

SciServer



Collaborative data-driven science

SciServer Compute Workshop



JOHNS HOPKINS
UNIVERSITY



idies

SciServer



Collaborative data-driven science

Bring Your Code to the Data

Alex Szalay



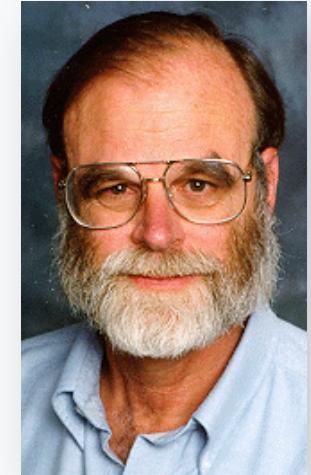
JOHNS HOPKINS
UNIVERSITY



idies

Motivation and History

- ▶ Started with the SDSS SkyServer
- ▶ Built very quickly in 2001
- ▶ **Goal:** instant access to rich content
- ▶ **Idea:** bring the analysis to the data
- ▶ Interactive access at the core
- ▶ Much of the scientific process is about data
 - Data collection, data cleaning, data archiving, data organization, data publishing, mirroring, data distribution, data analytics, data curation...
- ▶ 2012: NSF DIBBS to extend/reengineer SkyServer



Jim Gray

Where Are We Going?

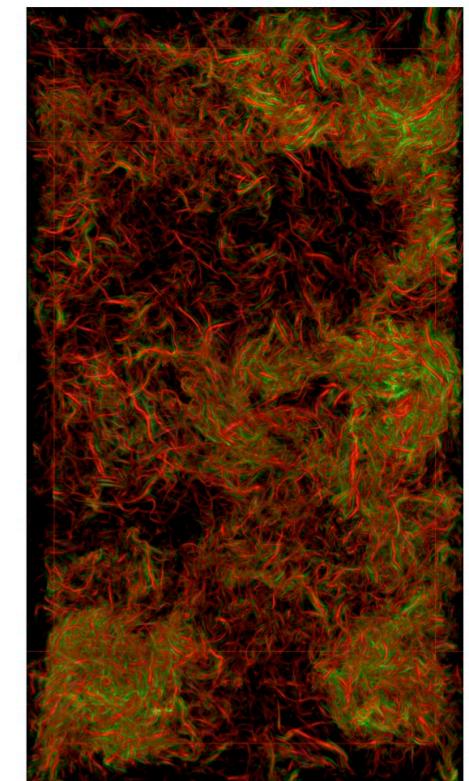
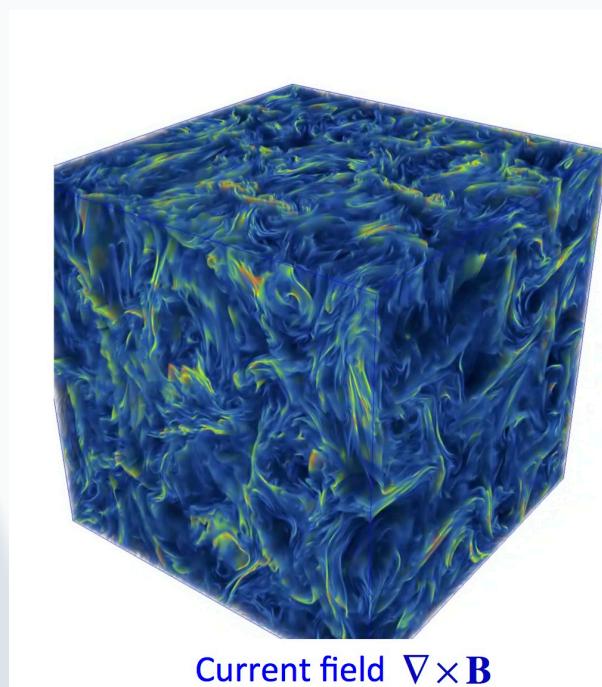
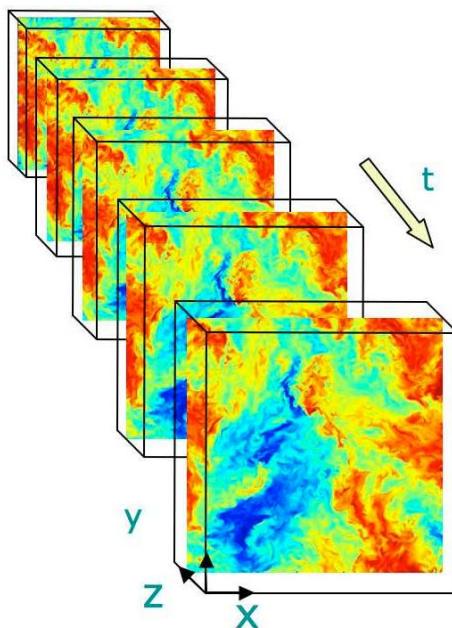
- ▶ Interactive science on petascale data
- ▶ Sustain and enhance our astronomy effort
- ▶ Grow a footprint into new disciplines
- ▶ Build scalable open numerical laboratories
- ▶ Scale system to several petabytes
- ▶ Deep integration with the “Long Tail”
- ▶ Use sharable, well-defined building blocks

Data in HPC Simulations

- ▶ HPC is an instrument in its own right
 - Largest simulations approach/exceed petabytes
- ▶ Need public access to the best and latest
- ▶ Also need ensembles of simulations for UQ
- ▶ Creates new challenges
 - How to access the data?
 - What is the data lifecycle?
 - What are the analysis patterns?
 - What architectures can support these?
- ▶ On Exascale everything is a Big Data problem

