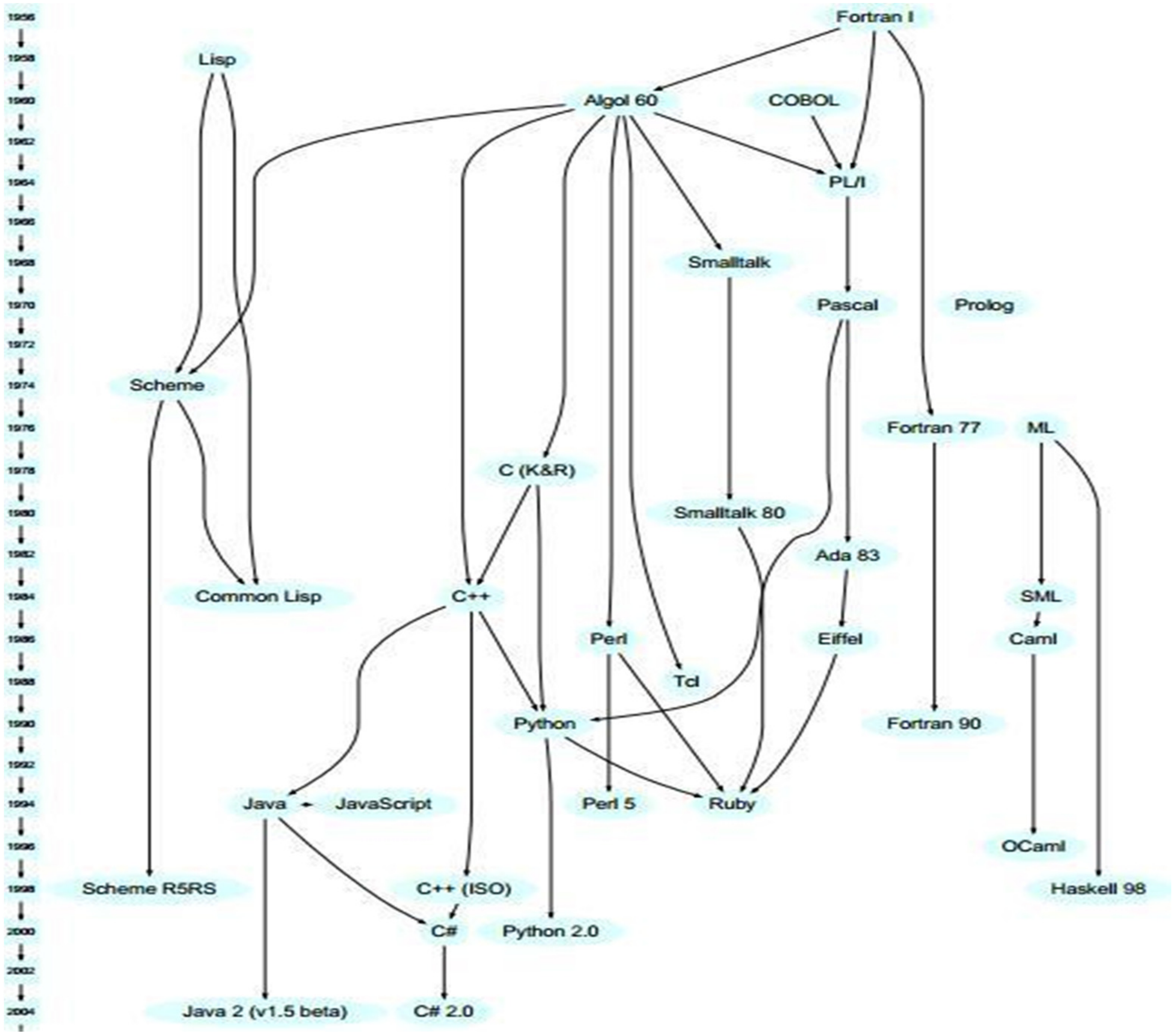
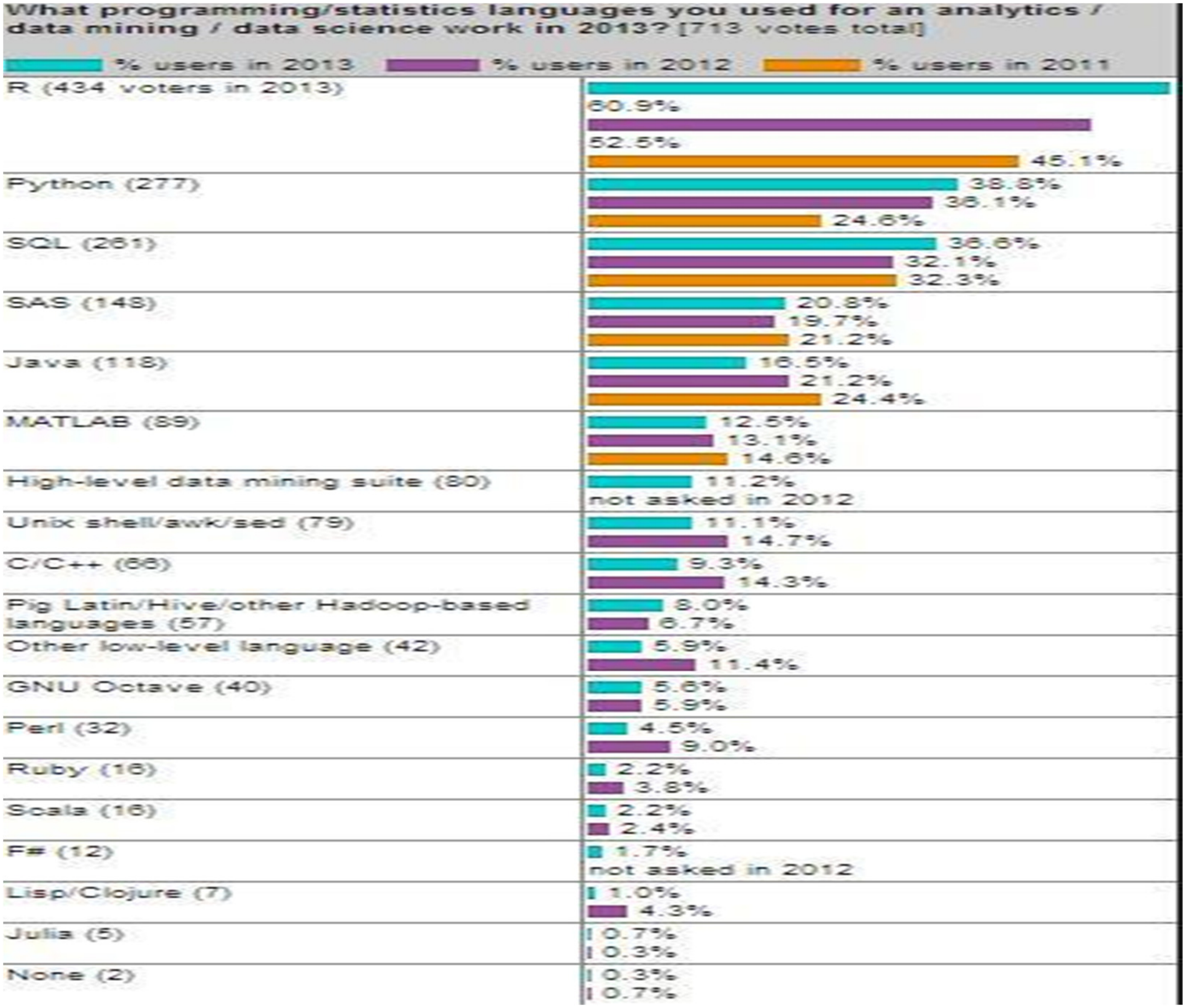
Future of HPC Software

