

R++ THE NEXT STEP?

Christophe Genolini

R++ THE NEXT STEP?

Big data
Parallélisme
IHM
Statistiques

BIG DATA

Statistique

- R → ~1 Giga
- R64 → ~10 Giga
- STATA → ~1 Giga
- Julia → ~0.3 Giga

Informatique

- Oracle → 50 Terra

PARALLÉLISME

Lecture

- Quelqu'un a-t-il du mal à lire cette phrase ?

Ecriture

- Deux c'est personnes difficile écrivent à en lire même non temps

STATISTIQUE ET PARALLÉLISME (THÉORIQUE)

	A	B	C	D	E	F
1	Nom	Prénom	Adresse	Code P.	Ville	Date naiss.
2	Angelin	Pierre	24 rue des acacias	92600	Asnières	2/01/37
3	Dupont	Gaston	2 rue des Alouettes	75013	Paris	23/01/38
4	Baudrin	Patricia	4 grande rue	75008	Paris	1/04/69
5	Dupont	Bernard	12 avenue des Armées	75016	Paris	13/04/59
6	Leclerc	Laurent	4 rue des Alouettes	78000	Paris	4/04/29
7	Dupont	Bernard	5 rue de Tanger	75019	Paris	29/12/49
8	Jouhaux	Laurence	2 bis rue Tierce	75020	Paris	18/08/55
9	Houral	Claude	4 rue du Lencheval	75020	Paris	15/06/54
10	Dupond	Doriane	5 rue Calmels	92270	Bois-Colombes	22/12/32
11	Zoalir	Gérard	120 rue du Fg. Poissonnière	75010	Paris	31/10/64
12	Dupond	Philippe	5 bd Montmartre	75009	Paris	1/11/63
13	Daniel	Christine	9 rue Haute	75001	Paris	23/05/65
14	Rain	Florence	8 avenue Jean Jaurès	75012	Paris	25/09/56
15	Claron	Carine	5 bd du Grand Duc	92270	Bois-Colombes	16/05/56
16	Maurin	Martine	6 rue St-Honoré	75008	Paris	17/05/56
17						

Lecture



Analyse univariée
1 à 250

Ecriture



Analyse univariée
250 à 500

Ecriture



Analyse univariée
501 à 750

Ecriture



Analyse univariée
751 à 1000

Ecriture

STATISTIQUE ET PARALLÉLISME (PRATIQUE)

	A	B	C	D	E	F
1	Nom	Prénom	Adresse	Code P.	Ville	Date nais.
2	Angelin	Pierre	24 rue des acacias	92600	Asnières	2/01/37
3	Dupont	Gaston	2 rue des Alouettes	75013	Paris	23/01/38
4	Baudrin	Patricia	4 grande rue	75008	Paris	1/04/69
5	Dupont	Bernard	12 avenue des Armées	75016	Paris	13/04/59
6	Leclerc	Laurent	4 rue des Alouettes	78000	Paris	4/04/29
7	Dupont	Bernard	5 rue de Tanger	75019	Paris	29/12/49
8	Jouhaux	Laurence	2 bis rue Tierce	75020	Paris	18/08/55
9	Houral	Claude	4 rue du Lencheval	75020	Paris	15/06/54
10	Dupond	Doriane	5 rue Calmels	92270	Bois-Colombes	22/12/32
11	Zoalir	Gérard	120 rue du Fg. Poissonnière	75010	Paris	31/10/64
12	Dupond	Philippe	5 bd Montmartre	75009	Paris	1/11/63
13	Daniel	Christine	9 rue Haute	75001	Paris	23/05/65
14	Parin	Florence	8 avenue Jean Jaurès	75012	Paris	25/09/56
15	Claron	Carine	5 bd du Grand Duc	92270	Bois-Colombes	16/05/56
16	Maurin	Martine	6 rue St-Honoré	75008	Paris	17/05/56
17						