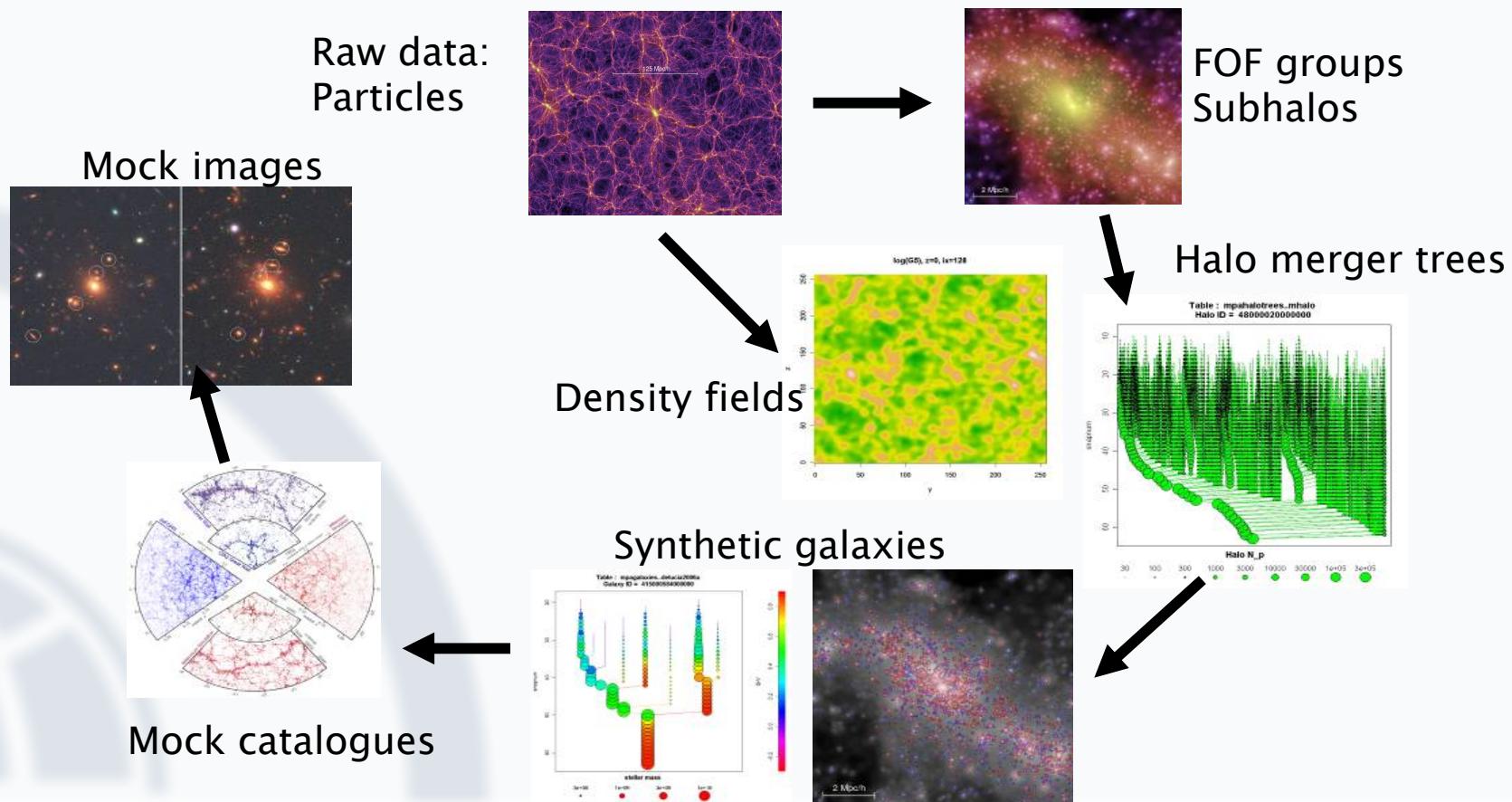
Turbulence databases (JHUTB)

► <http://turbulence.pha.jhu.edu/>



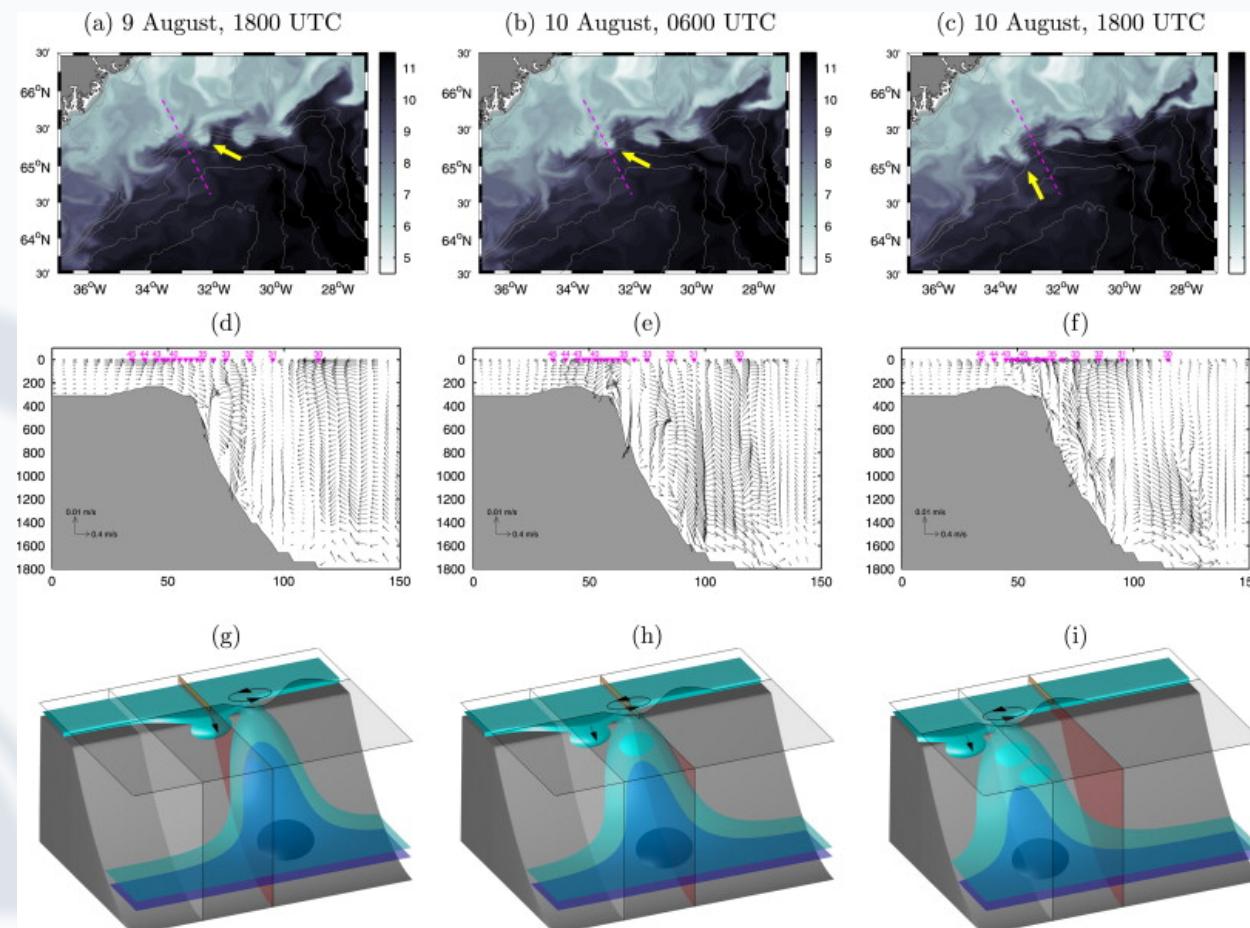
Cosmological Simulations

► Mirror of Millennium Database



Oceanography

Hydrostatic and non-hydrostatic simulations of dense waters cascading off a shelf:
The East Greenland case [Marcello G. Magaldi](#), [Thomas W.N. Haine](#)



Genomics

The screenshot shows the SciServer Genomics interface. At the top, there's a navigation bar with links like Dashboard, Home, RDefault-C, dices@2c87, Untitled, FITS File, Johns Hopkins, R:Install Page, TSE Query, and NSF 18 Mo. Below the navigation bar is a large green DNA helix graphic.