Moderator: Michael Walker

Panelists:

Monte Lunacek
Wesley Jones
Michael Malak
Cary Miller
Andrew Weekley


1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
Fortran I
Lisp
Algol 60
COBOL
PL/I
Smalltalk
Pascal
Prolog
Scheme
Fortran 77
ML
C (K&R)
Smalltalk 80
Ada 83
Common Lisp
C++
SML
Perl
Eiffel
Caml
Tcl
Python
Fortran 90
Java
JavaScript
Perl 5
Ruby
OCaml
Scheme R5RS
C++ (ISO)
Haskell 98
C#
Python 2.0
Java 2 (v1.5 beta)
C# 2.0


What programming/statistics languages you used for an analytics / data mining / data science work in 2013? [713 votes total]
% users in 2013
% users in 2012
% users in 2011
R (434 voters in 2013)
60.9%
52.5%
45.1%
Python (277)
38.8%
36.1%
24.6%
SQL (261)
36.6%
32.1%
32.3%
SAS (148)
20.8%
19.7%
21.2%
Java (118)
16.5%
21.2%
24.4%
MATLAB (89)
12.5%
13.1%
14.6%
High-level data mining suite (80)
11.2%
not asked in 2012
Unix shell/awk/sed (79)
11.1%
14.7%
C/C++ (66)
9.3%
14.3%
Pig Latin/Hive/other Hadoop-based languages (57)
8.0%
6.7%
Other low-level language (42)
5.9%
11.4%
GNU Octave (40)
5.6%
5.9%
Perl (32)
4.5%
9.0%
Ruby (16)
2.2%
3.8%
Scala (16)
2.2%
2.4%
F# (12)
1.7%
not asked in 2012
Lisp/Clojure (7)
1.0%
4.3%
Julia (5)
0.7%
0.3%
None (2)
0.3%
0.7%

