama for America is an equal opportunity employer and it is our policy to recruit, hire, train, promote and administer any and all personnel actions without regard to sex, race, age, color, creed, national origin, religion, economic status, sexual orientation, veteran status, gender identity or expression, ethnic identity or physical disability, or any other legally protected basis. Obama for America will not tolerate any unlawful discrimination and any such conduct is strictly prohibited.

Obama « Yes, I can rule Big Data ».





CCI PARIS ILE-DE-FRANCE

Obama un cas d'école éclatant

- 2008 – 2012 (Nate Silver, climat, réseaux puis big data)
- Datacrunching : levée de fonds, thèmes délaissés des indécis, registration / persuasion / turnout
- Fusion de base de données (giga base de données alimentée en continu) : Airwolf & Media Optimizer
- Fin des campagnes transterritoriales > une élection locale > personnalisation extrême de l'e-mailing > 40 variables > modélisation à niveau individuel > nanotargeting
- Les réseaux sociaux ? Encore de l'analytique derrière….

Nouveaux savoirs faire

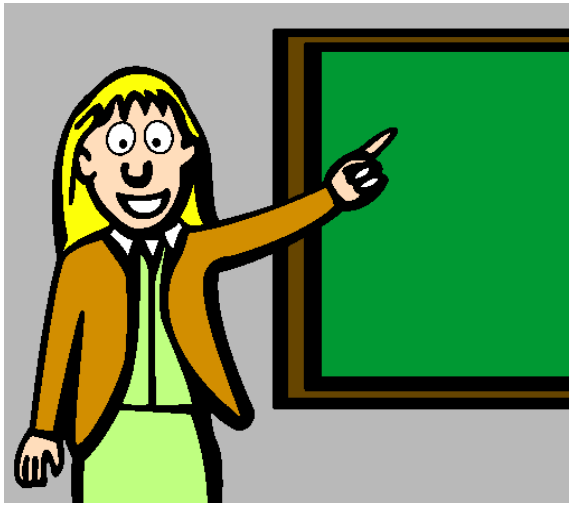
CCI PARIS ILE-DE-FRANCE

DATA SCIENTIST : LE TALON D'ACHILLE

- 3 compétences
- Ressources humaines issues de la BI ?
- Interaction omni-canal, omni-segment, verticale
- Bouleverse cycle de prise de décisions

DATA SCIENTIST : LA FILIÈRE FRANÇAISE

- Fort cloisonnement des enseignements
- Design, sémantique
- Culture de la statistique théorique (VS US)
- Compréhension multidisciplinaire, holistique
- Architecte, analyste, programmeur algorithmique, statisticien, designer, viz'
- 20 K à 200K
- Filière excellence française, master dédié
- ENSAE, ENSAI, Paris Tech
- Environnements techniques matures, l'industrie a posé les jalons (triptyque)



LA FILIÈRE FRANÇAISE

CCI PARIS ILE-DE-FRANCE

LA FILIÈRE FRANÇAISE

- Fleur Pellerin, AFDEL et quartiers numériques, Parsi Capitale Numérique
- Big Data Launchpad, milestone project 2013-2018, fenêtre opportunité
- Un train à ne pas louper, s'emparer de cette rupture technologique
- Les propositions : un écosystème à structurer, lequel ?
- Pénaliser par capacités de financement
- Objectifs quantitatifs de création de valeur

CCI PARIS ILE-DE-FRANCE

LA FILIÈRE FRANÇAISE

- Quels positionnement pour construire France Big Data Inc.

- Infra ?
- Intégration de données ?
- Etage applis métiers ?
- Quelles nouvelles places de marché créer ?

ACCEL® PARTNERS

CCI PARIS ILE-DE-FRANCE


ezakus
CAN YOU READ MINDS?
adomik
advertising intelligence for publishers
TRENDSBOARD
linkfluence
AT INTERNET
Online Intelligence Solutions
Dataiku
focus matic
ACCURATE DIGITAL REACH
bime
Captain DASH
cinequant
Cliris
DATA PUBLICA
SAFETY LINE
take control of your safety
neolane
marketing that delivers
FIFTY-FIVE.COM
55
MIND THE GAP
QWAM
CONTENT INTELLIGENCE
Tell Plus Me
iAdvize
nexedi
kxen
Predictive Power. Infinite Insight.
kyriba
PATHOQUEST
BIsquare
squid
SOLUTIONS
antidot
maporama
OpenDataSoft
open wide
ARCHITECTE OPEN SOURCE
viavoo
smarter feedbacks
yseop
(EASY · OP)
LOKAD
forecasting web services
mesagraph
mfg labs
tinyclues
qunb
quantities & numbers
netwave
YOUR FULL E-POTENTIAL
Conexance
G!
Gamned!
PolySpot
proxem
from text to data
Sirdata
SyLLabs
SINEQUA
CONNECT TO KNOWLEDGE™
hurence
idealanalytics
Visibrain
isthma
TalkMap
smarter together
trendybuzz
web live tracker

OPEN DATA : FAIRE PLUS AVEC MOINS

- Transparence, démocratie 2.0 VS régime monarchique
- Société collaborative, crowsourcée
- Création de valeur (applis) grâce à l'accès aux données publiques

OPEN DATA

- Transparence
- Société collaborative, crowsourcée

CNIL
Commission Nationale de l'Informatique et des Libertés

Un marché à épurer, consolider

> acteurs

Machine learning – Data Product

> psychographie
> Intelligence artificielle
> Modélisation comportementale

Outils intuitifs

> Abaisser barrières à l'entrée et expertise
> Freine adoption généralisation : besoin d'outils intuitifs, accélérer courbe apprentissage

CCI PARIS ILE-DE-FRANCE

Contacts

Matthias FILLE *Conseiller Développement International*

Filière TIC

Service actions sectorielles

Tél : +33 (0)1 55 65 35 42 - Fax : +33 (0)6 60 04 84 63

mfille@cci-paris-idf.fr

CCI PARIS ILE-DE-FRANCE