The main area has tabs for TSE, INTERVAL SEARCH, CUSTOM SEARCH, METADATA SEARCH, and ABOUT. A CONTACT US link is also present. The INTERVAL SEARCH tab is active, showing a search form:

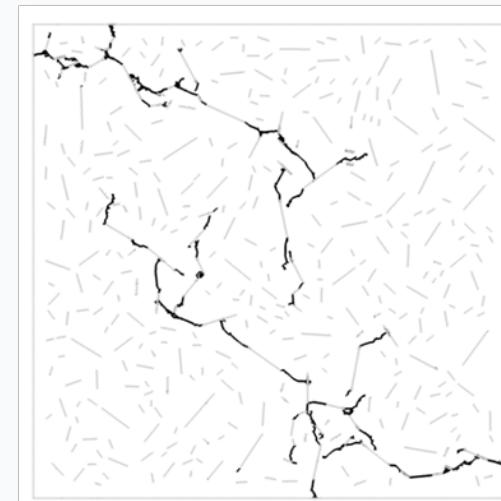
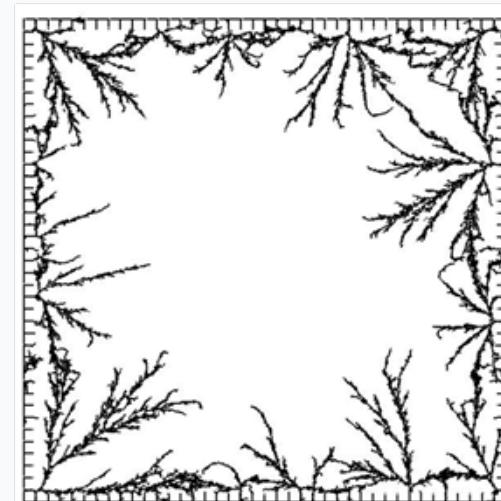
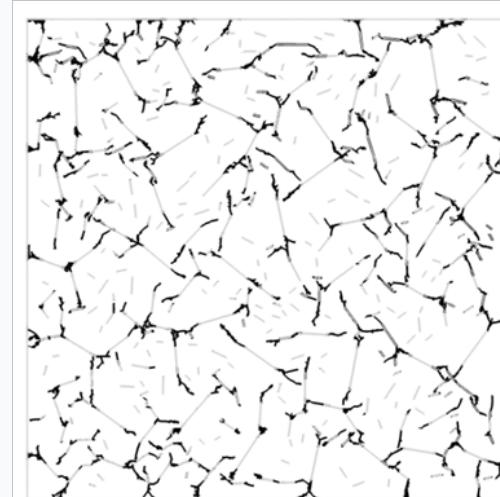
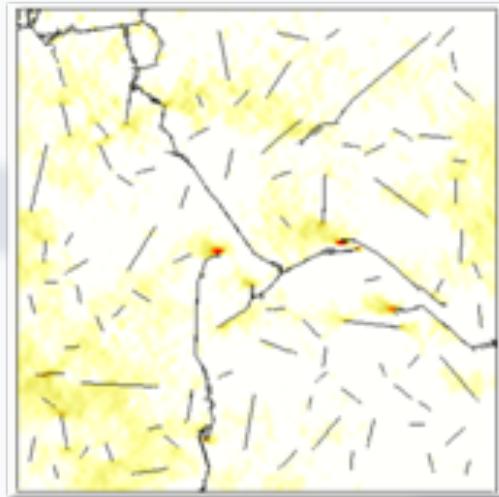
- Chromosome Number:** dropdown set to 6
- Additional Options:**
 - Only overlapping reads
 - Include Paired End
 - Concordant
 - Discordants
 - Ummapped
- From:** 121103500
- To:** 121103600
- Submit** button

Below the search form is a "User Interval" section with a "Paste your Sequence" input field and a "Submit" button. There's also an "Or Upload your Sequence" section with a "UploadFile" button and a "example.csv" file listed.

The results section displays a table with 8224 rows of data, each containing a unique ID, sequence, and various numerical values. To the right of the results table, a WordPad window is open, showing the same data in a plain text format.