Duplication

	A	B	C	D	E	F
1	Nom	Prénom	Adresse	Code P.	Ville	Date nais.
2	Angelin	Pierre	24 rue des acacias	92600	Asnières	2/01/37
3	Dupont	Gaston	2 rue des Alouettes	75013	Paris	23/01/38
4	Baudrin	Patricia	4 grande rue	75008	Paris	1/04/69
5	Dupont	Bernard	12 avenue des Armées	75016	Paris	13/04/59
6	Leclerc	Laurent	4 rue des Alouettes	78000	Paris	4/04/29
7	Dupont	Bernard	5 rue de Tanger	75019	Paris	29/12/49
8	Jouhaux	Laurence	2 bis rue Tierce	75020	Paris	18/08/55
9	Houral	Claude	4 rue du Lencheval	75020	Paris	15/06/54
10	Dupond	Doriane	5 rue Calmels	92270	Bois-Colombes	22/12/32
11	Zoalir	Gérard	120 rue du Fg. Poissonnière	75010	Paris	31/10/64
12	Dup	Philippe	5 bd Montmartre	75009	Paris	1/11/63
13	Dan	Christine	9 rue Haute	75001	Paris	23/05/65
14	Par	Florence	8 avenue Jean Jaurès	75012	Paris	25/09/56
15	Clar	Carine	5 bd du Grand Duc	92270	Bois-Colombes	16/05/56
16	Mau	Martine	6 rue St-Honoré	75008	Paris	17/05/56

Lecture



Analyse univariée
1 à 250

	A	B	C	D	E	F
1	Nom	Prénom	Adresse	Code P.	Ville	Date nais.
2	Angelin	Pierre	24 rue des acacias	92600	Asnières	2/01/37
3	Dupont	Gaston	2 rue des Alouettes	75013	Paris	23/01/38
4	Baudrin	Patricia	4 grande rue	75008	Paris	1/04/69
5	Dupont	Bernard	12 avenue des Armées	75016	Paris	13/04/59
6	Leclerc	Laurent	4 rue des Alouettes	78000	Paris	4/04/29
7	Dupont	Bernard	5 rue de Tanger	75019	Paris	29/12/49
8	Jouhaux	Laurence	2 bis rue Tierce	75020	Paris	18/08/55
9	Houral	Claude	4 rue du Lencheval	75020	Paris	15/06/54
10	Dupond	Doriane	5 rue Calmels	92270	Bois-Colombes	22/12/32
11	Zoalir	Gérard	120 rue du Fg. Poissonnière	75010	Paris	31/10/64
12	Dup	Philippe	5 bd Montmartre	75009	Paris	1/11/63
13	Dan	Christine	9 rue Haute	75001	Paris	23/05/65
14	Par	Florence	8 avenue Jean Jaurès	75012	Paris	25/09/56
15	Clar	Carine	5 bd du Grand Duc	92270	Bois-Colombes	16/05/56
16	Mau	Martine	6 rue St-Honoré	75008	Paris	17/05/56

Lecture



Analyse univariée
250 à 500

	A	B	C	D	E	F
1	Nom	Prénom	Adresse	Code P.	Ville	Date nais.
2	Angelin	Pierre	24 rue des acacias	92600	Asnières	2/01/37
3	Dupont	Gaston	2 rue des Alouettes	75013	Paris	23/01/38
4	Baudrin	Patricia	4 grande rue	75008	Paris	1/04/69
5	Dupont	Bernard	12 avenue des Armées	75016	Paris	13/04/59
6	Leclerc	Laurent	4 rue des Alouettes	78000	Paris	4/04/29
7	Dupont	Bernard	5 rue de Tanger	75019	Paris	29/12/49
8	Jouhaux	Laurence	2 bis rue Tierce	75020	Paris	18/08/55
9	Houral	Claude	4 rue du Lencheval	75020	Paris	15/06/54
10	Dupond	Doriane	5 rue Calmels	92270	Bois-Colombes	22/12/32
11	Zoalir	Gérard	120 rue du Fg. Poissonnière	75010	Paris	31/10/64
12	Dup	Philippe	5 bd Montmartre	75009	Paris	1/11/63
13	Dan	Christine	9 rue Haute	75001	Paris	23/05/65
14	Par	Florence	8 avenue Jean Jaurès	75012	Paris	25/09/56
15	Clar	Carine	5 bd du Grand Duc	92270	Bois-Colombes	16/05/56
16	Mau	Martine	6 rue St-Honoré	75008	Paris	17/05/56