Most popular languages:

R (61% of responders)
Python (39%)
SQL (37%)

Language with highest relative growth (2013 vs 2012) was Julia, which doubled in popularity, but still was used only by 0.7% in 2013.

Largest relative increases in share of usage from 2012 to 2013:

Pig Latin/Hive/other Hadoop-based languages, 19% growth, from 6.7% in 2012 to 8.0% in 2013

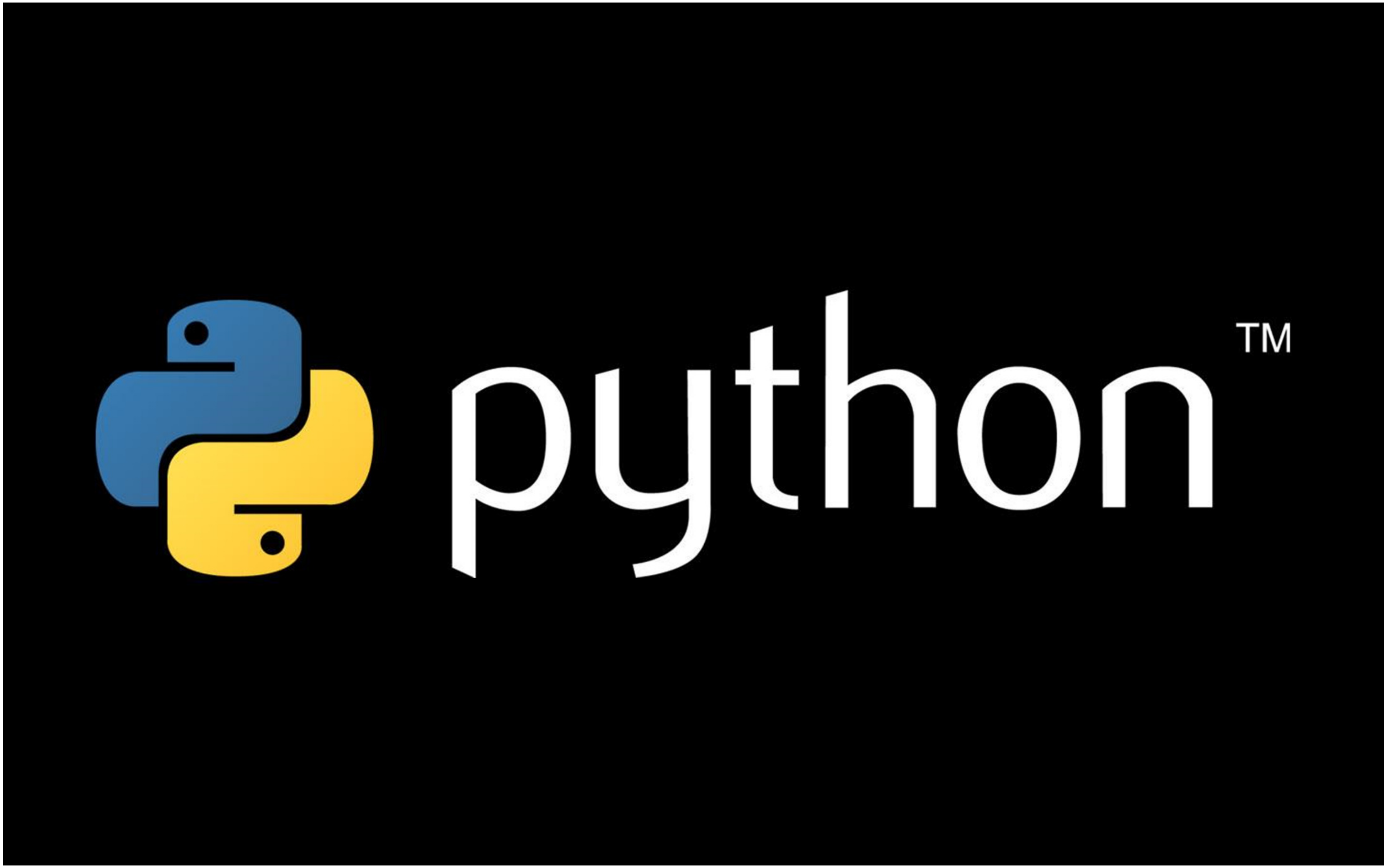
R, 16% growth

SQL, 14% growth (perhaps the result of increasing number of SQL interfaces to Hadoop and other Big Data systems?)