ID	Sequence	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value 9	Value 10	Value 11	Value 12	Value 13	Value 14	Value 15	Value 16	Value 17	Value 18	Value 19	Value 20	Value 21	Value 22	Value 23	Value 24	Value 25	Value 26	Value 27	Value 28	Value 29	Value 30	Value 31	Value 32	Value 33	Value 34	Value 35	Value 36	Value 37	Value 38	Value 39	Value 40	Value 41	Value 42	Value 43	Value 44	Value 45	Value 46	Value 47	Value 48	Value 49	Value 50	Value 51	Value 52	Value 53	Value 54	Value 55	Value 56	Value 57	Value 58	Value 59	Value 60	Value 61	Value 62	Value 63	Value 64	Value 65	Value 66	Value 67	Value 68	Value 69	Value 70	Value 71	Value 72	Value 73	Value 74	Value 75	Value 76	Value 77	Value 78	Value 79	Value 80	Value 81	Value 82	Value 83	Value 84	Value 85	Value 86	Value 87	Value 88	Value 89	Value 90	Value 91	Value 92	Value 93	Value 94	Value 95	Value 96	Value 97	Value 98	Value 99	Value 100	Value 101	Value 102	Value 103	Value 104	Value 105	Value 106	Value 107	Value 108	Value 109	Value 110	Value 111	Value 112	Value 113	Value 114	Value 115	Value 116	Value 117	Value 118	Value 119	Value 120	Value 121	Value 122	Value 123	Value 124	Value 125	Value 126	Value 127	Value 128	Value 129	Value 130	Value 131	Value 132	Value 133	Value 134	Value 135	Value 136	Value 137	Value 138	Value 139	Value 140	Value 141	Value 142	Value 143	Value 144	Value 145	Value 146	Value 147	Value 148	Value 149	Value 150	Value 151	Value 152	Value 153	Value 154	Value 155	Value 156	Value 157	Value 158	Value 159	Value 160	Value 161	Value 162	Value 163	Value 164	Value 165	Value 166	Value 167	Value 168	Value 169	Value 170	Value 171	Value 172	Value 173	Value 174	Value 175	Value 176	Value 177	Value 178	Value 179	Value 180	Value 181	Value 182	Value 183	Value 184	Value 185	Value 186	Value 187	Value 188	Value 189	Value 190	Value 191	Value 192	Value 193	Value 194	Value 195	Value 196	Value 197	Value 198	Value 199	Value 200	Value 201	Value 202	Value 203	Value 204	Value 205	Value 206	Value 207	Value 208	Value 209	Value 210	Value 211	Value 212	Value 213	Value 214	Value 215	Value 216	Value 217	Value 218	Value 219	Value 220	Value 221	Value 222	Value 223	Value 224	Value 225	Value 226	Value 227	Value 228	Value 229	Value 230	Value 231	Value 232	Value 233	Value 234	Value 235	Value 236	Value 237	Value 238	Value 239	Value 240	Value 241	Value 242	Value 243	Value 244	Value 245	Value 246	Value 247	Value 248	Value 249	Value 250	Value 251	Value 252	Value 253	Value 254	Value 255	Value 256	Value 257	Value 258	Value 259	Value 260	Value 261	Value 262	Value 263	Value 264	Value 265	Value 266	Value 267	Value 268	Value 269	Value 270	Value 271	Value 272	Value 273	Value 274	Value 275	Value 276	Value 277	Value 278	Value 279	Value 280	Value 281	Value 282	Value 283	Value 284	Value 285	Value 286	Value 287	Value 288	Value 289	Value 290	Value 291	Value 292	Value 293	Value 294	Value 295	Value 296	Value 297	Value 298	Value 299	Value 300	Value 301	Value 302	Value 303	Value 304	Value 305	Value 306	Value 307	Value 308	Value 309	Value 310	Value 311	Value 312	Value 313	Value 314	Value 315	Value 316	Value 317	Value 318	Value 319	Value 320	Value 321	Value 322	Value 323	Value 324	Value 325	Value 326	Value 327	Value 328	Value 329	Value 330	Value 331	Value 332	Value 333	Value 334	Value 335	Value 336	Value 337	Value 338	Value 339	Value 340	Value 341	Value 342	Value 343	Value 344	Value 345	Value 346	Value 347	Value 348	Value 349	Value 350	Value 351	Value 352	Value 353	Value 354	Value 355	Value 356	Value 357	Value 358	Value 359	Value 360	Value 361	Value 362	Value 363	Value 364	Value 365	Value 366	Value 367	Value 368	Value 369	Value 370	Value 371	Value 372	Value 373	Value 374	Value 375	Value 376	Value 377	Value 378	Value 379	Value 380	Value 381	Value 382	Value 383	Value 384	Value 385	Value 386	Value 387	Value 388	Value 389	Value 390	Value 391	Value 392	Value 393	Value 394	Value 395	Value 396	Value 397	Value 398	Value 399	Value 400	Value 401	Value 402	Value 403	Value 404	Value 405	Value 406	Value 407	Value 408	Value 409	Value 410	Value 411	Value 412	Value 413	Value 414	Value 415	Value 416	Value 417	Value 418	Value 419	Value 420	Value 421	Value 422	Value 423	Value 424	Value 425	Value 426	Value 427	Value 428	Value 429	Value 430	Value 431	Value 432	Value 433	Value 434	Value 435	Value 436	Value 437	Value 438	Value 439	Value 440	Value 441	Value 442	Value 443	Value 444	Value 445	Value 446	Value 447	Value 448	Value 449	Value 450	Value 451	Value 452	Value 453	Value 454	Value 455	Value 456	Value 457	Value 458	Value 459	Value 460	Value 461	Value 462	Value 463	Value 464	Value 465	Value 466	Value 467	Value 468	Value 469	Value 470	Value 471	Value 472	Value 473	Value 474	Value 475	Value 476	Value 477	Value 478	Value 479	Value 480	Value 481	Value 482	Value 483	Value 484	Value 485	Value 486	Value 487	Value 488	Value 489	Value 490	Value 491	Value 492	Value 493	Value 494	Value 495	Value 496	Value 497	Value 498	Value 499	Value 500	Value 501	Value 502	Value 503	Value 504	Value 505	Value 506	Value 507	Value 508	Value 509	Value 510	Value 511	Value 512	Value 513	Value 514	Value 515	Value 516	Value 517	Value 518	Value 519	Value 520	Value 521	Value 522	Value 523	Value 524	Value 525	Value 526	Value 527	Value 528	Value 529	Value 530	Value 531	Value 532	Value 533	Value 534	Value 535	Value 536	Value 537	Value 538	Value 539	Value 540	Value 541	Value 542	Value 543	Value 544	Value 545	Value 546	Value 547	Value 548	Value 549	Value 550	Value 551	Value 552	Value 553	Value 554	Value 555	Value 556	Value 557	Value 558	Value 559	Value 560	