Lecture



Analyse univariée
501 à 750

	A	B	C	D	E	F
1	Nom	Prénom	Adresse	Code P.	Ville	Date nais.
2	Angelin	Pierre	24 rue des acacias	92600	Asnières	2/01/37
3	Dupont	Gaston	2 rue des Alouettes	75013	Paris	23/01/38
4	Baudrin	Patricia	4 grande rue	75008	Paris	1/04/69
5	Dupont	Bernard	12 avenue des Armées	75016	Paris	13/04/59
6	Leclerc	Laurent	4 rue des Alouettes	78000	Paris	4/04/29
7	Dupont	Bernard	5 rue de Tanger	75019	Paris	29/12/49
8	Jouhaux	Laurence	2 bis rue Tierce	75020	Paris	18/08/55
9	Houral	Claude	4 rue du Lencheval	75020	Paris	15/06/54
10	Dupond	Doriane	5 rue Calmels	92270	Bois-Colombes	22/12/32
11	Zoalir	Gérard	120 rue du Fg. Poissonnière	75010	Paris	31/10/64
12	Dup	Philippe	5 bd Montmartre	75009	Paris	1/11/63
13	Dan	Christine	9 rue Haute	75001	Paris	23/05/65
14	Par	Florence	8 avenue Jean Jaurès	75012	Paris	25/09/56
15	Clar	Carine	5 bd du Grand Duc	92270	Bois-Colombes	16/05/56
16	Mau	Martine	6 rue St-Honoré	75008	Paris	17/05/56

Lecture



Analyse univariée
751 à 1000

écriture

écriture

écriture

écriture

INTERFACE HOMME MACHINE

EXPORT AUTOMATIQUE DE TABLEAUX

- Voir la vidéo jointe.

R++, THE NEXT STEP!

R++, THE NEXT STEP

- Création d'un logiciel Métier (statistique)
 - Hautes performances
 - Parallèle
 - Big data
 - IHM
 - Basé sur R
-

POURQUOI BASÉ SUR R ?