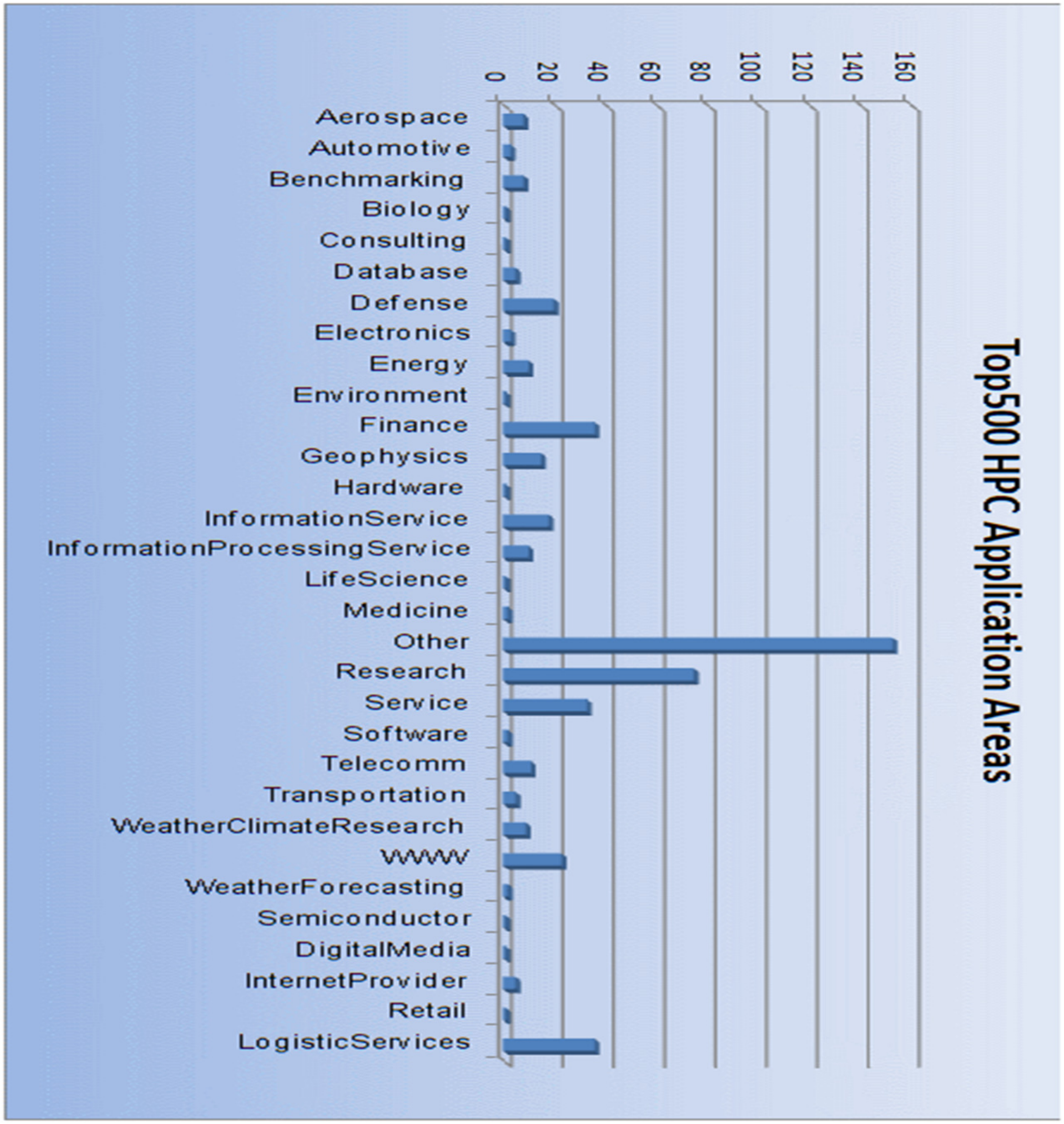
Languages with largest decline in share of usage:

Lisp/Clojure, 77% down
Perl, 50% down
Ruby, 41% down
C/C++, 35% down
Unix shell/awk/sed, 25% down
Java, 22% down