Value 561	Value 562	Value 563	Value 564	Value 565	Value 566	Value 567	Value 568	Value 569	Value 570	Value 571	Value 572	Value 573	Value 574	Value 575	Value 576	Value 577	Value 578	Value 579	Value 580	Value 581	Value 582	Value 583	Value 584	Value 585	Value 586	Value 587	Value 588	Value 589	Value 590	Value 591	Value 592	Value 593	Value 594	Value 595	Value 596	Value 597	Value 598	Value 599	Value 600	Value 601	Value 602	Value 603	Value 604	Value 605	Value 606	Value 607	Value 608	Value 609	Value 610	Value 611	Value 612	Value 613	Value 614	Value 615	Value 616	Value 617	Value 618	Value 619	Value 620	Value 621	Value 622	Value 623	Value 624	Value 625	Value 626	Value 627	Value 628	Value 629	Value 630	Value 631	Value 632	Value 633	Value 634	Value 635	Value 636	Value 637	Value 638	Value 639	Value 640	Value 641	Value 642	Value 643	Value 644	Value 645	Value 646	Value 647	Value 648	Value 649	Value 650	Value 651	Value 652	Value 653	Value 654	Value 655	Value 656	Value 657	Value 658	Value 659	Value 660	Value 661	Value 662	Value 663	Value 664	Value 665	Value 666	Value 667	Value 668	Value 669	Value 670	Value 671	Value 672	Value 673	Value 674	Value 675	Value 676	Value 677	Value 678	Value 679	Value 680	Value 681	Value 682	Value 683	Value 684	Value 685	Value 686	Value 687	Value 688	Value 689	Value 690	Value 691	Value 692	Value 693	Value 694	Value 695	Value 696	Value 697	Value 698	Value 699	Value 700	Value 701	Value 702	Value 703	Value 704	Value 705	Value 706	Value 707	Value 708	Value 709	Value 710	Value 711	Value 712	Value 713	Value 714	Value 715	Value 716	Value 717	Value 718	Value 719	Value 720	Value 721	Value 722	Value 723	Value 724	Value 725	Value 726	Value 727	Value 728	Value 729	Value 730	Value 731	Value 732	Value 733	Value 734	Value 735	Value 736	Value 737	Value 738	Value 739	Value 740	Value 741	Value 742	Value 743	Value 744	Value 745	Value 746	Value 747	Value 748	Value 749	Value 750	Value 751	Value 752	Value 753	Value 754	Value 755	Value 756	Value 757	Value 758	Value 759	Value 760	Value 761	Value 762	Value 763	Value 764	Value 765	Value 766	Value 767	Value 768	Value 769	Value 770	Value 771	Value 772	Value 773	Value 774	Value 775	Value 776	Value 777	Value 778	Value 779	Value 780	Value 781	Value 782	Value 783	Value 784	Value 785	Value 786	Value 787	Value 788	Value 789	Value 790	Value 791	Value 792	Value 793	Value 794	Value 795	Value 796	Value 797	Value 798	Value 799	Value 800	Value 801	Value 802	Value 803	Value 804	Value 805	Value 806	Value 807	Value 808	Value 809	Value 810	Value 811	Value 812	Value 813	Value 814	Value 815	Value 816	Value 817	Value 818	Value 819	Value 820	Value 821	Value 822	Value 823	Value 824	Value 825	Value 826	Value 827	Value 828	Value 829	Value 830	Value 831	Value 832	Value 833	Value 834	Value 835	Value 836	Value 837	Value 838	Value 839	Value 840	Value 841	Value 842	Value 843	Value 844	Value 845	Value 846	Value 847	Value 848	Value 849	Value 850	Value 851	Value 852	Value 853	Value 854	Value 855	Value 856	Value 857	Value 858	Value 859	Value 860	Value 861	Value 862	Value 863	Value 864	Value 865	Value 866	Value 867	Value 868	Value 869	Value 870	Value 871	Value 872	Value 873	Value 874	Value 875	Value 876	Value 877	Value 878	Value 879	Value 880	Value 881	Value 882	Value 883	Value 884	Value 885	Value 886	Value 887	Value 888	Value 889	Value 890	Value 891	Value 892	Value 893	Value 894	Value 895	Value 896	Value 897	Value 898	Value 899	Value 900	Value 901	Value 902	Value 903	Value 904	Value 905	Value 906	Value 907	Value 908	Value 909	Value 910	Value 911	Value 912	Value 913	Value 914	Value 915	Value 916	Value 917	Value 918	Value 919	Value 920	Value 921	Value 922	Value 923	Value 924	Value 925	Value 926	Value 927	Value 928	Value 929	Value 930	Value 931	Value 932	Value 933	Value 934	Value 935	Value 936	Value 937	Value 938	Value 939	Value 940	Value 941	Value 942	Value 943	Value 944	Value 945	Value 946	Value 947	Value 948	Value 949	Value 950	Value 951	Value 952	Value 953	Value 954	Value 955	Value 956	Value 957	Value 958	Value 959	Value 960	Value 961	Value 962	Value 963	Value 964	Value 965	Value 966	Value 967	Value 968	Value 969	Value 970	Value 971	Value 972	Value 973	Value 974	Value 975	Value 976	Value 977	Value 978	Value 979	Value 980	Value 981	Value 982	Value 983	Value 984	Value 985	Value 986	Value 987	Value 988	Value 989	Value 990	Value 991	Value 992	Value 993	Value 994	Value 995	Value 996	Value 997	Value 998	Value 999	Value 1000	Value 1001	Value 1002	Value 1003	Value 1004	Value 1005	Value 1006	Value 1007	Value 1008	Value 1009	Value 1010	Value 1011	Value 1012	Value 1013	Value 1014	Value 1015	Value 1016	Value 1017	Value 1018	Value 1019	Value 1020	Value 1021	Value 1022	Value 1023	Value 1024	Value 1025	Value 1026	Value 1027	Value 1028	Value 1029	Value 1030	Value 1031	Value 1032	Value 1033	Value 1034	Value 1035	Value 1036	Value 1037	Value 1038	Value 1039	Value 1040	Value 1041	Value 1042	Value 1043	Value 1044	Value 1045	Value 1046	Value 1047	Value 1048	Value 1049	Value 1050	Value 1051	Value 1052	Value 1053	Value 1054	Value 1055	Value 1056	Value 1057	Value 1058	Value 1059	Value 1060	Value 1061	Value 1062	Value 1063	Value 1064	Value 1065	Value 1066	Value 1067	Value 1068	Value 1069	Value 1070	Value 1071	Value 1072	Value 1073	Value 1074	Value 1075	Value 1076	Value 1077	Value 1078	Value 1079	Value 1080	Value 1081	Value 1082	Value 1083	Value 1
----	----------	---------	---------	---------	---------	---------	---------	---------	---------	---------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	---------