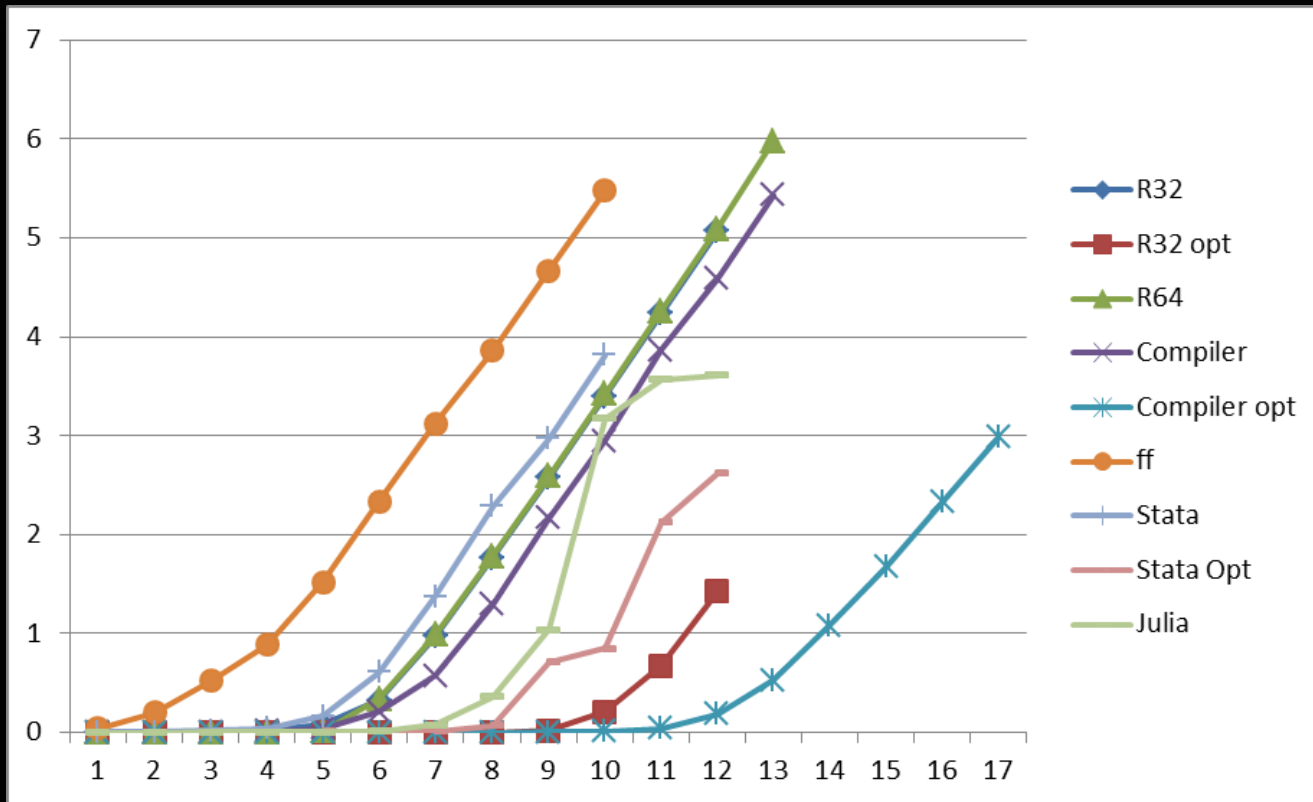
- Libre
 - Concepts (data.frame)
 - Chirurgiens, économistes, linguistes
 - Communauté
 - On garde R
 - On change « juste » le moteur
-

L'ÉQUIPE

- ParSys (LRI, Orsay)
 - Joel Falcou, Daniel Etiemble + 1 stagiaire M2 (→ thèse)
- SIG (IRIT, Toulouse)
 - Mohand Boughanem, Ronan Tournier + 2 stagiaires M2 (→ thèses)
- InSitu (Inria, Paris)
 - Michel Beaudouin-Lafon, Wendy Mackay + chef d'œuvre étudiant
- INSERM (1027, toulouse ; 669 Paris)
 - Christophe Genolini, Bruno Falissard
- Sympathisants
 - Jérôme Collet (EDF), Pierre Courtieux (CNAM), Abel Malek Haddad (INRA), Xavier Clerc (INRIA), Stéphane Ribas (INRIA), Scott Kostyshak (USA, économiste)

AVANCEMENT DU PROJET

- Big Data



AVANCEMENT DU PROJET

- Parallélisme

Time Size	R	Boot()	CPU	GPU	MC_6	MC_8	MC_10	MC_12
10 ²	0,05	0,22	1	0,04	1,41	0,60	0,23	1,03
10 ³	0,04	0,48	1	0,31	3,31	3,62	2,67	2,04
10 ⁴	0,04	0,69	1	2,29	4,85	6,37	5,74	7,38
10 ⁵	0,04	0,70	1	8,45	5,23	6,89	8,23	9,86
10 ⁶	0,04	0,58	1	8,81	5,27	6,75	8,44	9,89

AVANCEMENT DU PROJET

- Trois prototype vidéo
 - Export de tableau
 - Data management
 - Graphe interactif

MERCI DE VOTRE ATTENTION

Des questions ?