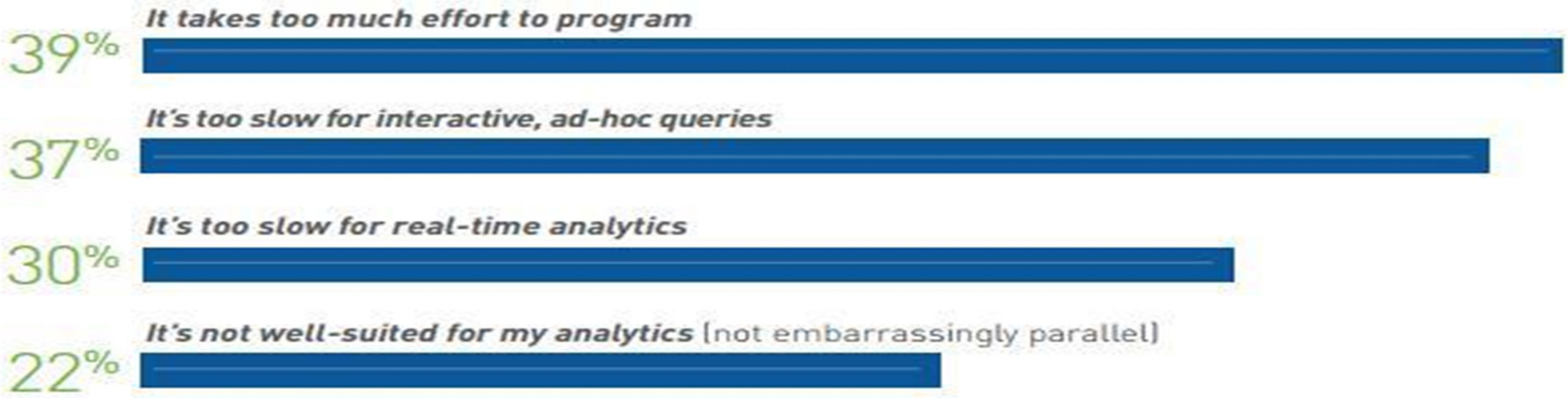
R

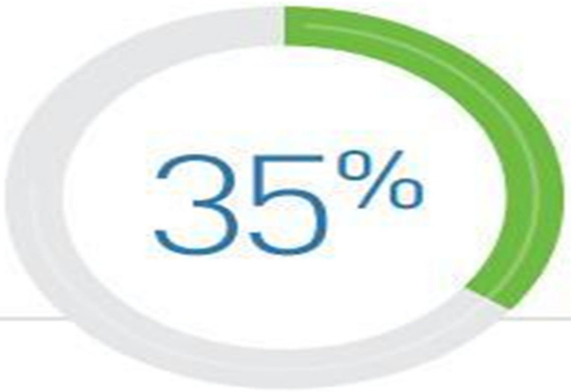

python™

julia


Top500 HPC Application Areas
0
20
40
60
80
100
120
140
160
Aerospace
Automotive
Benchmarking
Biology
Consulting
Database
Defense
Electronics
Energy
Environment
Finance
Geophysics
Hardware
InformationService
InformationProcessingService
LifeScience
Medicine
Other
Research
Service
Software
Telecomm
Transportation
WeatherClimateResearch
WWW
WeatherForecasting
Semiconductor
DigitalMedia
InternetProvider
Retail
LogisticServices

From the 76% reporting problems, what are the limitations of Hadoop / SPARK?





of data scientists who tried Hadoop or SPARK have stopped using it

New HPC language needed for performance at Petascale.

New HPC language needed for performance at Exascale.

Needed to improve software productivity.

New HPC language should:

- Address performance issues of large future systems
- Express well common HPC programming patterns
- Support well performance programming, with incremental code refinement
- Support OO
- Take advantage of advances in PL's: strong typing, type and memory safety, atomicity, efficient support for generic programming

New HPC language should:

- Take advantage of advances in compilers: dynamic compilation, heuristic search, telescoping languages
- Coexist with existing languages
- Provide state of the art ADE
- Largely built on language with large market

Non Technical Obstacles

- It takes money to make a good compiler; there is no market for HPC unique optimizations
- It takes time to make a good compiler; there is no funding mechanism for a sustained 5 years development effort
- It takes people to make a good compiler; there is no independent compiler company

Should Hardware vendors develop the HPC ADE?

Good ADE is more than language:

- Porting tools
- Good support for performance tuning
- Tools for refactoring
- Notation for capturing tuning decisions
- Good observability
- Integrated performance stream mining

Bottom Line

- HPC is hampered by lack of good software support
- Language is only part of the problem
- Most obstacles are not technological
- Key issues for petascale computing are just now being addressed