Materials Science



Daphalapurkar, Brady, Ramesh, Molinari. JMPS (2011)

Open Numerical Laboratories

- ▶ Create interactive Numerical Laboratories
- ▶ Analysis server-side through web service
- ▶ Use virtual sensor metaphor
- ▶ Many access patterns are local
- ▶ No need to download whole data sets
- ▶ Concept very successful in turbulence and cosmological N-body
- ▶ `turbulence.pha.jhu.edu:`
19 trillion points delivered!
- ▶ Total science data in SciServer currently ~2.5PB

New Analysis Patterns Emerging

- ▶ User written crawlers, inefficient
- ▶ Cutouts delivered to users, slow
- ▶ Scalability challenge (over 100TB scales)
- ▶ Requests for scripting access
- ▶ Need for easy joins with long-tail data
- ▶ Still expecting interactive response

Architectural Challenges

- ▶ Need to define sharp tradeoffs
 - Data Analytics system is different from supercomputer
 - What is the right balance between I/O and compute?
- ▶ Need high bandwidth to large data
 - Computations/visualizations must be on top of the data
 - Must support at least few 100TB per server
 - Petascale: 3 copies for production (or erasure code?)
 - Wide area data movement/backbone is hard
- ▶ Lessons from the database world:
 - It is nontrivial to schedule complex I/O patterns
 - For subsets we must use indexing, cache resilient storage

Directions

- ▶ Offer more computing resources server side
- ▶ Enhanced visualization tools (ParaView)
- ▶ **Augment and combine** SQL queries with easy-to-use scripting tools
- ▶ **Heavy use of virtual machines/ Docker**
- ▶ **Interactive portal via iPython/Matlab/R**

SciServer



Collaborative data-driven science

Workshop Overview

Mike Rippin

April 27, 2016



JOHNS HOPKINS
UNIVERSITY



idies

Introductions and Logistics

The Team

- ✧ Alex Szalay (PI)
- ✧ Mike Rippin (PM)
- ✧ Ani Thakar
- ✧ Jordan Raddick
- ✧ Bonnie Souter
- ✧ Gerard Lemson
- ✧ Jaiwon Kim
- ✧ Dmitry Medvedev
- ✧ Deoyani Heinis
- ✧ Manu Popp
- ✧ Victor Paul
- ✧ Sue Werner
- ✧ Jan Vandenberg

Agenda

8:30 AM	Continental Breakfast & Coffee	
9:00 AM	Welcome	Alex Szalay
9:05 AM	SciServer Overview	Mike Rippin
9:25 AM	Getting Started with SciServer	Jordan Raddick
9:40 AM	Technical Overview	Dmitry Medvedev
10:30 AM	Coffee	
10:45 AM	Demo Notebook #1	Gerard Lemson
12:00 PM	Lunch	
1:00 PM	Astronomy & Cosmology Examples	Gerard Lemson
2:30 PM	Break	
2:45 PM	Explore & Customize	Participants
3:30 PM	Q&A	
3:50 PM	Closing Remarks	Mike Rippin
4:00 PM	Adjourn	

Logistics

- ▶ Stay in this room all day
- ▶ Restrooms
- ▶ Coffee and Breaks – morning and afternoon
- ▶ Lunch – ‘Working Lunch’ if preferred
- ▶ Wrap up – 4pm

Structure of the Day

Test Environment

- ▶ Technology
 - Everyone should be able to connect to WIFI
 - Everyone will create an account
 - Everyone will create a Docker Container
- ▶ Workshop running in TEST Environment
- ▶ MyDB etc is temporary
- ▶ Jupyter Notebooks can be saved and taken away

Objectives of the Workshop

Participants	SciServer Team
<ul style="list-style-type: none">• Set up a SciServer Notebook• Authenticate with the SciServer Login Portal• Import and query SDSS with CasJobs• Save your data and graphics locally• Save your data and graphics on SciDrive• Save & Retrieve your data in MyDB• Learn the SciServer API	<ul style="list-style-type: none">• Test the Compute feature set• Test out the Architecture• Gain early feedback from participants• Implement this feedback before live release

We want this to be Interactive...

Structure of the Workshop

- ▶ Agenda sets the scene
- ▶ To start: *Structured*
 - First example workbooks cover the ‘building blocks’ and will be done in a structured way
- ▶ Subsequently: *Flexible*
 - Notebooks delve deeper into specifics
 - Timing and deviations are fine, Q&A, examples etc
 - Tune to the experiences and needs of participants

Emphasis on PRACTICAL exercises...

SciServer Project Background

SciServer Project Award

Award

- ▶ NSF DIBBs (Data Infrastructure Building Blocks)
- ▶ 5 years: 2013 – 2018
- ▶ Approx \$10M
- ▶ Cooperative Agreement

SciServer Project Objectives

Objectives

- ▶ Extend infrastructure for SDSS to support additional Science Domains
- ▶ Host and serve petabyte datasets
- ▶ Support custom user datasets
- ▶ Provide access and query services
- ▶ Provide scalable compute services
- ▶ Support analyses and data sets too large to handle locally
- ▶ Provide collaborative tools for shared analysis

Computations stay CLOSE to the DATA...

SciServer Project Components

Major Components	Supporting Technologies
<i>Core</i>	Microsoft SQL Server
• Login Portal	Open Stack
• CASJobs	Docker
• SciServer Compute	Jupyter
• SciDrive	
<i>Applications</i>	
• SkyQuery	
• SkyServer	
• GLUSEEN	
• Turbulence	

SciServer Project Timeline

Timelines

Year 1 (2013-2014)	Project Setup, Scoping, Planning, Begin Refactoring, SDSS Unification
Year 2 (2013-2014)	Architectural Refactoring – API, Single Sign-on, prototype Compute
Year 3 (2013-2014)	SciServer System Release, Interactive Compute, Scalable Job Management, Basic Dashboard, Initial Collaborative capabilities
Year 4 (2013-2014)	Implementation in Science Domains, Educational workbooks
Year 5 (2013-2014)	System Scale out, Data Analytics, Advanced Deployment Scenarios



SciServer Project Current Plans

Timelines – Year 3

Apr 2016	<ul style="list-style-type: none">• SciServer System Release
May 2016	<ul style="list-style-type: none">• Interactive Compute• SkyQuery• Gluseen
August 2016	<ul style="list-style-type: none">• Prototype Scalable Job Management• Basic Dashboard• Initial Collaborative capabilities
October 2016	<ul style="list-style-type: none">• Scalable Job Management• Turbulence• Cosmology
November 2016	<ul style="list-style-type: none">• Project 3 year Review

SciServer



Collaborative data-driven science

Getting Started with SciServer

Jordan Raddick
April 27, 2016



Resources

- ▶ Agenda:

[www.sciserver.org/outreach/spring-workshop/
detailed-agenda](http://www.sciserver.org/outreach/spring-workshop/detailed-agenda)

- ▶ Documentation and Support (go here now!):

[www.sciserver.org/outreach/spring-workshop/
documentation](http://www.sciserver.org/outreach/spring-workshop/documentation)



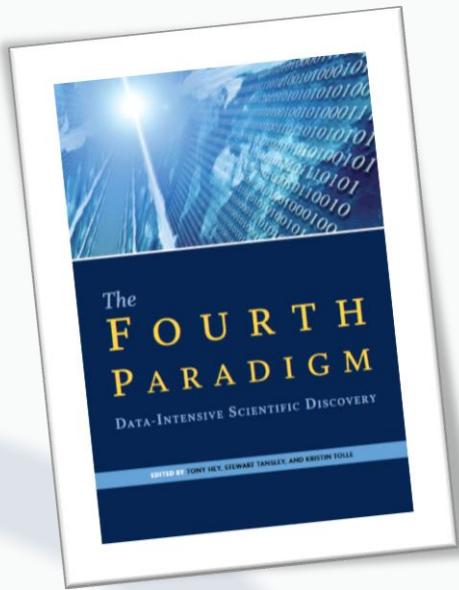
Collaborative data-driven science

Inside *SciServer Compute*

Dmitry Medvedev
Johns Hopkins University



Inspired by...



"For data analysis, one possibility is to move the data to you, but the other possibility is to move your query to the data... Often it turns out to be more efficient to move the questions than to move the data."

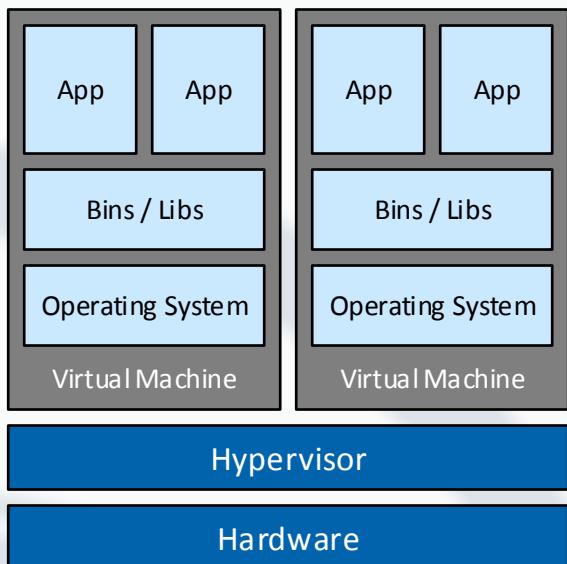
-- Jim Gray

Helen Shen's article for *Nature* – **Interactive notebooks: Sharing the code** – featured a live demo of IPython notebooks created on-demand using Docker containers, and made a strong case for using IPython notebooks in scientific data analysis.

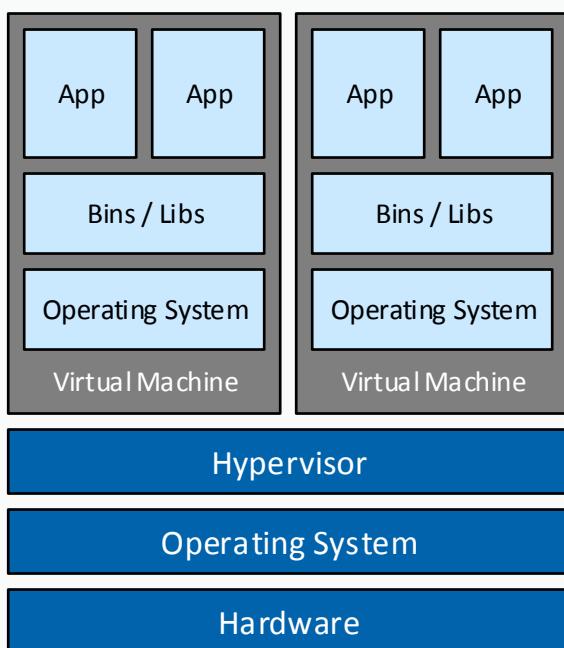
What is SciServer Compute?

- ▶ Interactive Jupyter notebooks hosted inside Docker containers.
- ▶ Pre-configured images to create new containers from (R, Python, MATLAB, ...).
- ▶ High-bandwidth, low-latency access to other SciServer services and data sources through the notebooks.
- ▶ Users manage their own containers.

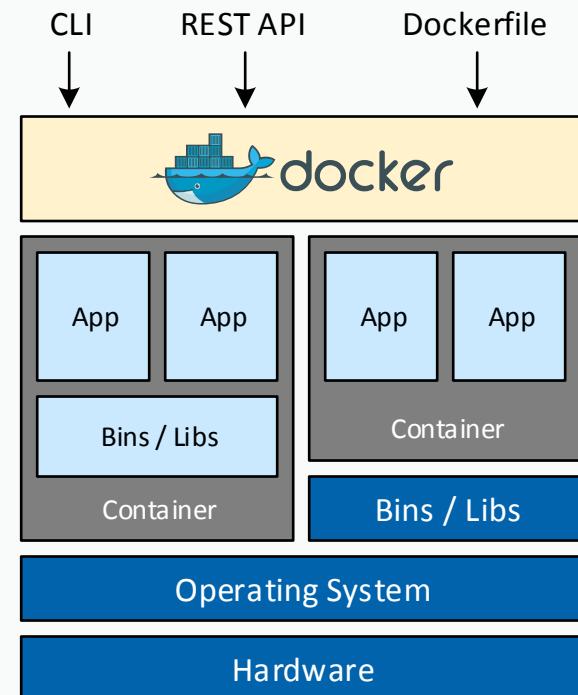
What are Docker Containers?



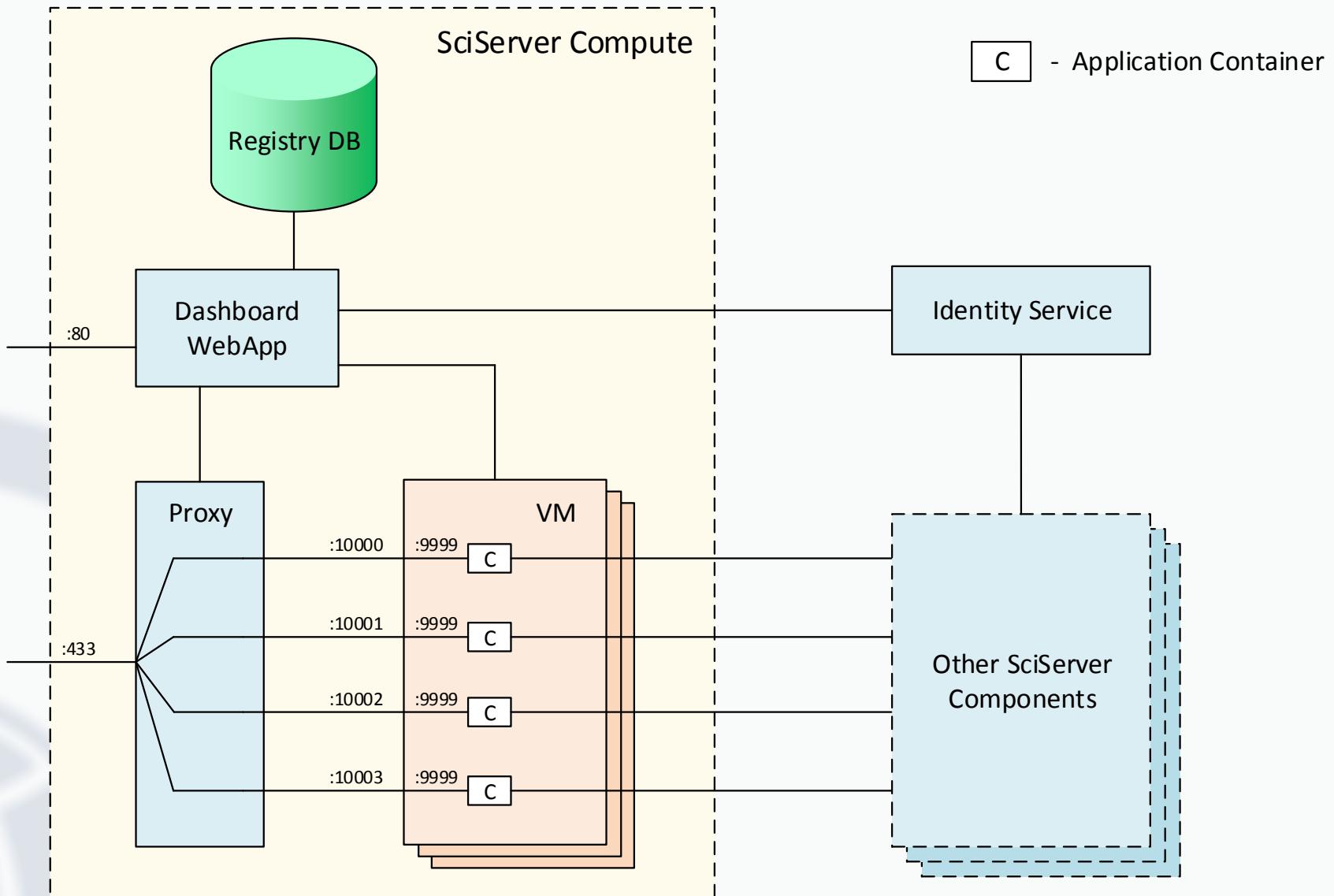
Type 1 Hypervisor



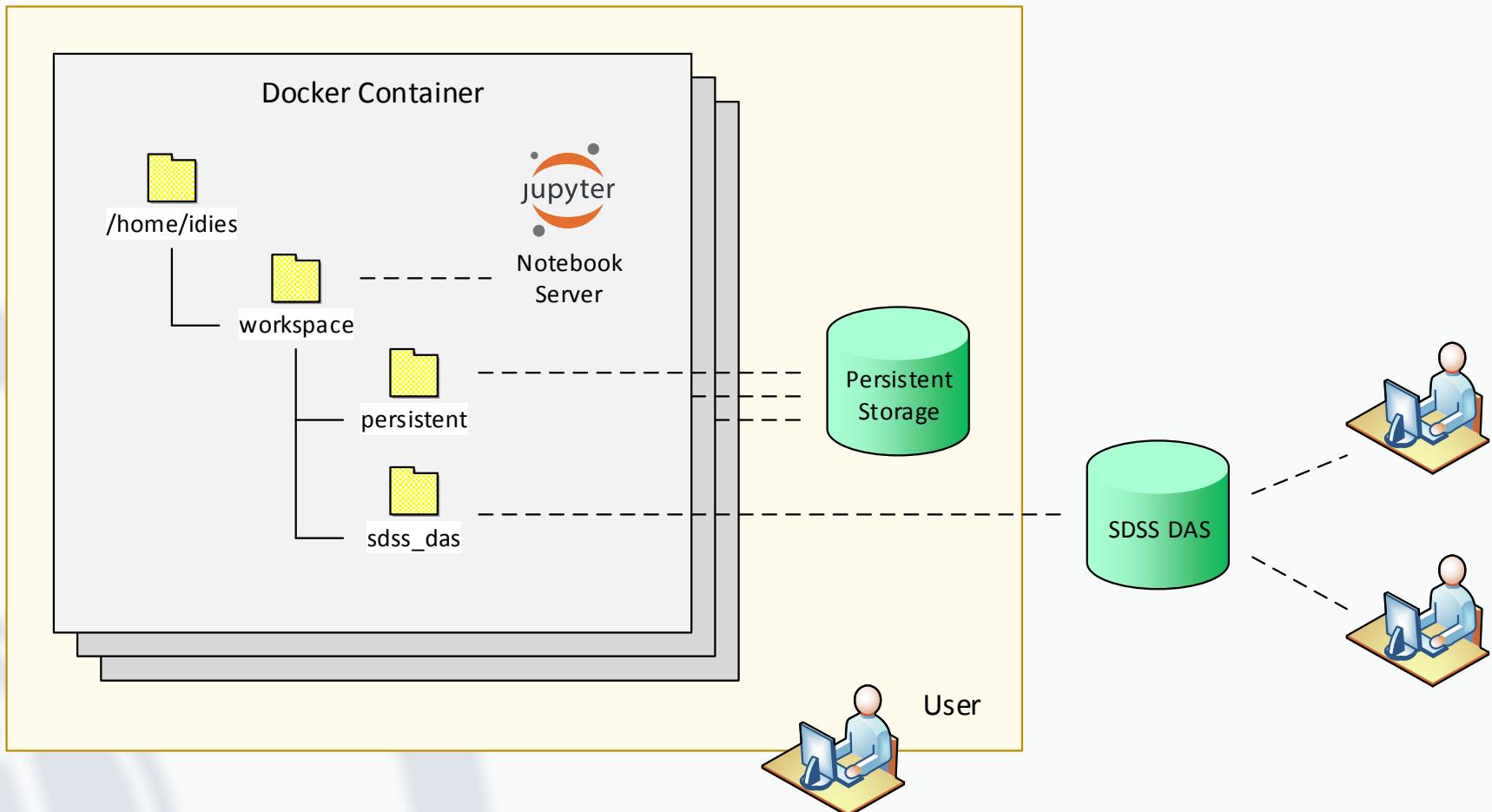
Type 2 Hypervisor



Linux Containers



Data Storage Configuration



Work in Progress

- ▶ Run asynchronous non-interactive jobs in separate Docker containers. It's meant to be more than just Jupyter notebooks!
- ▶ Create new VM nodes on-demand to accommodate growing number of users.
- ▶ Provide scratch (temporary) storage space for working with large amounts of data.
- ▶ Improve resource management.

Questions?